THE UNIVERSITY
of EDINBURGH

Entropy-based Nonlinear Analysis for Electrophysiological Recordings of Brain Activity in Alzheimer's Disease

Hamed Azami

A thesis submitted for the degree of

Doctor of Philosophy

The University of Edinburgh

February 2018

THE UNIVERSITY *of* EDINBURGH
School of Engineering

Dedication

To my dear mother and father, ***Fouzieh*** *and* ***Abbas***

for their love, support, and encouragement.

Hamed Azami

February 2018

Abstract

Alzheimer's disease (AD) is a neurodegenerative disorder in which the death of brain cells causes memory loss and cognitive decline. As AD progresses, changes in the electrophysiological brain activity take place. Such changes can be recorded by the electroencephalography (EEG) and magnetoencephalography (MEG) techniques. These are the only two neurophysiologic approaches able to directly measure the activity of the brain cortex. Since EEGs and MEGs are considered as the outputs of a nonlinear system (i.e., brain), there has been an interest in nonlinear methods for the analysis of EEGs and MEGs.

One of the most powerful nonlinear metrics used to assess the dynamical characteristics of signals is that of entropy. The aim of this thesis is to develop entropy-based approaches for characterization of EEGs and MEGs paying close attention to AD. Recent developments in the field of entropy for the characterization of physiological signals have tried: 1) to improve the stability and reliability of entropy-based results for short and long signals; and 2) to extend the univariate entropy methods to their multivariate cases to be able to reveal the patterns across channels.

To enhance the stability of entropy-based values for short univariate signals, refined composite multiscale fuzzy entropy (MFE - RCMFE) is developed. To decrease the running time and increase the stability of the existing multivariate MFE (mvMFE) while keeping its benefits, the refined composite mvMFE (RCmvMFE) with a new fuzzy membership function is developed here as well.

In spite of the interesting results obtained by these improvements, fuzzy entropy (FuzEn), RCMFE, and RCmvMFE may still lead to unreliable results for short signals and are not fast enough for real-time applications. To address these shortcomings, dispersion entropy (DispEn) and frequency-based DispEn (FDispEn), which are based on our introduced dispersion patterns and the Shannon's definition of entropy, are developed. The computational cost of DispEn and FDispEn is O(N) – where N is the signal length –, compared with the O(N^2) for popular sample entropy (SampEn) and FuzEn. DispEn and FDispEn also overcome the problem of

equal values for embedded vectors and discarding some information with regard to the signal amplitudes encountered in permutation entropy (PerEn). Moreover, unlike PerEn, DispEn and FDispEn are relatively insensitive to noise.

As extensions of our developed DispEn, multiscale DispEn (MDE) and multivariate MDE (mvMDE) are introduced to quantify the complexity of univariate and multivariate signals, respectively. MDE and mvMDE have the following advantages over the existing univariate and multivariate multiscale methods: 1) they are noticeably faster; 2) MDE and mvMDE result in smaller coefficient of variations for synthetic and real signals showing more stable profiles; 3) they better distinguish various states of biomedical signals; 4) MDE and mvMDE do not result in undefined values for short time series; and 5) mvMDE, compared with multivariate multiscale SampEn (mvMSE) and mvMFE, needs to store a considerably smaller number of elements.

In this Thesis, two restating-state electrophysiological datasets related to AD are analyzed: 1) 148-channel MEGs recorded from 62 subjects (36 AD patients vs. 26 age-matched controls); and 2) 16-channel EEGs recorded from 22 subjects (11 AD patients vs. 11 age-matched controls). The results obtained by MDE and mvMDE suggest that the controls' signals are more and less complex at respectively short (scales between 1 to 4) and longer (scales between 5 to 12) scale factors than AD patients' recordings for both the EEG and MEG datasets. The p-values based on Mann-Whitney U-test for AD patients vs. controls show that the MDE and mvMDE, compared with the existing complexity techniques, significantly discriminate the controls from subjects with AD at a larger number of scale factors for both the EEG and MEG datasets. Moreover, the smallest p-values are achieved by MDE (e.g., 0.0010 and 0.0181 for respectively MDE and MFE using EEG dataset) and mvMDE (e.g., 0.0086 and 0.2372 for respectively mvMDE and mvMFE using EEG dataset) for both the EEG and MEG datasets, illustrating the superiority of these developed entropy-based techniques over the state-of-the-art univariate and multivariate entropy approaches.

Overall, the introduced FDispEn, DispEn, MDE, and mvMDE methods are expected to be useful for the analysis of physiological signals due to their ability to distinguish different types of time series with a low computation time.

Lay Summary

Alzheimer's disease (AD) is a degenerative brain disease and the most common form of dementia. As AD progresses, there are differences in brain activity that can be recorded in electroencephalogram (EEG) and magnetoencephalogram (MEG) signals, which reflect the electrical waves of brain activity. Nonlinear entropy approaches may be able to detect some of these changes and analyze EEGs and MEGs, as the outputs of a nonlinear system (e.g., brain). This Thesis aims at developing entropy-based methods for characterization of EEGs and MEGs in AD.

Dispersion entropy (DispEn) and frequency-based DispEn (FDispEn) are introduced here. They are two metrics to address the problems of unreliable entropy values and high computation times in popular sample and fuzzy entropy on the one hand, and discarding some information regarding the amplitudes in widely-used permutation entropy on the other hand.

DispEn is not able to account for the multiple time scales inherent in biomedical recordings. To address this shortcoming, multiscale DispEn (MDE) is developed to quantify the complexity of univariate signals across temporal scales. To deal with the patterns shared across channels and time, multivariate MDE (mvMDE) is introduced as well.

The results show that MDE and mvMDE have the following advantages over the existing entropy-based methods: 1) they are faster to compute; 2) MDE and mvMDE lead to more stable results for short and long signals; and 3) they better discriminate different kinds of biomedical signals. Two EEG and MEG datasets are used to evaluate our methods for the discrimination of AD patients from healthy controls. MDE and mvMDE, compared with the state-of-the-are techniques, better differentiate the diseased from healthy age-matched individuals for both the EEG and MEG datasets.

On the whole, our introduced FDispEn, DispEn, MDE and mvMDE are expected to be useful for the analysis of physiological signals, such as EEGs and MEGs, thanks to their ability to distinguish various changes in such recordings with a low computation time.

Declaration

I hereby declare that the research recorded in this thesis and thesis itself have been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

Hamed Azami

February 2018

Acknowledgements

First, I would like to express my sincerest gratitude to my supervisor, Dr. Javier Escudero. This research would not have been accomplished without his constant support and great guidance. Javier has been beyond a supervisor to me and I will never forget the lessons he has taught me.

I would like to give my special thanks to Prof. Mike Davies, to whom I have gained so much knowledge in the field of signal processing.

I am grateful to Mr. Keith Smith, Dr. Daniel Abásolo, Dr. Alberto Fernández, and Prof. Anne Humeau-Heurtier for their invaluable help and contribution throughout my PhD study.

Not only has it been an honour to work in such a wonderful environment, it was a privilege to live here as well. So many people have enriched my life here. I am surely blessed to have met you all. Please forgive me if I have forgotten to thank you... The list is too long!

Finally, I am particularly indebted to my beloved parents, Fouzieh and Abbas, for all their unconditional devotion, endless care, support, and encouragement throughout my life. I am also very thankful to my sister, Rana, and brother, Vahid, for all their love and encouragement, and the great moments we have spent altogether.

Contents

List of Tables

List of Figures

List of Abbreviations

AAPerEn . . . . amplitude-aware permutation entropy

Aβ . . . . . amyloid beta

AD . . . . . Alzheimer's disease

ApEn . . . . . approximate entropy

APOE . . . . . apolipoprotein E

AR . . . . . autoregressive

BAR . . . . . bivariate autoregressive

ChebDist . . . . Chebyshev distance

ConEn . . . . . conditional entropy

CSF . . . . . cerebrospinal fluid

CV . . . . . coefficient of variation

D_2 . . . . . correlation dimension

DispEn . . . . . dispersion entropy

DistrEn$_{2D}$ . . . . two-dimensional distribution entropy

EcDist . . . . . Euclidean distance

EEG . . . . . electroencephalography

ERP . . . . . event-related potential

FD . . . . . fractal dimension

FDG . . . . . ^{18}F-flouroxyglucose

FDispEn . . . . frequency-based dispersion entropy

FS . . . . . focal signals

FuzEn . . . . . fuzzy entropy

IMPE . . . . . improved multiscale permutation entropy

KS . . . . . Kolmogorov-Sinai

L_1 . . . . . first Lyapunov exponent

logsig . . . . . log-sigmoid

LP . . . . . lumbar puncture

LZC . . . . . Lempel-Ziv complexity

MCI . . . . . mild cognitive impairment

MDE . . . . . multiscale dispersion entropy

MEG . . . . . magnetoencephalography

MEMD . . . . . multivariate empirical mode decomposition

MFE . . . . . multiscale fuzzy entropy

MMSE . . . . . mini-mental state examination

MPE . . . . . multiscale permutation entropy

MRI . . . . . magnetic resonance imaging

MSE . . . . . multiscale sample entropy

mvDE . . . . . multivariate dispersion entropy

mvFE . . . . . multivariate fuzzy entropy

mvMDE . . . . multivariate multiscale dispersion entropy

mvMFE . . . . . multivariate multiscale fuzzy entropy

mvMPE . . . . . multivariate multiscale permutation entropy

mvMSE . . . . . multivariate multiscale sample entropy

mvPE . . . . . multivariate permutation entropy

mvSE . . . . . multivariate sample entropy

NCDF . . . . . normal cumulative distribution function

NFS . . . . . non-focal signals

PerEn . . . . . permutation entropy

RCMDE . . . . refined composite multiscale dispersion entropy

RCMFE . . . . refined composite multiscale fuzzy entropy

RCMSE . . . . . refined composite multiscale sample entropy

RCmvMFE . . . refined composite multivariate multiscale fuzzy entropy

RCmvMSE . . . refined composite multivariate multiscale sample entropy

SampEn . . . . sample entropy

SD . . . . . standard deviation

ShEn . . . . . Shannon entropy

SNR . . . . . signal-to-noise ratio

SQUID . . . . . superconducting quantum interference device

tansig . . . . . tan-sigmoid

WGN . . . . . white Gaussian noise

Nomenclature

A . . . . . adjusting coefficient related to the mean value and difference between consecutive samples for amplitude-aware permutation entropy

$\mathbf{A}_\gamma$ . . . . . coefficient matrix of a bivariate autoregressive process corresponding to lag order γ

c . . . . . number of classes for dispersion entropy-based approaches

d . . . . . time dealy value for univariate complexity methods

$\mathbf{d}$ . . . . . time delay vector for multivariate complexity methods

$f(\eta_t)$ . . . . . number of occurrences of each of the order patterns η_t

H_φ . . . . . joint entropy for φ variables

$H(V_\kappa|V_{\kappa-1},\ldots,V_1)$ conditional entropy of V_κ given $V_{\kappa-1},\ldots,V_1$

h . . . . . rate at which the joint entropy grows

i . . . . . time variable

I_B . . . . . the dc bias current for SQUID

k . . . . . space variable

L . . . . . length of the univariate and multivariate time series

L_p . . . . . superconducting flux transformer for SQUID

L_s . . . . . superconducting flux transformer for SQUID

m . . . . . embedding dimension value for univariate complexity methods

$m!$ . . . . . number of potential ordinal patterns for permutation entropy

$\mathbf{m}$ . . . . . embedding dimension vector for multivariate complexity methods

$\boldsymbol{MIX}$ . . . . . signal obtained by the mix process

n . . . . . maximum value of $\max\{\mathbf{m}\}$ multiplied by $\max\{\mathbf{d}\}$, where $\mathbf{m}$ and $\mathbf{d}$ are respectively the embedding dimension and time delay vectors

n_f . . . . . fuzzy entropy power for fuzzy entropy-based approaches

N . . . . . length of a coarse-grained signal

p . . . . . number of channels

$Pr(v)$ . . . . . probability for *V* to take the value *v*

r . . . . . threshold for sample and fuzzy entropy-based approaches

$\mathbf{R}$ . . . . . correlation matrix of correlated multivariate noise signals

$\mathbf{u}$ . . . . . univariate signal used as the input of a univariate multiscale entropy method

$\mathbf{U}$ . . . . . multivariate signal used as the input of a multivariate multiscale entropy method

$\mathbf{w}$ . . . . . classified signal for dispersion entropy-based approaches

$\mathbf{x}$ . . . . . univariate coarse-grained signal obtained by multiscale entropy approaches

$\mathbf{x}_i^{m,d}$ . . . . . i^{th} template vector for sample entropy-based methods with embedding dimension *m* and time delay *d*

$\mathbf{X}$ . . . . . multivariate coarse-grained signal obtained by multivariate multiscale entropy approaches

$\mathbf{y}$ . . . . . mapped signal for dispersion-entropy based approaches

α . . . . . model parameter for the logistic map

β . . . . . system parameter of the Lorenz system

$\gamma_{\max}$ . . . . . maximum lag in the bivariate autoregressive model

δ . . . . . intervals of time in an Ω-dimensional dynamical system

$\Delta_{a,b}$ . . . . . Chebyshev distance between $\mathbf{x}_a^{m,d}$ and $\mathbf{x}_b^{m,d}$

η_t . . . . . order patterns for permutation entropy-based methods

λ . . . . . system parameter of the Lorenz system

Ω . . . . . number of dimensions in a dynamical system

τ . . . . . temporal scale factor

$\tau_{\max}$ . . . . . maximum temporal scale factor

$\aleph_*$ . . . . . time or index of the element in the reconstruction vector $\mathbf{x}_{\Lambda}^{m,d}$ with embedding dimeniosn *m* and time delay *d*

ρ . . . . . system parameter of the Lorenz system

$\pi_{v_0 \ldots v_{m-1}}$ . . . . potential dispersion patterns for dispersion entropy-based methods

θ . . . . . fuzzy membership function for fuzzy entropy-based methods

φ . . . . . number of random variables

Chapter 1

Introduction

1.1 Main Motivation of the Thesis

Dementia is a syndrome (a group of related symptoms) associated with a number of progressive disorders affecting memory, behaviour, thinking, speaking, and the ability to perform daily activities [1, 2]. There are around 36 million people with dementia disease worldwide in 2010. It is expected that this number rises to around 66 million by 2030 and 115 million by 2050. Approximately two-thirds of these patients live in low and middle income countries, where the sharpest increases in numbers are set to happen [1]. It is predicted that the number of individuals in the UK aged 65 and over with moderate and severe dementia will increase by about 80% from 2010 to 2030 [3]. The total cost of dementia to society in the UK is £26.3 billion: £4.3, £10.3, and £11.6 billion are respectively spent for healthcare costs, social care, and the work of unpaid carers of persons with dementia [4].

Alzheimer's disease (AD) as a neurodegenerative disease is the most common form of dementia, accounting for an estimated 60 to 80 percent of cases [2, 5]. A positive diagnosis of AD, especially in its early stages, allows the patient and his/her family time to be informed about the disease, to make life and financial decisions, and to plan for the future. In contrast, a negative diagnosis may reduce worry about memory loss associated with ageing. Moreover, it permits for early treatments of reversible conditions with similar symptoms (like depression and nutrition or medication problems) [1, 6, 7].

Medical-based diagnosis of AD is not fully reliable and symptoms are sometimes dismissed as normal consequences of healthy ageing [6–9]. To this end, there is a real need to do research in various fields of science and engineering to understand how AD affects the brain.

As a cortical dementia, AD changes the interaction between neurons in the brain during its progression. Consequently, it alters brain activity. Some of these changes may be recorded by the non-invasive electroencephalography (EEG) and magnetoencephalography (MEG) techniques [6, 8, 10]. Given this and the fact that EEG and MEG are the only two signal-based neurophysiologic approaches able to directly measure the activity of the brain cortex non-invasively, these types of signals have potential as useful research tools in AD [6, 8, 10]. Thus, additional studies to investigate the effect of AD on the EEG and MEG recordings are needed. Eventually, it could rise the possibility of clinical use of the EEG and MEG in the diagnosis and monitoring of AD in the future.

In recent years, because the brain signals are considered as the outputs of a nonlinear system [11–13], there has been an interest in nonlinear techniques to analyze the EEG and MEG recordings to help the diagnosis of AD [6, 8, 10, 14]. One of the most popular and powerful nonlinear concepts used to evaluate the dynamical characteristics of signals is that of entropy [15–18]. Shannon entropy and conditional entropy, which respectively denote the amount of information and rate of information production, are two important fundamental concepts in information theory widely used in characterizing physiological signals [16,17]. These concepts relate to the uncertainty or irregularity of a time series [16–19]. Higher entropy stands for higher irregularity, whereas smaller entropy values show lower irregularity in a signal [16, 18].

In spite of the previous findings in the field of entropy for the analysis of electrophysiological signals, there is room for the introduction and development of novel nonlinear approaches for the characterization of EEG and MEG recordings in AD. The recent trends include: 1) improvement of the stability and reliability of entropy-based results for short signals (e.g., 100 sample points for embedding dimension 2 - see 4.3) since the majority of existing entropy-based approaches are either undefined or unavailable for short time series; and 2) extension of the univariate entropy approaches to their multivariate cases to be able to reveal the patterns across channels. These two trends are really important for characterization of EEGs and MEGs because such recordings are often multi-channel and sometimes short.

1.2 Aims and Hypotheses

This Thesis aims to develop new entropy-based approaches for characterization of EEGs and MEGs in AD. This is based on the following hypotheses:

1. The differences of physiological signals may be detected by entropy-based nonlinear analysis [15, 16, 20, 21]. In particular, difference in the EEG and MEG recordings between AD patients and age-matched healthy subjects may be distinguished with statistical significance using the entropy-based approaches [6, 10].
2. Healthy conditions correspond to more complex states due to their ability to adapt to adverse conditions, exhibiting long range correlations, and rich variability at multiple scales, while aged and diseased individuals present complexity loss [21–23]. In particular, EEGs and MEGs for age-matched controls are more complex than those recorded from AD patients [6, 21, 24, 25].

