Multivariate multiscale complexity
analysis

MOSABBER UDDIN AHMED

SUPERVISOR: PROF. DANilo P. MANDIC

A Thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy and Diploma of the Imperial College London

Communications and Signal Processing Group
Department of Electrical and Electronic Engineering
Imperial College London
2012
Abstract

Established dynamical complexity analysis measures operate at a single scale and thus fail to quantifi

Abstract

Established dynamical complexity analysis measures operate at a single scale and thus fail to quantify inherent long-range correlations in real world data, a key feature of complex systems. They are designed for scalar time series, however, multivariate observations are common in modern real world scenarios and their simultaneous analysis is a prerequisite for the understanding of the underlying signal generating model. To that end, this thesis first introduces a notion of multivariate sample entropy and thus extends the current univariate complexity analysis to the multivariate case. The proposed multivariate multiscale entropy (MMSE) algorithm is shown to be capable of addressing the dynamical complexity of such data directly in the domain where they reside, and at multiple temporal scales, thus making full use of all the available information, both within and across the multiple data channels. Next, the intrinsic multivariate scales of the input data are generated adaptively via the multivariate empirical mode decomposition (MEMD) algorithm. This allows for both generating comparable scales from multiple data channels, and for temporal scales of same length as the length of input signal, thus, removing the critical limitation on input data length in current complexity analysis methods. The resulting MEMD-enhanced MMSE method is also shown to be suitable for non-stationary multivariate data analysis owing to the data-driven nature of MEMD algorithm, as non-stationarity is the biggest obstacle for meaningful complexity analysis. This thesis presents a quantum step forward in this area, by introducing robust and physically meaningful complexity estimates of real-world systems, which are typically multivariate, finite in duration, and of noisy and heterogeneous natures. This also allows us to gain better understanding of the complexity of the underlying multivariate model and more degrees of freedom and rigor in the analysis. Simulations on both synthetic and real world multivariate data sets support the analysis.
Acknowledgment

First and foremost, I am heartily thankful to my supervisor, Prof. Danilo P. Mandic, for his helpful suggestions, constant encouragement and patience in guiding me to finish this work. I am very grateful to him for his faith in me and I hope to have fulfilled his expectations. I would also like to express my gratitude to my colleagues for extending their helping hands to me in difficult situation. In particular, I would like to thank Dr. Naveed-ur-Rehman, Dr. David Looney, Dr. Ling Li and Dr. Clive Cheong Took with whom I share thoughts and exchange opinions regarding my work oftentimes. Next, I would like to extend my gratitude to Commonwealth Scholarship Commission (CSC) in the United Kingdom for sponsoring my study and University of Dhaka for granting study leave.

On a private note I would like to thank my parents, brothers and sister-in-laws for their endless support and insightful advice. And a special mention to my charming wife Adri and my beautiful daughter Mounia for their unconditional love and perpetual inspiration. Last but not least, I would like to express my deepest gratitude to God for His divine blessings in my life.
## Contents

Abstract .................................................. 2
Acknowledgment ........................................... 3
Contents .................................................... 4
List of Figures ............................................ 7
List of Tables ............................................. 10
Statement of Originality ................................. 11
List of Publications ...................................... 12
Abbreviations ............................................. 14

### Chapter 1. Introduction ................................ 16
1.1 Background and Motivation ........................... 16
1.2 Aim and Objectives .................................... 20
1.3 Original Contributions ................................ 22
1.4 Outline of the Thesis .................................. 23

### Chapter 2. Complexity and Its Measures ............. 25
2.1 What is Complexity? ................................... 25
2.2 Manifestation of Complexity .......................... 28
2.3 Classification of Complexity Measures ............... 33
2.4 Representative Complexity Measures ................ 35
  2.4.1 Information-theoretic Complexity Measures ....... 35
  2.4.2 Algorithmic Complexity Measures ................. 41
  2.4.3 Chaos-theoretic Complexity Measures ............... 43
  2.4.4 Random Fractal-theoretic Complexity Measures .... 44

### Chapter 3. Multiscale Entropy Analysis .............. 45
3.1 Introduction ................................................................. 46
3.2 Quantitative Complexity Estimators ................................. 48
  3.2.1 Sample Entropy ..................................................... 48
  3.2.2 Permutation Entropy .............................................. 50
  3.2.3 Lempel-Ziv Entropy ............................................... 52
3.3 Multiscale Entropy Framework ....................................... 53
3.4 Calibration with $1/f$ Noise and White Noise ....................... 53
3.5 Application to Characterize Brain Dynamics ........................ 56
  3.5.1 Data ................................................................. 60
  3.5.2 Statistical Analysis .............................................. 61
  3.5.3 Simulation Results ............................................... 61
  3.5.4 Classification .................................................... 63
  3.5.5 Discussion ....................................................... 65
3.6 Conclusion ............................................................... 67

Chapter 4. Multivariate Multiscale Entropy Analysis ................. 68
  4.1 Introduction ........................................................... 68
  4.2 Multivariate Embedding ............................................ 69
  4.3 Multivariate Sample Entropy ....................................... 71
    4.3.1 Geometric Interpretation of MSampEn .......................... 74
    4.3.2 Effect of Data Length on MSampEn ............................ 76
    4.3.3 Effect of Embedding Parameter on MSampEn .................. 78
    4.3.4 Effect of Threshold Parameter on MSampEn .................. 78
  4.4 Multivariate Multiscale Entropy ................................... 80
  4.5 Multivariate Complexity Analysis ................................ 82
    4.5.1 Validation on Synthetic Data ................................. 82
  4.6 Applications for Real World Multivariate Processes ............. 86
    4.6.1 Case Study: Stride Interval Characterization ............... 86
    4.6.2 Case Study: Postural Sway Analysis .......................... 90
    4.6.3 Case Study: Complexity Analysis of Physiological Signals .. 92
  4.7 Conclusion ........................................................... 94

Chapter 5. Multivariate Entropy Analysis with Data Adaptive Scales 95
  5.1 Introduction ........................................................... 96
  5.2 Multivariate Empirical Mode Decomposition ...................... 99
    5.2.1 Simulations of MEMD Properties ............................... 104
  5.3 MEMD-enhanced Multivariate Multiscale Entropy .................. 106
    5.3.1 Validation on Synthetic Data ................................. 108
  5.4 Applications for Real World Multivariate Processes ............ 110
### Chapter 6. Applications of Multivariate Multiscale Complexity Analysis

#### 6.1 Introduction
- Page 114

#### 6.2 Brain Consciousness Analysis
- Page 115

#### 6.3 Complexity Analysis of Meteorological Data
- Page 116
  - 6.3.1 Wind and Air Temperature Data During Different Cloud Covering
  - Page 118
  - 6.3.2 Wind and Air Temperature Data at Different Altitudes
  - Page 120
  - 6.3.3 Wind and Air Temperature Data from Various Wind Dynamics Regimes
  - Page 122
  - 6.3.4 Wind Data from Different Fan Systems
  - Page 125
  - 6.3.5 Discussion and scope of the work
  - Page 127

#### 6.4 Complexity Analysis of Eye Gaze Dynamics
- Page 128

#### 6.5 Complexity Analysis of Heart and Respiratory Function During Stress
- Page 131

#### 6.6 Complexity Analysis of Electrohysterogram (EHG) Records
- Page 132

#### 6.7 Conclusion
- Page 135

### Chapter 7. Conclusions and Future Work

#### 7.1 Conclusions
- Page 136

#### 7.2 Future work
- Page 139

### Bibliography
- Page 142

### Appendix A. MSampEn Calibration with Correlated and Uncorrelated Noises
- Page 155

### Appendix B. Hurst Parameter
- Page 158
List of Figures

2.1 An illustration to show what is complex ............................ 26
2.2 Two basic forms of fractals ............................................ 30
2.3 An illustration of self-similar vs. self-affine structure ......... 30
2.4 An illustration of ‘butterfly effect’ ................................. 31
2.5 Three dimensional sketch of a fitness landscape .............. 32
2.6 Classification of complexity measures ............................. 33

3.1 An illustration of SampEn calculation ............................... 48
3.2 An illustration of PermEn calculation .............................. 51
3.3 A schematic diagram of the coarse-graining procedure for scale 2 and 3 ... 54
3.4 Plot of MSE analysis for $1/f^\beta$ noises with $0 \leq \beta \leq 1$, that is, for noises with power-law correlations ................................. 55
3.5 Plot of MSE analysis for random white and $1/f$ noise .......... 56
3.6 Plot of multiscale PermEn curves for random white and $1/f$ noise ...... 57
3.7 Plot of multiscale LZEn curves for random white and $1/f$ noise .... 58
3.8 Multiscale entropy analysis using different entropy estimators. Symbols represent the mean values of entropy for each set and the bars represent the standard error ........................................... 64

4.1 An illustration of Embedding theorem ............................... 70
4.2 Geometry behind the calculation of MSampEn, for a two-dimensional wind signal ......................................................... 75
4.3 Sensitivity of MSampEn to data length ......................... 77
4.4 Sensitivity of MSampEn to embedding parameter ............. 79
4.5 Sensitivity of MSampEn to threshold parameter ............. 81
4.6 MMSE analysis for 3-channel data containing white and $1/f$ noise ...... 83
4.7 MMSE analysis for correlated vs uncorrelated trivariate white and $1/f$ noise 84
4.8 Standard univariate multiscale entropy analysis for white noise and scalar AR processes ................................................. 85
4.9 MMSE analysis for bivariate white noise and bivariate AR processes .... 86
4.10 MMSE analysis for self-paced vs metronomically-paced stride interval (human gait) time series ........................................ 87
4.11 MMSE analysis of COP time series for young and elderly subjects ....... 90
4.12 MSE analysis for the cardiac and respiratory dynamics .................. 93
4.13 MMSE analysis of the bivariate (R-R, Interbreath interval) signal ....... 94

5.1 Magnitude response of the equivalent FIR filter in the coarse graining process for different scale factor $\epsilon$ .............................................. 97
5.2 An example of real world wind signal being analyzed by MEMD ........ 102
5.3 Averaged spectra of IMFs obtained for a $N = 500$ realizations of 8-channel white Gaussian noise using MEMD (top) and the standard EMD (bottom) 103
5.4 The normalized IMF cross-correlation matrices from multiple channels of wind speed data, obtained via the univariate EMD channel-wise (a, c and e) and by direct multivariate EMD (MEMD) (b, d and f). .......................... 105
5.5 MEMD-enhanced MMSE analysis for bivariate white and $1/f$ noise .... 109
5.6 MEMD-enhanced MMSE analysis for 3-channel data containing white and $1/f$ noise ................................................................. 110
5.7 MEMD-enhanced MMSE analysis for self-paced vs metronomically-paced stride interval time series .............................................. 111
5.8 MEMD-enhanced MMSE analysis of COP time series for young and elderly subjects ................................................................. 113

6.1 MMSE analysis for EEG signals from coma vs QBD patients ............ 116
6.2 MMSE analysis of 3D wind data ............................................. 119
6.3 MMSE analysis of 2D wind data from different altitudes ................. 121
6.4 Magnitude of the 3D wind signal. The wind dynamics regimes are identified as ‘low’, ‘medium’, and ‘high’ ........................................... 123
6.5 MMSE analysis of 3D wind data from different wind dynamics regimes . 124
6.6 MMSE analysis of 4D wind data from different wind dynamics regimes . 125
6.7 MMSE analysis of 3D wind data from different fan systems ............. 126
6.8 The classic ‘Yarbus experiment’ illustration .................................... 129
6.9 Segment of raw gaze data, both horizontal and vertical components. .... 129
6.10 MMSE analysis of the classic ‘Yarbus experiment’ .......................... 130
6.11 MMSE analysis for ‘normal’ and ‘stressed’ states, based on heart and respiratory functions ........................................... 132
6.12 Coarse graining based MMSE analysis of filtered 3-channel UEMG signal . 133
6.13 MEMD-enhanced MMSE analysis of filtered 3-channel UEMG signal . . . . 134

A.1 Geometric interpretation of MSampEn calculation from uncorrelated bivariate white noise ................................. 156
A.2 Geometric interpretation of MSampEn calculation from correlated bivariate white noise ........................................ 157
List of Tables

3.1 Summary of different entropy statistic measured in different sets . . . . . . . 62
3.2 Result of the two statistical tests . . . . . . . . . . . . . . . . . . . . . . . 62
3.3 The classification accuracy for the different single scale entropy measures . 65
3.4 The classification accuracy for the different entropy measures in the multi-
    scale framework . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65

4.1 Statistical significance tests for the univariate, bivariate and trivariate hu-
    man stride interval analysis . . . . . . . . . . . . . . . . . . . . . . . . . . 88
4.2 Complexity indices and p values obtained using one-tailed Student’s t-test
    with unequal variance. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 91

B.1 Summary of Hurst exponents estimated for all input data sets using stan-
    dard methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 159
Statement of Originality

I declare that this is an original thesis and it is entirely based on my own work. I acknowledge the sources in every instance where I used the ideas of other writers and where I used any diagrams or visuals. This thesis was not and will not be submitted to any other university or institution for fulfilling the requirements of a degree.
List of Publications

The following publications support the material given in this thesis.

**Refereed Journals:**


**Refereed Conferences:**

Abbreviations

ANOVA: Analysis of Variance
ApEn: Approximate Entropy
AR: Autoregressive
ARIMA: Autoregressive Integrated Moving Average
CAS: Complex Adaptive System
COP: Center of Pressure
DFA: Detrended Fluctuation Analysis
EEG: Electroencephalography
EHG: Electrohysterography
EMC: Effective Measure Complexity
EMD: Empirical Mode Decomposition
IBI: Inter Beat Interval
IMF: Intrinsic Mode Function
LMS: Least Mean Squares
LOOC: Leave-one-out Classification
LZEn: Lempel-Ziv Entropy
MEG: Magnetoencephalography
MEMD: Multivariate Empirical Mode Decomposition
MMPE: Multivariate Multiscale Permutation Entropy
MMSE: Multivariate Multiscale Entropy
MSE: Multiscale Entropy
NNC: Nearest Neighbour Classification
0. Abbreviations

PermEn: Permutation Entropy

SampEn: Sample Entropy

SE: Sample Error

VAR: Vector Autoregressive

VARMA: Vector Autoregressive Moving Average

UEMG: Uterine Electromyography
Chapter 1

Introduction

1.1 Background and Motivation

We live in a complex world. From natural systems (e.g., brains, immune systems, ecologies, societies) to many artificial systems (e.g., parallel and distributed computing systems, artificial intelligence systems, Internet), complexity is ubiquitous and is characterized by the spontaneous emergent behavior arising from nonlinear spatio-temporal interactions among a large number of component systems at different levels of organization. Researchers from diverse disciplines such as physics, biology, economics, ecology and archaeology are trying to define the notion of complexity and measure it. As a result, the literature related to complexity science is extensive and typically each complexity analysis measure focuses on one particular aspect of the system, such as dimensionality, hierarchical organization, regularity or irregularity, randomness, predictability, self-similarity, compressibility, synchrony, long-range correlations etc. [1] [2] [3].

Recent advances in nonlinear dynamics theory have highlighted the need for more insight into the dynamical properties of the complex phenomena under observation. In nonlinear systems theory, the time delay embedded reconstruction [4] provides a general framework for the estimation of invariant quantities (in terms of smooth transformations of the state space of the attractor) of the original system, such as attractor dimensions, Lyapunov exponents and entropies [3] [5]. On the other hand, information theoretic mea-
sures of structural complexity include effective measure complexity (EMC) [6] [7], excess entropy [8], predictive information [9] etc.\(^1\)

However, there is neither a unique way of defining complexity rigorously, nor is it used in a consistent way in the literature. Some authors consider any signal that is not constant or periodic as complex, while most agree that neither strictly periodic nor completely random processes should be deemed complex [3]. For example, traditional entropy-based complexity measures, such as Shannon entropy [10], Kolmogorov-Sinai (KS) entropy [11], approximate entropy (ApEn) [12], and sample entropy (SampEn) [13], are maximized for random sequences - they thus suggest higher structural complexity for randomized surrogate time series than for the original one. This is misleading, especially when the signal comes from more complex systems, with pronounced correlation structures over multiple spatio-temporal scales.

There are many proposed measures of complexity based on different notions of entropy. Historically, correlation dimension/entropy is developed to classify deterministic systems by rates of information generation and was not developed for statistical applications. It is badly compromised by steady, small amounts of noise, and generally requires a prohibitively large number of data points to achieve convergence even for low-dimensional chaotic systems, usually infinite for stochastic and MIX processes. As a result, it is primarily applied by ergodic theorists to well-defined theoretical transformations, with no noise and for an infinite amount of data available.

In contrast, in [12] Pincus introduced approximate entropy as a regularity statistic to distinguish between finite, noisy, possibly stochastic (or composite deterministic), and truly stochastic data sets. It represents the conditional probability that sequences that are close (in the sense of some metric) to one another over \(m\) consecutive data points will still exhibit similarity when one more data point is added. Though approximate entropy was constructed along similar lines as the correlation entropy, it has a different aim: to provide a widely applicable and statistically valid entropy formula. The justification is that if joint probability measures for reconstructed dynamics that describe the two systems

\(^1\)For a detailed overview of complexity measures, see Chapter 2 and references therein.
are different, then their marginal probability distributions for a fixed partition, given by
aforementioned conditional probabilities, are likely to be different too. As a result, using
approximate entropy one needs orders of magnitude less points to accurately estimate
these marginal probabilities as compared to the number of points needed to accurately
reconstruct the attractor measure, as is the case with correlation entropy. Approximate
entropy is thus applicable to noisy, typically short, real-world time series and unlike the
correlation entropy, it can distinguish between correlated stochastic processes.

However, the limitation of ApEn measure is that the algorithm counts each sequence
as matching itself to avoid the occurrence of ln(0) in the calculations. As a result, it
presents a bias in the result and heavily depends on the time series length and lacks
relative consistency [13]. Afterwards, Richman & Moorman [13] proposed a modification
of approximate entropy, known as sample entropy (SampEn). It is based on the definition
of the distance between two vectors in a maximum norm sense, when self-matches are
excluded. As such, it represents an unbiased estimator which is largely independent of the
length of the time series and displays relative consistency over a wide range of operating
conditions.

Another important consideration is that above mentioned standard entropies are
based on a ‘one step difference’ (e.g., $H_{n+1} - H_n$) and hence do not account for features
related to structure and organization over a range of time scales, other than the shortest
one. Costa et al. noticed this discrepancy and argued that the dynamics of a complex
nonlinear system manifests in multiple inherent scales of the observed time series and,
thus, sample entropy estimates calculated on a single scale are not sufficient descriptors.
To that end, they proposed multiscale entropy (MSE) analysis which aimed at quantifying
the interdependence between entropy and scale. This was achieved by first extracting
multiple temporal scales of input data using the so-called coarse graining method and
sample entropy estimates were subsequently calculated for each scale separately [14] [15].
This facilitates the assessment of the dynamical complexity of a system.

Besides, MSE measure of complexity is small for both deterministic (predictable) and uncorrelated random (unpredictable) signals and large for correlated (lin-
ear/nonlinear) stochastic processes, when evaluated over larger scales. This is in agreement with the consensus that the notion of dynamical complexity spans a whole range between the properties of perfect regularity and total randomness.

Moreover, MSE is used along with a concept called ‘complexity loss’ hypothesis [16] which postulates that the complexity of a physiological or behavioral control system degrades with disease and aging. Generally, the output of healthy systems, under certain critical parameter conditions, reveals a type of complex variability associated with long-range (fractal) correlations. This complexity also associated with the ability of living systems to adjust to a changing environment. On the other hand, this multiscale complexity appears to degrade in characteristic ways with aging and disease. The loss of system complexity reflects either a loss or impairment of functional components and/or altered nonlinear coupling between the components.

In MSE, the underlying integrative multiscale functionality of a healthy system is interpreted by non-diminishing entropy values across increasing time scales whereas for pathologic systems, the entropy decreases monotonically. As MSE can quantify this fractal nonlinear breakdown in disease and aging, it has the potential to use for diagnostic and prognostic assessment. As a result, MSE has been successfully applied to analyze the fluctuations of the human heartbeat under pathologic conditions like erratic cardiac arrhythmia and congestive heart failure [14], study EEG and MEG recordings in patients with Alzheimer’s disease [17], and characterize the complexity of human gait dynamics from healthy subjects under different conditions [18], examine variations in EEG complexity in response to photic stimulation during aging [19], study complex dynamics of human red blood cell flickering and alterations with in vivo aging [20] etc. All the reported results strongly support the general ‘complexity-loss’ theory for systems under ‘stress’, for instance, through aging and disease [16].

The existing MSE algorithm has been designed for the analysis of scalar time series, and is not suited for multivariate time series that are routinely measured in experimental and biological systems. Standard MSE treats multivariate time series as a set of individual time series by considering each variable separately, however, this is only applicable if all the
data channels are statistically independent or uncorrelated at the very least (which is often not the case). In addition, there are substantial advantages in simultaneously analyzing several variables observed from the same phenomenon, especially if there is a large degree of uncertainty and coupling underlying the system dynamics or data acquisition. To that cause, the MSE method should be extended to multivariate case, which is the motivation of this work. The proposed extension will not only process simultaneously multiple (possibly heterogeneous) data channels, but also specifically cater for all the available information, both within and across the multiple channels.

1.2 Aim and Objectives

The main aim of this thesis is to develop a multivariate extension of MSE so that it can be applied to real world multivariate signals. Since real world signals are inherently multivariate, a direct multichannel processing of the signal is a prerequisite in the context of complexity analysis. This is because complexity of the data can not be solely attributed to long range temporal (auto-)correlations within each channel; the inter-dependence (or cross-correlation) among multiple channels also critically contributes to the complexity of the signal [3]. Focusing on this aim, several objectives of this thesis are next laid out.

The first objective is to propose a novel definition of multivariate sample entropy based on multivariate embedding. Afterwards, using the notion of multivariate sample entropy, multivariate multiscale entropy (MMSE) is introduced which can deal with the different embedding dimensions, time lags, and amplitude ranges of multiple data channels in a rigorous and unified way. The method is shown to cater for linear and/or nonlinear within- and cross-channel correlations as well as for complex dynamical couplings and various degrees of synchronization over multiple scale.

The second objective is to propose a method to overcome another limitation of the MSE algorithm which is also inherited by the multivariate extension. Namely, the MSE method is not a perfect match for processing real world non-stationary data due its deterministic way of generating scales via coarse graining of input data. Coarse graining, that
is, signal averaging over non-overlapping segments of increasing length, is unsuitable for the analysis of high frequency components and also results in aliasing, causing artifacts. More critically, it reduces the input data length by the scale factor for each successive data scale, thereby, imposing a limit on the length of input data which can be effectively processed via MSE. In [21] it was proposed to use a bank of Butterworth filters to circumvent the aliasing problem, however such an approach requires the a priori selection of filter parameters, making the analysis critically sensitive to slight changes in experimental conditions. It is therefore desirable to make the complexity analysis data-adaptive, so that it can be conducted across scales that occur naturally, as defined by the data.

Empirical mode decomposition (EMD) is a data-driven method that decomposes a given time series into a set of oscillatory functions (time-scales), known as intrinsic mode functions (IMFs). Unlike projection-based methods, such as those based on Fourier and wavelet theory, EMD obtains the oscillatory modes (scales) adaptively and considers the signal dynamics at the ‘local’ level, making it a natural choice for generating the data-scales required for entropy-based analysis [22]. The recently developed multivariate extension of EMD (MEMD) is able to process multichannel data [23] and crucially, aligns the decomposed components from different channels in similar frequency bands, a prerequisite for enabling a scale-by-scale analysis. Recent work [24] has combined the MEMD algorithm with univariate sample entropy estimation, whereby MEMD was used to examine dynamics across multiple trials. As a result, it was limited to single-channel analysis and did not exploit the full potential of MEMD. Besides, for a fair and meaningful comparison between complexity estimates obtained from different multivariate process, a fully multivariate scheme must be used to obtain multiple data-driven scales.

To fulfill the second objective and to develop a complete and robust framework for an entropy-based complexity analysis of multivariate data, we propose to use multivariate EMD in combination with the multivariate sample entropy statistic. This would ensure that the scales generated from each channel are same in number and similar in terms of spectral properties (belonging to same frequency bands), thus, making comparison of complexity estimates among multivariate data physically meaningful. In addition, in
this way the limitation of sufficient input data length due to coarse graining process is alleviated since EMD/MEMD generates temporal data scales of same length as the length of the input signal.

The final objective is to validate the proposed multivariate multiscale complexity analysis framework on different multivariate real world case studies. This will allow us to gain insight into the complexity of the underlying multivariate signal generating system, to design accurate multivariate models, or to use it as a complementary tool for medical diagnosis and prognosis.

1.3 Original Contributions

The specific contributions presented in this thesis can be summarized as follows:

1. Along with sample entropy (SampEn), two popular complexity metrics: permutation entropy (PermEn) and Lempel-Ziv entropy (LZEn) are also placed into the same multiscale framework leading to the development of multiscale PermEn and multiscale LZEn. All these methods are validated on synthetic signals and applied to characterize brain dynamics. Thus, the superiority of SampEn in multiscale framework is also established.

2. Multivariate sample entropy (MSampEn) is introduced and its sensitivity to different parameters is investigated. Afterwards based on MSampEn, the multivariate multiscale entropy (MMSE) is proposed which caters both within- and cross-channel correlations and thus produces meaningful complexity estimates of the underlying multivariate signal generating system.