The irregularity or complexity decrease in the EEG signals recorded from AD patients could be described by a reduction of dynamical complexity of part of the brain [6, 10]. Nevertheless, the pathophysiological implications of the reduction of EEGs complexity or irregularity are not quite clear. Among others, three mechanisms can account for it: neuronal death, a general effect of lack of neurotransmitter, and loss of connectivity of local neural networks as a consequence of nerve cell death [6, 26, 27].

1.3 Contributions

The contributions of this Thesis are listed as follows:

1. Univariate Entropy Methods (Chapter Three):
    - Proposing amplitude-aware permutation entropy (PerEn - AAPerEn) to make PerEn more sensitive to the change in the amplitude, besides the frequency, of signals [28].
    - Introducing dispersion entropy (DispEn) based on the Shannon's definition of entropy and dispersion patterns to address the shortcomings of unreliable sample entropy (SampEn) and fuzzy entropy (FuzEn) values, high commutation time of

SampEn and FuzEn, and high sensitivity of AAPerEn and PerEn to noise [18, 29].

- Developing frequency-based DispEn (FDispEn) on the basis of Shannon entropy and the differences between adjacent elements of dispersion patterns [29].

2. Univariate Multiscale Entropy Methods (Chapter Four):
   - Suggesting the improved multiscale PerEn (MPE) to increase the stability of MPE-based results [28].
   - Proposing refined composite multiscale FuzEn (MFE- RCMFE) to address the shortcomings of MPE and multiscale SampEn (MSE) at the same time for short or noisy times series [30].
   - Developing multiscale DispEn (MDE) based on our developed DispEn to address the limitations of all the existing univariate multiscale entropy approaches [31].
3. Multivariate Multiscale Entropy Methods (Chapter Five):
   - Proposing a new fuzzy membership function to decrease the computation time of the existing multivariate MFE (mvMFE) while maintaining its advantages [32].
   - Introducing multivariate MDE (mvMDE) to address the shortcomings of all the existing multivariate multiscale entropy methods [33].
4. Application of Entropy-based Approaches to AD (Chapter Six):
   - Comparing the existing and developed univariate and multivariate multiscale entropy approaches to characterize two resting-state EEG and MEG datasets in AD [34].
   - Investigating changes in entropy values for different frequency bands of EEGs and MEGs for AD patients vs. controls to understand the effect of AD and entropy-based methods on each frequency range [34].

1.4 Structure of the Thesis

The reminder of the Thesis is organized as follows:

- Chapter 2: Background. This Chapter first summarises some concepts associated with AD. Then, the MEG and EEG as two non-invasive useful techniques in AD are described. Due to the broad use of nonlinear methods for biomedical signals, a brief literature review in such techniques is carried out. Next, the state-of-the-art

entropy-based approaches for physiological time series are explained. A literature review of nonlinear analysis for EEGs and MEGs in AD is carried out. Finally, *irregularity* and *complexity*, as two key concepts used in the Thesis, are detailed.

- Chapter 3: Univariate Entropy Methods. A set of univariate synthetic time series based on several straightforward signal processing concepts are initially described. Afterwards, the algorithm of AAPerEn, as a modified version of PerEn, is explained. Then, the DispEn and FDispEn approaches are detailed. Finally, the existing and developed entropy methods are compared and their advantages and disadvantages are illustrated using the synthetic and real physiological signals.
- Chapter 4: Univariate Multiscale Entropy Approaches. The synthetic and real univariate signals used in this Chapter are first described. Next, MPE and IMPE, and their advantages and disadvantage are mentioned. Our proposed refined composite multiscale fuzzy entropy (MFE - RCMFE) is then explained. Afterwards, MDE and RCMDE are introduced. Finally, the results for the existing and proposed multiscale techniques are illustrated and compared.
- Chapter 5: Multivariate Multiscale Entropy Approaches. After describing the synthetic and real multivariate signals used in this Chapter, the algorithm of mvMPE and its benefits and shortcomings are explained. The existing and our developed multivariate MFE (mvMFE) are then detailed. Next, refined composite mvMFE (RCmvMFE) and its advantages over mvMFE are briefly explained. Afterwards, mvMDE is introduced. Then, the simulation results obtained by the mvMSE, mvMFE, and mvMDE methods are shown and discussed.
- Chapter 6: Illustration in Alzheimer's Disease. The EEG and MEG datasets used in this Thesis are briefly described in this Chapter. The usefulness of the developed and existing univariate and multivariate entropy-based techniques to characterize the EEG and MEG signals in AD is inspected. The results also are compared with the previous studies to understand if our results are in agreement with the literature and hypotheses of complexity or not. Finally, the changes in entropy values for different frequency bands are investigated to understand the effect of AD and entropy-based methods on each frequency range.
- Chapter 7: Summary, Conclusions, Limitations, and Future Research. This Chapter

summarises the main findings of the Thesis. In addition, the limitations and future research directions are discussed.

- Appendix A: Publications. This Appendix details the publications in journals and conferences in which the PhD student has collaborated thanks to his work in this Doctoral Thesis.
- Appendix B: Effect of Number of Channels on Multivariate Sample and Fuzzy Entropy Approaches. This Appendix investigates how multivariate multiscale sample and fuzzy entropy methods change when the number of channels of a multivariate signal decreases.

Chapter 2

Background

Alzheimer's disease (AD) is a progressive degenerative disease, causing loss of memory and other cognitive functions before death [2]. As the only currently clinically accepted approach for accurate diagnosis of AD is through necropsy [9], and due to increasing the number of effected patients worldwide [2], there is a real need to enhance diagnostic procedures for AD. The progress of AD leads to changes in brain activity [35]. Some of these changes can be recorded by the non-invasive electroencephalography (EEG) and magnetoencephalography (MEG) techniques, which are able to directly measure the activity of the brain cortex [35–37].

The hypothesis that brain signals are the output of a linear system (i.e., brain) may be rebutted based on the ability of the brain to perform sophisticated cognitive tasks due to its complex structure [35]. Additionally, since brain neurons are controlled by nonlinear phenomena, such as saturation processes and threshold, brain activity signals are considered as the output of a nonlinear system (i.e., brain) [35, 37]. Accordingly, nonlinear techniques have been widely used to analyze EEG and MEG time series [10, 35].

Entropy is a powerful and broadly-used nonlinear metric used to assess the dynamical characteristics of time series [16, 19]. Shannon entropy (ShEn) and conditional entropy (ConEn) respectively show the amount of information and the rate of information production [18, 21, 38]. They are two most common concepts used in the context of analysis of physiological signals [19, 38, 39].

There are several univariate and multivariate entropy methods to characterize biomedical signals in general and EEG and MEG in particular, and each has its own advantages and disadvantages. This Chapter carries out a literature review on AD, EEGs and MEGs, the existing non-linear approaches, especially entropy-based ones, for characterization of MEGs

and EEGs, and their ability to distinguish AD patients from healthy subjects.

2.1 Alzheimer's Disease (AD)

2.1.1 Introduction

Dementia is a clinical syndrome characterized by inexorably progressive deterioration in cognitive ability affecting the capability to perform daily activities [1, 2, 40]. AD is the most common form of dementia in the elderly population, accounting for an estimated 60% to 80% of cases [2, 5].

In 1907, Dr. Alois Alzheimer explained a surprising new pathology in the brain of a woman who recently died several years after developing a clinically unusual dementia at age 51 [41]. Later named AD, this neurodegenerative disease affects intellectual, behavioural, and functional abilities [5, 7]. In AD, neurons in several parts of the brain are eventually damaged or destroyed, including those that enable a person to carry out basic bodily functions such as walking and speaking. Patients in the final stages of AD are bed-bound and need around-the-clock care. Note that AD is ultimately fatal [2].

2.1.2 Symptoms of AD

Some of the early signs or symptoms of AD are as follows [2]:

- Memory loss disrupting daily life, especially forgetting recently learned information.
- Challenges in planning or solving problems.
- Difficulty completing daily tasks at home or work.
- Confusion with time or place.
- Trouble understanding visual images and spatial relationships.
- New problems with words in writing or speaking.
- Misplacing things and losing the ability to retrace steps.
- Changes in judgement or decision-making.

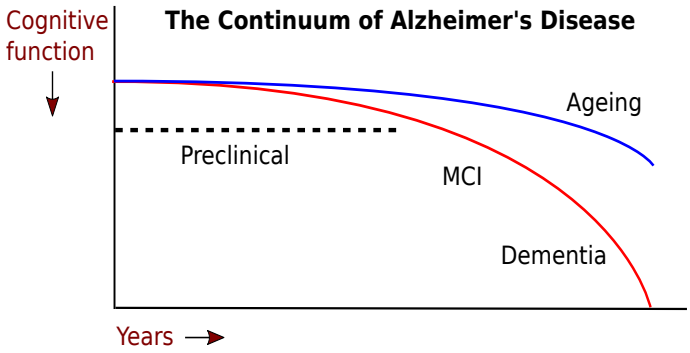


Figure 2.1: Model of the clinical trajectory of AD. The stage of preclinical AD precedes mild cognitive impairment (MCI) and encompasses the spectrum of presymptomatic autosomal dominant mutation carriers, asymptomatic biomarker-positive elderly subjects at risk for progression to MCI because of AD, as well as biomarker-positive subjects who have demonstrated subtle reduction of their own baseline that exceeds that expected in typical ageing, but would not yet meet criteria for MCI.

2.1.3 Evolution of AD

The preclinical stage of AD may show a continuum from completely asymptomatic subjects with biomarker evidence suggestive of AD-pathophysiological process at-risk for progression to AD to biomarker-positive subjects who are already showing very subtle decline but not yet meeting standardized criteria for mild cognitive impairment (MCI) [42]. This group of subjects might be classified as "not normal, not MCI" but would be included under the rubric of preclinical AD (see Figure 2.1).

Converging evidence from both genetic at-risk groups and clinically elderly control people showed that the pathophysiological process of AD starts years, if not decades, before the diagnosis of clinical AD [42, 43]. Recent developments and advances in cerebrospinal fluid (CSF) assays, neuroimaging, and other biomarkers have enabled us to detect evidence of the AD pathophysiological process *in vivo* [42]. Emerging data in clinically normal elderly people show that biomarker evidence of amyloid beta (Aβ) accumulation is associated with functional and structural brain changes, consistent with the patterns of abnormality observed in subjects with mild MCI and AD [2, 42].

Moreover, clinical studies suggested that there may be subtle cognitive changes that can be detected years before meeting the criteria for MCI diagnosis, and that predict progression to AD [2, 42]. As some older subjects with the pathophysiological process of AD may not get

symptomatic during their lifetime, it is essential to better define the biomarker and/or cognitive profile that best predicts progression from the preclinical to clinical stages of MCI and AD [2, 42].

2.1.4 Risk Factors

Medical scientists and experts believe that AD, like other common chronic diseases, develops as a consequence of a number of factors rather than a single reason. The main risk factors are as follows:

- Non-modifiable risk factors: The most important risk factors for late-onset AD are as follows:
  - Ageing: Age is the greatest risk factor for AD patients who are 65 years or older. The percentage of people with AD increases noticeably with age: 3% of people aged 65-74, 17% of people aged 75-84, and 32% of people aged 85 or older have AD [2, 44].
  - Family history: People who have a parent, brother, or sister with AD are more likely to develop the disease in comparison with those who do not have a first-degree relative with AD [2, 45].
  - Apolipoprotein E (APOE) $\varepsilon 4$ gene: The APOE $\varepsilon 4$ allele is a cholesterol transporter found in the brain [46]. Due to their effects on aggregation of Aβ and other neuropathological changes, APOE $\varepsilon 4$ alleles are the main genetic determinants of AD risk [47].
- Modifiable risk factors: Several risk factors such as age and family history cannot be changed or enhanced to decrease the risk of cognitive decline and dementia. However, studies assessing the state of evidence on impacts of modifiable risk factors on cognitive decline and dementia showed that there is sufficiently strong evidence that some causes decrease the risk of cognitive decline and may decrease the risk of dementia [48, 49]. Three main modifiable risk factors are as follows [2, 50]:
  - Cardiovascular factors: The brain is impacted by the health of the heart and blood vessels. A healthy heart ensures enough blood is pumped to the brain, whereas healthy blood vessels cause the oxygen- and nutrient-rich blood to reach the brain, leading to its normal function. A number of factors, such as smoking, obesity

in mid-life, and diabetes, increase the risk of cardiovascular disease, and so, are associated with a higher risk of dementia.

- Education: People with more years of formal education are at lower risk for AD and other types of dementia in comparison with those with a smaller number of years of formal education.
- Social and cognitive engagement: Social interaction has been shown to support brain health and decrease the risks of AD two times.

2.1.5 Diagnosis of AD

The only current clinically accepted approach for certain diagnosis of AD is through necropsy (the microscopical analysis of the subject's brain tissue after death) [9, 51]. There is not any single test to diagnose AD with 100% accuracy at the moment. Instead, physicians, in collaboration with neurologists and geriatricians, use various approaches and tools to help make a diagnosis [2]:

- Obtaining a medical and family history from the person, consisting of psychiatric history and history of cognitive and behavioral changes.
- Asking a family member or very close friend to provide some information about changes in thinking skills and behavior.
- Conducting cognitive tests and physical and neurologic examinations.
- Having the subject undergo blood tests and brain imaging to exclude other potential causes of dementia symptoms, like a tumor or certain vitamin deficiencies.

Diagnosis of AD needs a careful and comprehensive medical assessment. Although physicians can usually determine if an individual has dementia, it may be difficult to detect the exact cause. A few days or even weeks may be required for the person to complete the needed tests and examinations and for the physician to interpret the results and make a diagnosis [2].

2.1.6 Biomarkers

AD biomarkers are physiological, biochemical, and anatomic parameters that can be measured to indicate the presence or absence of AD, or the risk of developing it [2,52]. According to the literature available in 2011 [52], the most studied possible biomarkers were incorporated into

the new criteria for AD diagnosis. These biomarkers are CSF, positron emission tomography (PET) and ^{18}F-flouroxyglucose (FDG)-PET, and structural magnetic resonance imaging (MRI) [53].

CSF is an uncolored clear body fluid found in the spinal cord and brain [54]. The most common way to get a sample of CSF is by lumbar puncture (LP) or spinal tap [54]. As CSF directly interacts with the extracellular space in the brain, it is able to reflect the associated biochemical/pathologic changes. CSF concentration of Aβ42 is decreased in AD subjects with a sensitivity and specificity of about 85% compared with control subjects [55]. However, this decrease is also observed in several other diseases and cannot always be attributed to AD alone [55]. This is similar to the concentration of tau proteins in the CSF, albeit specifically hyperphosphorated tau concentrations increase in AD [56].

About the side effects of CSF, severe complications associated with LP are rare. Nevertheless, the acceptance and positioning of LP procedures and the use of CSF biomarkers are noticeably varies in countries. Therefore, a global standardization of LP procedures for CSF biomarker analysis is required, and the appropriate setting (primary care versus specialist center and neurologist versus geriatrician versus psychiatrist) should be established [57].

PET is a nuclear medicine functional imaging technique, which is useful to reveal the progressive decrease in glucose metabolism associated with AD [58, 59]. Thus, this kind of images may be considered as a biomarker [56, 59]. Hypometabolism in the temporal, parietal, and posterior cingulate cortex, identified by FDG-PET, may be used to distinguish with high sensitivity patients with AD from cognitively normal elderly individuals [59]. Temporoparietal hypometabolism on FDG-PET might predict the progression from MCI to severe AD with high accuracy [59]. However, these biomarkers are very costly and are not widely available [56].

As current research biomarkers, structural and functional MRI images are able to illustrate that the hippocampus and entorhinal cortex typically become smaller, whereas the ventricle increases in size with disease progression [59]. However, the considerable processing time for the structural and functional MRI-based approaches and, more importantly, their costs mean that these methods are not very broadly used [60]. Nevertheless, a limited number of studies have shown the discrimination of AD from other types of dementia [2, 59].

Slowing of the EEG recordings, through decreased high frequency components and increased low frequency components, has been illustrated to correlate with the severity of AD and therefore, may be considered as a biomarker [6]. Although AD diagnostic accuracy is around 80%, it has an overlap with other diseases [61].

The MEG is not broadly-used and considered as a biomarker of AD [2]. Nevertheless, the availability of whole-head MEG equipment has recently increased [62]. This brain recording technique has a few advantages over the conventional EEG (see Section 2.2; [62, 63]). Thus, MEG may provide a more accurate image of ongoing neural activity. Additionally, considerable advances have been made in neuroscience with regard to the understanding of oscillatory and synchronized brain activities and abnormal patterns associated with different kinds of brain disorders, including AD [10, 37, 62]. Therefore, it is possible that these changes showing abnormalities in specific networks and neurotransmitter systems may be beneficial in future AD diagnosis [37, 62].

2.2 Electroencephalograms (EEGs) and Magnetoencephalograms (MEGs)

There is an increasing interest in neurophysiological techniques that are appropriate to capture the macroscopic spatio-temporal dynamics of the electromagnetic fields of the brain [35]. The EEG and MEG are two neurophysiological approaches able to directly measure the activity of the brain cortex, without having to interpret the information on the basis of vascular or metabolic changes [36]. Both EEG and MEG are non-invasive techniques allowing the recording of electromagnetic fields produced by brain activity with high temporal resolution (millisecond-range) [36, 64].