3. Multivariate empirical mode decomposition (MEMD) is proposed for generating data adaptive scales instead of the existing coarse-graining method in MMSE, thus circumventing some limitations of MMSE. As a result, MEMD-enhanced MMSE yields a robust framework for physically meaningful complexity analysis of multivariate time series.
4. The novel framework is extensively applied to multivariate real world recordings from different application areas.

1.4 Outline of the Thesis

In Chapter 2, a brief overview of the definitional controversies of complexity as well as its different aspects and classification criteria are presented. It ends with the comprehensive survey of different complexity measures.

The third chapter lays down the groundwork for the understanding of the recently introduced multiscale entropy (MSE) method. Two popular established complexity measure, namely, the permutation entropy (PermEn) and Lempel-Ziv entropy (LZEn) are also placed in the same multiscale framework which leads to the development of multiscale PermEn and multiscale LZEn. Both the methods are calibrated with white and $1/f$ noise along with MSE. Afterwards, they are applied for the characterization and classification of EEG epochs recorded from different recording regions and for different brain states.

In Chapter 4, multivariate sample entropy (MSampEn) is introduced based on multivariate embedding. It is also shown that the proposed extension is non-trivial, as it deals with the different embedding dimensions, time lags, and amplitude ranges of data channels in a rigorous and unified way. Then, the parameter sensitivity of MSampEn is investigated, and based on MSampEn, multivariate multiscale entropy (MMSE) is presented. Finally, the MMSE method is validated on both synthetic and real world multivariate processes.

A critical algorithm enhancement of the MMSE method, that is, the use of multivariate empirical mode decomposition (MEMD) for generating data-adaptive scales instead of coarse-graining process is reported in the fifth chapter. This is of considerable importance since the data-adaptive scales make the approach robust to data non-stationarity and also the so produced scales are of same length as the length of input signal, thus, removing the limitations on input data length in the original MMSE method. Validation on synthetic and real world multivariate data are also conducted.

In Chapter 6, both the coarse-graining based MMSE and MEMD-enhanced MMSE
methods are extensively investigated in different application areas. These include brain consciousness analysis for coma and quasi-brain-death (QBD) patients, analysis of climatic variables like wind and air temperature belonging to different environmental conditions and dynamic regimes, detection of signatures caused by increased cognitive load and stress, and analysis of the uterine EMG records from the term and pre-term deliveries to bring new insights into the physiology of parturition.

Finally, in the seventh chapter the conclusions of the thesis are drawn and some directions for future work are outlined.
Chapter 2

Complexity and Its Measures

Complexity can be interpreted as a manifestation of the intricate entwining or inter-connectivity of elements within a system and between a system and its surroundings. Complex adaptive systems (CAS) are comprised of multiple subsystems that exhibit nonlinear deterministic & stochastic characteristics, and are regulated hierarchically. Examples of CAS include stock markets, human heart or brain, weather and climate systems, internet etc. A system’s complexity usually reflected in the dynamical fluctuations of the output generated by the free-running conditions. In this chapter, the definitional controversies for complexity are introduced and signal properties associated with complexity are reviewed. We then introduce some criteria used to classify complexity measures in the literature, and finally some representative complexity measures are described.

2.1 What is Complexity?

The concept of complexity and the study of complex adaptive systems originated from a whole chain of interdisciplinary approaches, from the theory of nonlinear dynamical systems to information theory, statistical mechanics, biology, sociology, ecology, economics and others. Researchers in those areas have been trying to define complexity, nonetheless, no all-encompassing definition has emerged as yet.
Murray Gell-Mann, the winner of 1969 Nobel Prize in physics, traces the meaning of complexity to the root of the word. The English word *complex* is derived from the Latin word *complexus*, meaning ‘entwined’ or ‘braided together’ [25]. Similarly, the Oxford Dictionary defines something as ‘complex’ if it is ‘made of (usually several) closely connected parts’. New England Complex Systems Institute (NECSI ¹) defines complexity as: (i) ‘the (minimal) length of a description of the system’. (ii) ‘the (minimal) amount of time it takes to create the system’. On the other hand, Grassberger [6] positioned complexity somewhere between order and disorder. He considered three patterns similar to those in Figure 2.1 and described that human identifies the pattern in the middle as the most complex since it seems to have more “structure.” On the contrary, the pattern on the right has no rules as it is generated using a random number generator though at the level of each pixel, however, the amount of information required to objectively and fully describe this pattern is greatest. This agrees with the notion that living organism should be more complex than, for example, both perfect crystals and ideal gases.

![Figure 2.1: An illustration of complexity: Left panel displays a completely ordered pattern, middle panel displays a complex pattern and right panel displays a completely random pattern [6].](image)

Besides, the following definitions are taken from the existing literature of complexity science featuring many key figures in this field:

1. ‘To us, complexity means that we have structure with variations.’ [26]

2. ‘Complexity starts when causality breaks down.’ [27]

¹http://www.necsi.edu/guide/concepts/complexity.html
3. ‘The term describe phenomena, structures, aggregates, organisms, or problems that share some common themes: (i) They are inherently complicated or intricate; (ii) they are rarely completely deterministic; (iii) mathematical models of the system are usually complex and involve non-linear, ill-posed, or chaotic behavior; (iv) the systems are predisposed to unexpected outcomes (so-called emergent behavior)’. [28]

4. ‘An ill-defined term that means many things to many people. Complex things are neither random nor regular, but hover somewhere in between. Intuitively, complexity is a measure of how interesting something is.’ [1]

5. ‘Complexity is that property of a model which makes it difficult to formulate its overall behavior in a given language, even when given reasonably complete information about its atomic components and their inter-relations.’ [2]

6. ‘Complexity = integration of observer and trinity of factors (size, variety, rules).’ [29]

7. ‘Complexity can then be characterized by lack of symmetry or “symmetry breaking”, by the fact that no part or aspect of a complex entity can provide sufficient information to actually or statistically predict the properties of the others parts.’ [30]

From the above definitions, it can be concluded that:

- There is still no agreed-upon definition, much less a theoretically rigorous formalization of complexity.

- Complexity is contextual and domain-specific. It is pretty much in the eye of the beholder.

- None of the definitions or approaches in the complexity literature are mutually exclusive.

- Neither perfect disorder nor perfect order are complex, rather complexity lies between order and disorder, or, using a recently fashionable expression, “on the edge of chaos”.
• Complexity is a comparative/relative measure rather than an absolute measure, as we want to be able to say frequently “X is more complex than Y which is, in turn, more complex than Z”.

2.2 Manifestation of Complexity

There are numerous attributes or characteristics associated with complexity. The following attributes are neither exhaustive list nor are they individually necessary and/or sufficient conditions for complexity:

Nonlinearity. Nonlinearity means that the superposition principle does not hold. Though often considered to be essential for complexity, it is not a necessary condition for complexity as there are complex systems studied in game theory and quantum dynamics which are in the domain of linear dynamics [31].

Feedback. Feedback is an important necessary condition for complexity. It ensures adaptability and self-organization in complex systems. For example, Malthusian population dynamics rely on positive feedback for explosive growth and negative feedback, due to resource shortages, for population decline [32]. However, the existence of feedback in a system is not sufficient for complexity as it requires the individuals to be part of a large enough group for exhibiting complexity [31].

Spontaneous order. This arises from the combined effect of a very large number of uncoordinated interactions between elements. Notions related to order include symmetry, organization, periodicity, determinism and pattern. Admitting the fact that complexity lies in between total order and absolute randomness, it is a necessary condition for a complex system that it exhibits some kind of spontaneous order. For example, spontaneous order can be easily seen in the flocking behavior of birds or in swarm behavior of shoals of fishes [33].

No central control. There is no single centralized control in a complex system. Control is distributed throughout the system and local decisions are made by parts or modules (elements) within overall system. As the control is distributed/self-organized, the system
is robust and stable under perturbations. For example, the order observed in the swarming behavior of birds (fishes) are stable and robust despite the individual and erratic motions of its members or the perturbations in the system by the wind (ocean current) or the random elimination of some of the members of the swarm. This is due to the lack of central control in the system. However, it is not sufficient for complexity as non-complex systems may have no control or order at all [31].

**Emergence.** It is associated with the properties of ‘wholes’ (or more complex situations) that cannot be defined through the properties of the ‘parts’ (or simpler situations) as concisely summarized by Aristotle in the Metaphysics: “The whole is more than the sum of its parts”. It is some kind of global property which arises from the aggregate behavior of individuals, something that could not have been predicted from understanding each particular individual. It is in contrast to the reductionist approach. Emergent behavior seems to be ubiquitous in nature. For example, in the brain, consciousness is an emergent phenomenon, which comes from the interaction of the brain cells. However, it is not sufficient, for example, an ideal gas exhibits emergent order but is not a complex system [33] [34].

**Hierarchical organization.** Complexity arises when many levels of organization forms a hierarchy of system and sub-system. The complex system exhibits a variety of levels of structure and various kinds of symmetry, order and periodic behavior through interaction within and across levels. Examples include ecosystem, brain, cosmos etc. [35].

**Numerosity.** System should consists of a large number of parts and the individual parts should be engaged in many linear/non-linear interactions to constitute complexity. All of the above properties emerges only if the system consists of a large number of parts [34].

**Self-similarity.** A special kind of scale invariance in which the parts closely resemble the whole. If the parts are identical re-scaled replica, the fractal is exact. If they show the same statistical properties at different scales, the fractal is statistical. Mathematical fractals such as the Cantor set or the Mandelbrot tree are exact whereas most natural fractal objects like Pial arterial tree running along the brain cortex of the cat or coastline are statistical as shown in Figure 2.2 [36]. The self-similarity can also be distinguished as spatial self-similarity and temporal self-similarity. Most of the complex fluctuations in
2.2 Manifestation of Complexity

respiration, heart rate variability, human gait etc. possesses temporal self-similarity and their self-similarity is statistical. However, self-similarity need to be differentiated from

![Mandelbrot tree and Pial arterial network](image)

**Figure 2.2:** Spatial fractals: Left panel displays Mandelbrot tree, an exact fractal and right panel displays Pial arterial network, a statistical fractal [36].

self-affinity. Self-similar objects are isotropic as the scaling is identical in all directions whereas for self-affine objects scaling is anisotropic. For example, physiological signals are self-affine temporal structures because the units of their amplitude is not time as shown in Figure 2.3 [36].

![Fractal tree and heart rate fluctuation](image)

**Figure 2.3:** Fractals: Left panel displays a tree-like fractal which is self-similar and right panel displays heart rate fluctuation process which is self-affine [37].

**Long-range correlation.** A hallmark of physiologic systems is their extraordinary complexity resulting from the long-range correlations extending exponentially over many spatio-temporal scales. For example, the inter beat interval (IBI) time series from healthy subjects shows long-range correlations extending to thousands of heartbeats. This multi-scale complexity appears to degrade in characteristic ways with aging and disease, reducing the adaptive capacity of the individual. As a result, this fractal breakdown and associ-
2.2 Manifestation of Complexity

Manifestation of Complexity loss can be quantified and have potential applications to diagnostic and prognostic assessment [16].

**Far from equilibrium.** This phenomenon illustrates how systems that are forced to explore their space of possibilities will create different structures and new patterns of relationships. This notion is also related to physiologic complexity. Under healthy conditions, living systems which are open dissipative systems (as there are mass, energy, entropy, and information fluxes across their boundaries) operates far from equilibrium. On the contrary, maximum equilibrium is reached when a living system approaches death. For example, the healthy heart rate dynamics exhibits irregular fluctuations far from equilibrium. Regularity in the heartbeat interval is a sign of disease indicating insensitivity and inflexibility and thus inability to adapt to the changing environment [36].

![Figure 2.4: Lorenz attractor](image)

**Sensitive dependence on initial conditions.** Complexity also arises due to the sensitive dependence on initial conditions. This means that small differences in the initial state of a system can lead to dramatic differences in its long-term dynamics (popularly known as the ‘butterfly effect’). Edward Lorenz was among the first to study this dependency by simulating the long term evolution of weather using a simplified version of the Navier-Stokes equations. By slightly varying initial values of temperature, pressures and other parameters, he found solutions showing new type of behavior patterns as shown in Figure 2.4. The consequences of this discovery were profound as limited precision in
measuring the initial state precludes accurate predictions of future states of such systems. Long-term prediction and control of complex systems are therefore believed to be very hard or impossible [33].

**State of paradox.** Complexity arises from dynamics combining both order and chaos which supports the idea of bounded instability or the edge of chaos characterized by a state of paradox: stability and instability, competition and cooperation, order and disorder [33].

**Co-evolution.** With co-evolution, elements in a system can change not only based on their interactions with one another but also with the environment. Additionally, patterns of behavior can change over time. For example, evolution of species occurs with respect to their environment as well as their relationship to other species. Stuart Kauffman described co-evolution with the concept of fitness landscapes [33]. The fitness landscape is an $n$-dimensional function made of many maxima/minima. Each of them corresponds to a potential of fitness/unfitness. The higher a maximum/minimum is, the greater the fitness/unfitness it represents. Figure 2.5, shows a three dimensional fitness landscape. The evolution of a system can be thought of as a voyage across the fitness landscape with the aim of discovering the global maximum. The system can get stuck on the first maximum it approaches if the strategy is incremental improvement. If the system changes its strategy, other interconnected systems will respond and thus changes the shape of the fitness landscape dynamically.

![Figure 2.5: Three dimensional fitness landscape](image-url)
2.3 Classification of Complexity Measures

As with the definitions, there are again several approaches regarding classification of complexity measures. Some of the approaches are described below:

**From the perspective contrasting complexity and randomness.** In this perspective, measures of complexity are grouped into two types: Type 1 measures monotonically increase with increasing disorder or randomness in the system whereas Type 2 measures behave as globally convex function of randomness and exhibit highest complexity for systems with intermediate order or regularity. This is depicted in Figure 2.6. Example of Type 1 measures include Algorithmic complexity [38], various generalized entropies, approximate entropy (ApEn) [12], sample entropy (SampEn) [13] etc. On the other hand, effective measure complexity (EMC) [6], $\epsilon$-machine complexity [39], fluctuation complexity [40], multiscale entropy (MSE) [14] etc. belongs to the Type 2 measures.

![Figure 2.6: Complexity versus randomness](image)

**From the perspective contrasting determinism and stochasticity.** In this view, complexity measures are classified either as deterministic or statistical. All deterministic measures are essentially measures of randomness and increases monotonically with the
increase of randomness whereas all statistical measures are convex function of randomness.

**From the perspective of dynamical systems.** In this view, complexity measures are classified in the context of dynamical systems so that the measures can distinguish between order and chaos. The four-fold scheme is based on the dichotomous notions of structure and dynamics as well as homogeneous partitions and generating partitions. Structural aspects are related with the appearance of state probabilities (with respect to the partition) in the definition of the measure while dynamical measures contain transition probabilities in addition [41].

**From philosophical perspective.** This scheme is epistemologically inspired and assigns ontic and epistemic levels of description to deterministic and statistical measures respectively. It turns out that this scheme distinguishes measures of complexity precisely in the same manner as Type 1 and Type 2. In other words, ontic measures are actually Type 1 measures whereas epistemic measures are Type 2 measures [42].

**From the perspective of the way statistics is implemented.** Complexity measures can be classified according to the way statistics is implemented in each of these measures. In this perspective, convex measures, in contrast to monotonic measures, are meta-statistically formalized. In other words, they use second-order statistics in the sense of ‘statistics of statistics’. For example, fluctuation complexity is the standard deviation (second-order) of a net mean information flow (first-order), effective measure complexity (EMC) is the convergence rate (second-order) of a difference of entropies (first-order) etc. [42].

**From the perspective of computability.** In this context, complexity measures can be classified as computable or non-computable. Example of computable measures are statistical complexity, effective measure complexity (EMC) etc. On the other hand algorithmic complexity, effective complexity [43] etc. are non-computable.

**From the perspective of difficulty.** In this scheme, 42 complexity measures are grouped according to the answer of three questions that are frequently posed when attempting to quantify complexity of something under study [44]:
• How hard is it to describe?

• How hard is it to create?

• What is its degree of organization?

In the category of ‘difficulty of description’, there are algorithmic complexity, generalized entropies etc. In the category of ‘difficulty of creation’, there are computational complexity, logical depth etc., whereas effective complexity, effective measure complexity (EMC) etc. belong to the category of ‘degree of organization’.

2.4 Representative Complexity Measures

As there are a great number of complexity measures in the literature and still new measures are enriching this list, it is almost impossible to review all these measures. In developing complexity measures, there have been contributions from a diverse set of fields including Thermodynamics, Information Theory, Statistical Mechanics, Control Theory, Applied Mathematics, Operations Research etc. In the following, we categorize complexity measures according to those broad field from which they are originated and describe some representative measures.

2.4.1 Information-theoretic Complexity Measures

Most of information-theoretic complexity measures are based on either Shannon entropy or its derivatives. Claude Shannon was the first to show the connection between thermodynamical entropy and information-theoretic entropy [10]. Though it was derived as a measure of randomness or uncertainty in a system, it has been used as a measure of complexity by subsequent authors. This type of measure requires data to be first categorized into binary or another fixed number of bins (‘symbols’). Complexity measures are then based on the probability \( p_x \) of observing symbol \( x \) in the data, as well as on second order probabilities such as the frequency of observing symbol \( x \) next to symbol \( y \) or of observing particular sequences of symbols of length \( L \) in the series (known as \( L \)-words).
2.4 Representative Complexity Measures

The measures can be of Type 1 or Type 2 as defined previously in Figure 2.6.

**Shannon Entropy**

Shannon entropy [10] is a measure of the average amount of information in a transmitted message or the difficulty of guessing a message passing through a channel, given the range of possible messages. Let \( X \) be a discrete random variable on a finite set \( X = \{x_1, x_2, \ldots, x_n\} \), with probabilities \( p(x_i) = Pr(X = x_i) \). Then the entropy \( H(X) \) of \( X \) is defined as \( H(X) = - \sum_{i=1}^{n} p(x_i) \log p(x_i) \). The logarithm is usually taken to the base 2, in which case the entropy is measured in ‘bits’, or to the base e, in which case \( H(X) \) is measured in ‘nats’. In the case of transmitted messages, these probabilities are the probabilities that a particular message was actually transmitted and the term \( - \log p(x_i) \) is defined as the information content of the message \( x_i \). For the case of equal probabilities (i.e. each message is equally probable), the Shannon entropy is highest.

**Kolmogorov-Sinai Entropy**

This is related to measures of partitions in dynamical systems. Suppose a phase space is divided into \( D \)-dimensional hypercubes of content \( \varepsilon^D \). Let \( P_{i_0,i_1,\ldots,i_n} \) be the probability that a trajectory is in hypercube \( i_0 \) at \( t = 0 \), \( i_1 \) at \( t = T \), \( i_2 \) at \( t = 2T \), etc. and define

\[
K_n = h_K = - \sum_{i_0,i_1,\ldots,i_n} P_{i_0,i_1,\ldots,i_n} \ln P_{i_0,i_1,\ldots,i_n} \tag{2.1}
\]

where \( K_{n+1} - K_n \) is the information needed to predict which hypercube the trajectory will be in at \((n+1)T\) given trajectories up to \( nT \). The Kolmogorov-Sinai (KS) entropy is then defined by

\[
k = \lim_{T \to 0} \lim_{\varepsilon \to 0^+} \lim_{n \to \infty} \frac{1}{nT} \sum_{n=0}^{n-1} (K_{n+1} - K_n) \tag{2.2}
\]

The KS-entropy can be interpreted as highest average amount of information that the system can produce per step relative to a coding and is thus related to the Shannon entropy. A positive KS-entropy is often linked to chaos. Moreover, the sum of all the positive Lyapunov exponents gives an estimate of the Kolmogorov-Sinai entropy according
2.4 Representative Complexity Measures

to Pesin’s theorem [45].

**Correlation Entropy**

Correlation entropy/K2 entropy is a lower bound of KS entropy. Historically, it is developed to classify deterministic systems by rates of information generation and was not developed for statistical applications.

Grassberger & Procaccia [46] first defined the correlation sum for a collection of delay embedding vectors $X_n$ of dimension $m$ as the fraction of all possible pairs of vectors which are closer than a given distance $\epsilon$ in the Euclidean norm sense. The basic formula is

$$C(m, \epsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta(\epsilon - \|X_i - X_j\|)$$

(2.3)

where $\Theta$ is a Heaviside step function. In the limit, for an infinite amount of data ($N \to \infty$) and for small $\epsilon$, $C(m, \epsilon)$ scales like a power law, that is $C(\epsilon) \propto \epsilon^{D_2}$, where the correlation dimension $D_2$ is defined by

$$d(N, \epsilon) = \frac{\partial \ln C(\epsilon, N)}{\partial \ln \epsilon}$$

$$D_2 = \lim_{\epsilon \to 0} \lim_{N \to \infty} d(N, \epsilon)$$

(2.4)

when the embedding dimension $m$ exceeds the box-dimension of the attractor. Since we do not know the box-dimension a priori, we need to check for convergence of the estimated values of $D_2$ in $m$.

While correlation dimension characterizes the dependence of $\epsilon$ on the correlation sum inside the scaling range, the correlation entropy characterizes the $m$-dependence. The number of $\epsilon$-neighbors of an $m$-dimensional delay vector is an estimate of the local probability density, and in fact it is a kind of joint probability - all $m$-components of the neighbor have to be similar to those of the actual vector simultaneously. Thus, when increasing $m$, joint probabilities covering larger time spans get involved. The scaling of these joint probabilities is related to the correlation entropy, $h_2$, such that $C(m, \epsilon) \propto$
$\epsilon^2 e^{-m^2}$, where the correlation entropy is calculated as:

\[
h_2(m, \epsilon) = \ln \frac{C(m, \epsilon)}{C(m + 1, \epsilon)}
\]  

(2.5)

As for the scaling in $\epsilon$, also the dependence on $m$ is valid only asymptotically for large $m$, which cannot be attained due to the lack of data points. So, a so called scaling range of $\epsilon$ values is sought for which $h_2$ is nearly constant (convergent) for large $m$ (which is extrapolated from different $h_2$ versus $m$ plot). As correlation entropy is defined by some limiting procedure, we are in trouble if we have a finite sample instead of a full distribution: $N$ is limited by the sample size, and the range of meaningful choices for $\epsilon$ is limited from below by the finite accuracy of the data and by the inevitable lack of near neighbors at small length scales [3].

**Approximate Entropy**

Pincus [12] introduced approximate entropy (ApEn) as a regularity statistic with fixed $m$ and $\epsilon$ to distinguish between finite, noisy, possibly stochastic, or composite deterministic and stochastic data sets. It represents the conditional probability that sequences that are close to each other for $m$ consecutive data points will still be close to each other when one more data point is known. Though approximate entropy was constructed along similar lines to the correlation entropy, it had a different focus: to provide a widely applicable, statistically valid entropy formula. The motivation behind approximate entropy was that, if joint probability measures for reconstructed dynamics that describe each of two systems are different, then their marginal probability distributions on a fixed partition, given by aforementioned conditional probabilities are probably different too. Typically, one need orders of magnitude fewer points to accurately estimate these marginal probabilities as compared to the number of points needed to accurately reconstruct the “attractor” measure defining the process. As a result, approximate entropy is applicable to noisy, typically short, real-world time series and, unlike the correlation entropy, it can distinguish between different correlated stochastic processes.
Sample Entropy

Sample entropy (SampEn) is a modification of the Approximate Entropy introduced by Richman and Moorman [13]. It reduces the bias of ApEn by excluding self-matches, i.e., vectors are not compared to themselves. Moreover, it is largely independent of time series length and displays relative consistency over a wide range of operating conditions.

Multiscale Entropy

Costa et al. [14] proposed multiscale entropy (MSE) which calculates SampEn across multiple time scales. As multiscale entropy is based on SampEn, it measures the degree of randomness (or inversely, the degree of orderliness) of a time series. But unlike KS entropy or sample entropy, which are based on a ‘one step difference’ (e.g., $H_{n+1} - H_n$) and hence do not account for features related to structure and organization of scales other than the shortest one, multiscale entropy analysis focuses on quantifying the interrelationship of entropy and scale thus associating complexity with the ability of living systems to adjust to an ever-changing environment which requires integrative multiscale functionality. The scales in MSE method are generated using the so called ‘coarse-graining’ method.

Fourier Entropy

To calculate Fourier entropy, first the power spectral density (PSD) of a time series is computed and then normalized to produce a probability-like distribution. Afterwards, the Shannon entropy of this normalized PSD is calculated which is known as spectral entropy or Fourier entropy. The Fourier entropy measures the extent to which the power spectrum of the signal is concentrated (or not) into a given (narrow) frequency range. Low entropy values correspond to narrow-band (mono-frequency) activity characterizing highly ordered (regularized) signal and thus lower complexity. On the contrary, high entropy values reflect a wide-band (multi-frequency) activity in the signal and correspond to higher complexity.
2.4 Representative Complexity Measures

Wavelet Entropy

This complexity measure is related to the number of wavelet components needed to fit a signal [47]. In contrast to spectral entropy, it is capable of detecting changes in a nonstationary signal due to the localization characteristics of the wavelet.

Effective Measure Complexity

Grassberger [6] introduced Effective Measure Complexity (EMC) of a pattern as the asymptotic behavior of the amount of information that must be stored in order to make an optimal prediction about the next symbol to the level of granularity. EMC can be seen as the difficulty of predicting the future values of a stationary series, as measured by the size of regular expression of the required model. It can capture structure in sequences that range over multiple scales. For a random string, EMC is zero while for structured or correlated string, it increases linearly with the correlation.