The EEG is a measurement of currents flowing during synaptic excitations of the dendrites of a large number of pyramidal neurons in the cerebral cortex. It reflects the electrical brain activity originated by the neurons with a set of electrodes located on the scalp [37]. Richard Caton used a galvanometer and placed two electrodes over the scalp of a person and therefore first recorded brain activity in the form of electrical time series in 1875. After that time, the concepts of electro-(referring to registration of brain electrical activities) encephalo- (referring to emitting the time series from the head), and gram (or graphy, meaning drawing or writing)

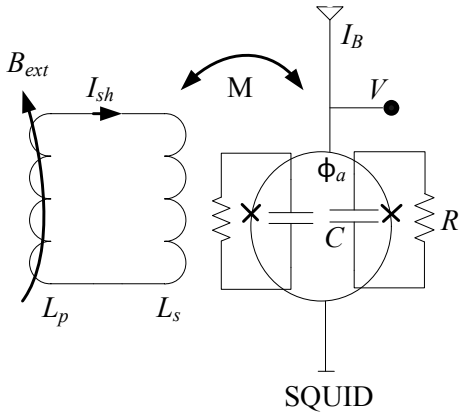


Figure 2.2: Schematic diagram of a dc superconducting quantum interference device (SQUID) magnetometer. The external field is connected via a superconducting flux transformer (L_p, L_s) to the SQUID. In a gradiometer, there is an additional compensation coil in series with L_p. The average voltage V across the SQUID relates to the dc bias current I_B and is a periodic function of the Φ_a coupled to the SQUID ring.

were combined in order that the term EEG was so forth used to denote electrical neural activity of the brain [37]. Of note is that the EEG systems can now record tens of channels [37].

The MEG technique is based on the superconducting quantum interference device (SQUID) as a sensitive detector of magnetic flux [36, 65]. Many MEG-based instruments are based on dc SQUIDs due to their readout electronics. The dc SQUID includes a superconducting loop interrupted by two Josephson junctions (see Figure 2.2) [36].

David Cohen was the first person who carried out SQUID measurement of magnetic fields of the brain in 1970. He recorded the spontaneous alpha activity of a healthy person and the abnormal activity of an epileptic patient [36,66]. In the first MEG measurements, a one-channel system was used. Therefore, the equipment had to be moved to measure activity for different parts of the scalp [36]. A few years later, novel instrumentation based on the gradiometers was used to simultaneously record the activity of a brain hemisphere with several channels [36]. It is worth noting that current MEG systems include hundreds of channels [62, 67].

One is sometimes interested in specific frequency bands in the EEG and MEG recordings since changes in different frequency ranges are associated with various diseases and disorders [6,35,37,62]. These frequency bands from low to high frequencies are respectively termed delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-40 Hz) [6, 37]. An

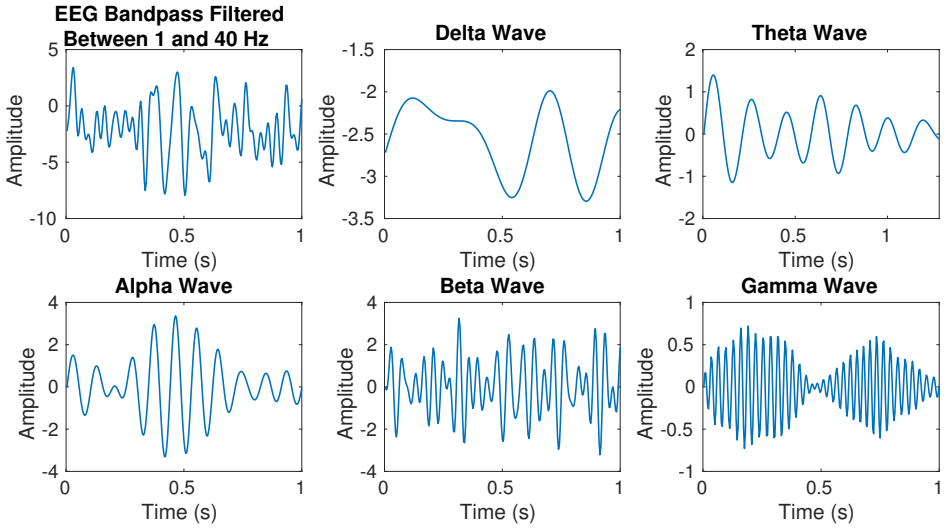


Figure 2.3: An EEG signal and its delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-40 Hz) waves.

EEG signal and its main frequency bands are illustrated in Figure 2.3.

1. Delta waves: They are high amplitude brain signals primarily associated with deep sleep [37, 68]. As an application of this band, the delta wave abnormalities are often used to detect brain dysfunction [37, 68].

2. Theta waves: They appear as consciousness slips towards drowsiness and are associated with access to meditative concentration, creative inspiration, and deep meditation [37, 68].

3. Alpha waves: They commonly appears as a round or sinusoidal shaped signal and reflect visual processing in the occipital brain region. These waves may be related to memory function. For instance, increasing mental effort leads to a suppression of alpha activity, particularly from the frontal areas [68, 69].

4. Beta waves: They are the usual waking rhythms of the brain associated with active attention, active thinking, focusing on the outside world, or solving problems [37].

5. Gamma waves: Although the amplitudes of gamma waves are very low, detection of these rhythms can be used for confirmation of certain brain diseases, such as AD [70, 71].

As the EEG and MEG techniques record the electromagnetic activity generated by the same primary currents in the brain, some similarities between their waveforms are to be expected [63]. To study the brain activity, EEG and MEG techniques can also complement each other.

Table 2.1: Comparison of the MEG and EEG techniques in terms of activity measured, temporal and spatial resolutions, safety, and portability.

Electrophysiological technique	Activity measured	Direct/Indirect measurement	Temporal resolution	Spatial resolution	Risk	Portability
EEG	electrical	direct	~ 0.05 s	~ 10 mm	invasive	portable
MEG	magnetic	direct	~ 0.05 s	~ 5 mm	invasive	non-portable

However, the MEG approach have the following advantages over the EEG one:

1. MEGs, unlike EEGs, are reference-free recordings, and are not distorted by the resistive properties of the skull [36, 72].
2. The spatial resolution of MEG is higher than that of EEG, as the number of channels in MEG is often larger than that in EEG [36].
3. Recording signals from a very large number of sensors is more easily performed by the use of MEG, compared with EEG, as the time-consuming application of electrodes on the scalp is not required [62, 73].

In spite of its aforementioned benefits, the MEG technique is subjected to several noticeable limitations:

1. The small amplitude of the brain magnetic fields requires the application of further approaches to attenuate the noisy times series [64]. This leads to considerable restrictions on the recording process and makes the MEG technique non-portable [63].
2. As magnetic time series recorded from the human brain have small values, SQUID sensors are essential for their detection and MEGs should be recorded in a magnetically shielded room [36, 64]. Thus, the MEG technique is associated with a high investment cost for the MEG system and the shielded room, which has prevented widespread use of this technique to quantify brain activity [74].

To summarize, the MEG and EEG techniques are compared in terms of activity measured, temporal and spatial resolutions, safety, and portability in Table 2.1 [68].

2.3 Nonlinear Dynamical Analysis for Physiological Signals

A system that satisfies both the homogeneity and additivity properties are considered to be a linear system [75]. In contrast, a non-linear system does not obey these properties. A nonlinear system reflects that the whole gets something greater than the sum of its individual parts due to feedback or multiplicative effects between its components [35, 76]. Dynamical shows that the system changes over time based on its current state. Approximately every nontrivial real-world system is a nonlinear dynamical system [76]. A dynamical system is described by two elements: a state and the dynamics. The state of a dynamical system is determined by the values of all the variables describing the system at a specific time. Therefore, the state of a system represented by Ω variables can be shown by a point in an Ω-dimensional space. This space is termed the state or phase space of the system. The dynamics of the system is the set of laws or equations describing how the state of a system changes during the time [35].

Physiologists and clinicians are often confronted with the problem of distinguishing different kinds of dynamics for biomedical signals, such as heart rate traces from infants who have an aborted sudden infant death syndrome versus control infants [20], and EEG signals for young subjects versus elderly people [77]. A number of physiological time series, such as cardiovascular and brain activity recordings, are considered as the outputs of a nonlinear system [12, 37, 78, 79]. Moreover, several studies suggested that physiological recordings from healthy subjects have nonlinear complex relationships with ageing and disease [21, 39]. Thus, there is an increasing interest in nonlinear techniques to analyse the dynamics of physiological signals. There are a number of nonlinear dynamical analysis techniques, such as fractal dimension (FD) [80], Lyapunov exponent [81], Lempel-Ziv complexity (LZC) [82], and entropy-based metrics [19, 83].

2.3.1 Fractal Dimension (FD)

FD refers to a noninteger or fractional dimension of a geometric object [80]. FD is a powerful nonlinear approach for describing the regularity of a signal [84]. FD algorithms are defined in the time and phase (state) space domains [80]. Time domain-based FD approaches are useful for transient detection, with the further advantage of fast computation. However, the phase

space representation of a nonlinear system can contain one or more attractors with generally fractional dimension, leading to a considerable computational burden [80].

The most popular phase space algorithm is correlation dimension (D_2) or FD of the attractor described in [85]. An embedding system is constructed from the original time series based on nonlinear ordinary differential equations, and the attractor of this system is untangled before estimating its FD [80, 86]. In fact, D_2 quantifies the number of independent variables which are needed to describe the dynamics of a system and thus, higher D_2 values represent more irregular or complex systems.

There are several algorithms in the time domain for the computation of FD. Among them, Katz's [87], Higuchi's [88], and Petrosian's [89] FD are the most common techniques applied to physiological signals [80, 90]. The Katz's method is the most consistent approach for distinguishing epileptic states from the intracranial EEGs, likely thanks to its exponential transformation of FD values and relative insensitivity to noise. However, the Higuchi's and Petrosian's FD algorithms are less suitable for EEG signal analysis due to their high sensitivity to noise [80].

2.3.2 Largest Lyapunov Exponents

Lyapunov exponents quantify the rate of separation of two infinitesimally close trajectories given several state spaces. Therefore, the largest Lyapunov exponent provides a metric of stability. This uses two points on a Lorenz attractor and measures the exponential decrease or increase in the vector between these two points over time [35]. This provides a measure that is more sensitive to the flexibility of underlying system to process varying information than D_2 [26]. It was demonstrated that positive largest Lyapunov exponents, once an indicator for chaotic systems, is also present with noise time series [91].

In fact, Largest Lyapunov exponents provide a qualitative and quantitative characterization of dynamical behavior and are related to the exponentially fast divergence or convergence of nearby orbits in phase space [81]. However, the techniques for estimating the largest Lyapunov exponents suffer from at least one of the following shortcomings [92]: 1) unreliable for small data sets; 2) computationally intensive; and 3) relatively difficult to implement.

2.3.3 Lempel-Ziv Complexity

As a powerful and widespread metric to analyze biological signals [82], LZC is associated with the number of distinct substrings and the rate of their recurrence along the analysed time series, with larger values corresponding to more complex data [82]. However, LZC is noticeably sensitive to noise and may not be able to discriminate some deterministic from stochastic signals [93].

2.3.4 Entropy-based Approaches in the Context of Information Theory and Biomedical Signal Processing

Nonlinear measures taken from information theory are of great interest for the evaluation of the degree of irregularity and complexity of physical, physiological, social, and econometric systems [94]. To this end, a wide range of techniques rooted in the concept of entropy [19] have been proposed [18, 21, 38, 94, 95].

2.3.4.1 Information-Theoretic Preliminaries

The development of the concept of entropy of data by Shannon [19] provided the foundations of information theory [95]. ShEn describes the amount of information of a univariate (or multivariate) random variable *V* as its average uncertainty [38]. Shannon entropy is formulated as follows:

$$H(V) = -\sum_{v} Pr(v) \cdot \log_e(Pr(v)), \tag{2.1}$$

where $Pr(v)$ is the probability for *V* to take the value *v*, and the sum is taken over all nonzero probability values (i.e., $Pr(v)\log_e(Pr(v)) = 0$ if $Pr(v) = 0$) [38]. It is worth mentioning that the entropy-based approaches are measured in *natural units* (or nats) in this Thesis.

When all the probability values are equal, the maximum entropy occurs, while if one probability is certain and the others are impossible, the minimum entropy is achieved [18, 19].

For an indexed sequence of φ random variables $\{V_1, V_2, \ldots, V_\kappa, \ldots, V_\varphi\}$, the joint entropy is

defined as [21, 95]:

$$H_{\varphi} = H(V_1, \ldots, V_{\varphi}) = -\sum_{v_1} \cdots \sum_{v_{\varphi}} Pr(v_1, \ldots, v_{\varphi}) \cdot \log_e(Pr(v_1, \ldots, v_{\varphi})), \tag{2.2}$$

where $Pr(v_1, \ldots, v_{\varphi})$ is the joint probability for the φ variables $V_1, \ldots, V_{\varphi}$ [21, 95].

Based on the chain rule and Equation (2.2), the joint entropy can be defined as a summation of ConEn values. Each of them is a non-negative quantity [21, 95]:

$$H_{\varphi} = \sum_{\kappa=1}^{\varphi} H(V_{\kappa}|V_{\kappa-1}, \ldots, V_1), \tag{2.3}$$

where $H(V_{\kappa}|V_{\kappa-1}, \ldots, V_1)$ is the ConEn of V_{κ} given $V_{\kappa-1}, \ldots, V_1$. Thus, the joint entropy can be described as an increasing function of φ.

The entropy rate *h* as the rate at which the joint entropy grows is defined as follows [21, 95]:

$$h = \lim_{\varphi \to \infty} \frac{H_{\varphi}}{\varphi}. \tag{2.4}$$

For stationary ergodic precesses, a number of studies have illustrated that the evolution of the entropy rate is a very useful measure for the analysis of physiological signals [15, 16, 21].

2.3.4.2 Kolmogorov-Sinai (KS) Entropy

Assume that the phase space of an Ω-dimensional dynamical system is partitioned into hypercubes of content ε^{Ω}. The state of the system is measured at intervals of time δ. Let us consider the joint probability $Pr(v_1, \ldots, v_{\varphi})$, where the state of the system is in the hypercube v_1 at time δ, in the hypercube v_2 at time 2δ, and in the hypercube v_{φ} at time $\varphi\delta$. The Kolmogorov-Sinai (KS) entropy is defined as follows [21, 95]:

$$H_{KS} = -\lim_{\delta \to 0} \lim_{\varepsilon \to 0} \lim_{\varphi \to \infty} \frac{1}{\varphi\delta} \sum_{v_1 \ldots, v_{\varphi}} Pr(v_1, \ldots, v_{\varphi}) \cdot \log_e(Pr(v_1, \ldots, v_{\varphi})) = \lim_{\delta \to 0} \lim_{\varepsilon \to 0} \lim_{\varphi \to \infty} \frac{H_{\varphi}}{\varphi\delta}. \tag{2.5}$$

When dealing with stationary precesses, the entropy rate can be defined as [96]:

$$h = \lim_{\varphi \to \infty} \frac{H_\varphi}{\varphi} = \lim_{\varphi \to \infty} H(V_\varphi | V_{\varphi-1}, \ldots, V_1). \tag{2.6}$$

Afterwards, according to the chain rule, the following equation can be proved [21, 95]:

$$H_{KS} = \lim_{\delta \to 0} \lim_{\varepsilon \to 0} \lim_{\varphi \to \infty} (H_\varphi - H_{\varphi-1}). \tag{2.7}$$

In fact, the KS entropy measures the mean rate of creation of information or the increase of certainty at a receiver by knowing the current state of the system given the past history [21].

Practically, only entropies of finite order φ can be calculated. When φ is large, the entropy H_φ is underestimated and decays toward 0. Nevertheless, several studies suggested a few formulas to estimate the KS entropy with appropriate precision. To this end, Grassberger and Procaccia, and Eckmann and Ruelles proposed respectively K_2 [97] and ER entropy [98] techniques to estimate the KS entropy directly from a time series.

2.3.4.3 K_2 and ER Entropy Methods

The algorithm for K_2 is as follows. Assume a univariate time series of length *N*: $\mathbf{x} = \{x_1, x_2, \ldots, x_j, \ldots, x_N\}$. First, the vectors $\mathbf{x}_\hbar^m = \{x_\hbar, x_{\hbar+1}, \ldots, x_{\hbar+m-1}\}$ $(1 \leq \hbar \leq N - m + 1)$ with length *m* are created, where m denotes the embedding dimension. Then, the number of vectors satisfying EcDist$[\mathbf{x}_\hbar^m, \mathbf{x}_\Upsilon^m] \leq r$ $(1 \leq \Upsilon \leq N - m + 1)$ over $(N - m + 1)$ - represented as $\vartheta_\hbar^m(r)$ - is calculated, where EcDist denotes the Euclidean distance and *r* is a predefined threshold [97]. In fact, $\vartheta_\hbar^m(r)$ shows the probability that any vector $\mathbf{x}_\Upsilon^m$ is close to $\mathbf{x}_\hbar^m$ within threshold *r*. The probability that any two vectors are within a Euclidean distance *r* of each other, shown as $\vartheta^m(r)$, is calculated as follows:

$$\vartheta^m(r) = \frac{1}{N - m + 1} \sum_{\hbar=1}^{N-m+1} \vartheta_\hbar^m(r). \tag{2.8}$$

Then, the dimension is increased to $m+1$ and then, $\vartheta^{m+1}(r)$ is calculated based on the number of $\mathbf{x}_\hbar^{m+1}$ within *r* of $\mathbf{x}_\Upsilon^{m+1}$, where Υ ranges from 1 to $N - m$ [97].