GeoEntropy

GeoEntropy [48] is based on the theory of regionalized variables. The concept of regionalized variables is central to geo-statistics which concerns with spatially distributed properties. Such a variable is usually a characteristic of certain phenomenon, as metal grades, for example, are characteristics of mineralization. The degree of spatial variability of a regionalized variable is usually expressed by a semi-variogram. Based on the semi-variogram, GeoEntropy yields an analytical procedure for estimating the parameter $m$ and $r$ that are used to compute SampEn/MSE.

Correntropy

Correntropy [49] is a nonlinear similarity measure between two random variables. It generalizes the autocorrelation function to nonlinear spaces. If $\{x_t, t \in T\}$ is a strict stationary stochastic process with an index set $T$, then the correntropy function is defined as $V(s,t) = E\{<\phi(x_s),\phi(x_t)>\}$, where $\phi$ is a nonlinear mapping from the input space
2.4 Representative Complexity Measures

2.4.1 Kernel Trick

It makes use of the ‘kernel trick’ to define the inner product of the nonlinear mappings as a positive-definite Mercer kernel: $K(x_s, x_t) = \langle \phi(x_s), \phi(x_t) \rangle$. Though it is not a direct complexity measure, it is used to detect nonlinearity in the signal and thus measures complexity.

2.4.2 Algorithmic Complexity Measures

Apart from the various information-theoretic measures of complexity, some measures are algorithmic in nature and are applied on strings or other data structures. For these measures, complexity relates to a generative process or description rather than to an actually existing physical system or dynamical process.

Algorithmic Information Complexity (AIC)

This measure was developed independently by Solomonoff, Kolmogorov, and Chaitin and also known as Kolmogorov-Solomonoff-Chaitin complexity. Though it is the most influential among algorithmic complexity measures, it is not computable. It is defined as the minimum possible length of a description in some language. For a string of symbols, it is the length of the shortest program to produce the string as an output. Highly regular, periodic or monotonic strings may be computed by programs that are short because the description can be greatly compressed without loss of meaning and as a result, they are less complex. On the other hand, the complexity of a truly random string is highest since it cannot be compressed at all and the shortest program that outputs that string will have to contain the string itself. However, incompressible strings (those whose programs are not shorter than themselves) are indistinguishable from random strings [50].

Logical Depth

Logical depth [51] is defined as the minimum necessary effort or computational cost (time and memory) required for creating and running a shortest program that can reproduce a given object. Thus it is a combination of both storage and computational power. A random string is incompressible and hence the shortest program that can reproduce it
is a simple copying program. Consequently, both random strings and very simple ones have a low logical depth as the computation time for both is very small. On the other hand, strings which have structure and regularity, in addition to some randomness, are logically deep as they require longer time to compute. Similar to AIC, this measure is not computable.

**Topological Complexity**

Crutchfield and Young extended the concept of algorithmic information complexity by defining complexity as the minimal size of a model representation of a system that can statistically reproduce the observed data within a specified tolerance. This definition takes into account both the minimal size and the fixed hierarchy or structural rules of a system. One disadvantage of this definition is that it could not provide a unique measure of complexity for a system because there may not be necessarily a minimal model for the system or users may construct different models of the same system [29].

**Computational Complexity**

It is the asymptotic difficulty in computing the output relative to the size of the input in terms of computational resources (typically time or storage space), given the specification of the problem. The degree of complexity of a problem depends on the level of detail used to describe it. As a result, it is a measure of the effort required to obtain an agreed amount of details of information [2].

**Arithmetic Complexity**

Arithmetic complexity is defined as the minimum number of arithmetic operations needed to complete a task. Though it is not a general concept of complexity, it is practically important for making computational algorithms more efficient [2].
2.4 Representative Complexity Measures

Lempel-Ziv (LZ) Complexity

It is a major compression scheme in Unix systems which approximates the algorithmic complexity of a given string. It asymptotically approaches Shannon entropy. This scheme is also the basis for compression of TIFF images and ZIP files, where it is referred to as LZW (Lempel-Ziv-Welch) compression [31].

Effective Complexity

Effective complexity was introduced by Gell-Mann [43] as a statistical measure of complexity based on Kolmogorov complexity. It is defined as the length of a concise description of the regularities (as contrasted to the degree of randomness) of a system or bit string. It is a statistical measure of complexity as it is the shortest description, not of the entity itself, but of the ensemble in which the entity is embedded as a typical member. It is a Type 2 measure and also not computable.

2.4.3 Chaos-theoretic Complexity Measures

These type of measures are based on the basic properties of chaotic system like the aperiodic and highly erratic behavior of trajectories of chaotic dynamical systems or their sensitive dependence on initial conditions.

Lyapunov Exponent

It characterizes the rate of divergence (or convergence) of two neighboring trajectories in the phase space and thus measures the instability (or stability) of generic trajectories against infinitesimal perturbations. Though it is not a direct complexity measure, it can be interpreted as a measure of the rate at which the system generates new information. Besides, the maximal Lyapunov exponent (the largest of the spectrum of Lyapunov exponents), is related to a notion of predictability for a dynamical system and thus can measure complexity [3].
2.4 Representative Complexity Measures

**Attractor Dimension**

Various attractor dimensions like information dimension, box-counting dimension, generalized dimension etc. can also be used as complexity measures. The higher the dimension of the attractor, the more complex the dynamical system associated with the attractor is.

**Permutation Entropy**

This complexity measure has aspects of both dynamical systems and information theory measures because complexity is estimated as the entropy of the distribution of permutations of groups of time samples. At first, each member of the group is given a sequence number from 1 to n. Then the n members of each group are placed in ascending order and the new order of the sequence numbers for each group is noted. The new order serves as a bin number into which the total matches of that order among all of the groups is stored. The result is a histogram of number of occurrences of each sequence order which is further normalized to a probability distribution and finally the entropy of that probability distribution is calculated. Permutation entropy (PermEn), thus, measures the local order structure of the time series in phase space. The use of ordinal statistics (rank) makes PermEn robust to noise embedded in phase space. It is an extremely fast and robust measure and used for measuring complexity for real-world time series [52].

2.4.4 Random Fractal-theoretic Complexity Measures

Complexity measures based on random fractal theory differ from those based on chaos theory in the sense that in chaos theory complexity is assumed to be generated by nonlinear deterministic interactions of the system components with only a few degrees of freedom, where noise or intrinsic randomness does not play an important role. On the contrary, random fractal theory assumes that the dynamics of the system are inherently random. These measures also relate complexity with the long-range correlation and scaling behavior of the system. Complexity measures from this field includes Hurst parameter, spectral index calculated from power spectral density, $\alpha$-parameter of detrended fluctuation analysis (DFA) method etc [36].
Chapter 3

Multiscale Entropy Analysis

Traditional complexity measures typically operate at a single scale and thus fail to quantify inherent long-range correlations in real-world data, a key feature of complex systems. This has led to the introduction of multiscale entropy (MSE) method in which multiple scales of input data are first extracted using so-called ‘coarse graining’ and sample entropy estimates are subsequently calculated for each scale separately. The recently introduced multiscale entropy (MSE) method thus has the ability to detect fractal correlations and has been used successfully to assess the complexity of univariate data. In this chapter, the MSE method is described briefly and its applicability is illustrated for electroencephalogram (EEG) signal characterization and classification. Moreover, two other complexity metrics: permutation entropy (PermEn) and Lempel-Ziv entropy (LZEn) are also placed into the same multiscale framework leading to the development of multiscale PermEn and multiscale LZEn. All the three multiscale entropies along with their single-scale counterparts are employed in the context of discriminating EEG epochs recorded from different recording regions and for different brain states. Experimental results demonstrate that the multiscale entropies outperformed their single-scale counterparts, whilst multiscale method with sample entropy as the complexity estimator performed best among the cases considered.
3.1 Introduction

Most real world signals exhibit complex dynamical behavior and a number of measures have been proposed to characterize the underlying signal generating mechanisms. The measures addressed in this thesis are based on complexity, a phenomenon which encompasses local predictability, dimensionality, regularity or irregularity, randomness, self-similarity, synchrony, etc. [3]. Time delay embedded reconstruction is particularly popular, as it allows for the estimation of invariant quantities (in terms of smooth transformations in state space) of the original system, such as attractor dimensions, Lyapunov exponents and various entropy measures\(^1\) [3].

The notion of entropy is commonly used to define signal complexity by effectively measuring the amount of structure in a time series by assessing its degree of regularity/irregularity [12] [13]. Moreover, traditional entropy measures, such as Shannon entropy [10], Kolmogorov-Sinai (KS) entropy [11], approximate entropy (ApEn) [54], and sample entropy (SampEn) [13], are maximized for completely random processes (Type 1 measure), and are used to quantify the regularity of univariate time series on a single scale, by e.g. evaluating repetitive patterns [14]. As a result, using entropy to measure complexity of physiological data, which exhibit high degree of structural richness, yields lower entropy than for their randomized surrogates, formed by shuffling the original data samples and thus destroying any structure present. This is counterintuitive for a measure of complexity, and the greater entropy of the uncorrelated surrogate series also highlights a lack of a straightforward correspondence between regularity and complexity. Neither completely predictable (e.g. periodic) nor completely unpredictable (e.g. uncorrelated random) signals are truly complex, since at a global level none is structurally rich. Instead, time series observed from dynamical physical and physiological systems generally exhibit long-range correlations at multiple spatial and temporal scales.

The multiscale entropy (MSE) method proposed by Costa et al. [14] is based on evaluating at multiple time scales the sample entropy, a refinement of approximate entropy

\(^1\)A practical implementation of most of the methods can be found in the free software package TISEAN [53], publicly available at http://www.mpipks-dresden.mpg.de/~tisean/.
which measures the degree of randomness (or inversely, the degree of orderliness) of a time series. Historically, correlation entropy was developed to distinguish between deterministic systems by rates of information generation and was not developed for applications on stochastic data. In contrast, in [12] Pincus introduced approximate entropy as a regularity statistic to distinguish between finite, noisy, possibly stochastic (or composite deterministic), and truly stochastic data sets. It represents the conditional probability that sequences that are close (in the sense of some metric) to one another over $m$ consecutive data points will still exhibit similarity when one more data point is added.

There exist several improvements of MSE, especially regarding the definition of the time scales [21] [22] [24], contributing to its theoretical foundations. A detailed analysis of MSE for correlated and uncorrelated noises with Gaussian and inverse Gaussian distributions can be found in [15] and [55]. The method has been successfully applied across the field of biomedical research, such as in fluctuations of the human heartbeat under pathologic conditions [14], EEG and MEG in patients with Alzheimer’s disease [17], complexity of human gait under different walking conditions [18], variations in EEG complexity related to aging [19], and human red blood cell flickering [20]. These results strongly support the general ‘complexity-loss’ theory for systems under ‘stress’, for instance, through aging and disease [16].

The aim of this chapter is to briefly describe the MSE method and apply it to characterize brain electrical activity. To that end, the process of estimating sample entropy is first reviewed. Moreover, as we proposed to further include permutation entropy (PermEn) and Lempel-Ziv entropy (LZEn) into the same multiscale framework, the estimation process of these measures are also described in this chapter. The so introduced multiscale PermEn and multiscale LZEn along with the original multiscale SampEn/MSE is then calibrated synthetically with $1/f$ noise and white noise. Afterwards, all three multiscale entropies along with their single-scale counterparts are employed in the context of discriminating EEG epochs recorded from different recording regions and for different brain states. Finally, the superiority of multiscale entropy in terms of classification accuracy with sample entropy as the complexity estimator is established.
3.2 Quantitative Complexity Estimators

In this section, a set of different entropy measures that will be applied in multiscale framework is described along with their estimation process. They are sample entropy (SampEn), permutation entropy (PermEn) and Lempel-Ziv entropy (LZEn). All the estimators are suitable for analysis of time series of limited length.

3.2.1 Sample Entropy

Pincus [12] introduced a family of statistics, named approximate entropy (ApEn) to quantify the regularity of typically short and noisy time series. As ApEn algorithm counts each sequence as matching itself to avoid the occurrence of ln(0) in the calculations, it introduces a bias in the result which causes ApEn to heavily depend on the time series length. Therefore, ApEn estimates lack relative consistency [13]. To overcome these limitations of ApEn, Richman & Moorman [13] introduced the sample entropy (SampEn) which represents the conditional probability that two sequences of $m$ consecutive data points, which are similar to each other within a tolerance level $r$ will remain similar when the next consecutive point is included, provided that self-matches are not considered in calculating the probability. It is largely independent of time series length and exhibits relative consistency over a wide range of operating conditions.

![Figure 3.1: An illustration showing sample entropy estimation procedure where the pattern length, $m$ is 2. Adapted from reference [56].](image)
Algorithm 1 The Sample Entropy (SampEn)

1. Form $N - m$ vectors $X_m(1), X_m(2), \ldots, X_m(N - m)$ defined by $X_m(i) = [x(i), x(i + 1), \ldots, x(i + m - 1)]$ where $i = 1, 2, \ldots, N - m$. These vectors represent $m$ consecutive $x$ values, commencing with the $i$-th point.
2. Define the distance between $X_m(i)$ and $X_m(j)$ as the maximum norm, $d[X_m(i), X_m(j)] = \max_{k=1, \ldots, m}\{ |x(i+k-1) - x(j+k-1)| \}$
3. For a given $X_m(i)$, count the number of $j$ ($1 \leq j \leq N - m, j \neq i$), denoted as $B_i$, such that $d[X_m(i), X_m(j)] \leq r$. Then, for $1 \leq i \leq N - m$, $B_i^m(r) = \frac{1}{N-m-1}B_i$
4. Define $B^m(r)$ as $B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r)$
5. Then increase the dimension of the vectors to $m + 1$ and calculate $A_i$ as the number of $X_{m+1}(i)$ within $r$ of $X_{m+1}(j)$, where $j$ ranges from 1 to $N - m (j \neq i)$.
6. Define $A_i^m(r) = \frac{1}{N-m-1}A_i$ and $A^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r)$
7. Thus, $B^m(r)$ is the probability that two sequences will match for $m$ points, whereas $A^m(r)$ is the probability that two sequences will match for $m + 1$ points.
8. Then SampEn is estimated by,

$$SampEn(r, m, N) = -\ln\left[ \frac{A^m(r)}{B^m(r)} \right]. \quad (3.1)$$

where $r$ is the tolerance level, $m$ is the pattern length and $N$ is the length of the time series.

For a time series $\{x(n)\} = x(1), x(2), \ldots, x(N)$, SampEn is outlined in Algorithm 1. Comparisons between series can only be done with fixed values of $r$, $m$ and $N$. $r$ is usually set as some percentage of the standard deviation of the original data sequence which gives SampEn scale invariance in that it remain unchanged under uniform process magnification, reduction, or constant shift to higher or lower values [13]. Various theoretical and clinical applications have shown that $m = 1$ or 2, and $r = 0.1 - 0.25$ of the standard deviation of the original data sequence provides good statistical validity for SampEn. Besides, it has been suggested that signal lengths of $10^m$ to $20^m$ points should be sufficient for reliable estimation of SampEn [13].

A simulated time series $u[1], u[2], \ldots, u[n]$ is shown in Figure 3.1 to illustrate the estimation process of sample entropy (SampEn). In this case, the pattern length, $m$, is 2, and the tolerance level, $r$, is 20. Dotted horizontal lines around data points $u[1]$, $u[2]$ and $u[3]$ represent $u[1] \pm r$, $u[2] \pm r$, and $u[3] \pm r$, respectively. Two data values are indistinguishable, if the absolute difference between them is $\leq r$. All green, red and blue points represent data points that are similar to the data point $u[1]$, $u[2]$, and
If we consider the 2-component green-red template sequence $(u[1], u[2])$ and the 3-component green-red-blue $(u[1], u[2], u[3])$ template sequence, then there are only two green-red sequences, $(u[13], u[14])$ and $(u[43], u[44])$, that are similar to the template sequence $(u[1], u[2])$ and only one green-red-blue sequence that is indistinguishable from the template sequence $(u[1], u[2], u[3])$ for the segment shown (provided the template sequences are not matched with themselves). As a result, the number of counts similar to the 2-component template sequences is two whereas the number of counts similar to the 3-component template sequence is one. These calculations are repeated for the next 2-component and 3-component template sequence, which are, $(u[2], u[3])$ and $(u[2], u[3], u[4])$, respectively and the number of counts in these case is added to the previous values. This procedure is then repeated for all other possible template sequences to find the total number of 2-component template matches and the total number of 3-component template matches. Thus, SampEn is the negative natural logarithm of this ratio and reflects the conditional probability that sequences that match each other for the first two data points will also match for the next point.

### 3.2.2 Permutation Entropy

Bandt and Pompe [52] took an ordinal approach to quantify complexity of time series. They introduced permutation entropy (PermEn) that measures the local order structure of the time series in phase space. The use of ordinal statistics (rank) makes PermEn less sensitive to noise embedded in phase space. And as the method is extremely fast and robust, it can be calculated for arbitrary length real-world time series.

For a time series $\{x(n)\} = x(1), x(2), \ldots, x(N)$, PermEn is computed as outlined in Algorithm 2. Theoretically, the normalized entropy is maximal when there is an equal distribution of order patterns. The selection of the embedding dimension $m$ is crucial for calculating PermEn and for practical purposes, $m = 3, \ldots, 7$ is recommended.

A segment of a simulated time series is shown as dotted line in the bottom panel of Figure 3.2. Suppose, the pattern length, $m = 3$ and lag, $\tau = 1$ in this case. First, the series is divided into 3-component patterns or motifs. Some examples are shown in grey and a
Algorithm 2 The Permutation Entropy (PermEn)

1. Form vectors $X_n = [x(n), x(n+1), \ldots, x(n+m-1)]$ where $n = 1, 2, \ldots, N - m + 1$ with the embedding dimension $m$ and lag $\tau = 1$.
2. Arrange any vector $X_n$ in increasing order such that $x(n) \leq x(n+1) \leq \ldots \leq x(n+m-1)$.
3. For $m$ different numbers, there will be $m!$ possible order patterns $\pi$, which are also called permutations. Find its frequency in the time series and denote it as $f(\pi)$.
4. Then the relative frequency is $p(\pi) = f(\pi)/(N - m + 1)$.
5. Thus PermEn is estimated by, $PermEn(m) = - \sum_{m=1}^{m!} p(\pi) \ln p(\pi)$.
6. And the normalized permutation entropy is

$$PermEn = PermEn(m)/\ln(m!)$$ (3.2)

Figure 3.2: An illustration showing permutation entropy estimation procedure where the pattern length, $m$ is 3. Adapted from reference [57].
motif when $\tau = 2$ is also shown as a paler dashed-line in the bottom panel of Figure 3.2. For 3-component sequence, there are six possible order patterns as shown in the middle panel of Figure 3.2. As the patterns are sequentially visited through the signal, they are identified as one of the six possible motif types. A histogram of the relative numbers of each motif in the signal is then computed as shown in the top panel of Figure 3.2 and normalized afterwards to get the PermEn.

### 3.2.3 Lempel-Ziv Entropy

Another approach to quantify complexity of time series is to use the theory of symbolic dynamics. To compute the Lempel-Ziv complexity [58], the original time series is first mapped onto a finite symbol sequence. One popular approach is to convert the time series into a binary (0 : 1) sequence by comparing with a threshold $X_{th}$, i.e., whenever the original time series samples are larger than $X_{th}$, it maps onto 1, otherwise to 0. Usually the median of the series is used as the threshold $X_{th}$ due to its robustness to outliers. The resulting symbol sequence $\{S_i\}$ with $i = 1, \ldots, N$ is then passed through the Lempel-Ziv parsing [58] algorithm to estimate the size $c(\{S_i\})$ of its vocabulary. In this algorithm, the symbol sequence $\{S_i\}$ is scanned from left to right and its complexity counter $c(\{S_i\})$ is increased by one unit every time a new subsequence of consecutive symbols is encountered and the scanning proceeds regarding the following symbol as the starting of the next symbol sequence. To obtain a complexity measure that is independent of the sequence length, $c(\{S_i\})$ should be normalized by the expected asymptotic value for a random sequence of symbols of length $N$ which is $N/\log_\alpha N$, where $\alpha$ is the number of symbols (for binary sequence it is 2). Thus, the normalized Lempel-Ziv complexity or Lempel-Ziv entropy is,

$$LZEn = \frac{\log_\alpha N \times c(\{S_i\})}{N}$$

This reflects the rate of new pattern occurrences or new information generated which is very much in the spirit of the KS entropy.
3.3 Multiscale Entropy Framework

Costa et al. [14] proposed multiscale entropy (MSE) as a meaningful physiologic complexity measure which evaluates the relative complexity of normalized time series across multiple scales. Briefly, the MSE methodology has two steps:

1. Multiple coarse-grained time series are generated from the original time series \( \{x_1, x_2, \ldots, x_N\} \) by averaging the data points within non-overlapping windows of increasing length \( \epsilon \), also known as scale factor. The elements of the coarse-grained time series of scale factor \( \epsilon \) are calculated as the equation below:

\[
y_j^\epsilon = \frac{1}{\epsilon} \sum_{i=(j-1)\epsilon+1}^{j\epsilon} x_i \quad \text{where}, \, 1 \leq j \leq \frac{N}{\epsilon} \tag{3.4}
\]

The length of each coarse-grained series is \( \epsilon \) times shorter than the length of the original series. For scale 1, the coarse-grained time series is simply the original one. An illustration of the coarse-graining process for scale 2 and 3 is shown in Figure 3.3.

2. \( \text{SampEn} \) is calculated\(^2\) for each coarse-grained time series, and then plotted as a function of the scale factor to yield the MSE complexity curves.

To compare the relative complexity of normalized time series, the MSE curves are interpreted as described in [15]: (i) a time series is considered more complex than another if for the majority of the scales its entropy values are higher than others; (ii) a monotonic decrease of the entropy values with scales represents that the signal only contains information in the smallest scale.

3.4 Calibration with 1/f Noise and White Noise

Using MSE analysis, it was shown in [15] that for random white noise (uncorrelated), the entropy values monotonically decreases whereas for 1/f noise (long-range correlated), it

\(^2\)In the original formulation of MSE, \( \text{SampEn} \) is used as the complexity estimator. In this chapter, \( \text{PermEn} \) and \( \text{LZEn} \) are also used and compared to establish the superiority of the \( \text{SampEn} \) in the multiscale framework.
remains constant over multiple scales. This indicates that $1/f$ noise is structurally more complex than uncorrelated signals. This result has been shown using $1/f^\beta$ noises with $0 \leq \beta \leq 1$, that is, for noises with power-law correlations in Figure 3.4. Of related note, for random white noise ($\beta = 0$), entropy monotonically decreases. As the spectral exponent increases the entropy value remains high even for large time scales. For $1/f$ noise ($\beta = 1$), it remains constant and higher than white noise for majority of scales. Figure 3.5 further illustrates this behavior of MSE curves for random white noise and $1/f$ noise.

As PermEn and LZEn are also used as a complexity estimator in multiscale framework and thus multiscale PermEn and multiscale LZEn measure are proposed in this chapter, so they are also calibrated with random noise and $1/f$ noise. Figure 3.6 and Figure 3.7 illustrate the behavior of the complexity curves for multiscale PermEn and multiscale LZEn estimator respectively. In case of multiscale PermEn curves, for both white noise and $1/f$ noise the complexity curves monotonically decreases with scales and the trend of decrease is linear. For multiscale LZEn, the complexity curve for white noise is almost constant for all the scales whereas complexity curve for $1/f$ noise is slightly increases with the increasing scales. Moreover, in both cases, white noise shows higher
3.4 Calibration with $1/f$ Noise and White Noise

Figure 3.4: Plot of MSE analysis for $1/f^\beta$ noises with $0 \leq \beta \leq 1$, that is, for noises with power-law correlations. The Fourier filtering method has been used to generate time series of 30,000 points. In this plot, each point represents the average of 50 independent realizations. The value of SampEn is given according to the color panel. Parameters of SampEn calculation are $m = 2$ and $r = 0.15$.

In general, entropy measures the degree of irregularity of a signal that cannot be completely characterized by the standard deviation (SD) or correlation measures, individually or in combination [59]. MSE measures complexity of a signal by measuring sample entropy that takes into account both its SD and its correlation properties. Specifically, SampEn estimation requires a threshold parameter, $r$ which is defined as a percentage of the SD of the normalized time series. On the other hand, the coarse-graining process used in MSE for probing system dynamics on different scales, is basically similar to averaging and decimation of the original sequences. As a result, SD will decrease at each increase of scale according to the statistical properties of the original time series. Besides, in the MSE...
Figure 3.5: MSE analysis for random white ($\beta = 0$) and $1/f$ noise ($\beta = 1$), each with 30,000 data points. The curves represent an average of 20 independent realizations and error bars the standard deviation (SD). Parameters of SampEn calculation are $m = 2$ and $r = 0.15$.

approach the same $r$ value is used for different scales and thus the scale-related changes in SD for $1/f$ noise are much smaller than for white noise [60]. All these contribute to the constant MSE curve for $1/f$ noise and monotonically decreasing MSE curve for white noise. As there is no such parameter $r$ in PermEn and LZEn estimation, the scale-related changes in SD is not reflected in their corresponding MSE curves and both the measures failed to demonstrate the similar results like SampEn in multiscale approach for synthetic white and $1/f$ noises.

In the next section, these measures are further applied to real-world physiological signal to test whether they are better than multiscale SampEn in terms of signal characterization and classification.