Finally, K_2 is calculated as follows [21, 97]:

$$K_2 = -\lim_{N\to\infty}\lim_{m\to\infty}\lim_{r\to 0}(\vartheta^{m+1}(r) - \vartheta^m(r)). \tag{2.9}$$

Subsequently, ER entropy was developed as another alternative of KS entropy [98]. The difference between the ER entropy and K_2 approaches is the former was defined based on

$$\varrho^m(r) = \frac{1}{N-m+1}\sum_{\hbar=1}^{N-m+1}\ln\vartheta_\hbar^m(r), \tag{2.10}$$

where the distance between two vectors $\mathbf{x}_\hbar^m$ and $\mathbf{x}_\Upsilon^m$ is calculated according to the Chebyshev distance [98]:

$$\text{ChebDist}[\mathbf{x}_\hbar^m, \mathbf{x}_\Upsilon^m] = \max\{|x_{\hbar+\zeta} - x_{\Upsilon+\zeta}| : 0 \le \zeta \le m-1\}. \tag{2.11}$$

Although the ER entropy is useful in classifying low-dimensional chaotic systems, it does not apply to experimental signals as its result tends to infinity for a process with superimposed noise of any magnitude [99].

2.3.4.4 Approximate Entropy (ApEn)

To take into account the limitation of ER entropy for noisy signals, Pincus developed ApEn [15]. In fact, ApEn is defined based on Equation (2.10) as follows:

$$ApEn(m,r) = \lim_{N\to\infty}(\varrho^m(r) - \varrho^{m+1}(r)). \tag{2.12}$$

Thus, ApEn can be estimated as follows [15]:

$$ApEn(m,r,N) = \varrho^m(r) - \varrho^{m+1}(r). \tag{2.13}$$

ApEn applies to "real-world" data, leading to its wide-use in physiological signals [21, 100, 101]. Nevertheless, ApEn counts each sequence as matching itself [15, 16]. This step biases

ApEn [16]. This results in ApEn being dependent on the length of time series and is uniformly smaller than that expected for short signals [16].

2.3.4.5 Sample Entropy (SampEn)

To overcome the data record length limitation of ApEn, SampEn was introduced [16]. Assume a univariate signal vector $\mathbf{x} = \{x_1, \ldots, x_N\}$. First, all the template vectors $\mathbf{x}_\Lambda^{m,d}$ ($\Lambda = 1, 2, \ldots, N-(m-1)d$) are created as follows [16]:

$$\mathbf{x}_\Lambda^{m,d} = \{x_\Lambda, x_{\Lambda+d}, \ldots, x_{\Lambda+(m-1)d}\}, \tag{2.14}$$

where *d* is the time delay. Then, the average of the number of vector pairs in template vectors of length *m* having ChebDist$[\mathbf{x}_\Lambda^{m,d}, \mathbf{x}_a^{m,d}] \leq r$ ($1 \leq a \leq N - md, a \neq \Lambda$) is calculated as follows:

$$\phi_\Lambda^{m,d}(r) = \frac{\#\{\mathbf{x}_a^{m,d} \mid \text{ChebDist}[\,\mathbf{x}_\Lambda^{m,d}, \mathbf{x}_a^{m,d}] \leq r\}}{N - md - 1}, \tag{2.15}$$

where $\#$ represents the cardinality (number of elements of the set). Then, the function $\phi^{m,d}(r)$ is calculated as follows [16]:

$$\phi^{m,d}(r) = \frac{1}{N - md} \sum_{\Lambda=1}^{N-md} \phi_\Lambda^{m,d}(r). \tag{2.16}$$

Then, the dimension is increased to $m + 1$ and subsequently, $\phi^{m+1,d}(r)$ is calculated based on the number of $\mathbf{x}_\Lambda^{m+1,d}$ within *r* of $\mathbf{x}_a^{m+1,d}$, where *a* ranges from 1 to $N - md$ ($a \neq \Lambda$) [16].

Finally, the function SampEn is defined as follows:

$$\text{SampEn}(\mathbf{x}, m, r, d) = -\ln\left(\frac{\phi^{m+1,d}(r)}{\phi^{m,d}(r)}\right). \tag{2.17}$$

2.3.4.6 Fuzzy Entropy (FuzEn)

In the ApEn and SampEn algorithms, the similarity of vectors is based on the Heaviside function as follows:

$$\theta(\text{ChebDist}[\mathbf{x}_\Lambda^m, \mathbf{x}_a^m], r) = \begin{cases} 1, & \text{ChebDist}[\mathbf{x}_\Lambda^m, \mathbf{x}_a^m] \leq r \\ 0, & \text{ChebDist}[\mathbf{x}_\Lambda^m, \mathbf{x}_a^m] > r \end{cases} \tag{2.18}$$

The Heaviside function can be considered as a conventional two-state classification method. However, in real world applications, boundaries between classes may be vague, and it is difficult to determine whether an input pattern belongs totally to a class. To deal with the problem, the concept of "fuzzy sets" was introduced [102]. Based on this concept, the fuzzy entropy (FuzEn) algorithm was proposed as follows [103].

Like SampEn, for the univariate time series $\mathbf{x}$, all the template vectors $\mathbf{x}_\Lambda^{m,d}$ ($\Lambda = 1, 2, \ldots, N - (m-1)d$) are first created according to Equation (2.14). Next, the distance between each of $\mathbf{x}_\Lambda^{m,d}$ and $\mathbf{x}_a^{m,d}$ is defined as $\Delta_{\Lambda,a} = \text{ChebDist}[\mathbf{x}_\Lambda^{m,d} - x0(\Lambda), \mathbf{x}_a^{m,d} - x0(a)], \Lambda \neq a$, where $x0(\Lambda)$ is the average of $\{x_\Lambda, x_{\Lambda+d}, \ldots, x_{\Lambda+(m-1)d}\}$ to remove the baseline [103]. Given a FuzEn power n_f and threshold *r*, the similarity degree is calculated through a fuzzy function as $\exp\left(-\frac{\Delta_{\Lambda,a}^{n_f}}{r}\right)$. The function $\psi^{m,d}(n_f, r)$ is then calculated as follows:

$$\psi^{m,d}(n_f, r) = \frac{1}{(N-md)(N-md-1)} \sum_{\Lambda=1}^{N-md} \sum_{a=1, a\neq\Lambda}^{N-md} \exp\left(-\frac{\Delta_{\Lambda,a}^{n_f}}{r}\right). \tag{2.19}$$

Finally, the FuzEn of the signal is defined as the negative natural logarithm of the ratio of $\psi^{m,d}(n_f, r)$ and $\psi^{m+1,d}(n_f, r)$ (computed following the same procedure for the embedding dimension $m+1$):

$$\text{FuzEn}(\mathbf{x}, m, r, n_f, d) = -\ln\left(\frac{\psi^{m+1,d}(n_f, r)}{\psi^{m,d}(n_f, r)}\right). \tag{2.20}$$

2.3.4.7 Parameters of Sample and Fuzzy Entropy Methods

There are three main parameters for the SampEn and FuzEn methods, including the embedding dimension *m*, threshold *r*, and the time delay d. It is advisable to set $d > 1$ for oversampled signals. However, some information regarding the frequency of time series may be ignored and the phenomenon of aliasing may also occur for $d > 1$ [18]. Thus, like previous studies about univariate entropy methods [16, 103], $d = 1$ is set for simplicity.

The parameter *r* is chosen to balance the quality of logarithmic likelihood estimates with the loss of signal information. When *r* is too small, poor conditional probability estimates are achieved. Furthermore, to avoid the effect of noise on data, larger *r* is recommended. In contrast, for a large *r* value, too much detailed data information is lost. Therefore, a trade-off between large and small *r* values is needed [16, 103]. Lake *et al.* proposed an approach to optimally select *r* [104]. However, as it is needed to calculate SampEn for a ranges of *r* and pick the value that optimizes an efficiency metric, this may be too time-consuming [105]. To alleviate this problem, a method based on the heuristic stochastic model was proposed to automatically determine *r* [105]. However, this approach still considers a number of *r* values leading to the computational burden. In the literature, it is common to set the threshold *r* as a constant (usually between 0.1 to 0.3) multiplied by the standard deviation (SD) of the original signal [16, 21, 103]. This strategy makes SampEn a scale-invariant measure [16, 104].

The embedding dimension *m* is the length of sequences to be compared. Larger *m* allows more detailed reconstruction of the dynamic process, while a large value of *m* is unfavourable because of the need of a very large number of sample points ($10^m - 20^m$), which is hard to meet for physiological data [16, 103]. For FuzEn, n_f illustrates the gradient of the boundary of the exponential function. For the sake of convenience, n_f is often equal to 2 [103].

2.3.4.8 Univariate Multiscale (Sample) Entropy (MSE)

The traditional entropy methods for quantifying irregularity of physiological signals may fail to account for the multiple temporal scales inherent in such series [21]. To this end, multiscale entropy (MSE) was proposed [39].

The algorithm of MSE includes two steps:

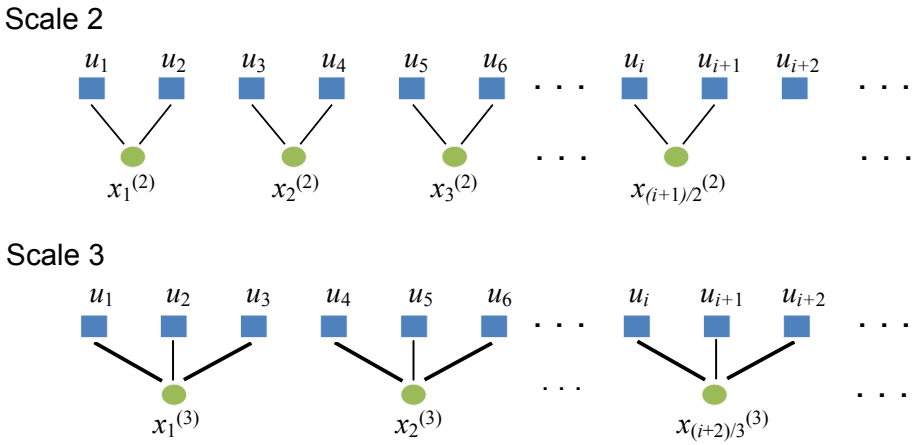


Figure 2.4: Demonstration of the coarse-graining of a sequence for scale factor $\tau = 2$ and $\tau = 3$.

1. Univariate coarse-graining process: Assume we have a univariate signal of length L: $\mathbf{u} = \{u_1, u_2, \ldots, u_i, \ldots, u_L\}$. In the coarse-graining process, as can be seen in Figure 2.4, the original signal $\mathbf{u}$ is first divided into non-overlapping segments of length τ, named scale factor. Then, the average of each segment is calculated to derive the coarse-grained signals as follows [21]:

$$x_j{}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} u_i, \quad 1 \leq j \leq \left\lfloor \frac{L}{\tau} \right\rfloor = N. \tag{2.21}$$

2. Calculation of SampEn: The SampEn value is calculated for each coarse-grained signal $\mathbf{x}^{(\tau)} = \{x_j^{(\tau)}\}$.

The MSE has been widely used in different research fields, including biomedical applications [106]. MSE has been successfully employed to, for example, diagnose depression using physiological signals, including heart rate, speech recordings, and EEGs [107], detect Parkinson's disease using EEGs [108], and characterize AD by the use of MEGs [109].

2.3.4.9 Multivariate Multiscale (Sample) Entropy (mvMSE)

MSE does not take into account the spatial domain for multi-channel signals [23]. To this end, multivariate entropy techniques were developed to deal with both the time and spatial patterns of multi-channel time series [23, 110, 111].

Many physiological activities and non-physiological signals include interactions among various recording sources. Therefore, it is expected that measures with different origins are considered in a multivariate way [112, 113]. Recent developments in sensor technology enabling routine recordings of multi-channel signals have led to an increasing popularity of this kind of analyses on data [110–112].

Multivariate multiscale entropy (mvMSE) is a powerful non-linear complexity measure taking into account both the spatial and time domains [23]. The mvMSE algorithm includes two steps:

1. Multivariate coarse-graining process: Assume we have a p-channel time series $\mathbf{U} = \{u_{k,i}\}_{k=1,2,\dots,p}^{i=1,2,\dots,L}$ of length L. In the mvMSE algorithm, for each channel, the original signal is first divided into non-overlapping segments of length τ. Next, for each channel, the average of each segment is calculated to derive the coarse-grained signals as follows [23]:

$$x_{k,j}{}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} u_{k,i},\ 1 \leq j \leq \left\lfloor \frac{L}{\tau} \right\rfloor = N\ ,\ 1 \leq k \leq p, \tag{2.22}$$

   where N denotes the length of the coarse-grained signal. The procedure of multivariate coarse-graining process is shown in Figure 2.5.

2. Calculation of multivariate SampEn (mvSE): For each scale factor τ, the mvSE of the multivariate coarse-grained signal is calculated. To take into account both the spatial and time domains, multi-channel embedded vectors are generated according to the multivariate embedding theory [114]. The multivariate embedded reconstruction of $\mathbf{X} = \{x_{k,j}\}_{k=1,2,\dots,p}^{j=1,2,\dots,N}$ is defined as [23]:

$$\begin{aligned} X_{\mathbf{m}}(b) = [&x_{1,b}, x_{1,b+d_1}, \dots, x_{1,b+(m_1-1)d_1}, \\ &x_{2,b}, x_{2,b+d_2}, \dots, x_{2,b+(m_2-1)d_2}, \dots, \\ &x_{p,b}, x_{p,b+d_p}, \dots, x_{p,b+(m_p-1)d_p}], \end{aligned} \tag{2.23}$$

   where $\mathbf{m} = [m_1, m_2, \dots, m_p]$ and $\mathbf{d} = [d_1, d_2, \dots, d_p]$ denote the embedding dimension and the time delay vectors, respectively. Note that the length of $X_{\mathbf{m}}(b)$ is $\sum_{k=1}^{p} m_k$. For simplicity, we assume $d_k = d$ and $m_k = m$, that is, all the embedding dimension values and all the delay values are equal.

   For p-variate time series $\{\mathbf{x}_k\}_{k=1}^{p}$, the mvSE algorithm, as a natural extension of standard

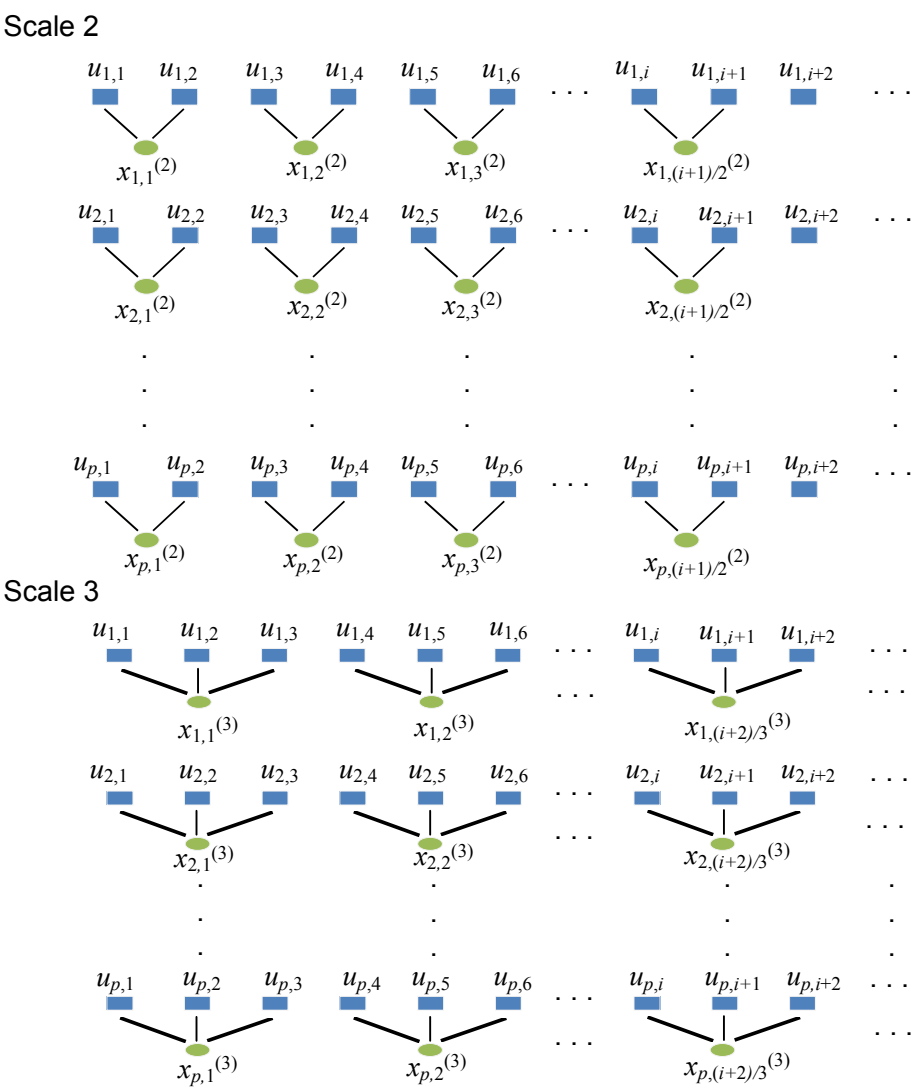


Figure 2.5: Demonstration of the multivariate coarse-graining of a multivariate sequence for scale factor $\tau = 2$ and $\tau = 3$.

SampEn, is described as follows [111]:

(a) Form multivariate embedded vectors $X_{\mathbf{m}}(b) \in R^m$, where $b = 1, 2, ..., N - n$ and $n = \max\{\mathbf{m}\} \times \max\{\mathbf{d}\}$.

(b) Calculate the Chebyshev distance (ChebDist) between any two composite delay vectors $X_m(b)$ and $X_m(\beta)$ as the maximum norm.

(c) For a given $X_{\mathbf{m}}(b)$ and a threshold *r*, count the number of instances Γ_b, where ChebDist$[X_{\mathbf{m}}(b), X_{\mathbf{m}}(\beta)] \leq r,\ b \neq \beta$. Next, calculate the frequency of occurrence as $\Theta_b^m(r) = \frac{1}{N-n}\Gamma_b$. Afterwards, define the global quantity as follows:

$$\Theta^m(r) = \frac{1}{N-n}\sum_{b=1}^{N-n}\Theta_b^m(r). \tag{2.24}$$

(d) Extend the dimensionality of the multivariate delay vector in Equation (2.23) from *m* to $(m+1)$ (while keeping the dimension of the other variables unchanged). This is done in *p* different ways, as from a space defined by the embedding vector $\mathbf{m} = [m_1, \ldots, m_k, \ldots, m_p]$ the system evolves to any space for which the embedding vector is $[m_1, \ldots, m_k + 1, \ldots, m_p]$ $(k = 1, \ldots, p)$. Therefore, a total of $p(N - n)$ vectors $X_{\mathbf{m}}(b + 1)$ are calculated, where $X_{\mathbf{m}}(b + 1)$ shows any embedded vector upon increasing the embedding dimension from m_k to m_k+1, while the embedding dimension of the other data channels is kept unchanged.

(e) Repeat steps (a) to (c) to find $\Theta_b^{(m_k+1)}(r)$ for $1 \leq k \leq p$. Next, calculate $\Theta_b^{(m+1)}(r)$, which denotes the average over all the *p* representations of $\Theta_b^{(m_k+1)}(r)$. Afterwards, find $\Theta^{(m+1)}(r)$, which stands for the average over all $N - n$ of $\Theta_b^{(m_k+1)}(r)$ in an $(m + 1)$-dimensional space.

(f) Finally, mvSE is defined as:

$$\text{mvSE}(\mathbf{X}, \mathbf{m}, \mathbf{d}, r) = -\ln\left(\frac{\Theta^{(m+1)}(r)}{\Theta^m(r)}\right). \tag{2.25}$$

mvMSE has a growing appeal and broad use in the context of biomedical signal analysis [23, 32,115,116]. It has been successfully used in a number of applications, such as, to characterise EEG signals in AD [24, 115], to analyze the multivariate cardiovascular time series [116], to characterize focal and non-focal EEG time series [32], to analyze the complexity of interbeat

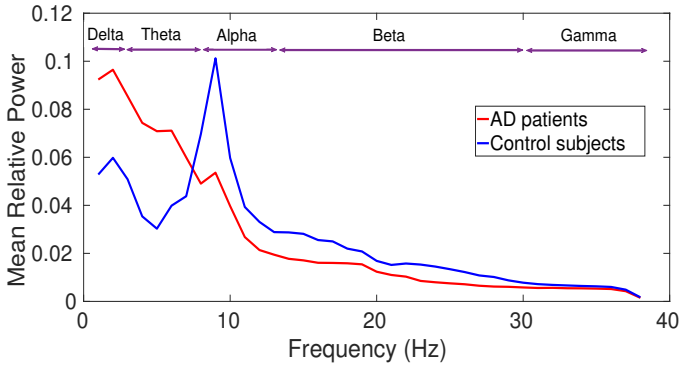


Figure 2.6: Mean relative power spectra for 11 AD patients' vs. 11 controls' EEG recordings.

interval and interbreath signals [23], and to analyze the postural fluctuations in fallers and non-fallers older adults [117].

2.4 Selected Previous Literature on EEG and MEG Analysis in AD

According to the review [6], there are three main effects of AD on EEG (or MEG) signals, including slowing of EEG, perturbations in EEG synchrony, and reduction of EEG complexity. These effects are described as follows:

2.4.1 Slowing of the EEG and MEG Signals of AD Patients

A sizeable number of studies suggested that MCI and AD cause EEG signals to slow down (see, e.g., [6, 118–123]). AD affects different frequency bands of EEGs and MEGs in various ways. An increase of spectral, total, or relative power for delta and theta waves, and a decrease of power for alpha, beta, and gamma waves were reported in AD patients in comparison with healthy age-matched control subjects [6, 119, 120, 124, 125].

To show the slowing of EEGs for AD patients, the relative power of each frequency band, averaged for all 16 channels, were calculated for the EEG dataset described in Subsection 6.1.1. Averaged power spectra for 11 AD patients vs. 11 controls are demonstrated in Figure 2.6. The results suggest that the AD patients have more relative power in lower frequency bands (delta and theta), while the relative power values for controls' signals are larger in higher frequency bands (alpha, beta, and gamma). Therefore, it demonstrates the slowing of EEGs of AD patients.

Furthermore, a number of studies have shown slowing of the EEG recordings by the use of spectral-based approaches, such as mean frequency [126, 127], peak frequency [126, 128], and transition frequency [126, 128]. The MEG spectral patterns in AD have also been studied based those metrics [125, 129, 130]. The results are in agreement with the previous EEG-based studies illustrating a decrease in characteristic frequencies in AD [125, 129, 130], which was accentuated by the severity of the disease.

2.4.2 Decreased EEG and MEG Synchrony in AD Patients

Synchrony measures seek to quantify the relationship between two time series or sensors [6]. A sizeable number of studies showed a decreased EEG synchrony in MCI and AD patients under rest conditions (spontaneous EEG signals): the statistical dependence between spontaneous EEG time series recorded from different channels looks to be lower in MCI and AD patients in comparison with age-matched controls [71]. In this field, a wide range of linear and nonlinear synchrony metrics in time and/or frequency domains have been applied, e.g., the Pearson correlation coefficient [71, 131], coherence [71, 132, 133], Granger causality [71, 134], information theoretic-based approaches [71, 135], phase synchrony indices [71, 119, 136], stochastic event synchrony [71, 136], and state space-based synchrony techniques [71, 119].

It was recently illustrated that a number of classical and recently developed similarity methods, although proposed from different perspectives, lead to similar results [71]. The results are strongly correlated (or anti-correlated) with the correlation coefficient, and thus, provide little complementary information regarding EEG synchrony. The phase synchrony indices, Granger causality, and stochastic event synchrony measures are those that are only weakly correlated with the correlation coefficient. Furthermore, these three families of synchrony measures are mutually uncorrelated, and hence, each of them looks to capture a specific kind of interdependence. Therefore, it is not essential to apply multiple synchrony measures. That is, to help AD diagnosis with higher accuracy, it is sufficient to combine one measure from each of the three main families [71, 71].

Furthermore, a number of studies also illustrated decreased MEG synchrony in AD patients under rest conditions [137–139]. Nevertheless, a comprehensive study on various similarity approaches for MEGs in AD is needed.

2.4.3 Reduced Complexity and Irregularity of the EEG and MEG Signals of AD Patients

Various studies inspected whether the irregularity or complexity of EEG recordings is perturbed by MCI or AD [10, 109, 140–143]. The following approaches were applied in this context: approximate entropy (ApEn) [140], sample entropy (SampEn) [10], Tsallis entropy [141], MSE [109, 144], auto-mutual information [10, 140], LZC [145], universal compression [141], FD [142], correlation dimension [143], and largest Lyapunov exponent [143].

The first nonlinear methods used to analyze EEGs were D_2 [85] and L_1 (first Lyapunov exponent) [81]. A number of studies showed that D_2 may provide interesting information from mental diseases using both time-delay [146–148] and spatial embedding approaches [149] to reconstruct the attractor. In AD, loss of synapses and neurons may lead to smaller D_2 values [27, 147, 149]. While D_2 deals with the statistical characteristic of a system, L_1 is a relatively dynamic metric, as it defines the divergence of trajectories that begin at similar initial states [81]. In fact, L_1 is frequently interpreted as a metric of the flexibility of a system to reach different states from almost identical initial states. Several studies [147, 149] represented the potential usefulness of L_1 in the context of AD characterization.

In spite of the interesting findings obtained by D_2 and L_1, the application of these nonlinear approaches to physiological time series has two main problems [92, 150, 151]:

1. The length of signals needed to obtain meaningful results is usually beyond the length of physiological data that can be collected experimentally [150]. Additionally, the D_2 algorithm needs stationary signals, something that is almost impossible to ensure when dealing with physiological signals [151].
2. These techniques are very sensitive to noise [92].

These shortcomings caused researchers to investigate other nonlinear approaches for the characterization of brain signals. Several of these alternatives are based on the concepts of irregularity and complexity [15–17, 39]. Among them, LZC is a widespread metric to analyze biological signals to evaluate the complexity of finite sequences, with larger values corresponding to more complex signals data [82]. The results obtained by LZC-based approaches showed that the EEG and MEG signals in AD patients are less complex than

those in healthy controls [10, 145, 152, 153]. It is worth noting that although LZC led to some interesting findings, they return high values when they are applied to random time series [21]. Therefore, such kind of statistics would not be strictly complexity measures, rather irregularity estimators [21, 154]. From this view point, a complexity measure should vanish for both completely regular and completely random time series [21]. In this Thesis, we adopt this view point of *complexity* (see Section 2.5).

To quantify the complexity of time series according to this view, MSE was proposed based on the calculation of SampEn over multiple temporal scales [39]. The MSE technique was employed to characterize EEGs in AD [155, 156]. For theses studies, it was found that the EEGs for AD patients have smaller MSE values at lower scale factors, whereas the AD patients' signals have larger MSE values at higher temporal scales. Low scale factors mean the time scales that are smaller or equal to the scale of crossing point of the curves for AD patients vs. controls. In contrast, high scale factors denote the temporal scales that are larger than the scale of crossing point of the curves for AD patients vs. controls. For more information please see Section 6.2.1.2.

Similarly, Escudero *et al.* proposed an approach based on MSE to characterize MEGs in AD [109]. It was illustrated that the MEGs recorded from the age-matched healthy subjects are more irregular at lower scale factors, while the recordings for AD patients are more irregular at higher scale factors [109]. The multivariate multiscale permutation entropy (mvMPE) and multivariate MSE (mvMSE) techniques were used for three different groups of individuals: MCI subjects, AD patients, and age-matched healthy controls in two studies [24, 25]. It was found that the EEG signals for AD patients are less complex than those for MCI. The latter ones are less complex than the EEGs for controls. However, since the dataset used in these studies included few subjects and channels, the results may not be completely reliable [24, 25]. Of note is that the exact number of subjects for each group was not mentioned.

2.5 Concepts of Irregularity and Complexity

A complex system denotes a system with a number of components intricately entwined altogether (e.g., the subway network of the New York City). In the analogy of human physiology, complexity refers to many components of the body interacting at levels ranging

from molecules, cells, to organs [157]. The neural networks in the brain, with their structure intermediate between order and randomness are considered as an example of complexity in the physiological area [35]. In fact, the human brain is a complex network of coupled and interacting subsystems [35, 158]. This complexity is due to the interactions of huge number of neurons operating over a wide range of temporal and spatial scales. These interactions enable the brain to adapt to the constantly changing environment and to perform different types of mental functions [157]. Numerous complex phenomena in nature are due to nonlinear phenomena. Therefore, nonlinear approaches have been broadly used in the analysis of complex systems [35].

In the context of the analysis of physiological signals following Costa's framework [21, 39], the complexity concept stands for "meaningful structural richness", which may be in contrast with regularity measures defined from classical entropy approaches such as LZC, SampEn, permutation entropy (PerEn), and dispersion entropy (DispEn) [18, 21, 22, 159]. In fact, these classical entropy techniques assess repetitive patterns and return maximum values for completely random processes [21, 22, 160]. However, a completely ordered signal with a small entropy value or a completely disordered signal with maximum entropy value is the least complex [21, 22, 161]. For example, white Gaussian noise is more irregular than $1/f$ noise although the latter is more complex because $1/f$ noise contains long-range correlations and its $1/f$ decay produces a fractal structure in time [21, 22, 161].

Some diseased individuals' time series, compared with those for healthy subjects, are associated with the emergence of more regular behavior, leading to lower entropy values [21, 154]. For example, the entropy values for focal EEG recordings are smaller than those for non-focal ones (for more information, please see Subsection 3.3.2.1) [162, 163]. In contrast, certain pathologies, such as cardiac arrhythmias, are associated with highly erratic fluctuations with statistical characteristics resembling uncorrelated noise. For instance, the EEG signals recorded from Parkinson's disease patients in stage slow-wave sleep are more irregular than those for controls [108]. The entropy values of these signals are higher than those of healthy individuals, even though the healthy individuals' time series show more physiologically complex adaptive behavior [21, 164]. To provide a unified framework for the evaluation of impact of diseases in physiological signals, MSE [21] was proposed to quantify the complexity of signals over multiple temporal scales.

In brief, the concept of complexity for univariate physiological signals builds on the following three hypotheses [21, 22]:

- The complexity of a biological or physiological time series indicates its ability to adapt and function in ever-changing environment.
- A biological time series requires to operate across multiple temporal and spatial scales and so, its complexity is similarly multiscaled and hierarchical.
- A wide class of disease states, in addition to ageing, decrease the adaptive capacity of the individual either more deterministic or more random, thus reducing the information carried by output variables. Therefore, the multiscale-based methods focus on quantifying the information expressed by the physiologic dynamics over multiple temporal scales.

Followed by the univariate complexity hypotheses, the multivariate multiscale entropy-based analysis is interpreted based on [23, 110]:

- The multivariate time series $\mathbf{X}$ is more complex than the multivariate time series $\mathbf{Y}$, if for the most temporal scales, the mvSE measures for $\mathbf{X}$ are larger than those for $\mathbf{Y}$.
- A monotonic fall in the multivariate entropy values along the temporal scale factors shows that the signal only includes useful information at the smallest scale factors.
- A multivariate signal illustrating long-range correlations and complex creating dynamics is characterized by either a constant mvSE or this demonstrates a monotonic rise in mvSE with the temporal scale factor.

2.6 Summary

Entropy-based approaches, FD techniques, and LZC quantify the irregularity of time series. FD methods are static metrics of attractors that do not provide any information on the evolution of trajectories over time. In contrast, Lyapunov exponents, entropy metrics, and LZC are considered as dynamic measures of an attractor [35]. Though SampEn and Lyapunov exponents were independently defined, it was shown that these two metrics are related to some extent. SampEn is based on KS entropy. The Margulis-Ruelle inequality demonstrates that the KS entropy is less than or equal to the sum of positive Lyapunov exponents [165]. The

Table 2.2: Comparison of nonlinear approaches in terms of measurement quantified, sensitivity to noise, and ability for characterization of short signals.

Methods	Measurement quantified	Sensitivity to noise	Short signals
SampEn [16] and FuzEn [103]	irregularity	no	Unreliable or undefined
MSE [39] and mvMSE [111]	irregularity and complexity	no	undefined or unreliable
LZC [82]	irregularity	no	yes
FD methods	irregularity	yes	reliable
Lyapunov exponents [81]	rate of separation of infinitesimally close trajectories	yes	unreliable

opposite inequality was proved by Pesin under some restricted conditions as well [166]. These two formulae demonstrate that, under some conditions, the two different techniques determine the same type of chaotic behaviour [165]. Therefore, all the nonlinear methods described in this Chapter are relevant to irregularity of signals.

The characteristics and limitations of the main nonlinear methods used for physiological signals are demonstrated in Table 2.2. The only techniques quantifying both the complexity and irregularity of time series are MSE and mvMSE. Accordingly, the first purpose of this Thesis is to address the shortcoming of MSE and mvMSE for characterization of short signals. Another aim of this Thesis is to introduce faster entropy-based methods that can be used for real-time applications.

Chapter 3

Univariate Entropy Methods

Sample entropy (SampEn) and fuzzy entropy (FuzEn), which are based on the concept of conditional entropy (ConEn), are metrics to quantify the irregularity of physiological signals [16, 103]. However, they have two deficiencies: 1) SampEn and FuzEn values are respectively undefined and unreliable for short time series (e.g., 100 sample points for embedding dimension 2 - see Section 4.3); and 2) their computational cost is O(N^2). To alleviate these deficiencies, permutation entropy (PerEn) was proposed based on the permutation patterns, or order relations, among amplitudes of a time series and Shannon entropy (ShEn) [17].

However, PerEn has three key shortcomings. First, the original PerEn assumes a signal has a continuous distribution. Therefore, equal values would be rare and they could be ignored by ranking them based on the order of their emergence [28, 167]. However, while dealing with digitized signals with coarse quantization levels, it may not be appropriate to simply ignore them [28, 167]. Second, when a time series is symbolized based on the permutation patterns (Bandt-Pompe procedure), only the order of amplitude values is taken into account and some information with regard to the amplitudes may be ignored [28, 167]. Third, PerEn is considerably sensitive to noise, since a small change in amplitude value may vary the order relations among amplitudes [18]. To address the first and second shortcomings, amplitude-aware PerEn (AAPerEn) is proposed in this Thesis [28].

To reduce the sensitivity of AAPerEn to noise, to address the undefined SampEn values, and to decrease the computation time of SampEn and FuzEn at the same time, dispersion entropy (DispEn) on the basis of our introduced dispersion patterns and the Shannon's definition of entropy is proposed [18, 29].

As another entropy approach taking into account the frequency of signals, like PerEn, frequency-based DispEn (FDispEn) is introduced as well [29]. FDispEn is based on ShEn and the difference between adjacent elements of dispersion patterns, named frequency-based dispersion patterns. To compare the existing and proposed techniques, a set of univariate synthetic time series based on several straightforward signal characteristics and two publicly-available datasets are used herein. The descriptions of theses data are explained in the next Section.

3.1 Univariate Signals for Evaluation

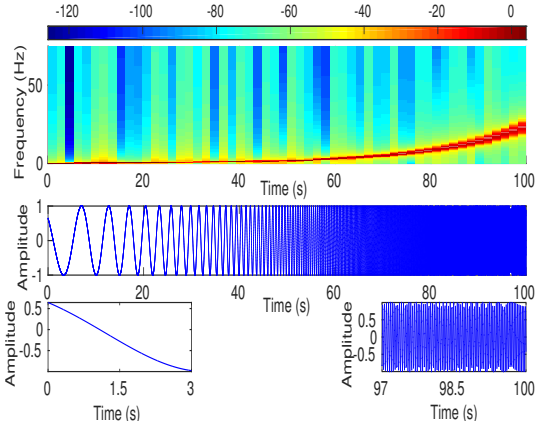
In this Section, the synthetic and real signals to investigate the behaviour of univariate entropy approaches are explained.