### 3.5 Application to Characterize Brain Dynamics

The brain is an extremely complex biological structure consisting billions of interconnected neurons. The electroencephalography (EEG) records the collective spontaneous electrical
activity of a large population of radially oriented pyramidal neurons of the cortex subsequently filtered through the skull and scalp. The filtered brain electrical signals are nonstationary, non-Gaussian and non-linear in nature, also have characteristic frequency ranges, spatial distributions, and are associated with different states of brain functions. As a result, EEG has found a widespread use in diagnostic application in many neurological diseases such as epilepsy, encephalopathies, tumors, stroke, and sleep disorders, in addition, for monitoring the depth of anesthesia in surgical operation, for brain function monitoring in intensive care units and in neuroscience, cognitive science, cognitive psychology, and psycho-physiological research [61].

To understand the neurophysiological mechanisms underlying normal and disturbed higher brain functions, traditionally linear analysis methods such as Fourier transforms and spectral analysis are used to characterize EEG. Although these methods can identify the rhythmic oscillations in the EEG signal that fall primarily within five frequency bands: delta(< 4Hz), theta(4-8Hz), alpha(8-14Hz), beta(14-30Hz) and gamma(> 30Hz) and the results from these linear methods are quite easy to interpret in physiological terms, they
Figure 3.7: Multiscale LZEn analysis for random white and $1/f$ noise, each with 30,000 data points. The curves represent an average of 20 independent realizations and error bars the standard deviation (SD).

are unable to yield information of the brain’s inherent nonlinear complex dynamics such as the dynamical behavior of individual neurons which exhibits nonlinear phenomena such as threshold and saturation [62]. Moreover, the initial assumption that EEG signals are generated by nonlinear deterministic processes with nonlinear coupling interactions between neuronal populations [17] also gave further impetus and as a result, nonlinear dynamics measures were increasingly used to analyze EEG signal in order to better characterize and understand brain functions [5].

The literature related to nonlinear dynamics analysis is very extensive and each nonlinear analysis measure is generally defined for quantifying different characteristics of time series data such as dimensionality, randomness, predictability, self similarity, synchrony, etc. Babloyantz [63] et el. first analyzed the EEG data of the human sleep cycle using correlation dimension and suggested the existence of chaotic attractors for sleep stages two and four. In another paper [64], they reported that the correlation dimension was substantially lower in absence seizure than that in normal waking EEG. In agreement with this finding, Iasemidis [65] et el. reported that the largest Lyapunov exponent decreased
3.5 Application to Characterize Brain Dynamics

during an epileptic seizure. These early attractor based methods were based on the phase space reconstruction procedure according to Taken’s theorem and using these methods the early researchers demonstrated low-dimensional deterministic chaos in various types of EEG signals. It created a strong controversy due to lack of sufficient statistical analysis and the method of ‘surrogate data testing’ was introduced later on to verify the previous results [66]. As a result, the initial claims for ‘chaos’ in brain dynamics were rigorously re-examined and rejected in subsequent studies [67]. Afterwards, nonlinear EEG analysis changed its focus and concentrated in two directions: “(i) the detection, characterization and modeling of nonlinear dynamics rather than strict deterministic chaos and (ii) the development of new nonlinear measures that are more suitable to be applied to noisy, non-stationary and high-dimensional EEG data” [61].

Recent advancements in nonlinear brain dynamics are extremely useful to provide more insight into the underlying important dynamical properties of the physiological and pathological phenomena. Andrzejak et al. [62] used measures, such as nonlinear prediction error and effective correlation dimension in combination with the method of iterative amplitude adjusted surrogate data to demonstrate strongest indications of nonlinear dynamics for seizure activity. Kannathal et al. [68] used chaotic measures like correlation dimension, largest Lyapunov exponent, Hurst exponent and Kolmogorov entropy to characterize EEG signals from normal, epileptic and alcoholic subjects and showed that the dynamical behavior is less random for alcoholic and epileptic compared to normal. They also had evaluated those nonlinear parameters for EEG signals under different mental states and had found that the measures are significantly lower when subjected to music and reflexological stimulus compared to the normal state [69]. Gautama et al. [5] introduced delay vector variance method to distinguish among EEG segments recorded in healthy subjects, in epilepsy patients during seizure-free interval and during epileptic seizure thus indicating different dynamical properties of brain electrical activity. In [68], different entropy estimators like spectral entropy, Renyi’s entropy, Kolmogorov entropy and approximate entropy were compared to show that the entropy measures are lesser for epileptic EEG as compared to normal EEG and further classified using an adaptive neuro-fuzzy classifier to achieve a classification accuracy of more than 90%. Ouyang et
al. [70] used a determinism measure based on recurrence plot to analyze the non-linear deterministic structure in EEG data of different absence seizure states. They showed that the determinism of EEG epochs gradually increases from seizure-free to seizure intervals. In another report, they used ordinal statistics based forbidden ordinal patterns of the EEG series of genetic absence epilepsy rats from Strasbourg to demonstrate evidence of deterministic dynamics during epileptic states [71]. A new scheme has been proposed recently by Ocak [72] based on approximate entropy and discrete wavelet transform to demonstrate that epileptic EEG is more predictable or less complex and has significant nonlinearity whereas normal and interictal EEG is similar to a Gaussian linear stochastic process. Kumar [73] et al. characterized normal, interictal and ictal EEGs using wavelet entropy, sample entropy and spectral entropy and then classified using two neural network models, namely, recurrent Elman network and radial basis network. Chua et al. [74] used different higher order spectra parameters, namely, bicoherence patterns and bispectrum entropies for classifying normal, pre-ictal and epileptic EEG signals.

It is pointed out in [61] that the neural networks in the brain possess a structure which is intermediate between complete randomness (for example, gas) and perfect order (for example, crystal) and long-range correlations are present in brain dynamics. So, it is intuitive to apply MSE method to characterize brain electrical activity. The major aim of the study presented here is to apply MSE (and the two other variations proposed) to characterize EEG signals recorded extracranially in healthy subjects with eyes open and closed, and intracranially in epilepsy patients both during seizure-free intervals and epileptic seizures and thus compare the classification of different brain states.

3.5.1 Data

The EEG data used in this study are obtained from the EEG database available publicly from the University of Bonn and described in detail in [62]. In brief, the complete data set was comprised of five sets denoted A-E, each consisting of 100 single-channel EEG epochs of duration 23.6 seconds each recorded with a sampling rate of 173.61 Hz. The first two sets A and B were obtained from EEG recorded extracranially using a standardized electrode
placement scheme from five healthy volunteers with eyes open and closed respectively. The last three sets were obtained from EEG recorded intracranially using depth electrodes from five epileptic patients undergoing presurgical evaluations. Epochs in sets C and D were taken during seizure free intervals respectively from the hippocampal formation of the opposite hemisphere of the brain and within the epileptogenic zone. The epochs of set E were selected from all recording sites exhibiting ictal activity and thus contain seizure activity. As a pre-processing step, the data had been filtered using a band-pass filter with settings 0.53-40 Hz (12 dB/octave).

### 3.5.2 Statistical Analysis

To evaluate the statistical difference of the calculated entropy statistics of different sets, Student’s t-test as well as the Mann-Whitney U test (also known as Wilcoxon rank sum test) has been applied. For more than two groups, one-way ANOVA (analysis of variance) which is a generalization of the Student’s t-test for more than two groups is used as a parametric test whereas Kruskal-Wallis test, an extension of the Mann-Whitney U test for three or more groups is used as nonparametric one. A multiple comparison procedure followed the ANOVA and Kruskal-Wallis test to find the information about which pairs of means/medians are significantly different, and which are not. In all cases, differences are considered statistically significant and the null hypothesis is rejected if the p value is lower than 0.01.

### 3.5.3 Simulation Results

Unless otherwise specified, the values of the parameters used to calculate SampEn were $N = 4097$, $m = 2$, and $r = 0.15$ which were chosen on the basis of various previous studies indicating good statistical reproducibility [13]. For MSE, as the length of each coarse-grained sequence is $\epsilon$ (scale factor) times shorter than the length of the original series, so the highest scale factor calculated for analysis was 20. In that case the coarse-grained sequence had more than 200 data points which was well within the SampEn signal length requirement ($10^m$ to $20^m$) [13]. For PermEn to satisfy the condition of $m! < 200$
at the highest scale factor, the value of \( m = 5 \) was chosen. SampEn, PermEn and LZEn values were computed for the entire 500 EEG epochs in the five data sets described earlier. Table 3.1 shows the different entropy statistic for the different EEG sets.

Table 3.1: Summary of different entropy statistic measured in different sets

<table>
<thead>
<tr>
<th>Entropy statistic</th>
<th>Set A</th>
<th>Set B</th>
<th>Set C</th>
<th>Set D</th>
<th>Set E</th>
</tr>
</thead>
<tbody>
<tr>
<td>SampEn</td>
<td>Mean 1.0591</td>
<td>0.9348</td>
<td>0.7084</td>
<td>0.6528</td>
<td>0.6102</td>
</tr>
<tr>
<td></td>
<td>Median 1.0106</td>
<td>0.9415</td>
<td>0.6981</td>
<td>0.6703</td>
<td>0.6179</td>
</tr>
<tr>
<td></td>
<td>Max 1.4785</td>
<td>1.2725</td>
<td>1.0514</td>
<td>0.8880</td>
<td>1.0153</td>
</tr>
<tr>
<td></td>
<td>Min 0.8274</td>
<td>0.7192</td>
<td>0.5003</td>
<td>0.2010</td>
<td>0.3123</td>
</tr>
<tr>
<td></td>
<td>STD 0.1409</td>
<td>0.1320</td>
<td>0.1142</td>
<td>0.1418</td>
<td>0.1281</td>
</tr>
<tr>
<td>PermEn</td>
<td>Mean 0.7619</td>
<td>0.7049</td>
<td>0.7038</td>
<td>0.6727</td>
<td>0.6449</td>
</tr>
<tr>
<td></td>
<td>Median 0.7558</td>
<td>0.7099</td>
<td>0.7089</td>
<td>0.6655</td>
<td>0.6469</td>
</tr>
<tr>
<td></td>
<td>Max 0.8509</td>
<td>0.8131</td>
<td>0.7993</td>
<td>0.7773</td>
<td>0.8268</td>
</tr>
<tr>
<td></td>
<td>Min 0.6955</td>
<td>0.6243</td>
<td>0.5572</td>
<td>0.5331</td>
<td>0.5216</td>
</tr>
<tr>
<td></td>
<td>STD 0.0314</td>
<td>0.0471</td>
<td>0.0505</td>
<td>0.0514</td>
<td>0.0536</td>
</tr>
<tr>
<td>LZEn</td>
<td>Mean 0.5414</td>
<td>0.5327</td>
<td>0.3480</td>
<td>0.3347</td>
<td>0.3877</td>
</tr>
<tr>
<td></td>
<td>Median 0.5287</td>
<td>0.5404</td>
<td>0.3456</td>
<td>0.3427</td>
<td>0.3837</td>
</tr>
<tr>
<td></td>
<td>Max 0.6942</td>
<td>0.6356</td>
<td>0.5477</td>
<td>0.4657</td>
<td>0.5800</td>
</tr>
<tr>
<td></td>
<td>Min 0.4335</td>
<td>0.4218</td>
<td>0.2080</td>
<td>0.1347</td>
<td>0.2343</td>
</tr>
<tr>
<td></td>
<td>STD 0.0575</td>
<td>0.0525</td>
<td>0.0589</td>
<td>0.0720</td>
<td>0.0773</td>
</tr>
</tbody>
</table>

From the table, it is clear that epileptic EEG (set E) is the least complex among the five data sets as it yields the least value of SampEn and PermEn statistic. To find the statistical significance of the differences among the different entropy values of the five sets, one-way ANOVA and Kruskal-Wallis test were used. Afterwards, multiple comparison procedure was applied on the result of the previous two tests to find the significant differences among the different pairs. The result is shown in Table 3.2 where the entries represent the pairs which cannot be significantly differentiated according to the above two statistical tests.

Table 3.2: Result of the two statistical tests

<table>
<thead>
<tr>
<th>Entropy statistic</th>
<th>One-way ANOVA test</th>
<th>Kruskal-Wallis test</th>
</tr>
</thead>
<tbody>
<tr>
<td>SampEn</td>
<td>C-D, D-E</td>
<td>A-B, C-D, D-E</td>
</tr>
<tr>
<td>PermEn</td>
<td>B-C</td>
<td>B-C, D-E</td>
</tr>
<tr>
<td>LZEn</td>
<td>A-B, C-D</td>
<td>A-B, C-D, C-E, D-E</td>
</tr>
</tbody>
</table>

Next, the multiscale procedure was applied with three entropy estimates, and the
result is shown in Figure 3.8. For the MSE analysis using SampEn (Figure 3.8(a)), the MSE curves of sets A and B have similar patterns: a local maximum at scale factor 5 followed by decreasing entropy values. On the other hand, the MSE curves of sets C and D have a similar shape: steep increase followed by a smoother increase whereas MSE curve of set E increases until scale factor 8 and then decreases slightly. And for all scale factors, the curve for set E remains quite below from the others which prove the fact that EEG complexity decreases in seizures. From the statistical point of view, for most of the scale factors the SampEn values are significantly different among the sets for at least four sets. It was found by using one-way ANOVA and Kruskal-Wallis test followed by multiple comparison procedure. However, the LZEn measures (Figure 3.8(b)) cannot differentiate in terms of complexity curves between sets A (healthy volunteer, eyes open) and B (healthy volunteer, eyes closed), and sets D (epileptogenic zone) and C (hippocampal formation of opposite hemisphere) and the PermEn measures (Figure 3.8(c)) cannot discriminate any sets after scale factor 10.

3.5.4 Classification

In this section, the ability and effectiveness of the entropy measures studied above are examined in the context of EEG classification. Every EEG epoch in the five different sets mentioned earlier is assigned a label of the set from which it is taken and the classifier aims to correctly label each of the 500 EEG epochs. Two cases are considered here: the five-class case and the simplified three-class case, which groups classes A and B, and classes C and D. The classification was performed on the feature vectors generated from the EEG epochs using SampEn, PermEn and LZEn individually and then with the multi-scale framework where the vector contained the entropy values for all the scale factors considered. Though there exists a great deal of classification methods, nearest neighbor classification (NNC) and leave-one-out classification (LOOC) from the supervised classification (as the desired labels are known a priori) paradigms were chosen for this investigation. These two methods can be interpreted as a lower and upper bound of possible classification systems. In NNC, the class prototypes are determined as the average of the feature vectors of all EEG epochs
Figure 3.8: Multiscale entropy analysis using different entropy estimators. Symbols represent the mean values of entropy for each set and the bars represent the standard error.
belonging to a certain set. The EEG epoch is classified to the class of its nearest neighbor, with the epoch being assigned to the class label of the prototype which is nearest in Euclidean norm to its feature vectors. In LOOC, the label of an EEG epoch (testing epoch) is determined by leaving out that epoch from the set and considering the remaining set of feature vectors as the set of labeled prototypes (training epochs). The label of the testing epoch is set equal to that of the nearest neighbor prototype in Euclidean norm sense. The classification accuracy is expressed as the fraction of correct classifications and is shown in Table 3.3 for single scale entropy metrics and in Table 3.4 for multiple scale framework for the different case setup. The classification performances for the MSE method with SampEn as the complexity measure outperform all the other entropy measures in single scale for 5-class case setup, whereas for multiple scales it outperforms for both case setups.

Table 3.3: The classification accuracy for the different single scale entropy measures

<table>
<thead>
<tr>
<th></th>
<th>NNC</th>
<th>LOOC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SampEn</td>
<td>LZEn</td>
</tr>
<tr>
<td>3-class case</td>
<td>65.4</td>
<td>72</td>
</tr>
<tr>
<td>5-class case</td>
<td>41.2</td>
<td>40.20</td>
</tr>
</tbody>
</table>

Table 3.4: The classification accuracy for the different entropy measures in the multiscale framework

<table>
<thead>
<tr>
<th></th>
<th>NNC</th>
<th>LOOC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SampEn</td>
<td>LZEn</td>
</tr>
<tr>
<td>3-class case</td>
<td>84.8</td>
<td>77.60</td>
</tr>
<tr>
<td>5-class case</td>
<td>67.2</td>
<td>49.80</td>
</tr>
</tbody>
</table>

3.5.5 Discussion

In this study, the recently introduced multiscale entropy (MSE) method with sample entropy as the complexity statistic is compared with other entropy estimators like permutation entropy (PermEn) and Lempel-Ziv (LZEn) in context of EEG signal characterization and classification. All of the entropy estimators provide a quantitative metric for the complexity measurement of different brain state. The complexity of EEG recordings of healthy volunteers with eyes open (set A) are found significantly ($p < 0.01$) higher than
the recordings with eyes closed (set B) using SampEn (in t-test only) and PermEn (both t-test and Mann-Whitney U test) measures while LZEn measure could not significantly differentiate between these two sets in any test. Moreover, MSE method with sample entropy as the complexity estimator was able to detect the high degree of complexity in set A, compared to set B, for most of the scale factors.

These results are in agreement with Hebb’s concept of cell assemblies in brain function. The brain electrical activity (i.e., EEG) is generated from individual cell assembly activities. In the “eyes closed” state, as there are no cognitive tasks involved, the number of independent, parallel functional processes active in the brain is less and brain goes into a passive state of relaxation. As a result, the neuronal networks display a state of synchrony and most of the neuronal groups within a cortex area oscillate at a certain frequency. This is associated with an increase in alpha frequencies in the brain waves making EEG signal structure more regular and thus reducing its complexity. On the other hand, the task of processing large amount of visual information in the “eyes open” state require more oscillating cell assemblies to change their oscillation patterns away from the preceding predominant one. This makes the EEG signal more irregular and thus increases complexity [19, 61].

All entropy measures were able to find significant (t-test and Mann-Whitney U test, \( p < 0.01 \)) decrease of complexity in seizure activity (set E) compared to normal EEG (eyes open or closed) recordings. This indicates a reduction in the intra-cortical information flow and lower neuronal process in the brain. The result is in agreement with the previous studies on dimensional analysis of EEG that epileptic seizures are emergent states with reduced dimensionality compared to normal state. It was observed in [75] that neuronal hyper-synchrony, a phenomenon during which the number of independent variables required to describe the dynamical system is smaller than other times, underlies seizures. This reduction of the system’s degree of freedom indicates either a strongly coupled system or the inactivation of previously active networks or a loss of dynamical brain responsiveness to the environmental conditions. Besides, the findings of our study support the more general concept of multiscale complexity loss with aging and disease.
which also reduces the adaptive capacity of biological organization at all levels [16].

Moreover, the classification results show that the MSE method with sample entropy as complexity statistic provides a sufficiently detailed characterization of the EEG to discriminate among the different sets. The performance of a classifier is expected to lie between 84% and 90% in the 3-class case. However, performance degrades for a more detailed classification in the 5-class case which further dissociates between sets A and B and sets C and D. Among the single scale entropy measures, LZEn performs better (72%-73.8%) in the 3-class case.

3.6 Conclusion

To conclude, the recently introduced multiscale entropy analysis (MSE) method has been examined against two other entropy statistics to reveal the hidden characteristics of the EEG signals. Compared with other traditional single scale entropy metrics, our study suggests that MSE with sample entropy as the complexity estimator can detect the long-range correlation and multiscale complexity in the brain and thus efficiently distinguish between different dynamical properties of brain electrical activity. This method has also been shown to have the potential to be used to extract features for more sophisticated EEG signal classification and hence characterize different brain conditions.
Chapter 4

Multivariate Multiscale Entropy Analysis

MULTIVARIATE physical and biological recordings are becoming increasingly common and their simultaneous analysis is a prerequisite for the understanding of the complexity of underlying signal generating mechanisms. To that end, in this chapter we generalize the recently introduced univariate multiscale entropy (MSE) analysis to the multivariate case. This is achieved by introducing multivariate sample entropy (MSampEn) in a rigorous way, in order to account for both within- and cross-channel dependencies in multiple data channels, and by evaluating it over multiple temporal scales. The so introduced multivariate MSE (MMSE) method is shown to provide an assessment of the underlying dynamical richness of multichannel observations, and more degrees of freedom in the analysis than standard MSE. The benefits of the proposed approach are illustrated by simulations on complexity analysis of multivariate stochastic processes and on real world multichannel signals.

4.1 Introduction

Recent developments in sensor technology have enabled routine recording of multivariate time series from both physical and biological systems, yet when assessing their complexity
the standard MSE analysis can only consider every data channel separately. This is only appropriate if the components of a multivariate signal are statistically independent or at the very least uncorrelated (which is usually not the case). For instance, the embedding theorem [4] establishes that the dynamics of a multivariate system can be reconstructed from sufficiently long time delay embedded vectors of a single time series (seen as a one-dimensional projection of the multivariate system trajectory). However, in practice this is not reliable for systems exhibiting more than a few degrees of freedom. Indeed, based on measurements of the $z$-coordinate of the Lorenz system we cannot reconstruct its dynamics, as embedding based on the $z$-coordinate does not resolve the $x$-$y$ symmetry [76] (Figure 4.1). In addition, there are substantial advantages in simultaneously analyzing several variables observed from the same phenomenon, especially if there is a large degree of uncertainty and coupling underlying the system dynamics or data acquisition.

The multivariate extension of MSE in this work is based on our novel definition of multivariate sample entropy (MSampEn). The proposed multivariate MSE (MMSE) evaluates MSampEn over different time scales and deals with the different embedding dimensions, time lags, and amplitude ranges of data channels in a rigorous and unified way. The method is shown to cater for linear and/or nonlinear within- and cross-channel correlations as well as for complex dynamical couplings and various degrees of synchronization over multiple scales, thus allowing for direct analysis of multichannel data. The advantages of the proposed multivariate approach, in contrast to analyzing each data channel separately, are illustrated for both synthetic stochastic processes and real world multivariate data.

4.2 Multivariate Embedding

As said earlier, multivariate sample entropy (MSampEn) is a prerequisite for performing multivariate multiscale entropy (MMSE) analysis simultaneously over a number of data channels. However, such approaches are still missing in the open literature. To that end, we will first propose MSampEn which is based on the notion of multivariate embedding.
Figure 4.1: The Lorenz attractor is created in phase space using the $x$, $y$ and $z$ coordinate, as shown in Top left panel. In Top right, it is created following the embedding theorem using only $x$ component. The bottom left panel shows the attractor created using only $y$ component and bottom right panel shows the same created using only $z$ component. Note that, $z$-component fails to reconstruct the attractor dynamics as it cannot resolve the $x-y$ symmetry.

Recall from multivariate embedding theory [76] that the multivariate embedded reconstruction for a $p$-variate time series $\{x_{k,i}\}_{i=1}^N$, $k = 1, 2, \ldots, p$ generated from the same system and observed through $p$ measurement functions $h_k(y_i)$, is based on the composite delay vector

$$X_m(i) = [x_{1,i}, x_{1,i+\tau_1}, \ldots, x_{1,i+(m_1-1)\tau_1}, x_{2,i}, x_{2,i+\tau_2},$$
$$\ldots, x_{2,i+(m_2-1)\tau_2}, \ldots, x_{p,i}, x_{p,i+\tau_p}, \ldots, x_{p,i+(m_p-1)\tau_p}],$$

where $M = [m_1, m_2, \ldots, m_p] \in \mathbb{R}^p$ is the embedding vector, $\tau = [\tau_1, \tau_2, \ldots, \tau_p]$ the time lag vector, and $X_m(i) \in \mathbb{R}^m$ ($m = \sum_{k=1}^p m_k$).

For illustration, consider a wind signal recorded using a 2D anemometer, where
4.3 Multivariate Sample Entropy

the wind speed time series from the east-west and north-south directions are denoted respectively by \(x(t) = \{x_1, x_2, \ldots, x_N\}\) and \(y(t) = \{y_1, y_2, \ldots, y_N\}\), with each time series of \(N\) data points. For the time lag vector \(\tau = [2, 1]\) and the embedding vector \(M = [2, 2]\), some of the composite delay vectors are \([x_1, x_3, y_1, y_2]\), \([x_2, x_4, y_2, y_3]\), \([x_3, x_5, y_3, y_4]\), etc. Next section evaluates this issue further and provides a geometric interpretation.

4.3 Multivariate Sample Entropy

Richman & Moorman [13] introduced sample entropy (SampEn) as a conditional probability that two sequences of \(m\) consecutive data points, which are similar to within a tolerance level \(r\), will remain similar when the next data point is included, provided that self-matches are not considered in calculating the probability. When extending univariate sample entropy to the multivariate case, we need to consider the following issues.

First, it is important to notice that multivariate data do not necessarily have the same amplitude range among the data channels, so that the distances calculated on embedded vectors may be biased towards the variates with largest amplitude ranges. To that end, we propose to scale all the data channels to the same amplitude range, and choose the range \([0, 1]\) as a preferred choice. As shown later, other choices will not affect the multivariate sample entropy calculation.

Secondly, for a fixed embedding dimension \(m\), sample entropy calculates the average number of delay vector pairs that are within a fixed threshold \(r\), and repeats the same for dimension \((m+1)\). There are \(p\) ways in which we can evolve from the space of dimension \(m\) described by the embedding vector \([m_1, m_2, \ldots, m_k, \ldots, m_p]\) to any space of dimension \((m+1)\) described by the embedding vector \([m_1, m_2, \ldots, m_k+1, \ldots, m_p]\) where \(k = 1, \ldots, p\) is the index of data channel. It is important to notice that the average number of delay vector pairs that are within a fixed threshold for dimension \((m+1)\) can be calculated in two ways. A naive approach would be, if for each of the \(k\)-th subspaces of the \((m+1)\)-dimensional space, we calculate the average number of delay vector pairs that are within a fixed threshold, and then average over all the \(p\) subspaces. A rigorous approach would be
to take into account all the delay vectors in all the subspaces and then compare the delay vectors both within and across the $p$ subspaces. This way, along with time correlations, we have means to cater for linear and/or nonlinear spatial correlations. This is crucial in situations where the measurements come from similar physical quantities, simultaneously recorded at different positions in a spatially extended system, like in the case of geophysical sensors or scalp EEG. For heterogeneous data channels (i.e. heart beat interval series and respiratory signals), this makes it possible to cater for the dynamical couplings and various degrees of synchronization over multiple scales.

Thirdly, as sample entropy is a relative measure and the threshold parameter is set as some percentage of the standard deviation of the observed time series, we also need a multivariate generalization of the univariate notion of variance. Though the covariance matrix, $S$, is one such generalization, we still need a single number to measure the multivariate scatter in the data. Two such common measures are the generalized variance, $|S|$, and the total variation, $tr(S)$; in this work we use total variation. To maintain the same total variation for all the multivariate series under consideration, we normalize each data channel to unit variance so that the total variation becomes equal to the number of channels/variables. This way, differences in the variance among the multivariate time series under consideration do not influence the calculation of multivariate sample entropy.

Finally, in univariate sample entropy methods, the time lag, $\tau$, is not used as a parameter ($\tau = 1$ is assumed), thus assuming that the phase space representation of a time series is independent of the value of the time lag $\tau$. However, this is only the case for an infinite amount of data. In digital signal processing, both embedding dimension, $m$, and time lag, $\tau$, ought to be considered for determining the optimal tap input length of an adaptive filter or a time-delay neural network. For instance, if the temporal span ($m \times \tau$) is too small, the signal variation within the delay vector is largely governed by noise and either $m$ or $\tau$ should be increased. There is no established criterion for choosing which of the two parameters to modify, and it is common to have a fixed time lag $\tau$ (sampling rate) and to adjust the embedding dimension $m$ (length of a filter) accordingly. In the multivariate case, different observed variables are likely to have different embedding
dimensions and time lags, and we need to use different \( m_k \) and \( \tau_k \) values for different channels/variables. In our approach, both \( M \) and \( \tau \) are varying (vector parameter) and can be optimized either separately or simultaneously [77].

For a \( p \)-variate time series \( \{x_{k,i}\}_{i=1}^{N} \), \( k = 1, 2, \ldots, p \), the MSampEn is outlined in Algorithm 3.

**Algorithm 3 The Multivariate Sample Entropy (MSampEn)**

1: Form \((N - \delta)\) composite delay vectors \( X_m(i) \in \mathbb{R}^m \), where \( i = 1, 2, \ldots, N - \delta \) and \( \delta = \max\{M\} \times \max\{\tau\} \).

2: Define the distance between any two composite delay vectors \( X_m(i) \) and \( X_m(j) \) as the maximum norm [78], that is, \( d[X_m(i), X_m(j)] = \max_{l=1, \ldots, m}\{|x(i+l-1) - x(j+l-1)|\} \).

3: For a given composite delay vector \( X_m(i) \) and a threshold \( r \), count the number of instances \( P_i \) for which \( d[X_m(i), X_m(j)] \leq r \), \( j \neq i \), then calculate the frequency of occurrence, \( B^m_i(r) = \frac{1}{N-\delta-1} P_i \), and define

\[
B^m(r) = \frac{1}{N-\delta} \sum_{i=1}^{N-\delta} B^m_i(r). \tag{4.2}
\]

4: Extend the dimensionality of the multivariate delay vector in (4.1) from \( m \) to \( (m + 1) \). This can be performed in \( p \) different ways, as from a space with the embedding vector \( M = [m_1, m_2, \ldots, m_k, \ldots, m_p] \) the system can evolve to any space for which the embedding vector is \([m_1, m_2, \ldots, m_k + 1, \ldots, m_p]\) (\( k = 1, 2, \ldots, p \)). Thus, a total of \( p \times (N - \delta) \) vectors \( X_{m+1}(i) \) in \( \mathbb{R}^{m+1} \) are obtained, where \( X_{m+1}(i) \) denotes any embedded vector upon increasing the embedding dimension from \( m_k \) to \( (m_k + 1) \) for a specific variable \( k \). In the process, the embedding dimension of the other data channels (\( \neq k \)) is kept unchanged, so that the overall embedding dimension of the system undergoes the change from \( m \) to \( (m + 1) \).

5: For a given \( X_{m+1}(i) \), calculate the number of vectors \( Q_i \), such that \( d[X_{m+1}(i), X_{m+1}(j)] \leq r \), where \( j \neq i \), then calculate the frequency of occurrence, \( B^{m+1}_i(r) = \frac{1}{p(N-\delta-1)} Q_i \), and define

\[
B^{m+1}(r) = \frac{1}{p(N-\delta)} \sum_{i=1}^{p(N-\delta)} B^{m+1}_i(r). \tag{4.3}
\]

6: Finally, for a tolerance level \( r \), estimate MSampEn as

\[
MSampEn(M, \tau, r, N) = -\ln \left[ \frac{B^{m+1}(r)}{B^m(r)} \right]. \tag{4.4}
\]
4.3.1 Geometric Interpretation of MSampEn

To gain a better understanding of the multivariate sample entropy (MSampEn) method, its geometric interpretation is given for a real world wind speed signal and is shown in Figure 4.2. Consider a two-dimensional recording of wind speed, where the eastward component is denoted by $x(n)$ and the northward component by $y(n)$. Assume the time lag vector $\tau = [1, 1]$ and the embedding vector $M = [1, 1]$; then as shown in Figure 4.2(b), the composite delay vectors will be $[x(n), y(n)]$, where $n$ denotes the sample index. In the process of calculation of MSampEn, for any such vector (e.g. $[x(64), y(64)]$), we need to count the number of neighbors which are within a distance $r$ (tolerance level), illustrated by a circle\(^1\) centered at $[x(64), y(64)]$ with radius $r$ in Figure 4.2(b). The average number of composite delay vectors that are within a fixed threshold $r$ (so called $r$-neighbors) in this two-dimensional space is next calculated, which serves as an estimate of the probability density, $B^m(r)$, that any two composite delay vectors in a 2-dimensional space are similar to within a tolerance level $r$.

For the example given in Figure 4.2(b), upon increasing the embedding dimension from two to three, we have two possible subspaces of dimension three: (i) the subspace of all the vectors $[x(n), x(n+1), y(n)]$ shown in Figure 4.2(c) and (ii) the subspace of all the vectors $[x(n), y(n), y(n+1)]$ shown in Figure 4.2(d). A naive approach would be to calculate the number of vectors that are within a fixed threshold $r$ in each three-dimensional subspace and then to average over both subspaces. Instead, we employ a rigorous approach and compare composite delay vectors (to find the neighbors) not only within each subspace but also across all the subspaces. Figure 4.2(e) shows both the subspaces within the three dimensional space. Note that here the points $[x(64), x(65), y(64)]$ and $[x(64), y(64), y(65)]$ are compared which is not the case for the naive method. This allows us to calculate an estimate of the probability density, $B^{m+1}(r)$, that any two composite delay vectors in a 3-dimensional space are similar to within a tolerance level $r$. Thus, we can calculate the conditional probability that two sequences of $m$ data points (or two composite delay vec-

\(^1\)For an $(m+1)$-dimensional space, the set of neighboring vectors would be enclosed by an $m$-sphere if the distance is calculated using the Euclidean norm and by an $m$-cube if we use a maximum distance norm.
Figure 4.2: Geometry behind the calculation of MSampEn, for a two-dimensional wind signal.
tors in $m$-dimensional space), which are similar to within a tolerance level $r$, will remain similar in the same sense, when the next data point is included (or equivalently the dimension of the composite delay vector is increased by one), provided that self-matches are not considered. A negative logarithm of this conditional probability, $-\ln[B_{m+1}(r)/B_{m}(r)]$, defines the MSampEn measure.

From the above analysis, it is worth mentioning that since MSampEn compares composite delay vectors across subspaces generated from all combinations of input data channels, it caters for both within-channel and cross-channel correlations among multivariate data. For the same reason, MSampEn is expected to yield higher complexity estimates for correlated noises than for uncorrelated noises; a geometric interpretation of results obtained by applying the method on bivariate correlated and uncorrelated noises is given in Appendix A.

### 4.3.2 Effect of Data Length on MSampEn

It has been empirically found in [54] that $10^m - 20^m$ data samples are sufficient to robustly estimate univariate approximate entropy or sample entropy. To assess the sensitivity of the proposed multivariate sample entropy to the data length parameter, we evaluated multivariate sample entropy of 2, 4, 6 and 8-channel white as well as $1/f$ noise as a function of sample size $N$, where for each channel the embedding dimension $m_k = 2$ and the threshold $r = 0.20$ was taken. To show that the behavior of MSampEn with respect to data length parameter is similar to that of SampEn, the univariate SampEn was also calculated for single channel white as well as $1/f$ noise. Figure 4.3 shows that for both the white and $1/f$ multichannel noise, MSampEn estimates were consistent for data length $N \geq 300$, illustrating suitability of MSampEn for the analysis of real world data. This way, standard deviation of multivariate sample entropy estimates (error bars) calculated from the simulated series is likely to be smaller than the deviations related to experimental errors as well as inter- and intra-subject variability for most practical applications. Also note that the greater the value of $N$, the more robust the MSampEn estimates, as seen from the errorbars in Figure 4.3.
Figure 4.3: MSampEn as a function of data length $N$ where $r = 0.20$ and $m_k = 2$ in each data channel. Shown are the mean values for 30 simulated multichannel time series containing white and $1/f$ noise.
4.3.3 Effect of Embedding Parameter on MSampEn

Physically, for the univariate sample entropy, the increase in sample entropy values with an increase in embedding dimension $m$ is due to progressively fewer delay vectors to compare as $m$ increases. On the contrary, for MSampEn the increase in $m$ does not reduce the number of the available delay vectors, as the composite multivariate embedded vectors are constructed in parallel. Figure 4.4 presents an MSampEn vs $m_k$ plot for 2, 4, 6 and 8-channel white and $1/f$ noise with 40,000 points in each channel, and illustrates that the result does not vary a lot depending on $m_k$ up to the value of 4. This behavior is similar to the univariate sample entropy as shown in Figure 4.4(a). For a large $m_k$, the length of the time series required for a valid entropy estimate would be prohibitively large (for $m_k=5$, we need at least $10^5$ data points in each channel), and therefore for practical purposes, in the literature $m_k$ is selected in the range of $1-2$. Besides, in MSampEn calculation we do not give preference to any particular $m_k$ over $m_j$ when increasing the embedding dimension from $m$ to $m+1$. Instead, we create embedded vectors in all the $p$ subspaces and compare them within and across the subspaces to find the $r$-neighbors. This also effectively increases the total number of available delay vectors $p$ times, and makes the proposed MSampEn robust to the variation in both the parameter $m$ and data length $N$.

4.3.4 Effect of Threshold Parameter on MSampEn

In the sample entropy (SampEn) calculation, the $r$ value defines the similarity criterion used to compare delay vectors. If the absolute difference between any two matched vector components is larger than $r \times SD$ (standard deviation of the time series), then the vectors are different; otherwise, they are considered equal. Thus, the tolerance $r$ can be treated as a filter that discards uncorrelated noise or determines the level of accepted noise. A large tolerance level acts as a coarse filter where templates are easily matched, that is, fewer vectors are distinguishable. This results in a relatively low entropy value. On the other hand, a small tolerance level operates as a fine filter and yields relatively large entropy values. In the multivariate sample entropy (MSampEn) calculation, the multivariate generalization of the standard deviation - the total variation $tr(S)$ is used, where $S$ is the
Figure 4.4: MSampEn as a function of embedding parameter $m_k$, where each channel has 40,000 samples and $r = 0.20$. Shown are the mean values for 30 simulated multichannel time series containing white and $1/f$ noise.
covariance matrix. To maintain the same total variation for all the multivariate series, the individual data channels were normalized to unit variance so that the total variation equals the number of channels/variables. Then, \( r \) is taken as some percentage of \( tr(S) \). This way, taking \( r \) as a percentage (say 15\%) of \( tr(S) \) is similar to taking \( r = 0.15 \) in each channel.

Figure 4.5 shows the dependency of MSampEn on \( r \) for multi-channel white and 1/f noise. Observe that as the \( r \) value increases, the MSampEn value for both simulated 1/f and white noise time series decreases. Similarly, as \( r \) decreases, the MSampEn increases. For some very small \( r \) values, the MSampEn are not defined and as a result they are not shown in the corresponding curves of Figure 4.5(a). It is to be noted that the consistency of MSampEn value (i.e., higher for white noise than 1/f noise) is preserved for the value of \( r \) from 0.1 to 0.3. This range of \( r \) is also suggested for the practical application of sample entropy in the literature as shown in Figure 4.5(a).

4.4 Multivariate Multiscale Entropy

The multivariate multiscale entropy (MMSE) method is described in detail in Algorithm 4.

**Algorithm 4 The Multivariate Multiscale Entropy**

1: Multiple coarse-grained time series are generated from the original time series \( \{x_{k,i}\}_{i=1}^{N} \), \( k = 1,2,\ldots,p \), where \( p \) denotes the number of variates (channels) and \( N \) the number of samples in each variate.

2: The elements of the coarse-grained time series of scale factor \( \epsilon \) are calculated as:

\[
y_{k,j}^\epsilon = \frac{1}{\epsilon} \sum_{i=(j-1)\epsilon+1}^{je} x_{k,i},
\]

\[ j = 1,2,\ldots,N_\epsilon \] and \( k = 1,\ldots,p \).

3: Calculate the multivariate sample entropy, MSampEn (described in Algorithm 3) for each coarse-grained multivariate \( \{y_{k,j}^\epsilon\}_{i=1}^{N} \), and plot MSampEn as a function of the scale factor \( \epsilon \).
Figure 4.5: MSampEn as a function of threshold parameter $r$, where each channel has 10,000 samples and $m_k = 2$ in each data channel. Shown are the mean values for 30 simulated multichannel time series containing white and 1/f noise.
4.5 Multivariate Complexity Analysis

The multivariate MSE (MMSE) plots, that is, multivariate sample entropy represented as a function of the scale factor, are next used to assess relative complexity of normalized multichannel temporal data. The interpretation of the MMSE curves is as follows:

- The multivariate time series $X$ is considered more dynamically complex than the multivariate time series $Y$, if for the majority of time scales the multivariate sample entropy values for signal $X$ are higher than those for signal $Y$.

- A monotonic decrease in the multivariate entropy values with the scale factor indicates that the signal in hand only contains useful information at the smallest scales, this is typical for both completely random and fully predictable signals.

- A multivariate system exhibiting long range correlations and complex generating dynamics is characterized by either a constant multivariate sample entropy or it exhibits a monotonic increase in multivariate sample entropy with the scale factor.

4.5.1 Validation on Synthetic Data

The univariate MSE analysis has shown [14] [15] that for random white noise (uncorrelated), the sample entropy values decrease monotonically with scale, whereas for a $1/f$ noise (long-range correlated) sample entropy remains constant over multiple time scales. This indicates that the univariate $1/f$ noise is structurally more complex than uncorrelated random signals.

To illustrate the corresponding behavior for the multivariate case, we generated a trivariate time series, where originally all the data channels were realizations of independent white noise. We then gradually decreased the number of variates that represent white noise (from 3 to 0) and simultaneously increased the number of data channels that represent independent $1/f$ noise (from 0 to 3), so that the total number of variates was always three. Figure 4.6 shows the MMSE curves for the cases considered; notice that as the number of variates representing $1/f$ noises increased, MSampEn at higher scales
also increased, and when all the three data channels contained $1/f$ noise, the complexity at larger scales was the highest. The analysis in Figure 4.6 therefore confirms that, as desired, the more variables/channels within a multivariate time series exhibit long range correlations, the higher the overall complexity of the underlying multivariate system.

Recall that the original univariate MSE algorithm accounts for long term correlations within a single data channel, however, due to its univariate nature, it cannot model the cross-channel information present in multivariate recordings. On the other hand, MMSE is designed for multivariate data. To illustrate this difference, we first generated independent realizations of white and $1/f$ noise, and the three channels of trivariate white and $1/f$ noise were constructed using combinations of those independent realizations, thus making the channels correlated. Figure 4.7 illustrates the ability of MMSE to model both within- and cross-channel properties in multivariate data. Figure 4.7(a) shows that the naive multivariate approach accounts for within-channel correlations but not for cross-channel correlations, and was not able to distinguish between uncorrelated and correlated trivariate random and $1/f$ noises. Figure 4.7(b) shows that, as desired, the proposed multivariate MSE fully caters for both within- and cross-channel correlations. Indeed, based
Figure 4.7: Multivariate multiscale entropy (MMSE) analysis for correlated vs uncorrelated trivariate white and $1/f$ noise, each with 10,000 data points. The curves represent an average of 20 independent realizations and error bars the standard deviation (SD).
Figure 4.8: Standard univariate multiscale entropy analysis for white noise and scalar AR processes, each with 10,000 data points. The curves represent an average of 20 independent realizations and error bars the standard deviation (SD).

on MMSE the complexity of the correlated trivariate $1/f$ noise at large scales was the highest, followed by the uncorrelated $1/f$ noise, and correlated and uncorrelated white noise. This conforms with the underlying physics and validates the proposed MMSE method, as the complexity of the considered multivariate processes exhibiting both within- and cross-channel correlations is higher than that of uncorrelated multivariate white noise and uncorrelated multivariate $1/f$ noise (where long range correlations only exist within single channels).

The usefulness of the MMSE analysis is further illustrated for the analysis of scalar and vector autoregressive (AR) processes [79]. The dependence of current values on past values and the error terms in AR and vector AR processes is assumed to be linear. The AR processes were designed so as to have an increasing correlation span with the model order. Figure 4.8 shows the standard univariate MSE analysis for the scalar AR processes considered and Figure 4.9 the MMSE analysis for the corresponding bivariate vector AR (VAR) processes. As desired, in both cases, as the model order increased, the complexity of the corresponding signals measured by MSE and MMSE increased too.
4.6 Applications for Real World Multivariate Processes

The multivariate multiscale entropy analysis is next evaluated for some multivariate real world recordings: human stride interval analysis, postural sway dynamics analysis, and bivariate physiological data (breathing and heart beats) from young and elderly subjects. More real world applications are described in Chapter 6.

4.6.1 Case Study: Stride Interval Characterization

The data used were from [80], where stride interval fluctuations were recorded from ten healthy subjects who walked for 1 hour at their usual, slow, and fast paces. The participants were further asked to walk following a metronome which was set to each participant’s mean stride interval.

To assess the differences in relative complexity between the unconstrained (slow, normal, fast) and the corresponding constrained (metronomically-paced) conditions, we considered the three walking paces as different variables from the same system, and used MMSE to discriminate between the ‘self-paced’ and ‘metronomically-paced’ walk.
To test the hypothesis that the complexity of such time series is encoded into the sequential ordering of the samples of stride intervals, we also produced the corresponding surrogate time series by shuffling (randomly reordering) the sequence of data points. In this way, in surrogates the correlations among the data samples were destroyed, while preserving statistical properties of the distributions (particularly the first and second moment), and the complexity of the surrogates is lower or equal (if the original is completely random) than that of the original signal.

![Image](image_url)

Figure 4.10: Multivariate multiscale entropy (MMSE) analysis for self-paced (solid red square) vs metronomically-paced (solid cyan circle) stride interval (human gait) time series and for their corresponding randomized surrogates (dashed line). Top: univariate MSE analysis; Middle: bivariate MMSE analysis; Bottom: Trivariate MMSE analysis. The curves represent an average of trials from 10 subjects and error bars the standard deviation (SD).
The values of the parameters used to calculate MSampEn were $m_k = 2$, $\tau_k = 1$ and $r = 0.15 \times$ (standard deviation of the normalized time series) for each data channel; these parameters were chosen on the basis of previous studies indicating good statistical reproducibility for SampEn [13]. For MSE/MMSE, the length of each coarse-grained sequence was $\epsilon$ (scale factor) times shorter than the length of the original series which was around 2000 samples, so the highest scale factor considered in the analysis was $\epsilon=7$.

The top panel in Figure 4.10 shows the results obtained by the univariate MSE performed for the normal speed time series [18] - the univariate MSE was not able to perform statistically significant discrimination between self-paced and metronomically-paced walk as the error bars overlapped. The middle and bottom panels in Figure 4.10 show that when the walking conditions are considered within the multivariate approach (bivariate for any two walking conditions or trivariate for all the three walking conditions), the proposed MMSE was able to discriminate between self-paced and metronomically-paced walk. This opens completely new analysis possibilities, since the MMSE method was able to consider all the walking conditions within one unifying model, directly benefiting from the multivariate approach. Figure 4.10 also indicates the presence of persistent serial correlations, which are long-range dependent in self-paced walking, and the lack of any correlations in metronomically-paced walking; in this case the shape of the MMSE curve is similar to that for multivariate white noise. As expected, the surrogate series (randomly shuffled) showed similar pattern to that for white noise (dashed line in Figure 4.10).

Table 4.1: Statistical significance tests for the univariate, bivariate and trivariate human stride interval analysis. Shown are scales for which the differences are statistically significant

<table>
<thead>
<tr>
<th>Conditions taken</th>
<th>Student’s t-test</th>
<th>Mann-Whitney U test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>No scale</td>
<td>No scale</td>
</tr>
<tr>
<td>Fast</td>
<td>No scale</td>
<td>No scale</td>
</tr>
<tr>
<td>Slow</td>
<td>No scale</td>
<td>No scale</td>
</tr>
<tr>
<td>Fast and Normal</td>
<td>6,7</td>
<td>6,7</td>
</tr>
<tr>
<td>Fast and Slow</td>
<td>4,5,6,7</td>
<td>4,5,6,7</td>
</tr>
<tr>
<td>Normal and Slow</td>
<td>5,6,7</td>
<td>5,6,7</td>
</tr>
<tr>
<td>Fast, Normal and Slow</td>
<td>3,4,5,6,7</td>
<td>3,4,5,6,7</td>
</tr>
</tbody>
</table>

To evaluate the statistical significance of the difference of the entropy statistics be-
tween self-paced and metronomically-paced sets, Student’s t-test and the Mann-Whitney U test (also known as Wilcoxon rank sum test) were applied. Student’s t-test is a parametric approach that tests the null hypothesis that the means of normally distributed populations are equal. On the other hand, the Mann-Whitney U test is a nonparametric test where the null hypothesis, that independent samples come from identical (similar shape) continuous (not necessarily normal) distributions with equal medians, is tested against the alternative that they do not have equal medians.

Entries in Table 4.1 represent the scales at which MSampEn measures between self-paced and metronomically-paced walking are significantly different according to the above two statistical tests. When all three walking conditions were simultaneously considered, both the statistical significance tests revealed significant differences ($p < 0.01$) in MSampEn measures at all scales except for $\epsilon = \{1, 2\}$ and the corresponding null hypothesis (equal mean or median) was rejected. On the other hand, for the univariate MSE there were no statistically significant differences between self-paced walking and metronomically-paced walking at any scales.

This result also indicates that for metronomically-paced walking in slow, normal and fast conditions, the time series share uncorrelated random underlying dynamics both within- and cross-channel, whereas for free walking, the time series for slow, normal and fast conditions are correlated both within- and cross-channel. That explains why at larger scales the complexity for the multivariate measurements was highest for self-paced walking (cf. metronomically-paced walking), and the separation was statistically significant over more scales when we considered all the available walking conditions. The MMSE therefore offers a significant improvement over the previous studies [18] [80] and also supports the more general concept of multiscale complexity loss with ageing and disease or when a system is under constraints (metronomically-paced), which all reduce the adaptive capacity of biological organization at all levels [16].
4.6.2 Case Study: Postural Sway Analysis

Multivariate complexity analysis of real world postural sway dynamics (time series of the center of pressure (COP) displacement) is next performed for young and elderly subjects during quiet standing, and has been compared with the existing univariate complexity analysis [81]. COP displacement was recorded simultaneously for the mediolateral (side-to-side) and anteroposterior (front-to-back) direction [82] for 15 healthy young and 12 healthy elderly volunteers at 60 Hz.

Figure 4.11: Multiscale entropy analysis of COP time series for young (red - solid line) and elderly (green - dotted line) subjects. Top: Univariate MSE for the mediolateral component; Middle: Univariate MSE for the anteroposterior component; Bottom: Bivariate MMSE analysis. The plots represent mean values of SampEn/MSampEn for all subjects and error bars represent standard error.
Since the postural sway time series exhibits high frequency fluctuations superimposed on low frequency trends, the data was first detrended using the multivariate empirical mode decomposition technique [23]. The values of the parameters used to calculate MSampEn were $m_k = 2$, $\tau_k = 1$, and $r = 0.15 \times$ (standard deviation of the normalized time series) for each data channel; these parameters were chosen on the basis of previous studies indicating good statistical reproducibility for the univariate SampEn [13] [81]. Since the original time series had $1.8 \times 10^3$ data points, the highest achievable scale factor ($\epsilon = 6$) had 300 data points; this was sufficient for accurate analysis, as shown in Figure 4.3.