3.1.1 Univariate Synthetic Signals

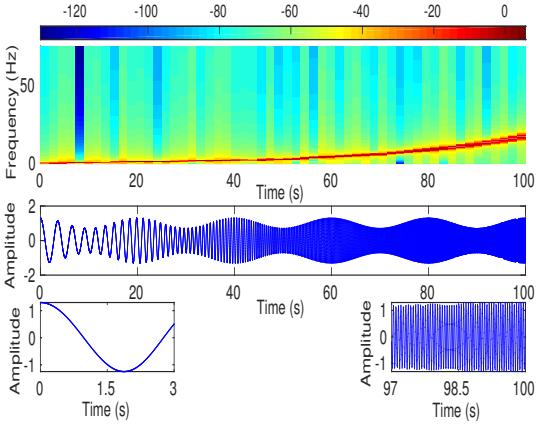
A number of time series and their interpretability in terms of classical signal processing concepts such as frequency, amplitude, noise power, and signal bandwidth are described. These time series have been employed to evaluate the Lempel-Ziv complexity measure [168], DispEn [18], improved multiscale PerEn [169], and auto-mutual information [170]. All the synthetic signals have a sampling frequency (f_s) of 150 Hz and a length of 100 s. Therefore, they have 15,000 sample points. The time plots of these synthetic signals, and their corresponding spectrograms, and two zooms (for each kind of signal) on their start and end, to demonstrate the changes in their characteristics, appear in Figure 3.1.

3.1.1.1 Univariate Entropy Methods vs. Frequency

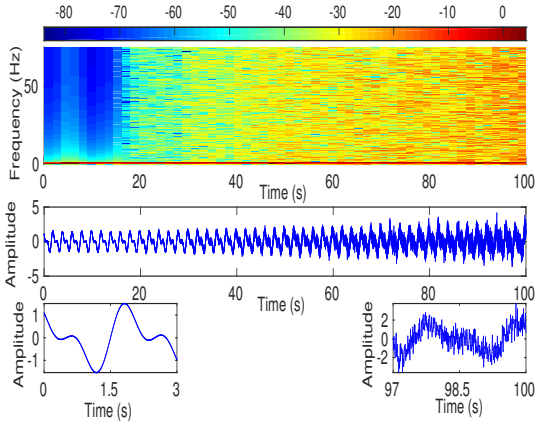
Change in amplitude and frequency of biomedical time series is broadly used to diagnose some diseases [37, 171]. Thus, to clarify the bahavior of entropy methods when the amplitude or frequency of sinusoidal signals are changed, two kinds of non-stationary synthetic signals were created. The first one is a constant amplitude chirp signal whose frequency is swept logarithmically from 0.1 Hz to 30 Hz in 100 s. The second kind of signal, whose frequency is swept logarithmically from 0.25 Hz to 5 Hz in 100 s, was generated by modulating the amplitude of the chirp signal by a pure sinusoid. Figures 3.1(a) and 3.1(b) demonstrate the



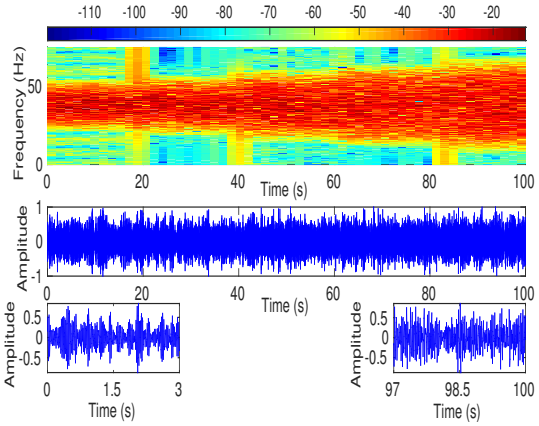
(a) Chirp signal with constant amplitude



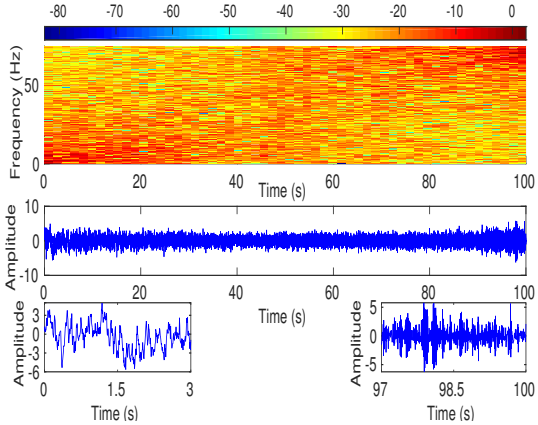
(b) Amplitude modulated chirp signal



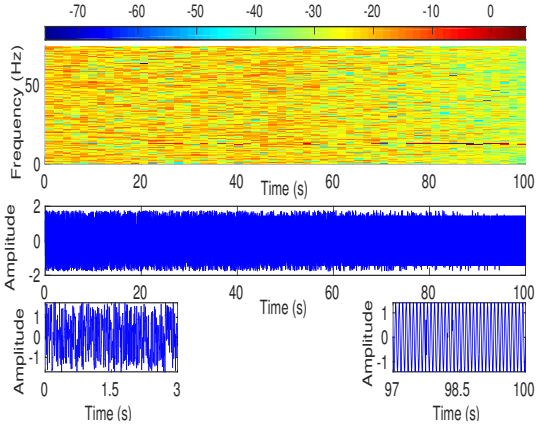
(c) Periodic signal with increasing additive noise power



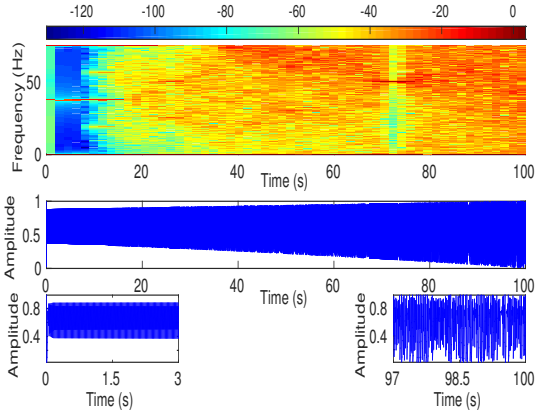
(d) Colored noise with increasing bandwidth



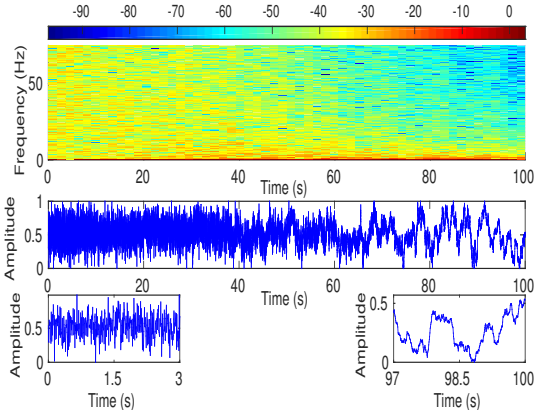
(e) AR(1) process with variable parameter



(f) MIX process evolving from randomness to periodic oscillations



(g) Logistic map signal



(h) Signal with spectral density changing from $\frac{1}{f^0}$ (white noise) to $\frac{1}{f^2}$ (brown noise)

Figure 3.1: Spectrograms, time plots and zoom views on the first and last time intervals of the synthetic signals used in this study. (a) Chirp signal with constant amplitude (frequency changes from f =0.1 Hz to f =30 Hz). (b) Amplitude modulated chirp signal (frequency changes from f =0.25 Hz to f =5 Hz). (c) Periodic signal with increasing additive noise power. (d) Colored noise with increasing bandwidth. (e) AR(1) process with variable parameter ρ changing linearly from $+0.9$ to -0.9. (f) MIX process evolving from randomness to periodic oscillations. (g) Logistic map with parameter α changing from 3.5 to 3.99. (h) Signal including 12 segments of different kinds of noise with spectral density $\frac{1}{f^\Gamma}$, which Γ changes from 0 (white noise) to 2 (brown noise) from the first to twelfth segment, respectively. Red corresponds to high power, and blue corresponds to low power.

constant and amplitude modulated chirp signals, respectively.

3.1.1.2 Univariate Entropy Methods vs. Noise Power

Physiological signals are frequently corrupted by different kinds of noise, such as additive white Gaussian noise (WGN) [172]. Additive WGN is also considered as a basic statistical model used in information theory to mimic the effect of random processes that occur in nature [173]. In order to understand the relationship between univariate entropy methods and the level of noise affecting periodic time series, we generated an amplitude-modulated periodic signal with additive WGN with diverse power. First, we created signal as an amplitude-modulated sum of two cosine waves with frequencies at 0.5 Hz and 1 Hz. The first 20 s of this series (100 s) does not have any noise. Then, WGN was added to the time series [169]. Figure 3.1(c) shows this time series.

3.1.1.3 Univariate Entropy Methods vs. Bandwidth of Colored Noise

In order to determine the relationship between univariate entropy methods and noise bandwidth, a 100 s time series composed of five segments of colored noise with increasing bandwidth is employed. The frequency spectra of the colored noises are all centered at $f_s/4$ and their bandwidth increases from $f_s/15$ to $f_s/3$ in five equal steps. Figure 3.1(d) depicts this signal.

3.1.1.4 Univariate Entropy Methods vs. Spectral Content of Colored Noise

Autoregressive (AR) methods have been used in many studies to model biomedical signals, such as electroencephalograms (EEGs) and electrocardiograms (ECGs), by representing the time series at each channel as a linear combination of the signal at previous time points [174, 175]. In order to investigate the dependence between the univariate entropy techniques and the spectral content of colored noise, an AR process of order 1, AR(1), was generated varying the model parameter, ρ, linearly from $+0.9$ to -0.9. Its energy therefore moved from low to high frequencies. When ρ was equal to 0, the sequence corresponded to WGN, in the center of the synthetic signal. Figure 3.1(e) depicts the signal's corresponding spectrogram, time plot and zoom views.

3.1.1.5 Univariate Entropy Methods vs. Changes from being Non-deterministic to Deterministic

Signals created by biological systems most likely include deterministic and stochastic components [39]. Hence, to inspect how univariate entropy methods change when a stochastic sequence progressively turns into a periodic deterministic signal, we generated a MIX process employed by [15, 144, 176]. It is defined as follows:

$$MIX_j = (1 - z_j)x_j + z_j y_j, \quad 1 \leq j \leq N \tag{3.1}$$

where N is the length of the signal vectors $\mathbf{z} = \{z_j\}$, $\mathbf{MIX} = \{MIX_j\}$, and $\mathbf{y} = \{y_j\}$. $\mathbf{z}$ denotes a random variable which equals 1 with probability p and equals 0 with probability $1 - p$. $\mathbf{x}$ shows a periodic time-series created by $x_j = \sqrt{2}\sin(2\pi j/12)$, and $\mathbf{y}$ is a uniformly

distributed series on $\left[-\sqrt{3}, \sqrt{3}\right]$ [144, 176]. The synthetic time series was based on a MIX process that *p* varied from 1 to 0 linearly. Therefore, this signal, depicted in Figure 3.1(f), evolved from randomness to orderliness.

3.1.1.6 Univariate Entropy Methods vs. Changes from Periodicity to Non-periodic Nonlinearity

Studies on physiological time series frequently involve relatively short epochs of signals containing informative periodic or quasi-periodic components [21, 177, 178]. Moreover, empirical evidence identifies nonlinear, in addition to linear, behavior in some physiological signals [10, 20, 35, 171]. Accordingly, to find the dependence of univariate entropy approaches with changes from periodicity to non-periodic nonlinearity, a logistic map is used. This analysis is relevant to the model parameter α as: $x_j = \alpha x_{j-1}(1 - x_{j-1})$, where the signal $\mathbf{x} = x_j\,(j = 1, \ldots, N)$ was generated varying the parameter α from 3.5 to 3.99. When α is equal to 3.5, the signal oscillates among four values. For $3.5 < \alpha < 3.57$, the time series is periodic and the number of values doubles progressively. For α between 3.57 and 3.99, the time series is chaotic, although it has windows of periodic behavior (e.g., $\alpha \approx 3.8$) [170, 176, 179]. Figure 3.1(g) depicts the spectrogram of the series, its time plot and zoom views.

3.1.1.7 Univariate Entropy Methods vs. Noise Signals

Noise is frequently considered as an unwanted component or disturbance to a system or data, whereas recent studies have shown that noise can play a beneficial role in systems [180, 181]. In any case, it has been evidenced that noise is an essential ingredient in systems [180, 181]. White, pink, and brown noise are three well-known noise [180, 181]. White noise is a random signal having equal energy across all frequencies. The power spectral density of white noise is as $S(f) = C_w$, where C_w is a constant [181]. Pink and brown noise are random processes suitable for modelling evolutionary or developmental systems [182]. The power spectral density $S(f)$ of pink and brown noise are as $\frac{C_p}{f}$ and $\frac{C_b}{f^2}$, respectively, where C_p and C_b are constants [181, 182].

To evaluate the ability of existing and introduced entropy methods, a 100 s time series composed of 12 equal length segments with power spectral density $\frac{1}{f^\Gamma}$, where Γ linearly

increases from 0 to 2 was created (*i.e.,* $\Gamma = \frac{2k}{11}$, $(0 \leq k \leq 11)$ respectively for the 1^{st} to 12^{th} segment). As Γ changes from 0 to 2, the signal includes white, pink, and brown noise. Figure 3.1(h) illustrates the spectrogram of the series, and its time plot and zoom views.

3.1.2 Real Biomedical Datasets

Entropy methods are widely used to characterize physiological signals, such as EEG, ECG, and blood pressure recordings [15, 18, 20, 167, 183]. To this end, two non-invasive EEGs [162] and blood pressure datasets [184] are used in this Chapter to distinguish different kinds of dynamics of signals.

3.1.2.1 Dataset of Focal and Non-focal Brain Activity

Focal connectivity deficits were shown in Alzheimer's disease (AD) [185]. Accordingly, the ability of univariate and multivariate entropy methods to discriminate focal from non-focal signals is evaluated by the use of an EEG dataset (publicly-available at http://ntsa.upf.edu/) [162]. The dataset includes 5 patients and, for each patient, there are 750 focal and 750 non-focal bivariate time series. The length of each signal was 20 s with sampling frequency of 512 Hz (10240 samples). For more information, please, refer to [162]. All subjects gave written informed consent that their signals from long-term EEG might be used for research purposes [162]. Before computing the entropies, the time series were digitally filtered using a Hamming window FIR band-pass filter of order 200 and cut-off frequencies 0.5 Hz and 40 Hz, a band typically used in the analysis of brain activity.

3.1.2.2 Fantasia Dataset

The association between blood pressure and AD was shown in a number of studies [186]. To this end, and to investigate the ability of entropy-based methods to characterize very long physiological recordings, the Fantasia dataset (publicly-available at http://www.physionet.org) is used to distinguish elderly from young subjects [184]. The dataset includes 10 young (21-34 years old) and 10 old (68-85 years old) rigorously-screened healthy individuals who underwent about 120 minutes of continuous supine resting while uncalibrated non-invasive blood pressure signals were recorded. Each group consisted of 5 women and 5 men [184]. All 20 individuals remained in an inactive state in sinus rhythm when watching the movie Fantasia (Disney, 1940)

to help to maintain wakefulness. For each subject, the time series were digitized at 250 Hz. Detailed information can be found in [184].

All 20 individuals provided written informed agreement and underwent a screening history, physical examination, routine blood count and biochemical analysis, electrocardiographic, and exercise tolerance test. Only healthy, nonsmoking individuals with normal exercise tolerance tests, no medical problems, and taking no medications were included to the research [184].

3.2 Entropy Methods based on Shannon Entropy (ShEn)

In this Section, the entropy methods based on ShEn [19], including PerEn, and our developed AAPerEn, DispEn and FDispEn are described.

3.2.1 Permutation Entropy (PerEn)

Assume we have a given time series vector of length N $\mathbf{x} = \{x_1, x_2, \ldots, x_j, \ldots, x_N\}$. For each time Λ, we embed the signal $\mathbf{x}$ in an m-dimensional space to obtain the reconstruction vectors $\mathbf{x}_{\Lambda}^{m,d} = \{x_{\Lambda}, x_{\Lambda+d}, \ldots, x_{\Lambda+(m-2)d}, x_{\Lambda+(m-1)d}\}$ for $\Lambda = 1, 2, \ldots, N-(m-1)d$, where m and d denote the embedding dimension and time delay, respectively. Next, each $\mathbf{x}_{\Lambda}^{m,d}$ is arranged in an increasing order, with integer indices from 0 to $m-1$, as follows:

$$\left\{x_{\Lambda+(\aleph_1-1)d}, x_{\Lambda+(\aleph_2-1)d}, \ldots, x_{\Lambda+(\aleph_{m-1}-1)d}, x_{\Lambda+(\aleph_m-1)d}\right\} \tag{3.2}$$

where $\aleph_*$ is the (time) index of the element in the reconstruction vector. There are $m!$ potential ordinal patterns or symbol sequences η_t ($1 \leq t \leq m!$), termed "motifs". Then, the occurrence of each of the order patterns η_t denoted as $f(\eta_t)$ is counted. For each η_t, the relative frequency $Pr(\eta_t)$ is calculated as follows:

$$Pr(\eta_t) = \frac{f(\eta_t)}{N-(m-1)d}. \tag{3.3}$$

Finally, the PerEn value is computed as follows [17]:

$$PerEn(\mathbf{x}, m, d) = -\sum_{t=1}^{m!} Pr(\eta_t) \cdot \ln Pr(\eta_t), \tag{3.4}$$

where $\ln$ denotes the natural logarithm. When all motifs have equal probability, the largest value of PerEn is obtained, which has a value of $\ln(m!)$. In contrast, if there is only one $Pr(\eta_t)$ different from zero, which illustrates a completely regular signal, the smallest value of PerEn is obtained [17, 167].