The top and middle panel in Figure 4.11 show the results obtained by the univariate MSE performed for the mediolateral (ML) and anteroposterior (AP) component of the COP time series, respectively. The bottom panel shows that when both the ML and AP component were considered within the multivariate approach, the proposed MMSE was able to discriminate between young and elderly subjects more efficiently, as indicated by the better separation of the MMSE curves and the complexity index (area under the MSE curve) being higher for healthy young subjects than for their elderly counterparts. Table 4.2 shows the statistical significance of the results obtained by the one-tailed t-test with unequal variances. From the table, it is evident that the difference in complexity between the young and elderly subjects was statistically significant in the AP direction for 5% significance level and not significant in the ML direction. If, for rigor, we take a 1% significance level, then none of the univariate analyses gave statistically significant results.

On the contrary, in the bivariate case we observed significant differences, even for the significance level of $p = 6.65 \times 10^{-4}$. Thus, using the multivariate approach we

---

**Table 4.2: Complexity indices and p values obtained using one-tailed Student’s t-test with unequal variance.**

<table>
<thead>
<tr>
<th></th>
<th>AP(mean±SD)</th>
<th>ML(mean±SD)</th>
<th>AP &amp; ML bivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young subjects</td>
<td>7.75 ± 1.67</td>
<td>8.90 ± 1.52</td>
<td>3.13 ± 1.53</td>
</tr>
<tr>
<td>Elderly subjects</td>
<td>7.07 ± 2.23</td>
<td>8.66 ± 1.62</td>
<td>2.48 ± 1.20</td>
</tr>
<tr>
<td>p value</td>
<td>0.0181</td>
<td>0.1598</td>
<td>6.65 × 10^{-4}</td>
</tr>
</tbody>
</table>

Due to ageing and the associated constraints, the complexity of postural sway for the elderly is lower than for the young.
obtained a narrower confidence interval on the population mean corresponding to better estimates. This illustrates significant advantages of using MMSE when assessing relative complexity of real world multivariate data, and supports the more general concept of multiscale complexity loss with ageing and disease or when a system is under constraints, as those factors reduce the adaptive capacity of biological organization at all levels [16].

### 4.6.3 Case Study: Complexity Analysis of Physiological Signals

We shall now apply the MMSE method to the Fantasia database [83] to simultaneously analyze the complexity of interbeat interval (R-R) and interbreath interval series. The presence of long-range correlations in both cardiac and respiratory dynamics was previously established using detrended fluctuation analysis (DFA) in [84] and [85].

A subset of the Fantasia database was chosen consisting of ten young (21 - 34 years old) and ten elderly (68 - 85 years old) rigorously-screened healthy subjects who underwent 120 minutes of continuous supine resting while continuous electrocardiographic (ECG) and respiration signals were collected. Each subgroup of subjects included seven women and three men. The continuous ECG and respiration signals were digitized at 250 Hz, and the Interbeat interval (R-R) time series and interbreath interval time series were generated; for more details see [84] and [85]. The values of the parameters used to calculate MSampEn were $m_k = 2$, $\tau_k = 1$ and $r = 0.15 \times$ (standard deviation of the normalized time series) for each variate.

First, the univariate MSE was applied separately to the interbeat interval (R-R) series (Figure 4.12(a)) and interbreath interval series (Figure 4.12(b)). For rigor, the corresponding surrogate time series were also produced by shuffling (randomly reordering) the sequence of data points. In both cases, although, as desired, for some scales physiological signals from healthy young subjects exhibited higher complexity than those of healthy elderly subjects, the complexity values were lower than those of the randomized surrogates. This behavior wrongly suggests a lack of long term correlations in both cardiac and respiratory dynamics, illustrating that the univariate approach was not able to produce robust estimates.
Next, MMSE was applied to a bivariate time series consisting of the R-R and interbreathing intervals. Figure 4.13 reveals long range correlations in both cardiac and respiratory dynamics, illustrated by the fact that the MSampEn values for larger scales were higher than those of the randomized surrogates, which have no temporal structure.
Figure 4.13: Multivariate multiscale entropy (MMSE) analysis of the bivariate (R-R, Interbreath interval) signal. The curves represent an average of 10 subjects and error bars the standard deviation (SD).

Figure 4.13 also indicates lower complexity of physiological responses of elderly subjects than the young ones, conforming with the complexity loss theory with aging [84].

4.7 Conclusion

In this chapter, we have generalized the recently introduced multiscale entropy (MSE) method to the multivariate case, to suit real world biological and physical systems which are typically of multivariate, correlated and noisy natures. The inherent complexity of such structures and their coupled dynamics also make the proposed multivariate multiscale entropy (MMSE) method naturally suited to reveal the long range within- and cross-channel correlations present. The MMSE method has been validated on both illustrative benchmark data and on real world multivariate gait, postural sway, and physiological data.
Chapter 5

Multivariate Entropy Analysis with Data Adaptive Scales

In the previous chapter, the complexity of real world multivariate data is addressed by employing our proposed multivariate extension of multiscale entropy (MSE) method, also known as multivariate multiscale entropy (MMSE) analysis. Both (univariate) MSE and MMSE methods are shown to perform better than traditional complexity detection techniques since they operate on multiple scales of the signals and are thus able to extract information regarding inherent long range correlations in the data. MMSE, in addition, can also quantify inter-channel correlations in multivariate data and is thus suited for signals containing multiple channels.

This chapter aims to propose a data-adaptive algorithm for the entropy-based analysis of structural regularities (complexity) in multivariate signals. This is achieved by combining multivariate sample entropy with the multivariate extension of empirical mode decomposition, a data-driven multiscale technique. The proposed analysis across data-adaptive scales makes the approach robust to nonstationarity, a critical issue with information theoretic measures. Simulations on synthetic and real-world physiological data support the approach and validate the hypothesis of increased complexity for unconstrained as compared to constrained (due to e.g. ageing or illness) biological systems.
5.1 Introduction

While the MSE measure has been successfully applied to distinguish between different real-world physiological time series based on their dynamical complexity [17–20], it also has some limitations stemming from the deterministic way of generating multiple scales of input data. The method uses the so-called coarse graining process which, owing to its low-pass filtering characteristics, is unsuitable for the extraction of high frequency components and also results in aliasing (see also Figure 5.1), causing potential systemic artifacts which inhibit the analysis. More critically, the coarse graining process reduces the input data length to half its original size for each successive data scale and the ‘deterministic’ temporal scales do not necessarily match the intrinsic dynamical scales defined by the signal-generating system. As a result, only input data of ‘sufficient’ length and ‘regular’ scales can be reliably processed by the MSE method; in turn, for real-world data we can evaluate the MSE only over a limited range of temporal scales. To alleviate these problems, a class of Butterworth filters were used to circumvent the aliasing observed in the original coarse graining process of MSE [21]. However, this does not circumvent the need for data driven scales, over which to perform the analysis.

It was recently proposed to employ a data-driven method, the empirical mode decomposition (EMD) [86], to generate intrinsic multiple data scales from input data, to be used for the subsequent MSE analysis [22,24]. The resulting EMD-based MSE method produced improved results owing to the fully data-driven nature of EMD and also due to the fact that it operates locally based on the extrema of the (univariate) input signal, yielding well defined narrowband scales intrinsic to the input data. Another benefit of using EMD in conjunction with the MSE method is that the standard MSE fails to cater for nonstationary signals, that is, if a signal contains one or more pronounced trends, little can be inferred from sample entropy - trends tend to dominate other interesting features. From the statistical perspective, it is therefore imperative to remove any trend before meaningful interpretation can be made from MSE analysis. Though moving/sliding window method is extensively used for nonstationary signal analysis, it is very difficult to capture the long-range correlations present in the signal by a fixed window length. Since
EMD decomposes data into narrow-band *quasi-stationary* signals \[86\], subsequent MSE analysis on each EMD output component promises to facilitate MSE based complexity analysis. For instance, EMD naturally captures a trend in the input data in its residue (last extracted component) which can be removed prior to the MSE analysis.

Recently, advances in sensor and data acquisition technologies have made it possible to record in a coherent way real-world signals containing multiple data channels, with possibly large differences in the dynamics across the channels. For such signals, assessment of cross-statistical properties between multiple input channels is vital for a complete understanding of the underlying signal-generating system. This calls both for the development of multivariate extensions of existing signal processing algorithms in order to directly process multiple channels of input data and, in the process, employ both within- and cross-channel information, and to design data-adaptive algorithms which define multiple intrinsic time scales in multivariate data, which is the main focus of this chapter.

Recent multichannel extensions of both EMD [87] and MSE algorithm [88–90] (pro-
posed in the previous chapter) have been shown to outperform their standard, univariate, counterparts in the analysis of real-world multivariate signals. The availability of these extensions provides an opportunity to develop a robust framework for the complexity analysis of multivariate data. More specifically, we propose to employ recently developed multivariate EMD (MEMD) to generate intrinsic data scales for the subsequent multivariate MSE (MMSE) analysis of input multichannel data, and to benefit from the mode alignment property of MEMD, yielding the following advantages:

- This ensures that the scales generated for each data channel are same in number and belong to the same frequency band (monocomponent), which makes the multiscale complexity analysis meaningful;

- This way, the limitation of sufficient input data length due to coarse graining process is alleviated since EMD/MEMD generates temporal data scales of same length as the length of the input signal;

- The so-generated time scales are data adaptive and fully suited to the dynamics of the signal in hand, unlike the currently used coarse graining techniques;

- The proposed multivariate MSE complexity assessment method operates on nonstationary data thus bypassing the main limitations of current methods - requirement of stationary data sources.

The presented MEMD-enhanced MMSE method is fully multivariate, unlike e.g. the method given in [24] where MEMD was used to generate multiple scales from the input data and univariate sample entropy was subsequently applied for complexity analysis. The implication of employing univariate sample entropy was that MEMD had to be operated across multiple trials rather than multiple channels, thus, not making use of the full potential of MEMD. Owing to the multivariate nature of the introduced multivariate multiscale entropy measure in conjunction with MEMD, our proposed approach is fully multivariate - it analyzes both the within- and cross-channel information simultaneously and uniquely, it operates on nonstationary multivariate data with large discrepancies in
channel dynamics, as demonstrated on the complexity analysis of real-world multivariate biological data.

5.2 Multivariate Empirical Mode Decomposition

The empirical mode decomposition (EMD) algorithm decomposes an input signal into a finite number of narrow-band amplitude/frequency modulated (AM/FM) components known as intrinsic mode functions (IMFs) [86]. These IMFs are designed such that the subsequent application of the Hilbert transform to the IMFs yields physically meaningful frequency estimates, resulting in an accurate time-frequency representation of the signal [91]. The data-driven nature of the algorithm also ensures a more compact and physically meaningful signal representation as compared to the standard time frequency algorithms such as short-time Fourier and wavelet transform.

In its original formulation EMD is univariate, that is, it can only process single-channel data. Recent developments in the field of sensor and engineering technologies, however, have given rise to a new class of multivariate signals containing multiple data channels. To obtain data-driven bases common to all channels within multivariate signals, a generalized multivariate extension of EMD (MEMD) has been recently developed [23]. MEMD has also been shown to align data corresponding to similar frequency bands from multiple channels thus providing an assessment of their possible interdependence; this is referred to as the mode alignment property of MEMD [87]. It is worth mentioning that standard EMD applied separately to each channel of a multivariate signal cannot achieve mode alignment, due to its empirical and univariate nature [92]. Since in multivariate signal processing, mode alignment is a necessary requirement for a ‘fair’ comparison between data channels at multiple scales [93], MEMD has found a number of real world applications involving multichannel data in a very short span of time [24] [94] [95].

The MEMD algorithm operates by estimating the local mean of input multivariate signals by taking multiple signal projections along different directions in $p$–dimensional spaces. Notice the difference from standard (univariate) EMD where the local mean is
5.2 Multivariate Empirical Mode Decomposition

taken by averaging only upper and lower envelopes, which are obtained by simply interpolating between signal maxima and minima. However, the notion and locations of extrema (maxima and minima) cannot be defined for multivariate signals directly [96] and to that end, MEMD calculates the local mean by first generating uniform direction vectors on $p-$sphere\(^1\) by using low-discrepancy Hammersley sequences [97]. Signal projections along the generated set of vectors are taken and their extrema are interpolated component-wise to yield multidimensional envelopes of a multivariate signal. These envelope curves, each corresponding to a particular direction vector, are then averaged to obtain the multivariate signal mean.

More specifically, consider a sequence of $p-$dimensional vectors $s(t) = \{s_1(t), s_2(t), \ldots, s_p(t)\}$, representing a multivariate signal with $p$ variates (channels), and the symbol $x_{\theta_v} = \{x_1^v, x_2^v, \ldots, x_p^v\}$ denoting a set of $v = 1, 2, \ldots, V$ direction vectors along the directions given by angles $\theta_v = \{\theta_{v1}, \theta_{v2}, \ldots, \theta_{v_p}\}$ in $\mathbb{R}^p$. Then, the steps for obtaining the signal decomposition via multivariate extension of EMD (MEMD) are summarized in Algorithm 5.

**Algorithm 5 Multivariate EMD**

1: **Sampling criterion:** Choose a suitable point set for sampling a $(p-1)$-sphere (uniform, equiangular, etc.);
2: **Univariate spatial projections:** Calculate a multidimensional projection, denoted by $q_{\theta_v}(t)$, of the $p$-variate input signal $s(t)$ along the direction vector $x_{\theta_v}$ for all $v \in \theta_v$ (the whole set of direction vectors), giving the set of projections $q_{\theta_v}(t)_{V=1}^V$;
3: **Extrema finding:** Find the time instants $\{t_{\theta_v}^i\}_{V=1}^V$ corresponding to the maxima for every member of the set of projected signals $q_{\theta_v}(t)_{V=1}^V$;
4: **Envelope detection:** Interpolate $[t_{\theta_v}, s(t_{\theta_v})]$ to obtain multivariate envelope curves $e_{\theta_v}(t)_{V=1}^V$;
5: **Local mean calculation:** For a set of $V$ direction vectors, the mean $m(t)$ of the envelope curves is calculated as:

$$m(t) = \frac{1}{V} \sum_{v=1}^{V} e_{\theta_v}(t) \quad (5.1)$$

6: **Sifting process:** Extract ‘detail’ $d(t)$ using $d(t) = s(t) - m(t)$. If $d(t)$ fulfills the stoppage criterion for a multivariate IMF, apply the above procedure to $s(t) - d(t)$, otherwise apply it to $d(t)$.

The sifting process in MEMD can be stopped when all the projected signals fulfill

\(^1\)A $p$-sphere, or equivalently a $p$-dimensional hypersphere, can be considered as an extension of the ordinary sphere to an arbitrary dimension.
any of the stoppage criteria adopted in standard EMD. One popular stopping criterion
used in EMD stops the sifting when the number of extrema and the zero crossings differ
at most by one for $S$ consecutive iterations of the sifting algorithm [98]. However, caution
must be taken while using this criterion for multivariate cases as it has been found to be
computationally very expensive for long signals.

Another criterion introduces an evaluation function based on the envelope ampli-
tude, which is given by

$$a(t) = \frac{1}{V} \sum_{v=1}^{V} |e_{\theta_v}(t) - m(t)|$$

(5.2)

The sifting process is continued until the value of the evaluation function, defined as

$$f(t) = \frac{|m(t)|}{a(t)}$$

where $m(t)$ is the local mean signal, is less than or equal to some predefined
thresholds $[\theta_1 \theta_2 \alpha]$ [99]. In MEMD, both the above criteria are used in conjunction: the
conditions of the above stopping criteria are imposed on $V$ multiple projections and the
sifting process is stopped when stopping conditions are fulfilled for all projections. Finally,
we obtain the following decomposition containing $J$ IMFs for a given multivariate signal
$s(t)$ containing $p$ variates:

$$s(t) = \sum_{j=1}^{J} c_j(t) + r(t)$$

(5.3)

where $c_j(t)$ represents the $j$th IMF of $s(t)$, also containing $p$ variates like $s(t)$ and $r(t)$
denotes the multivariate residue signal.

Since the multichannel algorithm, MEMD, operates directly on multivariate signals
with scales defined by the data, it provides intrinsic information regarding interaction
between multiple data channels. This offers us physical insight into the structure of the
real-world data and a convenient interpretation:

- The first extracted IMF is the highest frequency component in a signal, containing
  plenty of detail;
- The subsequent IMFs are ideally narrowband and monocomponent, whereby the
  characteristic frequency decreases with the IMF number;
• The last IMF - the trend - can often contain the signal power and little signal detail, so that it is typically omitted from analysis.

• Even if the original signal is non-stationary, the IMFs are much better conditioned and are typically quasi-stationary;

• The IMFs are locally orthogonal [86], providing a parsimonious representation with minimum artifacts in decomposition and reconstruction; the local orthogonality facilitates the ‘processing’ of non-stationary signals.

Figure 5.2: Trivariate wind data (plotted in first row) decomposed by multivariate EMD. The north-south, east-west and upward/downward wind speed components are depicted in the left, right and middle panels respectively.

Advantages offered by MEMD over the univariate (single-channel) EMD include:
1. Direct processing of multichannel data gives the same number of IMFs for all data channels facilitating the analysis of their properties at each (orthogonal) scale, independently;

2. MEMD automatically \textit{aligns} common scales, present across multiple channels, within its multivariate IMFs; a desirable property that is hard to achieve by applying univariate EMD channel-wise on multivariate data, as shown later in Figure 5.4.

3. If one has several multivariate signals to compare, all the multivariate signals can be combined to form a single multichannel signal. Subsequently, MEMD can be used to decompose that multichannel signal and can align common scales across multiple signals, all at once.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figures/5.3.png}
\caption{Averaged spectra of IMFs obtained from \(D = 500\) realizations of 8-channel independent white Gaussian noise via MEMD (top) and the standard EMD (bottom). Overlapping of the frequency bands corresponding to the same-index IMFs is improved in both cases after averaging, with MEMD bands showing much better alignment.}
\end{figure}
5.2 Multivariate Empirical Mode Decomposition

5.2.1 Simulations of MEMD Properties

To demonstrate the mode alignment property of MEMD, we considered real world trivariate wind speed data\(^2\) and the resulting decomposition is shown in Figure 5.2. MEMD generated equal number\(^3\) of IMFs, \(J = 11\), for all input channels, thereby, associating physical meaning to all IMFs. It can be noticed from the decomposition that the residue signal in the last row of Figure 5.2 captures the overall dynamic trend of the data, whereas the first extracted IMF corresponds to the highest frequency component in wind speed data; in subsequent IMFs, the characteristic frequency decreases with the IMF index.

Similarly, the property of alignment of common scales in multivariate data via MEMD can be illustrated by its quasi-dyadic filter bank structure for multivariate white Gaussian noise (WGN) inputs. For that purpose, simulations were carried out on multiple independent realizations of an 8-channel mutually independent WGN process. The average spectra of the IMFs obtained from \(D = 500\) realizations of 8-channel WGN are plotted in Figure 5.3, both for standard EMD (bottom plot) and multivariate EMD (top plot). Observe that statistically, for a given number of noise realizations \(D\), standard EMD failed to accurately align the band-pass filters associated with the sequence of multivariate IMFs shared across the eight noise channels. Although for univariate EMD this alignment is expected to become better with an increase in the number of noise realizations, MEMD-based spectra achieved much better results with the same number of ensembles.

A quantitative evaluation of the mode alignment observed in (M)EMD-based decomposition of trivariate wind speed data of Figure 5.2 is shown in Figure 5.4. It shows plots of normalized cross-correlation coefficients, \(\rho\), calculated between the IMFs obtained from applying standard univariate EMD to each wind speed channel separately (left column), together with MEMD applied on trivariate data directly (right column). The normalized cross-correlation measure \(\rho(j, j')\) between the \(j\)th and \(j'\)th IMFs from (M)EMD, denoted respectively by \(c_j\) and \(c_{j'}\), can be given by

\(^2\)Components of trivariate wind data are wind speed in north-south, east-west and upward-downward direction. The data is sampled at 50Hz and was provided by Prof. Aihara from Tokyo University in Japan.

\(^3\)Only 4 IMFs and the trend are plotted in the figure for convenience.
5.2 Multivariate Empirical Mode Decomposition

Figure 5.4: The normalized IMF cross-correlation from multiple channels of wind speed data, obtained via the univariate EMD channel-wise (a, c and e) and by direct multivariate EMD (MEMD) (b, d and f). The higher values of the cross-correlation measure along the diagonal line in the case of MEMD (right column) indicates the mode-alignment between IMFs from multiple channels. IMFs from standard univariate EMD, on the other hand, give off-diagonal spurious correlation estimates, indicating poor alignment between the corresponding scales (IMFs). The IMF indices grow from left to right and from top to the bottom in all subfigures.
\[
\rho(j, j') = \frac{\Upsilon(j, j')}{\sqrt{\Upsilon(j, j') \Upsilon(j', j')}}
\]  \hspace{1cm} (5.4)

where

\[
\Upsilon(j, j') = \frac{1}{N} \sum_{n=1}^{N} (c_j(n) - \mu_{c_j})(c_{j'}(n) - \mu_{c_{j'}})
\]  \hspace{1cm} (5.5)

while \(c_j(n)\) is the \(j\)th IMF of the input signal, \(\mu_{c_j}\) denotes its mean value, and \(N\) represents the data length.

Using equation (5.4), cross-correlation matrices are calculated and subsequently plotted for all pairs of input data channels (XY, YZ and XZ respectively). As expected, due to the ability of MEMD to align spectra corresponding to the same-indexed IMFs from multiple channels, the correlation matrices obtained from MEMD have significantly higher values along the diagonal as compared to off-diagonal. On the other hand, correlation coefficients from standard EMD are spread non-uniformly with pronounced off-diagonal values, thus exhibiting poor mode alignment between the corresponding scales (same-index IMFs) from different channels. These observations regarding wind speed data analysis via (M)EMD are in complete agreement with the mode alignment property observed for MEMD-based filter banks for multivariate white Gaussian noise (WGN) inputs [87].

**5.3 MEMD-enhanced Multivariate Multiscale Entropy**

In standard multivariate multiscale entropy (MMSE), multiple data scales are generated by applying the same coarse graining process used for the univariate MSE to each input channel in parallel (see listing of Algorithm 4). Performing coarse graining separately for each channel is inherently not a data-adaptive process and also lacks direct multivariate approach; therefore, it fails to generate ‘aligned’ and ‘intrinsic’ temporal scales from the data - a prerequisite for high fidelity multiscale analysis. Moreover, the output of the coarse graining process reduces the length of each subsequent scale to the length of the original time series divided by the corresponding scale factor, \(\epsilon\). The method, thus, imposes the
constraint that the highest scale should have enough data points to be able to calculate valid entropy estimates. This somewhat limits the applicability of the coarse graining based MMSE method for short real-world data.

To alleviate the above problems, we propose to use multivariate empirical mode decomposition (MEMD) to generate multiple scales of a given multivariate data and subsequently perform multivariate entropy analysis on cumulative\(^4\) IMFs (scales). As mentioned in Section 5.2, MEMD is a fully data-driven method which gives intrinsic and correlated (aligned) scales from multiple channels of input data, thus, fulfilling the fundamental requirements of a multivariate data adaptive analysis. To verify the above properties of MEMD on real world data, Figure 5.4 (right column) shows plots of cross-correlation coefficients of IMFs obtained from different combinations \((XY, YZ, XZ)\) of wind speed channels. These plots illustrate the mode alignment property of MEMD on real world wind speed data as the correlation coefficients were found to be significantly higher along the diagonal, which correspond to same-indexed IMFs (scales) from multiple channels, as compared to the off-diagonal ones. It should be mentioned that Figure 5.4 only shows comparison between scales generated from EMD and MEMD. Although, a similar comparison between scales from MEMD and the coarse-graining process would have been illustrative, this was not possible due to the fact that the scales obtained by coarse-graining are of different lengths and therefore cannot be compared. We did, however, calculate the cross-correlation coefficients among the different variates of same coarse-grained scales (corresponding to the diagonal of the correlation matrices shown in Figure 5.4), obtained from wind speed data used in Figure 5.2, and found their values to be significantly lower than the MEMD-based correlation coefficients. This further illustrated the superiority of MEMD over coarse graining process in terms of mode alignment. Moreover, each IMF (scale) obtained via MEMD has the same length as that of input data, thereby, overcoming the problem of successive shortening of data at higher scales in standard approaches employing coarse graining.

The proposed MEMD enhanced MMSE method thus uses a more data adaptive

---

\(^4\)Due to their narrow-band nature, an alternative option is to additionally apply coarse graining to the IMF-scales themselves, with minimal risk of aliasing.
Algorithm 6 MEMD-enhanced multivariate multiscale entropy

1: Generate multiple scales from \( J \) IMFs obtained by applying MEMD to a given multivariate time series \( \{x_{k,i}\}_{i=1}^{N} \) for \( k = 1, 2, \ldots, p \), where \( p \) denotes the total number of variates (channels) and \( N \) represents the total number of samples in each variate which does not change across MEMD-based scales.

2: Define data-driven ‘scales’ of \( x \) as the cumulative sum of IMFs either by \( c_n = \sum_{j=n}^{J} c_j \) (Approach 1) or by \( c_n = \sum_{j=n+1}^{J} c_j \) (Approach 2), where \( n \in [1, J] \) denotes the cumulative IMF index, and \( c_j \) denotes the \( j \)th IMF. Only Approach 1 is used in the sequel.

3: Calculate and plot multivariate sample entropy measure, given in (4.4), for each scale \( n \).

5.3 MEMD-enhanced Multivariate Multiscale Entropy

5.3.1 Validation on Synthetic Data

To illustrate the performance of the proposed MEMD based MMSE method, it was applied to synthetically generated bivariate white noise and bivariate 1/f noise. The 1/f noise possesses long-range correlations and its standard entropy (at scale 1) is lower than that of white noise, however, the 1/f noise is structurally complex whereas the bivariate white noise is not, and any complexity measure should be higher for 1/f noise at increasing scales. Observe from Figure 5.5(b) that though bivariate white noise has higher complexity than 1/f noise for the first scale, the complexity becomes lower than 1/f noise for higher scales. This example on synthetic data illustrates, that by design, 1/f noise is structurally more complex than uncorrelated random noise, a result consistent with standard MSE/MMSE [14, 89, 90] as shown in Figure 5.5(a).

Note that a direct comparison is often not possible between the scales of MEMD-enhanced MMSE and those of standard MMSE as, by design, the frequency ranges of the cumulative IMFs adapt to the data in hand. In the case of white noise, however, the dyadic filter bank property of MEMD is well established [87] as shown in Figure 5.3. Disregarding
5.3 MEMD-enhanced Multivariate Multiscale Entropy

![Graphs showing multiscale entropy analysis for bivariate white and 1/f noise](image.png)

**Figure 5.5**: Multivariate multiscale entropy (MMSE) analysis for bivariate white and 1/f noise, each with 5,000 data points using: (a) coarse graining based standard multivariate MSE, and (b) MEMD-enhanced multivariate MSE. The curves represent an average of 20 independent realizations and error bars the standard deviation (SD).

Certain elements of coarse graining, the averaging operation at scale $\epsilon$ is equivalent to low pass filtering with a cutoff frequency (normalized) of $f_c = 0.5/\epsilon$. Thus for the $n$th cumulative IMF index (Approach 1 in Algorithm 6) of white noise, the equivalent coarse grained scale factor is given by $\epsilon \approx 2^{n-1}$. For insight, the equivalent scale factors for white noise are shown for cumulative IMF indexes in Figure 5.5(b).

To further illustrate the benefits of MMSE in multichannel scenarios where different channels contain different noise realizations, we next generated a trivariate time series, where originally all the data channels were realizations of mutually independent white noise. We then gradually decreased the number of variates that represent white noise (from 3 to 0) and simultaneously increased the number of data channels that represent independent 1/f noise (from 0 to 3), so that the total number of variates was always three. Figure 5.6(a) shows the standard coarse graining based MMSE curves and Figure 5.6(b) MEMD-enhanced MMSE curves for the cases considered; notice that as the number of variates representing 1/f noises increased, MSampEn at higher scales also increased, and when all the three data channels contained 1/f noise, the complexity at

---

5. The filtering operation equivalent to coarse graining is characterized by a very slow roll-off as well as large sidelobes which introduce aliasing artifacts [21]. The equivalent relationship between scale factor and cumulative IMF index given in the paper assumes a considerably faster roll-off as well as the absence of sidelobes.
larger scales was the highest. The analysis in Figure 5.6 confirms that, as desired, the more variables/channels within a multivariate time series exhibit long range correlations, the higher the overall complexity of the underlying multivariate system.

5.4 Applications for Real World Multivariate Processes

The MEMD-enhanced MMSE is next evaluated for two previously described multivariate real world recordings: human stride interval analysis and postural sway dynamics analysis. In this chapter, we have revisited those two examples and compared the MEMD-enhanced MMSE analysis results with Standard MMSE results. More real world applications are described in Chapter 6.

5.4.1 Case Study: Stride Interval Characterization

In order to reveal long-range correlations in stride interval dynamics, a signature that suggests cooperation within the different bodily subsystems at different time scales, stride intervals from human gait [80] data (introduced in the previous chapter) were analyzed. Three walking conditions (from the data available from [80]) were considered as different
variables from the same system, and MSampEn values were calculated for different scales (cumulative IMFs) generated using MEMD and in this way we were able to discriminate between the ‘self-paced’ and ‘metronomically-paced’ walk. Figure 5.7(a) shows the results obtained by the standard coarse-graining based MMSE method and Figure 5.7(b) for the proposed MEMD-enhanced method. Both methods found that self-paced ‘unconstrained’ walk has higher complexity, and thereby exhibits greater long-range correlations compared to constrained ‘metronomically-paced’ walk.

The statistical difference of the entropy statistics of self-paced and metronomically-paced sets were evaluated using the Student’s t-test and the Mann-Whitney U test. Both these tests revealed significant differences ($p < 0.01$) at all scales except the first two for standard coarse graining as well as MEMD-enhanced MMSE method. Observe that the first scale corresponds to the raw signal and MSampEn measures cannot discriminate between self-paced and metronomically-paced walk in either method. Moreover, as desired the separation between the MMSE curves of unconstrained and metronomically-paced walk was higher for the MEMD-enhanced method (Figure 5.7(b)), as indicated by much smaller error bars. Thus, using cumulative IMFs as data-adaptive scales offers a significant improvement over the coarse-graining based MMSE.
5.4.2 Case Study: Postural Sway Analysis

The data used in the simulations was the measured COP displacement from 12 young and 12 elderly, all healthy, volunteers at 60 Hz as described in the previous chapter; it was recorded simultaneously for the mediolateral (side-to-side) and anteroposterior (front-to-back) direction [82] to form a bivariate time series. Since the postural sway time series exhibits high frequency fluctuations superimposed on low frequency trends, the data need first to be detrended using methods such as wavelet transform and/or empirical mode decomposition [86]. It was shown in the literature [81] [90] and also in the previous chapter that only after detrending (removing last few IMFs), a valid complexity analysis could be performed using multivariate multiscale entropy method.

To illustrate this point, we first present the multivariate complexity analysis of the COP data by applying coarse graining-based MMSE without any detrending, as shown in Figure 5.8(a). It is evident from Figure 5.8(a) that the standard coarse graining-based MMSE, without detrending, was not able to discriminate between young and elderly subjects as the error bars overlapped. Moreover, the mean MMSE curve was lower for healthy young subjects than for their elderly counterparts, which is contrary to the intuition and the underlying physics\(^6\), and is attributed to the shortcomings of the scale generation method (coarse graining) of standard MMSE.

Next, Figure 5.8(b) shows the complexity analysis results obtained by applying MEMD-enhanced MMSE method taking only the first five IMFs. From the Figure, it is evident that MEMD-enhanced MMSE was able to discriminate between young and elderly subjects more effectively, as indicated by a better separation of their MMSE curves. Moreover, the complexity was higher for healthy young subjects than for their elderly counterparts which is physically intuitive, as for young subjects there exist correlation between mediolateral and anteroposterior components of COP time series which degrades in elderly subjects. Moreover, the difference in complexity between the young and elderly subjects was found to be statistically significant \((p < 0.01)\) over scales 1–4 using one-tailed

\(^6\)Due to ageing and the associated constraints, the complexity of postural sway for the elderly should be lower than for the young.
t-test with unequal variances.

This illustrates significant advantages of using MEMD-enhanced MMSE when assessing relative complexity of real-world multivariate data, and supports the more general concept of multiscale complexity loss with ageing and disease or when a system is under constraints, as those factors reduce the adaptive capacity of biological organization at all levels [16].

5.5 Conclusion

By combining adaptive scale estimation with multivariate entropy theory, a robust structural complexity descriptor for multivariate time series has been developed. Unlike standard techniques, the proposed algorithm is both suitable for nonstationary data and can measure complex coupled dynamics within the vector data channels, pre-requisites for the analysis of real-world systems which are typically of a multivariate, coupled and noisy nature. Simulations for biological systems illustrate conclusively how the approach can be used to reveal long-range spatio-temporal correlations present in their time series, signatures of the underlying complex signal generating mechanism.
Chapter 6

Applications of Multivariate Multiscale Complexity Analysis

In the previous chapter, a data adaptive multivariate framework for complexity analysis of real-word data has been introduced. This has been achieved by combining adaptive scale estimation through MEMD with multivariate sample entropy. In this chapter, the multivariate data adaptive complexity analysis method will be further validated on different real world scenarios.

6.1 Introduction

By introducing two critical algorithm enhancements - multivariate sample entropy and rigorous account of data adaptive scales, a robust framework for complexity analysis of multivariate time series has been developed in the previous chapter. The proposed method has been shown to alleviate the stationarity requirements of the current multiscale entropy methods by defining data-adaptive scales through multivariate empirical mode decomposition (MEMD) thus making full use of cross-channel information based upon multivariate sample entropy estimate and MEMD. As a result, the proposed methodology has the ability to produce robust and physically meaningful complexity estimates for real-world systems, which are typically multivariate, finite in duration, and of noisy and heteroge-
neous natures. In this chapter, the method has been validated on EEG signal for brain consciousness analysis, several case studies based on real-world wind data for analyzing underlying dynamical complexity of climatic variables, uterine EMG signal to characterize term and pre-term delivery records, and physiological signals for detecting signatures caused by increased cognitive load and stress.

6.2 Brain Consciousness Analysis

In this section, we evaluated standard as well as MEMD-enhanced multivariate multiscale entropy (MMSE) for the characterization of brain consciousness, particularly, the coma and quasi-brain-death state. The legal definition of brain death is an irreversible loss of forebrain and brainstem functions [100], however, brain death diagnosis procedures are complicated, and some tests require temporary disconnection from medical support. An initial prognosis of quasi-brain-death (QBD) is given based on various methods used for studying brain states using electroencephalogram (EEG) [101]. Studies have shown that large activity in the alpha band reflects the alertness of a patient [102], however, standard spectral analyses are unable to yield information of the brain’s inherent nonlinear complex dynamics [103], an important feature for brain states diagnosis. It is natural to assume that a brain in the states of coma and quasi-brain-death would have different degrees of complexity, and that the more stressed the system (QBD) the lower the complexity. As a result, methods from nonlinear dynamics theory such as MMSE are a natural choice in this context [104].

The EEG data were recorded in the intensive care unit in Hua Shan Hospital, Shanghai, China using a standardized 10-20 system. The measured voltage signal was digitized via a portable EEG recording instrument with a sampling frequency of 1000 Hz. The data was then bandpass filtered (FIR filter) to retain frequencies within the range 1 - 40 Hz and then downsampled by a factor of 10. Experimental data were obtained from 10 patients in coma, and 10 in the quasi-brain-death (QBD) state. For each of the patients, 50s segments are taken from the EEG signal. The values of the parameters used to calculate MSampEn are $N = 5000$, $m_k = 2$, $\tau_k = 1$ and $r = 0.15$ for every channel,
6.3 Complexity Analysis of Meteorological Data

In this section, we apply multiscale complexity analysis on several multivariate meteorological data sets obtained under different environmental conditions. Since the notion of
complexity of a system is closely related to long-range temporal correlations in a data set, we first quantified the amount of long-range correlations in our input data using Hurst exponent $H$, which has been historically employed as a measure of assessing long-term memory in a time series; please refer to Appendix B for the results obtained by applying several existing Hurst exponent estimators on the meteorological data sets used in this section. For wind speed time series, Hurst parameters and detrended fluctuation analysis (DFA) had been used effectively in the past to gauge its temporal long-range correlations [105] [106] [107]. However, while different methods are available in the literature for estimating $H$ from time series data, there is generally a lack of consistency among the estimates obtained from those methods, highlighting potential difficulties in applying Hurst estimators to real-world time series data.\(^1\)

The above mentioned difficulties with Hurst estimators mainly arise because these methods implicitly assume stationarity in the data. Hence, they are susceptible to misinterpretation for nonstationarities arising from slowly moving trends such as seasonal cycles [106] which manifest long-range correlation. In other estimators, the very nature of their construction or design can pose additional constraints; for instance, the R/S method for Hurst parameter estimation cannot be used to detect values of $H > 1$. Problems may also arise if a given process contains multiple scaling regions [106]. Moreover, the majority of Hurst estimators are applicable to univariate data, thus not catering for cross-correlations present in multivariate recordings. It is therefore an imperative to explore multivariate nonlinear measures that quantify long-range correlations at multiple scales in terms of structural complexity underlying the real-world data. To that cause, we employed multivariate multiscale entropy analysis (MMSE) to perform multivariate complexity analysis on several real-world multivariate wind speed and temperature data sets collected under different environmental conditions. The data sets included:

- wind speed and temperature data collected under different cloud-cover conditions;
- wind speed and temperature data collected at different heights at a single location;

\(^1\)The discrepancies associated with existing Hurst exponent estimators are also apparent in the results shown in Table B.1 in Appendix B.
• wind speed data corresponding to ‘high’, ‘medium’ and ‘low’ wind regimes which were classified based on their speed variations;

• wind speed data from one, two and three fan systems in an indoor controlled environment.

In this manuscript, we refer to them as Cloud-cover data, Mast data, Variance data and Fan data, respectively. Further detail of the data sets is provided in the respective subsections where the results of complexity analysis performed via original MMSE and MEMD-enhanced MMSE on those data is given.

For rigor, we also performed the complexity analysis on a set of multivariate surrogates generated from the above data sets via random shuffling; this provided a reference for a suitable comparison of complexity estimates obtained from different physical systems. Randomized shuffling of the input data channels effectively destroyed temporal and cross-channel correlations among their samples, while preserving their first and second order statistical properties. This way, significant difference between observed complexity estimates from input data sets and their respective (randomly shuffled) surrogates, over a range of scales, would reject the null hypothesis of both temporal and cross-channel independence, implying a higher complexity and nonlinear coupling in the considered data sets.

6.3.1 Wind and Air Temperature Data During Different Cloud Covering

The energy balance of the Earth’s climate is greatly influenced by clouds because of the cooling effect of albedo (as it reflects solar radiations) and the greenhouse warming effect (as it absorbs and re-emits terrestrial radiations back to the earth surface). This also depends on a number of factors, including the size of the droplets, the density of the clouds, their thickness, altitude and temperature, among others. The interaction of clouds with radiation thus alters the surface-atmosphere heating distribution, which in turn drives atmospheric motion that is responsible for the redistribution of clouds. Due to the complexity of the multiscale nature of cloud formation and cloud-radiation inter-
actions, its total affect on the climate system is still unclear and thus provides one of the major uncertainties in climate modeling and prediction [108].

To analyze the underlying complexity of different cloud coverage regimes, we examined air temperature, wind speed and direction data taken from the Iowa Environmental Mesonet (IEM), Iowa State University Department of Agronomy’s website\(^2\). The data contained two groups, one where more than 90% of the sky was covered with cloud (termed as overcast) and another where there was no clouds below 12000 feet (termed as clear). The wind speed, direction and air temperature data were all collected in one minute intervals from the Washington station of the Automated Weather Observing System (AWOS) stations network. Each group consisted of 20 trials with 3000 samples. We also generated random shuffled surrogates of the above data to provide a suitable reference for performing a comparison between complexity curves from different systems. The values of the parameters used to calculate MSampEn were \(m = 2\), \(\tau = 1\) and \(r = 0.20 \times \text{(standard deviation of the normalized time series)}\), for all the three data channels.

The data and the set of random surrogates were first analyzed using the standard multivariate multiscale entropy method and the results are shown in Figure 6.2(a). Next, the proposed MEMD-enhanced multivariate multiscale entropy method was applied, and

\(^2\)http://mesonet.agron.iastate.edu/request/awos/1min.php
Figure 6.2(b) shows the results. It can be noticed that both the methods clearly differentiated between the clear and overcast cloud coverage regimes and detected higher complexity in the overcast cloud coverage. This confirms that the formation of cloud increases the complexity of the underlying environmental system. This is intuitive, since the interaction between temperature and wind increases during cloud formation which is absent in clear sky conditions [108].

Moreover, it is evident that although surrogates showed greater complexity than the considered data at higher (lower indexed) scales ($\epsilon$) and cumulative IMF index ($n$); their complexity decreased at lower (higher indexed) scales implying an absence of significant ‘complex’ structures (correlations) at multiple scales. The Cloud-cover data set, on the other hand, exhibited higher complexity on all scales as evident by approximately flat and high MMSE curves for cloud-cover and clear-sky regimes, using both coarse graining and MEMD-enhanced MMSE methods.

6.3.2 Wind and Air Temperature Data at Different Altitudes

Wind speeds are generally lowest near the ground and increase with height up to several hundred feet above the ground, a phenomenon known as wind shear. There are also some particular heights at which higher wind speeds are common - these special areas of higher wind speeds are called jet streams. Wind near the earth’s surface, however, encounters obstacles that reduce its speed. The reduction in wind speed near the surface is a function of surface roughness: irregular, rough ground and man-made obstructions inhibit movement of the air near the surface, reducing wind velocity. Accordingly, wind speeds do not increase as much with height above sea level as they do on land due to the regular and smooth surface of water. For renewable energy, it is important to measure wind shear in order to accurately predict the performance of a wind power plant. Generally, the shear can be measured by monitoring wind speeds at two or three heights on a wind mast. Since wind turbines produce much more power in stronger winds, wind turbine designers try to put turbines on the tallest possible towers; at some point, however, the increased cost of towers outweighs the benefits.
6.3 Complexity Analysis of Meteorological Data

Multivariate complexity analysis of wind speed data at different heights above ground level is, therefore, relevant as roughness of terrain which reduces wind speeds at lower heights can be seen as a constraint on a system. On the other hand, effects of terrain roughness decrease with increasing heights, and the height above ground where surface friction has a negligible effect on wind speed is called the ‘gradient height’; the wind speed above this height is assumed to be a constant. Hence, wind speeds at greater heights above ground level could be considered as being generated from a relatively unconstrained system (due to negligible wind shear effect). According to complexity science, unconstrained systems are generally more complex than their constrained counterparts [44]; it is therefore expected that MMSE analysis of wind data close to the surface will yield lower complexity as compared to the data collected at a greater height.

To verify the above assumption, wind profile data was collected from a single location but from two different heights and was subsequently analyzed using the MMSE method. The data set contained 10 minutes averaged wind speed and temperature data collected for 12 months in the Jhimpur area in Pakistan. For data collection, a wind mast was erected 81.5m above ground level on which optically scanned Thies First class cup anemometers were mounted to collect the wind speed data; temperature data was also collected via thermohygro sensors. Though the above data was collected at multiple

Figure 6.3: Multivariate multiscale entropy (MMSE) analysis of 2D wind data using: (a) coarse graining based standard multivariate MSE, and (b) MEMD-enhanced multivariate MSE. The curves represent an average of 10 trials and error bars standard error (SE).
heights (80m, 60m, 30m and 10m) on the mast, we combined only the speed and temperature data at 80m and 10m to form a bivariate model. To generate error bars on MMSE complexity curves, for rigorous analysis, we divided 12-months bivariate data sets into 12 segments, with each segment corresponding to a 1-month interval (with data length of \( N = 4320 \)). Moreover, for each data segment, the corresponding random surrogate series were also produced for comparative complexity analysis. The parameters used to calculate MSampEn were \( m = 2, \tau = 1 \) and \( r = 0.20 \times \) (standard deviation of the normalized time series).

Figure 6.3(a) and Figure 6.3(b) show the results obtained by applying standard MMSE and MEMD-enhanced MMSE methods, to the input data, respectively. Observe that the complexity curve corresponding to the wind speed at 80m above the ground level shows higher complexity than the one corresponding to 10m height which agrees with our assessment. Furthermore, complexity curves are significantly separated from each other implying a major difference in the complexity of the two systems. The complexity estimates of surrogate series follow a similar pattern observed in the previous cases: though surrogates showed higher entropy values than that of the original series for higher (low-indexed) scales, their entropy dropped at lower scales. From these results, it can be concluded that the degree of disorder or irregularity is higher at greater heights above the ground level. Moreover, it can be concluded that the correlations between wind speed and temperature are more pronounced at higher altitudes, verifying a well-known height-dependent behavioral trend in wind speed and air temperature data.

### 6.3.3 Wind and Air Temperature Data from Various Wind Dynamics Regimes

We shall now show that the MMSE method allows us to characterize different wind regimes based on the variance of their speeds. For this purpose, we identified and analyzed ‘high’, ‘medium’ and ‘low’ wind regimes respectively, corresponding to high, medium and low variance of their corresponding speeds. The data set used in our simulation was recorded using a 3D ultrasonic anemometer (measurements taken in the north-south, east-west
and vertical direction) at a sampling frequency of 50Hz in the courtyard of Institute of Industrial Science (IIS) of the University of Tokyo. Figure 6.4 shows a plot of the magnitude of wind speed recordings used in the simulation; it can be noticed that the wind dynamics was changing with time allowing us to define three wind regimes, labeled as ‘low’, ‘medium’, and ‘high’. Besides wind speed data, the corresponding air temperature was also measured as it indirectly affects the wind speed; namely temperature affects the atmospheric pressure which in turn affects the wind speed. To generate the mean and variance (error bars) of the calculated sample entropy estimates, we divided the data set into 18 segments, with 6 segments each containing high, medium and low dynamics wind speed data. To reduce the effects of high frequency noise, the data was preprocessed by a moving average filter. The values of the parameters used to calculate MSampEn were $m = 2$, $\tau = 1$ and $r = 0.20 \times$ (standard deviation of the normalized time series), for each of the four data channels. We also performed the complexity analysis of corresponding random-shuffled surrogate time series, generated for each 18 data segments, to provide reference complexity curves for a suitable comparison with the original data.

Figure 6.5(a) shows the standard multivariate multiscale entropy analysis, per-
formed by considering the three wind directions as variables in a trivariate model whereas Figure 6.6(a) shows the result while the air temperature is also included as another variable in a quadrivariate model. Observe that the multivariate approach was capable of detecting long-range correlations in the wind speed for all the wind regimes as the MMSE curve was similar to that of 1/f noise (cf. Figure 5.5(a)), conforming with the existing results [105] [106] [107]. Both Figure 6.5(a) and Figure 6.6(a) show that, as desired, the medium dynamics regime had higher complexity than either high or low dynamics regime. This is also intuitively clear, as medium wind dynamics has fewest constraints, and is thus most complex as mild winds come from a wide range of directions [96] [107]. Also observe that, as the number of variables increased (when temperature is included in Figure 6.6(a)), the MMSE curves for the low- and high- dynamics regions became separated. Surrogate series, as expected, showed higher MMSE values than that of input data only at highest temporal scales (lower index scales), their complexity decreasing with increasing scale indexes; this behavior is similar to that of random uncorrelated multivariate noise (see Figure 5.5).

Figure 6.5(b) and Figure 6.6(b) show the corresponding MEMD-enhanced multivariate multiscale entropy analyses. Observe that MEMD-enhanced MMSE for the quadrivariate model not only showed a comparatively higher complexity in the medium wind
6.3 Complexity Analysis of Meteorological Data

Figure 6.6: Multivariate multiscale entropy (MMSE) analysis of 4D wind data (3D wind and temperature) using: (a) coarse graining based standard multivariate MSE, and (b) MEMD-enhanced multivariate MSE. The curves represent an average of 6 trials and error bars the standard error (SE).

...dynamics regime, but was also able to differentiate among the low, medium and high dynamics regimes at higher indexed scales, as the error bars did not overlap, hence, a clear improvement over the standard MMSE method. Moreover, as we can consider the wind with medium dynamics as the least constrained system, as opposed to the high or low dynamics regimes which are constrained [96], this interpretation is also in agreement with the general complexity loss theory with constraints [16].

6.3.4 Wind Data from Different Fan Systems

We next performed complexity analysis of air speed data recorded from a simple indoor setting consisting of three table fans. Three different sets of air speed data were collected depending on the number of fans used. In the first data set, a single circular moving table fan was placed around 3 meters away from a 3D ultrasonic Gill Instruments anemometer which recorded 3D wind speed data. For two other data sets, we used two and three table fans respectively with the 3D anemometer kept at a distance of around 3 meters from the fans. The speed of three fans were kept at different levels. The sampling frequency of the collected data was 50 Hz. The values of the parameters used to calculate MSampEn in this simulation were \( m = 2, \tau = 1 \) and \( r = 0.15 \times \) (standard deviation of the normalized time series), for each three data channels. Similarly to the previous examples, surrogates...
were also generated and subsequently their complexity results were compared with those obtained from the original data.

The standard MMSE results for the above data sets are shown in Figure 6.7(a). Observe that the three-fan system had the highest complexity followed by the two-fan system, while the one-fan system was found to have the lowest complexity. The results of MEMD-enhanced MMSE, for the first seven IMFs, showed a similar behavior with the three-fan system showing the highest complexity, as illustrated in Figure 6.7(b). The complexity of the system increased with the number of added fans since by including each fan we effectively added another scale in the air speed data: the blades of the three fans were moving at different speeds. Though the three fans were moving independently of each other with no cross-channel correlation among speeds from different data sets, the multi-scale nature of two and three fans systems made them more complex as compared to the single fan system. Notice that the MEMD-enhanced method was able to differentiate among the three systems more robustly and thus outperformed the standard MMSE method. This time, however, we can notice that the complexity of the Fan data for all the three systems is significantly higher than its corresponding surrogates for only very few scales. This was expected, since there was very little long-range temporal correlation present among data, due to the constrained nature of the indoor experiment.
6.3 Complexity Analysis of Meteorological Data

6.3.5 Discussion and scope of the work

The results presented in the previous sections suggest the presence of long range (auto- ) and cross-channel correlations in the chosen meteorological data as MMSE curves for those data were relatively flat; note that relatively higher and constant MMSE curves imply similar complexity in multiple scales of the data, a characteristic common to data sets containing long range correlations such as $1/f$ (fractal) noise time-series. The flat MMSE curves were significantly more prominent for data from the real world outdoor mast, variance, and cloud cover data sets, as compared to synthetic indoor fan data; this is due to the constrained indoor environment. It was also observed that the analyzed data sets collected under different environmental conditions were easily distinguished based on their complexity curves, suggesting a direct influence of environmental conditions, e.g cloud covering, on meteorological data characteristics.