3.2.2 Amplitude-aware Permutation Entropy (AAPerEn)

The original PerEn has three main drawbacks:

1. PerEn considers only the order of amplitude values, and so, some information regarding the amplitude values themselves may be ignored [167]. For example, $\{1, 10, 2\}$ and $\{1, 3, 2\}$ have similar permutations, leading to the same motif "021" ($m = 3$) because the extent of the differences between sequential samples is not considered in the original definition of PerEn. Another example is the fact that both $\{1, 3, 2\}$ and $\{11, 13, 12\}$ have the same ordinal pattern "021" because the mean value of these samples is not considered in the original PerEn method.

2. When there are equal values in the vector of values, Bandt and Pompe [17] proposed ranking the possible equalities based on their order of emergence or solving this condition by adding noise. Considering the first alternative, for instance, the permutation pattern for both $\{1, 2, 4\}$ and $\{1, 4, 4\}$ are "012" ($m = 3$). As another example, assume $z1 = \{1, 2, 2, 2\}$ and $z2 = \{1, 2, 3, 4\}$. The PerEn with $m = 3$ of $z1$ is exactly the same as $z2$, both equalling 0 although, unlike $z1$, $z2$ is strictly ascending. Adding noise may not lead to a precise answer because, for example, $\{1, 4, 4\}$ has two possible permutation patterns as "012" and "021" and there are not any differences between them. It should be noted that this problem is particularly relevant for digitized signals with large quantization steps.

3. PerEn is sensitive to noise, since a small change in amplitude value may vary the order relations among amplitudes. For instance, small noise on $z3 = \{1, 2, 2.01\}$ may change

the motif from "012" to "021".

To address the first problem, we suggest adding a variable contribution, depending on amplitude, instead of a constant number to each level in the histogram representing the probability of each motif. That is, in PerEn, when a vector is assigned to a motif, the histogram bin corresponding to the motif is incremented by one. In contrast, in AAPerEn, the following relative normalized probability is added to the corresponding motif:

$$\frac{A}{m}\sum_{k=1}^{m}\left|x_{\Lambda+(k-1)d}\right| + \frac{1-A}{m-1}\sum_{k=2}^{m}\left|x_{\Lambda+(k-1)d} - x_{\Lambda+(k-2)d}\right| \tag{3.5}$$

where A is the adjusting coefficient related to the mean value and difference between consecutive samples to make the AAPerEn algorithm more flexible. A is in the range $[0, 1]$. Finally, the amplitude-aware version of Equation (3.3) is normalized by the total sum of the contributions. In brief, the pseudo code of the AAPerEn algorithm is shown in Algorithm 1.

begin $p(\eta_t^{m,d}) = 0$;
for $\Lambda = 1$ *to* $N - (m-1)d$ **do**
  for $t = 1$ *to* $m!$ **do**
    if $(\boldsymbol{x}_{\Lambda}^{m,d}) = \eta_t^{m,d}$ **then**
      $p(\eta_t^{m,d}) =$
      $p(\eta_t^{m,d}) + \left(\frac{A}{m}\sum_{k=1}^{m}\left|x_{\Lambda+(k-1)d}\right| + \frac{1-A}{m-1}\sum_{k=2}^{m}\left|x_{\Lambda+(k-1)d} - x_{\Lambda+(k-2)d}\right|\right)$
    end
  end
end
$p(\eta_t^{m,d}) = \frac{p(\eta_t^{m,d})}{\sum_{\Lambda=1}^{N-(m-1)d}\left(\frac{A}{m}\sum_{k=1}^{m}\left|x_{\Lambda+(k-1)d}\right| + \frac{1-A}{m-1}\sum_{k=2}^{m}\left|x_{\Lambda+(k-1)d} - x_{\Lambda+(k-2)d}\right|\right)}$;
$AAPerEn(\mathbf{x}, m, d) = -\sum_{\eta_k=1}^{\eta_k=m!} p(\eta_k)\cdot\ln p(\eta_k)$;
end

Algorithm 1: Pseudo code of the AAPerEn algorithm to take into account the mean value of amplitudes and differences between amplitude values.

The second drawback of PerEn is dealt with by the following approach. To start with, all potential permutations of similar states are considered. Then, all possible contributions coming from motifs with a tie (i.e., same state) are divided by the number of potential permutations of those equal states. For example, assume $\mathbf{x} = \{1, 2, 3, 2, 2\}$ and $m = 2$, leading to

"01" (full contribution associated with $\{1, 2\}$), "01" (full contribution associated with $\{2, 3\}$), "10" (full contribution associated with $\{3, 2\}$, "01" (associated with half the contribution of $\{2, 2\}$), and "10" (associated with the other half of the contribution of $\{2, 2\}$). In this case, the entropy value considering equal (i.e., half) contributions of the motif $\{2, 2\}$ is $AAPerEn(\mathbf{x}, 2, 1) = -(2.5/4)\ln(2.5/4) - (1.5/4)\ln(1.5/4) = 0.6616$. As an additional example, consider **x** and $m = 3$, leading to "012" (full contribution associated with $\{1, 2, 3\}$), "201" (associated with half the contribution of $\{2, 3, 2\}$), "021" (associated with the other half of the contribution of $\{2, 3, 2\}$), "012" (associated with half the contribution of $\{3, 2, 2\}$), "021" (associated with the other half of the contribution of $\{3, 2, 2\}$). In this case, the entropy value considering equal (i.e., half) contributions of the motifs $\{2, 3, 2\}$ and $\{3, 2, 2\}$ are $AAPerEn(\mathbf{x}, 3, 1) = -(1.5/3)\ln(1.5/3) - (1/3)\ln(1/3) - (0.5/3)\ln(0.5/3) = 1.0114$. In this way, the proposed method is able to discriminate the strictly ascending/descending from only ascending/descending sequences. It is worth noting that the combination of the first and second proposed algorithms makes to the proposed AAPerEn method.

3.2.3 Parameters of PerEn and AAPerEn

PerEn has two parameters, including the embedding dimension and time delay. For the time delay, as recommended in [17], $d = 1$ is set in this Thesis. The value of embedding dimension plays a key role in characterizing signals based on PerEn. In order to work with reliable statistics when calculating PerEn, it is highly recommended $(m + 1)! \leq N$ [167, 187]. In addition, when *m* is too large, the computation time will be higher. On the other hand, when *m* is high, the number of accessible states will be large, and the value of the PerEn will probably be more reliable. Thus, we should make a trade-off between the aforementioned cases.

For the AAPerEn, the embedding dimension and time delay values are set the same as PerEn. To select an appropriate value of *A* for AAPerEn (see Equation (3.5)), since the importance of the mean value of amplitudes and the differences of the amplitude values are equal, it is advisable to choose $A = 0.5$. Nevertheless, it is recommended to change the value of *A* in case we want to emphasize more on either the amplitude values change or average of amplitude values. For example, in spike detection, because the difference between two successive sample points is much more important than the mean value of amplitudes, it is recommended that $A < 0.5$. In contrast, in signal segmentation applications, the mean of amplitude values and

the differences of successive sample points are equally important and, therefore, it is advisable to choose $A = 0.5$ [28].

3.2.4 Discussion of AAPerEn vs. PerEn and its Developments

Our proposed AAPerEn, compared with PerEn, was applied to synthetic signals and EEG data for signal segmentation and realistic synthetic and real neuronal data for spike detection application. For both the applications, AAPerEn outperformed PerEn to detect the boundaries of segments and spikes [28].

The weighted PerEn was proposed in [188] addresses the limitation of PerEn related to the fact that it disregards the information contained in the amplitude values. This method is only dependent on the variance, which measures the spread of the signal amplitudes, and its importance in the metric is always kept constant. In contrast, in AAPerEn, we can adapt the importance of mean values of amplitudes and differences between samples to different applications by means of *A*. Therefore, AAPerEn is more flexible. Unlike the weighted PerEn and PerEn, AAPerEn can also discriminate between an original signal and a constant number added to the original signal. Furthermore, the weighted PerEn method does not address the equal values' problem. For this shortcoming, the algorithm proposed in [189] considerably increases the number of potential motifs. Hence, the algorithm may yield unreliable results for short signals (e.g., 100 points for embedding dimension 2 - see Section 4.3).

AAPerEn is a powerful tool to segment signals and detect spikes [28]. Moreover, AAPerEn was successfully employed to distinguish different emotional states using EEG signals [190]. On the whole, AAPerEn can be applied to different applications where the mean values of neighboring samples and change in amplitude values are important. Our modification also deals with the equal amplitude values' problem of PerEn. The AAPerEn's running time is slightly higher than the PerEn's one and it may be used in various entropy-based applications. Nevertheless, AAPerEn is noticeably sensitive to noise [29].

3.2.5 Dispersion Entropy (DispEn)

To deal with the sensitivity of AAPerEn to noise, improve the reliability of SampEn and FuzEn, and reduce the computation time of SampEn and FuzEn, DispEn is introduced in this Thesis

[18, 29]. Given a univariate signal $\mathbf{x} = \{x_1, x_2, \ldots, x_j, \ldots, x_N\}$ with length *N*, the DispEn algorithm is as follows:

1) First, $x_j (j = 1, \ldots, N)$ are mapped to c classes with integer indices from 1 to c. The classified signal is $w_j (j = 1, 2, \ldots, N)$. A number of linear and non-linear mapping techniques, introduced in Subsection 3.2.7, can be used in this step.

2) Time series $\mathbf{w}_j^{m,c}$ are made with embedding dimension m and time delay d according to $\mathbf{w}_\Lambda^{m,c} = \{w_\Lambda^c, w_{\Lambda+d}^c, \ldots, w_{\Lambda+(m-1)d}^c\}$, $\Lambda = 1, 2, \ldots, N-(m-1)d$ [17, 18]. Each time series $\mathbf{w}_i^{m,c}$ is mapped to a dispersion pattern $\pi_{v_0 v_1 \ldots v_{m-1}}$, where $w_\Lambda^c = v_0$, $w_{\Lambda+d}^c = v_1$,..., $w_{\Lambda+(m-1)d}^c = v_{m-1}$. The number of possible dispersion patterns assigned to each vector $\mathbf{w}_\Lambda^{m,c}$ is equal to c^m, since the signal $\mathbf{w}_\Lambda^{m,c}$ has m elements and each can be one of the integers from 1 to c [18].

3) For each of c^m potential dispersion patterns $\pi_{v_0 \ldots v_{m-1}}$, relative frequency is obtained as follows:

$$Pr(\pi_{v_0 \ldots v_{m-1}}) = \frac{\#\{\Lambda \,|\Lambda \leq N-(m-1)d, \mathbf{w}_\Lambda^{m,c} \text{ has type } \pi_{v_0 \ldots v_{m-1}}\}}{N-(m-1)d} \tag{3.6}$$

where $\#$ means cardinality. In fact, $Pr(\pi_{v_0 \ldots v_{m-1}})$ shows the number of dispersion patterns of $\pi_{v_0 \ldots v_{m-1}}$ that is assigned to $\mathbf{w}_\Lambda^{m,c}$, divided by the total number of embedded signals with embedding dimension m.

4) Finally, based on the Shannon's definition of entropy, the DispEn value is calculated as follows:

$$DispEn(\mathbf{x}, m, c, d) = -\sum_{\pi=1}^{c^m} Pr(\pi_{v_0 \ldots v_{m-1}}) \cdot \ln\left(Pr(\pi_{v_0 \ldots v_{m-1}})\right) \tag{3.7}$$

As an example, let's have a series $\mathbf{x} = \{3.6, 4.2, 1.2, 3.1, 4.2, 2.1, 3.3, 4.6, 6.8, 8.4\}$, shown on the top left of Figure 3.2. We want to calculate the DispEn value of $\mathbf{x}$. For simplicity, we set $d = 1$, $m = 2$, and $c = 3$. The $3^2 = 9$ potential dispersion patterns are depicted on the right of Figure 3.2. x_j $(j = 1, 2, \ldots, 10)$ are linearly mapped into 3 classes with integer indices from 1 to 3, as can be seen in Figure 3.2. Next, a window with length 2 (embedding dimension)

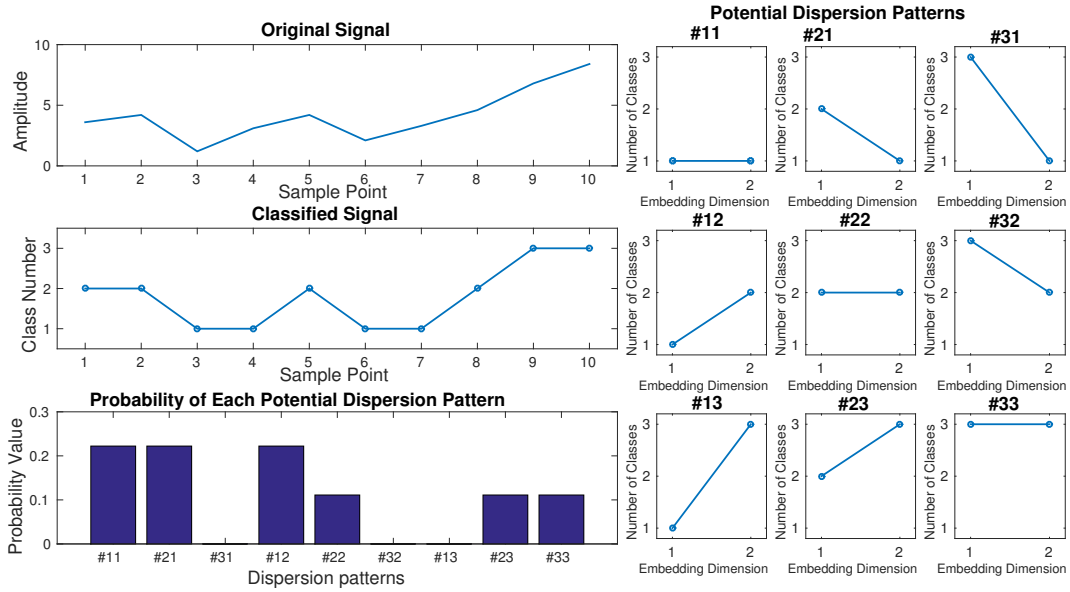


Figure 3.2: Illustration of the DispEn algorithm using linear mapping of $\boldsymbol{x} = \{3.6, 4.2, 1.2, 3.1, 4.2, 2.1, 3.3, 4.6, 6.8, 8.4\}$ *with three classes and embedding dimension equal to two.*

moves along the signal and the number of each of dispersion patterns is counted. The relative frequency is shown on the bottom left of Figure 3.2. Finally, using Equation (3.7), the DispEn value of **x** is equal to $\left(\frac{2}{9}\ln(\frac{2}{9}) + \frac{2}{9}\ln(\frac{2}{9}) + \frac{2}{9}\ln(\frac{2}{9}) + \frac{1}{9}\ln(\frac{1}{9}) + \frac{1}{9}\ln(\frac{1}{9}) + \frac{1}{9}\ln(\frac{1}{9})\right) = 1.7351$.

If all possible dispersion patterns have equal probability value, the DispEn reaches its highest value, which has a value of $ln(c^m)$. In contrast, when there is only one $p(\pi_{v_0 \ldots v_{m-1}})$ different from zero, which demonstrates a completely regular/predictable time series, the smallest value of DispEn is obtained [18]. Note that the normalized DispEn is calculated as $\frac{DispEn}{\ln(c^m)}$ in this study [18].

3.2.6 Frequency-based Dispersion Entropy (FDispEn)

When only the frequency of a signal is relevant (or the amplitude can be disregarded), there is no difference between dispersion patterns $\{1, 3, 4\}$ and $\{2, 4, 5\}$ or $\{1, 1, 1\}$ and $\{3, 3, 3\}$. To take into account only the frequency of signals, we introduce FDispEn in this Chapter. In fact, FDispEn considers the differences between adjacent elements of dispersion patterns, termed frequency-based dispersion patterns. In this way, we have vectors with length $m - 1$ which each of their elements changes from $-c + 1$ to $c - 1$. Thus, there are $(2c - 1)^{m-1}$ potential frequency-based dispersion patterns. The only difference between DispEn and FDispEn

algorithms is the potential patterns used in these two approaches.

As an example, let's have a signal $\mathbf{x} = \{3, 4.5, 6.2, 5.1, 3.2, 1.2, 3.5, 5.6, 4.9, 8.4\}$. We set $d = 1$, $m = 3$, and $c = 2$, leading to have $3^2 = 9$ potential frequency-based dispersion patterns $(\{(-1,-1),(-1,0),(-1,1),(0,-1),(0,0),(0,1),(1,-1),(1,0),(1,1)\})$. Then, x_j $(j = 1, 2, \ldots, 10)$ are linearly mapped into 2 classes with integer indices from 1 to 2 $(\{1,1,2,2,1,1,1,2,2,2\})$. Afterwards, a window with length 3 moves along the time series and the differences between adjacent elements are calculated $(\{(0,1),(1,0),(0,-1),(-1,0),(0,0),(0,1),(1,0),(0,0)\})$. Afterwards, the number of each frequency-based dispersion pattern is counted. Finally, using Equation (3.7), the FDispEn value of **x** is equal to $-\left(\frac{1}{8}\ln(\frac{1}{8}) + \frac{1}{8}\ln(\frac{1}{8}) + \frac{2}{8}\ln(\frac{2}{8}) + \frac{2}{8}\ln(\frac{2}{8}) + \frac{2}{8}\ln(\frac{2}{8})\right) = 1.5596$.