This study provides the conclusive evidence that both within and across channel correlations contribute to the multiscale complexity of the wind signal and thus has great potential in helping to accurately modeling wind speed for short-term wind power forecasting\(^3\) which is extremely important for wind turbine operation and efficient energy harvesting. In the literature, different techniques have been used to forecast wind speed: a) methods employing numerical weather prediction models which incorporate physical properties of environment such as pressure, terrain etc; b) statistical methods using autoregressive integrated moving average (ARIMA) based models [109], least mean squares (LMS) filters and their respective multivariate extensions e.g. vector autoregressive moving average (VARMA) models [110] and quaternion LMS [111]; c) methods using artificial neural networks [112]; d) spatio-temporal methods integrating information of wind speed at neighboring sites [113].

The methods and results presented in this work could be directly integrated into the statistical methods employed for wind forecasting. For instance, higher and constant complexity (MMSE) values at multiple scales of wind data would suggest the use of multi-scale

\(^3\) Accurate long term forecasting would require more complex and computationally expensive numerical weather prediction (NWP) models employing several environmental variables in a single complex model.
statistical ARIMA, VARMA and quaternion LMS models for wind speed data forecasting, that is, applying statistical models to each scale separately and then performing prediction. For data sets exhibiting different complexity at multiple scales (varying complexity of MMSE curves), different model parameters could be tuned for each scale depending on their complexity values, for improved speed forecasting. MEMD based MMSE could be of great significance in this multiscale framework as MEMD inherently generates quasi-stationary scales (components) which could qualify for the stationarity requirements of these statistical methods.

Moreover, the generic multivariate nature of the algorithms and analysis presented in the work makes it possible to be easily combined with existing spatio-temporal methods for wind forecasting. For this cause, wind speed information from neighboring sites could be combined with the original speed data to create a multivariate signal. Multivariate complexity analysis on the resulting signal would then exploit the speed information from neighboring sites and is expected to yield better forecasting estimates.

### 6.4 Complexity Analysis of Eye Gaze Dynamics

Psychologist Alfred L. Yarbus famously illustrated the impact of cognitive load on scanning eye patterns by presenting subjects with an image (see Figure 6.8(a)) and recording gaze trajectories in response to different instructions [114]. To re-investigate this classic study from a completely novel perspective, we set out to examine whether cognitive load is reflected in the complexity of the gaze dynamics. Seven healthy, naive subjects were asked to both examine the image in Figure 6.8(a) freely and to complete six different instructions over 100 s trials (see [114] for more details), while bivariate (vertical and horizontal) eye gaze was recorded (a segment of which is shown in Figure 6.9). The whole experiment was conducted locally in the Smart Environment Lab (SEL) of EEE department of Imperial College London during PhD study of the author. The values of the parameters used to calculate MSampEn were $m = 2$, $\tau = 1$ and $r = 0.15 \times$ (standard deviation of the normalized time series), for each two data channels.
6.4 Complexity Analysis of Eye Gaze Dynamics

Figure 6.8: The classic ‘Yarbus experiment’: (a) The presented image, and (b) Gaze intensity map for the instruction relating to the ages of the people.

Figure 6.9: Segment of raw gaze data, both horizontal and vertical components.
Figure 6.10(a) shows the average gaze complexity curves over all subjects, for both constrained (only two instructions is shown for clarity) and free examination computed by standard MMSE method, and Figure 6.10(b) shows the same computed by MEMD-enhanced MMSE method. Both figures illustrate that the cognitive instructions can be uniquely identified in the gaze complexity space. Compared to all instruction trials, the gaze complexity curve of free examination computed by standard MMSE method was the highest over high scale factors (>10) as shown in Figure 6.10(a). On the other hand, in Figure 6.10(b), the MEMD-enhanced MMSE curve for free examination was highest at all scales. This supports the general ‘complexity-loss’ theory, that is, the less constrained the cognitive task the higher the complexity. Moreover, the MEMD-enhanced method can clearly distinguishes the two other curves generated from two constrained instructions, thus illustrating the advantage of using MEMD-enhanced method.

![Figure 6.10: MMSE analysis of the classic ‘Yarbus experiment’ illustrating that induced cognitive load reduces the gaze complexity: (a) coarse graining based standard multivariate MSE, and (b) MEMD-enhanced multivariate MSE. The curves represent average complexity results for ‘free examination’ and instructions (i) estimate the material circumstances of the family and (ii) remember the clothes worn by the people, where the error bars denote the standard error (SE).](image-url)
6.5 Complexity Analysis of Heart and Respiratory Function During Stress

Stress-induced illnesses are a major concern in modern mankind, the American Institute of Stress\(^4\) estimates that 75-90% of all visits to primary care physicians are for stress related problems. Stress is manifested by changes in several psycho-physiological modalities [115] [116] - a perfect match for the multivariate nature of the MMSE method.

Three naive, healthy subjects participated in a study (two parts each lasting 20 mins), in which respiration waveforms and electrocardiography (ECG) were recorded while the subject was seated comfortably and instructed not to talk or move unnecessarily. The baseline physiological response was established ('normal state') by engaging the subject in a relaxing task - watching a movie. Next, the subject was presented with a series of demanding mathematical logic questions and was instructed to respond via a keypad as quickly and accurately as possible ('stressed state'). An increased level of background noise and verbal interference from the experiment coordinator were used to increase the level of subject engagement. The whole experiment was conducted locally in the Smart Environment Lab (SEL) of EEE department of Imperial College London during PhD study of the author.

For each subject and sub-experiment, the data was divided into 100 s segments and at least 5 artifact-free segments were extracted and analyzed. The ECG was bandpass filtered to occupy the frequency range (0.5 - 20) Hz and the respiration data was bandpass filtered to occupy the frequency range (0.05 - 3) Hz. All data was recorded at 1200 Hz and downsampled to 120 Hz. The values of the parameters used to calculate MSampEn were \(m = 2, \tau = 1\) and \(r = 0.15 \times \text{(standard deviation of the normalized time series)}\), for each two data channels.

Figure 6.11(a) shows the average complexity results obtained for the bivariate data [ECG, respiration] for both the 'normal' and 'stressed' states using standard MMSE method and Figure 6.11(b) shows the results found using MEMD-enhanced MMSE\(^4\) http://www.stress.org/americas-1-health-problem/
6.6 Complexity Analysis of Electrohysterogram (EHG) Records

method. Both the MMSE approach can clearly able to separate the two states in the complexity-space.

![Graph](image)

(a) Multivariate MSE  
(b) MEMD-enhanced Multivariate MSE

Figure 6.11: MMSE analysis for ‘normal’ and ‘stressed’ states, based on heart and respiratory functions: (a) coarse graining based standard multivariate MSE, and (b) MEMD-enhanced multivariate MSE. The curves represent average complexity results, where the error bars denote the standard error (SE).

6.6 Complexity Analysis of Electrohysterogram (EHG) Records

Finally, multivariate multiscale entropy analysis is applied on the uterine EMG records to differentiate the degree of complexity underlying the term and pre-term deliveries and to bring new insights into the physiology of parturition. The Electrohysterogram records (uterine EMG records) used in this study are included in the Term-Preterm Electrohysterogram Database (TPEHG DB) and publicly available from PhysioNet [83]. The records were obtained during regular check-ups either around the 22nd week of gestation or around the 32nd week of gestation. The database contains 300 uterine EMG records from 300 pregnancies (one record per pregnancy) of which:

- 262 records were obtained during pregnancies where delivery was on term (duration of gestation at delivery > 37 weeks):
  - 143 records were obtained before the 26th week of gestation and
Figure 6.12: Coarse graining based MMSE analysis of filtered 3-channel Uterine EMG (UEMG) signal in order to bring new insights into the physiology of parturition. The curves represent average complexity results underlying the term and pre-term deliveries recorded either in early or late pregnancies, where the error bars denote the standard error (SE).

- 119 were obtained later during pregnancy, during or after the 26th week of gestation;
- 38 records were obtained during pregnancies which ended prematurely (pregnancy duration ≤ 37 weeks), of which:
  - 19 records were obtained before the 26th week of gestation and
  - 19 records were obtained during or after the 26th week of gestation.

Each record is composed of three channels, recorded from 4 electrodes, sampled at 20 Hz and 30 minutes in duration. Besides, each signal was digitally filtered using 3
6.6 Complexity Analysis of Electrohysterogram (EHG) Records

different 4-pole digital Butterworth filters with a double-pass filtering scheme to ensure zero phase shift. A detail discussion about the database is given in Reference [117]. We have chosen two different m values, that is m=3 and m=2; and found that m=3 yields slightly better separation result than m=2 for each cases. The result also shows better separation when the band-pass filter cut-off frequencies were from 0.08Hz to 4Hz. As a result, only the results for m=3 and of cut-off frequency 0.08Hz-4Hz are reported here. In all the cases, r is taken as 0.15 times the total variation of the 3-channel UEMG signal.

Figure 6.13: MEMD-enhanced MMSE analysis of filtered 3-channel Uterine EMG (UEMG) signal in order to bring new insights into the physiology of parturition. The curves represent average complexity results underlying the term and pre-term deliveries recorded either in early or late pregnancies, where the error bars denote the standard error (SE).

There are significant differences between the UEMG signals recorded early and late (row1-column1 panel, row1-column2 panel and row3-column1 panel of Figure 6.12). This
means as the time of gestation progresses, the average multivariate sample entropy values for both term and pre-term delivery records drop indicating higher predictability or less complexity of the signals as the delivery approaches. On the other hand, the MSampEn values are lower for pre-term delivery records (row2-column1 panel, row2-column2 panel and row3-column2 panel of Figure 6.12) regardless of the gestation duration at the time of recording which confirms that the pre-term delivery records are less complex or more predictable than the signals of term delivery records. Besides, in all the cases, the separation is better if we consider the multiscale MMSE curves than the measures in scale 1. This also confirms that the original signal not only contains information in the smallest scale but also reveals new information at all scales.

Figure 6.13 reports the MEMD-enhanced MMSE results. From the figure, it can be concluded that the previous interpretation still holds for MEMD-enhanced MMSE curves though the separation between cases are not much improved for MEMD-enhanced method.

6.7 Conclusion

The recently introduced multivariate multiscale entropy (MMSE) method with data-driven scales has been illuminated as an enabling tool for the complexity analysis of real-world multivariate data. It has been shown to model the dynamical couplings between physiological variables, giving an insight into the underlying system complexity, a feature not achievable using standard univariate measures. Simulations on different multivariate real world case studies have also shown the effectiveness of the proposed approach in revealing long-range spatio-temporal within- and cross-channel correlations.
Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this thesis, two critical algorithm enhancement of the recently introduced multiscale entropy (MSE) have been proposed:

1. Multivariate multiscale complexity;
2. Data-driven temporal scales and the ability to operate on non-stationary data.

These enhancements are crucial for following reasons:

Firstly, the recent advances in vector sensor and data acquisition systems have made it possible to routinely measure multivariate real world data from experimental, biological and meteorological systems. This in turn, has highlighted the need for addressing the complexity of such data directly in the domain where they reside. Assessing the complexity of such multivariate systems by measuring the complexity of channel-wise data would be suboptimal and biased as that would not consider the inter-channel correlations within the multivariate data. It is also natural to associate complexity of a multivariate system not only to long range temporal (auto-)correlations within each channel but also to the inter-dependence (or cross-correlation) among multiple channels. As complexity of a system is considered by its ‘emergent’ behavior, so it is intuitive that a meaningful complexity measure should use a holistic approach rather than a reductionist approach.
Secondly, the MSE is not a perfect match for processing real world non-stationary data due its deterministic way of generating scales via coarse graining of input data. More critically, it reduces the input data length by the scale factor for each successive data scale, thereby, imposing a limit on the length of input data which can be effectively processed via MSE. Besides, MSE fails to cater for signals containing one or more pronounced trends; in such cases little can be inferred from entropy estimates as the trends tend to dominate other interesting features. From the statistical perspective, it is therefore imperative that any trends be removed before meaningful interpretation can be made from sample entropy values. All this above mention issues highlight the need for some nonstationary data-driven techniques to be used instead of coarse-graining for generating multiple intrinsic scales of the data.

Finally, the multiple scales generated from the multiple channels should not only be data-adaptive, but also they need to be same in number and similar in terms of spectral properties (belonging to same frequency bands). As complexity is a relative measure, this is critical for a physically meaningful comparison.

To successfully resolve all the above mentioned issues, a robust framework for the complexity analysis of multivariate signal has been proposed in this thesis. More specifically, the proposed framework for multivariate data analysis generates data driven scales from multivariate extension of EMD (MEMD) which are subsequently analyzed by multivariate sample entropy (MSampEn) estimate. Owing to its mode alignment property, MEMD generates comparable scales from multiple data channels and the so produced scales are of same length as the length of input signal, thus, removing the limitation on input data size in the original MMSE method. The resulting method is also suitable for non-stationary multivariate data analysis owing to the data-driven nature of MEMD algorithm as opposed to deterministic coarse graining process used in MMSE.

In this thesis, a brief overview of the notion of complexity as well as the state-of-the-art complexity measures were first reported in Chapter 2. This was followed by the detail description of multiscale entropy (MSE) method in Chapter 3, as its extension to multivariate case was the main focus of the thesis. The multiscale framework was
also tested with two popular entropy estimates, namely, PermEn and LZEn and the so introduced multiscale variants were calibrated with white and 1/$f$ noise afterwards. In addition, all the methods were applied for EEG characterization and classification. Finally, it had been established that multiscale entropy with sample entropy as the complexity estimator was superior than other two measures in terms of classification accuracy. This also motivated us to extend sample entropy statistics for multivariate case.

In Chapter 4, an extension of SampEn for multivariate data was first presented. This was achieved based on multivariate embedding. The proposed extension was not straightforward as it dealt with the different embedding dimensions, time lags and amplitude ranges of data channels in a rigorous and unified way. Based on MSampEn, the MMSE method was next proposed which catered for linear and/or nonlinear within- and cross-channel correlations as well as complex dynamical couplings and various degrees of synchronization present among multiple channels and over multiple scales, in contrast to analyzing each channel separately as performed by MSE. The method was also applied to analyze stride interval dynamics, postural sway dynamics and cardiorespiratory dynamics and could discriminate between normal and constrained (either through age, disease or stress) system which was in agreement with the ‘complexity loss’ hypothesis.

Next, the so-called coarse-graining method was replaced with the recently introduced multivariate time-frequency analysis method, known as, multivariate empirical mode decomposition (MEMD) in Chapter 5. As a result, the so generated scales were data adaptive and fully suited to the dynamics of the signal in hand, unlike the currently used coarse graining techniques. This was also validated on real world signals such as stride interval time series and postural sway series in quiet standing and it was shown that the MEMD-enhanced MMSE method better discriminated the constrained vs unconstrained state than the coarse-graining based MMSE method.

Finally, in Chapter 6, both the coarse-graining based MMSE and MEMD-enhanced MMSE methods were applied in different application areas and their results were compared and contrasted. The study included characterization of brain consciousness between coma and quasi-brain-death (QBD) patients, analysis of underlying dynamical complexity of
climatic variables on several case studies based on real-world wind and air temperature data, analysis of eye gaze dynamics, analysis of cardiorespiratory function during stress and analysis of electrohysterogram (EHG) records from the term and pre-term deliveries. All the reported results strongly supported the general ‘complexity loss’ hypothesis for constrained systems.

7.2 Future work

Following the studies in this thesis, some avenues for future work are next highlighted.

First, as mentioned earlier, MSE can quantify the fractal breakdown in terms of reduced complexity in pathological states. As a result, it has become very popular in the biomedical community. Since MMSE is the extension of MSE to the multivariate case and captures all the necessary within- and cross-channel information for quantifying the underlying complexity, it has greater potential as a complementary tool for medical diagnosis and prognosis. Future endeavor should explore different application areas to make full use of its multichannel capability.

Second, there are several parameters to be chosen in MSampEn calculation, each introducing their own constraints. For instance, for multi-modal data coming from heterogeneous data channels, the individual channels are likely to exhibit different embedding parameters $m$ and $\tau$. There are several methods [77] for determining the optimal embedding parameters $m$ and $\tau$ simultaneously, for a single channel. Future developments should optimize for all the $m$ and $\tau$ parameters from the different data channels simultaneously, but are not yet available. Besides, the threshold parameter, $r$, is currently chosen in an empirical basis, calling for a more systematic approach for choosing its optimal values.

Third, quite recently, motivated by the same spirit, Morabito et al. [118] have extended permutation entropy to the multivariate cases. Their method, known as multivariate multiscale permutation entropy (MMPE), has been applied for complexity analysis of EEG from Alzheimers Disease. Although, in Chapter 3 we showed that sample entropy is superior to permutation entropy in a multiscale framework, it would be interesting to
Future work

compare their multivariate extensions in this regard.

Fourth, the proposed MSampEn is not rotation invariant. Therefore, its absolute value varies depending on the ordering of the channels in a multi-channel signal. As MSampEn has been used as a relative entropy measure in multiscale framework, so the comparative results drawn from the multivariate signals are consistent. In other words, if complexity of a multivariate signal A is greater than B for some channel ordering, the same conclusion still holds if we change the initial position of the channels in the multi-channel signal though their absolute entropy value might be different. Therefore, if one wants to use it as an absolute complexity measure, it should be made rotation invariant. One approach to solve this issue might be to take all the permutations of the channels first and then calculate MSampEn for each permutations and afterwards take the average of all. But this is computationally intensive. In future, research should be sought for other alternatives to make MSampEn rotation invariant.

Fifth, in machine learning the MMSE can be used as a feature extraction method. It is very difficult to significantly extract representative feature based on single time or frequency scale for some signal. For example, physiologic signal possesses long-range correlation on multiple spatio-temporal scales which can not be accounted by traditional entropy measures. So the traditional single scale entropy metrics will not provide a discriminative feature in this scenario. As a result, feature extraction based on multi-scale framework like the MMSE method has great potential in machine learning.

Sixth, as shown in chapter 4, the MMSE method can differentiate between correlated and uncorrelated noises. This can be used to detect non-circularity in complex and quaternion signals specifically if the non-circularity arises due to the correlation between the real and imaginary parts of a complex or quaternion signal.

Finally, the potential for using MSampEn independently in event monitoring or change point detection for multichannel signal can be further explored. To that cause, sliding/moving window method can be used in which MSampEn is first calculated in a pre-defined window of the signal, then MSampEn of the subsequent window of the signal is computed and so forth. Thus, the time profile of the multivariate entropy measure
can be rendered for all the existing channels or different channel combinations. This has potential applications in areas such as evaluation of depth of anesthesia, seizure prediction for epilepsy, online patient monitoring system etc.
Bibliography


Appendix A

MSampEn Calibration with Correlated and Uncorrelated Noises

MSampEn yields higher complexity for correlated multichannel signals than for uncorrelated signals. Figure A.1 shows the geometric interpretation for uncorrelated bivariate white noise and Figure A.2 for correlated bivariate white noise. In Figure A.1(b), we see that there is no correlation between the channels of uncorrelated white noise as indicated by the circular shape of the scatter plot, whereas in Figure A.2(b), the correlation between the channels of a correlated bivariate noise is seen in the elliptical shape of the scatter plot. In Figure A.1(e), we see no perceived structure in the 3-dimensional space and as a result both the quantity $B^m(r)$ and $B^{m+1}(r)$ are very similar. In other words, the probability of finding any two composite delay vectors which are similar within a tolerance level $r$ in the 2-dimensional space and in the 3-dimensional space are higher and of the same order. As a result, MSampEn is lower. For correlated bivariate noise (Figure A.2), some structures (non-circular shape) are seen both in the 2-D space (Figure A.2(b)) and in the 3-D space (Figure A.2(c), A.2(d), A.2(e)). Moreover, this time the quantity $B^{m+1}(r)$ is much smaller than the quantity $B^m(r)$. In other words, the probability of finding any two composite delay vectors which are similar within a tolerance level $r$ in 2-dimensional
space is much higher than in the 3-dimensional space. As a result, MSampEn estimate is relatively higher.

Figure A.1: Geometric interpretation of MSampEn calculation from uncorrelated bivariate white noise
A. MSampEn Calibration with Correlated and Uncorrelated Noises

Figure A.2: Geometric interpretation of MSampEn calculation from correlated bivariate white noise
Appendix B

Hurst Parameter

Hurst exponent, $H$, is a classical parameter that characterizes long-range correlations in a time series. A value $H$ in the range $0.5 < H < 1$ indicates a time series with long-range positive autocorrelation whereas its value within the range $0 < H < 0.5$ indicates a time series with long-term negative correlation (switching between high and low values in adjacent pairs). A value of $H = 0.5$ can indicate uncorrelated data, but in practice it is the value applicable to series for which the autocorrelations decay exponentially quickly to zero. This is in contrast to a typically power law decay for the cases corresponding to $H \neq 0.5$ (positive or negative long-range dependencies).

As the value of $H$ is dependent on long range correlations in the data, we can give a relation between an autocorrelation sequence $r_H[k]$ of a time series $x_H[k]$ and its corresponding $H$ value. Let’s consider a specific case of a fractional Gaussian noise (fGn) which can be seen as a special case of a fractional Brownian motion (fBm). More specifically, \{x_H(k), k = \cdots - 1, 0, 1, \ldots \} is a fGn of index $H$ (with $0 < H < 1$) if and only if it is a zero-mean Gaussian stationary process with autocorrelation sequence, which is given by

$$r_H(k) = \frac{\sigma^2}{2} (|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}). \quad (B.1)$$

It is well understood that the sequence $x_H(k)$ for $H = 0.5$ corresponds to white noise, whereas, it exhibits negative correlations for $0 < H < 0.5$ and positive correlations
Table B.1: Summary of Hurst exponents estimated for all input data sets using standard methods

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Higuchi</th>
<th>R/S</th>
<th>MP</th>
<th>DSOD</th>
<th>Wavelet DSOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>1.0002</td>
<td>0.9737</td>
<td>1.3234</td>
<td>0.3551</td>
<td>0.3571</td>
</tr>
<tr>
<td>Medium</td>
<td>0.9991</td>
<td>0.9700</td>
<td>1.3706</td>
<td>0.3374</td>
<td>0.3376</td>
</tr>
<tr>
<td>Low</td>
<td>1.0002</td>
<td>1.0602</td>
<td>1.3999</td>
<td>0.2359</td>
<td>0.2361</td>
</tr>
<tr>
<td>Cloud cover data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clear sky</td>
<td>0.9997</td>
<td>1.0328</td>
<td>1.3140</td>
<td>0.6119</td>
<td>0.5880</td>
</tr>
<tr>
<td>Overcast sky</td>
<td>1.0005</td>
<td>1.0748</td>
<td>1.2598</td>
<td>0.6671</td>
<td>0.6334</td>
</tr>
<tr>
<td>Mast data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80m</td>
<td>1.0015</td>
<td>0.8677</td>
<td>1.2033</td>
<td>0.5645</td>
<td>0.5631</td>
</tr>
<tr>
<td>10m</td>
<td>1.0012</td>
<td>0.8624</td>
<td>1.1953</td>
<td>0.4902</td>
<td>0.4857</td>
</tr>
<tr>
<td>Fan data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-fan</td>
<td>0.9962</td>
<td>1.0456</td>
<td>1.3886</td>
<td>0.4631</td>
<td>0.4685</td>
</tr>
<tr>
<td>Two-fan</td>
<td>1.0020</td>
<td>0.8968</td>
<td>1.1252</td>
<td>0.4231</td>
<td>0.4377</td>
</tr>
<tr>
<td>Three-fan</td>
<td>0.9994</td>
<td>0.8771</td>
<td>0.9545</td>
<td>0.3892</td>
<td>0.3922</td>
</tr>
</tbody>
</table>

for 0.5 < H < 1. By taking the Fourier transform of equation (B.1), the power spectral density (PSD) of fGn is obtained, which can be written as

\[ S_H(f) = K\sigma^2 |e^{2\pi if} - 1|^2 \sum_{k=-\infty}^{\infty} \frac{1}{|f + k|^{2H+1}}. \] (B.2)

with \( f \leq 0.5 \). In the limiting case of \( f \to 0 \), and \( H \neq 0.5 \), the PSD can be written as

\[ S_H(f) \approx K\sigma^2 |f|^{1-2H}, \]

showing that fGn can be used as a model of power law spectra at low frequencies. In addition, for \( 0 < H < 0.5 \) (short-range correlations), we have \( S_H(0) = 0 \), and the spectrum is effectively high-pass, while for \( 0.5 < H < 1 \) (long-range correlations), we have \( S_H(0) = \infty \). In both cases, the power law form of the spectrum is approximately held and, in log-log coordinates, we have a quasi-linear relation which is given by

\[ \log S_H(f) = (1 - 2H) \log f + C. \] (B.3)

There are several methods for the estimation of the Hurst exponent; these include Higuchi method, Re-scaled Range (R/S) analysis, Modified Periodogram (MP) method, Discrete Second Derivative (DSOD) estimator, and Wavelet based DSOD method [119]. Table B.1 shows the Hurst exponents estimated by the above methods for all four data sets used in Section 3 of Chapter 6 of this thesis. The Hurst exponents are estimated from the modulus of wind speed time series in each case. Observe from Table 3.1 that, for
a given method the Hurst parameter estimation was consistent across the conditions of four different data sets. However, there was no consistency in the estimation of the Hurst index across the different methods: while the Higuchi, R/S and MP methods yielded Hurst exponent estimates in the range of 0.9 – 1.4, the DSOD estimator and its Wavelet based variant produced exponents in the range of 0.2 – 0.7. Such discrepancy in the results indicate potential difficulties in applying Hurst exponent estimators to real world time series data.