3.2.7 Mapping Approaches used in DispEn and FDispEn

A number of linear and nonlinear methods can be used to map the original signal $x_j (j = 1, \ldots, N)$ to the classified signal $w_j (j = 1, 2, \ldots, N)$. A fast algorithm is to sort $x_j (j = 1, 2, \ldots, N)$ and then, divide them into *c* classes in which each of them includes equal number of x_j.

We can also use several non-linear mapping techniques. Many natural processes show a progression from small beginnings that accelerates and approaches a climax over time [191]. When there is not a detailed description, a sigmoid function is frequently used [192]. The well-known log-sigmoid (logsig) and tan-sigmoid (tansig) transfer functions are suitable to represent this kind of phenomena. Logsig and tansig are respectively defined as:

$$y_j = \frac{1}{e^{-\frac{x_j - \mu}{\sigma}}} \tag{3.8}$$

$$y_j = \frac{2}{1 + e^{-2\frac{x_j - \mu}{\sigma}}} - 1 \tag{3.9}$$

where σ and μ are the SD and mean of time series **x**, respectively.

The cumulative distribution functions (CDFs) for many common probability distributions are sigmoidal. The most well-known such example is the error function, which is related to

the CDF of a normal distribution, termed normal CDF (NCDF). NCDF of **x** is calculated as follows:

$$y_j = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_j} e^{\frac{-(t-\mu)^2}{2\sigma^2}} \, \mathrm{d}t \tag{3.10}$$

Each of the aforementioned techniques maps **x** into $\mathbf{y} = \{y_1, y_2, \ldots, y_N\}$, ranged from α to β. Then, we use a linear algorithm to assign each y_j to a real number z_j from 0.5 to $c + 0.5$. Then, for each element of the mapped signal, we use $u_j^c = \text{round}(z_j)$, where u_j^c denotes the j^{th} element of the classified signal and rounding involves either increasing or decreasing a number to the next digit [18].

In [29], we evaluated the ability of DispEn and FDispEn with different mapping techniques to distinguish changes from periodicity to non-periodic non-linearity with different levels of noise [29]. The results showed that the DispEn with sorting method and linear mapping lead to the most stable results. Although DispEn with sorting method, unlike PerEn, takes into account repetitions, it considers only the order of amplitude values and thus, some information regarding the amplitudes may be discarded. For instance, it was found that DispEn with sorting method cannot detect the outliers or spikes, which are noticeably larger or smaller than their adjacent values [29]. Furthermore, for DispEn with linear mapping, when maximum or minimum values are noticeable larger or smaller than the mean/median value of the signal, the majority of x_j are mapped to only few classes [18]. Thus, it was illustrated that mapping based on NCDF, tansig and logsig addresses this problem [29]. For simplicity, we use DispEn and FDispEn with NCDF for all the simulations in this Thesis. It is worth noting the results obtained by DispEn and FDispEn based on logsig (see [29]) and NCDF (see the results in this Chapter) are similar.

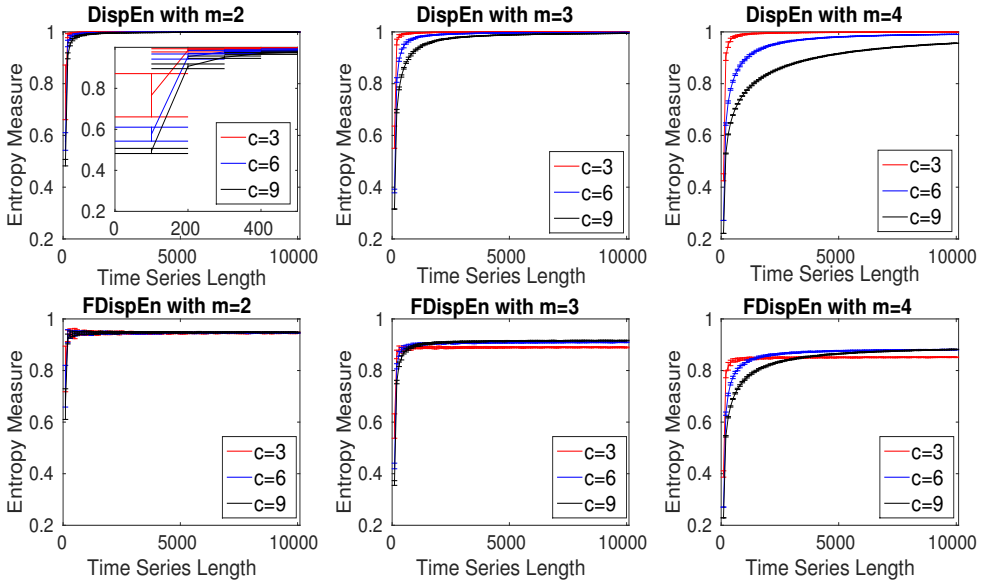


Figure 3.3: Mean and SD of results obtained by the DispEn and FDispEn with NCDF and different values of embedding dimension and number of classes for 40 realizations of univariate WGN.

3.2.8 Parameters of DispEn and FDispEn

3.2.8.1 Effect of Number of Classes, Embedding Dimension, and Signal Length on DispEn and FDispEn

To assess the sensitivity of DispEn and FDispEn with NCDF to the signal length, embedding dimension m, and number of classes c, we use 40 realizations of univariate WGN. The mean and SD of results, depicted in Figure 3.3, show that DispEn and FDispEn need a smaller number of sample points to reach their maximum values for a smaller number of classes or smaller embedding dimension. This is in agreement with the fact that we need at least $\ln(c^m)$ [18] and $\ln((2c-1)^{m-1})$ sample points to reach the maximum value of DispEn and FDispEn, respectively.

3.2.8.2 Effect of Number of Classes and Noise Power on DispEn and FDispEn

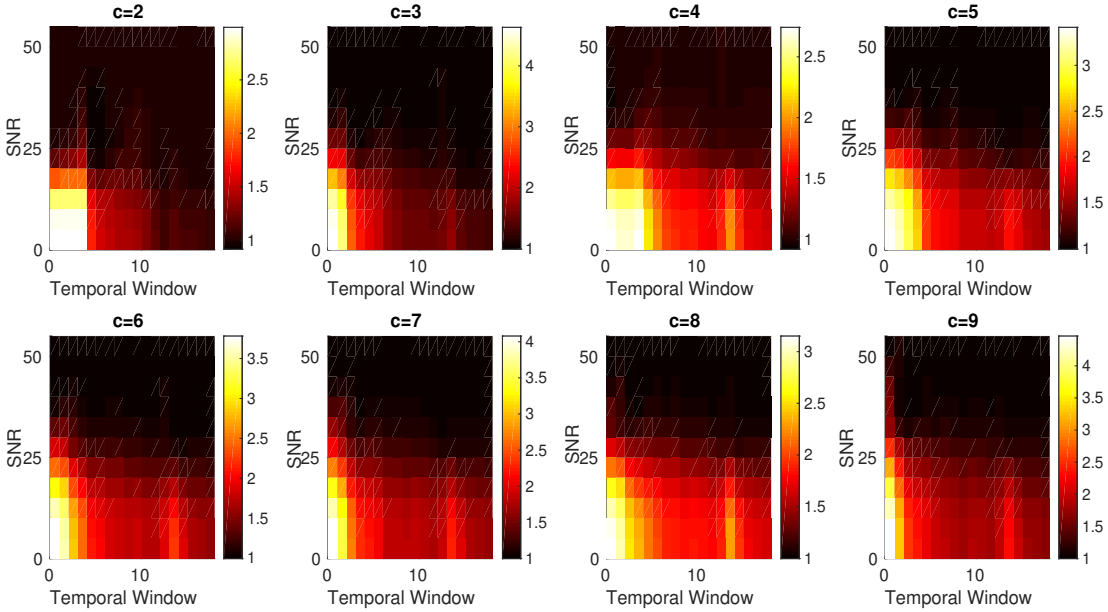
We also inspect the relationship between noise power levels and DispEn with different number of classes. To this end, we use a logistic map added with different levels of noise power. The logistic map was described in Subsection 3.1.1. We added 40 independent realizations of WGN

with different signal-to-noise ratios (SNRs) per sample, ranging from 0 to 50 dB, to the logistic map. We then employed a sliding window with length 1,500 sample points and 50% overlap moving along the signal to show the effect of noise power on each segment (window) of the signal. To compare the sensitivity of each method to WGN, we calculate *NrmEntN* as the entropy value of each segment with noise over the entropy value of its corresponding segment without noise ($NrmEntN = \frac{\texttt{entropy of a series with noise}}{\texttt{entropy of a series without noise}}$). It should be mentioned that $NrmEntN \approx 1$ means better results, especially when dealing with a high SNR.

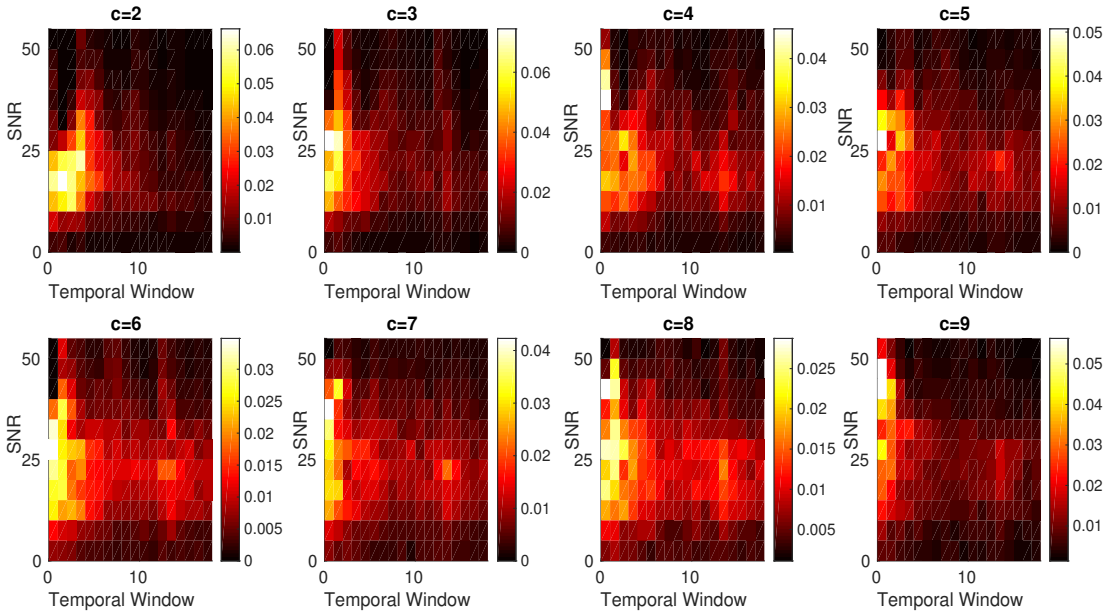
The average and SD values of results obtained by the DispEn using NCDF with different number of classes computed from the logistic map whose parameter (α) varies from 3.5 to 3.99 with additive 40 independent realizations of WGN with different noise powers are shown in Figure 3.4(a) and 3.4(b), respectively. $m = 2$ was set for DispEn [18]. Figure 3.4 suggests that the larger the number of classes, the larger the value of *NrmEntN*, as expected. Thus, when dealing with a low SNR, it is recommended to have a small c. In fact, when c is too large, small changes may alter the class of a sample and, therefore, the DispEn method might be sensitive to noise. On the other hand, if SNR is large, we can choose a large c. When c is too small, two amplitude values that are far from each other may be assigned to a similar class, leading to unreliable entropy values. Thus, we need to have a trade-off between large and small number of classes. As the SD and average of results are appropriate for $c = 6$ (Figure 3.4) and for simplicity, we set $c = 6$ for all the simulations below.

Compared with DispEn, in the FDispEn algorithm, we have vectors with length $m - 1$ where each of their elements changes from $-c + 1$ to $c - 1$. Thus, we set $m = 3$ here. Likewise DispEn, we changed c from 3 to 9 for FDispEn. We found that $c = 5$ leads to stable results when dealing with noise (results are not shown herein). Thus, we set $c = 5$ for all simulations using FDispEn, although the range $2 < c < 9$ leads to similar conclusions.

Overall, the parameter c is chosen to balance the quantity of entropy estimates with the loss of signal information. To avoid the impact of noise on signals, a small c is recommended. In contrast, for a small c, too much detailed data information is lost, leading to poor probability estimates. Thus, a trade-off between large and small c values is needed.



(a) Mean



(b) SD

Figure 3.4: (a) Mean and (b) SD of $NrmEntN$ values obtained by the DispEn using NCDF with different number of classes computed from the logistic map with additive 40 independent realizations of WGN with different noise power. Darker means better results in Figures that NrmEntN is used.

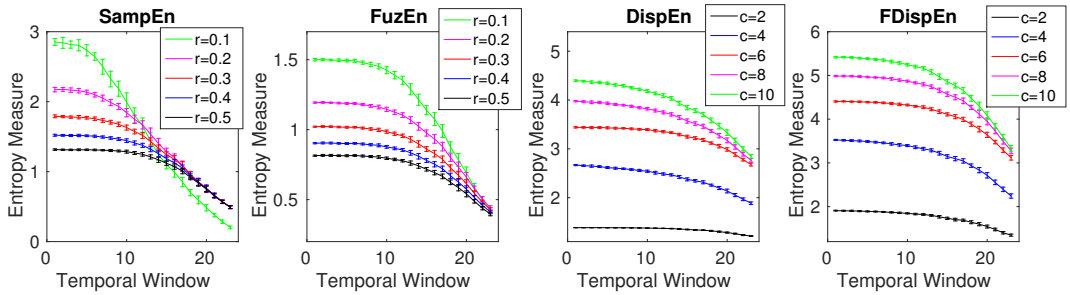


Figure 3.5: (a) Mean and (b) SD of entropy values obtained by the DispEn, FDispEn, SampEn, and FuzEn with different number of classes (for DispEn and FDispEn) and different threshold values (SampEn and FuzEn) using a MIX process evolving from randomness to periodic oscillations. We used a window with length 1,500 samples moving along the MIX process (temporal window).

3.2.9 Threshold *r* (SampEn and FuzEn) vs. Number of Classes *c* (DispEn and FDispEn)

The dependence of the number of classes c (DispEn and FDispEn) and threshold r (SampEn and FuzEn) is inspected by the use of the MIX process described in Subsection 3.1.1. The SampEn, FuzEn, DispEn, and FDispEn techniques are applied to 20 realizations of the MIX process using a moving window of 1,500 samples (10 s) with 50% overlap. We used different threshold values $r = 0.1, 0.2, 0.3, 0.4$, and 0.5 of SD of the signal [16] for SampEn and FuzEn, and $c = 2, 4, 6, 8$ and 10 for DispEn and FDispEn.

The results, depicted in Figure 3.5, show that the mean entropy values using all these approaches are the least in higher temporal windows, in agreement with the previous studies [15, 170]. The results also evidence that the number of classes (c) in DispEn and FDispEn is inversely related to the threshold value r used in the SampEn and FuzEn algorithms. It is worth noting that SampEn, unlike DispEn and FDispEn, is not consistent as $r = 0.1$ crosses the other lines. We set m =2, 2, and 3, for respectively SampEn, DispEn, and FDispEn, as recommended before.

To compare the results obtained by the entropy algorithms, we used the coefficient of variation (CV) defined as the SD divided by the mean. We use such a metric as the SDs of signals may increase or decrease proportionally to the mean. We inspect the MIX process with length 1,500 samples and $p = 0.5$ as a trade-off between random ($p = 1$) and periodic oscillations ($p = 0$). The CV values, depicted in Table 3.1, show that FuzEn leads to smaller CVs in comparison with

Table 3.1: CVs of DispEn, FDispEn, SampEn, and FuzEn values for the MIX process with $p = 0.5$ *and length 1000 samples.*

	$c=2$	$c=4$	$c=6$	$c=8$	$c=10$
DispEn	0.0097	0.0082	0.0069	0.0083	0.0027
FDispEn	0.0078	0.0075	0.0077	0.0095	0.0162
	$r=0.1\times SD$	$r=0.2\times SD$	$r=0.3\times SD$	$r=0.4\times SD$	$r=0.5\times SD$
SampEn	0.0683	0.0350	0.0285	0.0213	0.0203
FuzEn	0.0212	0.0153	0.0130	0.0117	0.0111

SampEn. It is also found that DispEn- and FDispEn-based CV values for different number of classes are noticeably smaller than those for SampEn with different threshold values, showing an advantage of DispEn and FDispEn over SampEn. Overall, the smallest CVs are obtained by DispEn.

3.3 Performance Results

In this Section, the ability of the proposed and existing univariate entropy approaches is inspected by the use of several synthetic and real univariate time series, described in Section 3.1. SampEn and FuzEn are based on ConEn [16, 103], whereas PerEn, DispEn, and FDispEn are based on ShEn [17,29]. This means that the methods work on different principles. However, the comparison of DispEn, FDispEn, and PerEn with SampEn and FuzEn is meaningful because the latter three are the most common entropy algorithms.

3.3.1 Univariate Synthetic Signals

We demonstrate the dependency of the FuzEn, PerEn, DispEn, and FDispEn on several straightforward signal processing concepts using a set of synthetic signals. As the SampEn and FuzEn lead to similar findings and the latter is more stable and reliable for short signals [30, 103], only FuzEn is used in this Subsection. We employed a sliding window of 1,200 sample points with 80% overlap moves along the signals with a sampling frequency of 150 Hz and a length of 100 s (15,000 sample points).

The FuzEn, PerEn, DispEn, and FDispEn values for the chirp signal with constant amplitude are shown in Figure 3.6(a). The results suggest that all the methods detect the changes in frequency of the signal, although the FuzEn values slightly decrease when dealing a very high

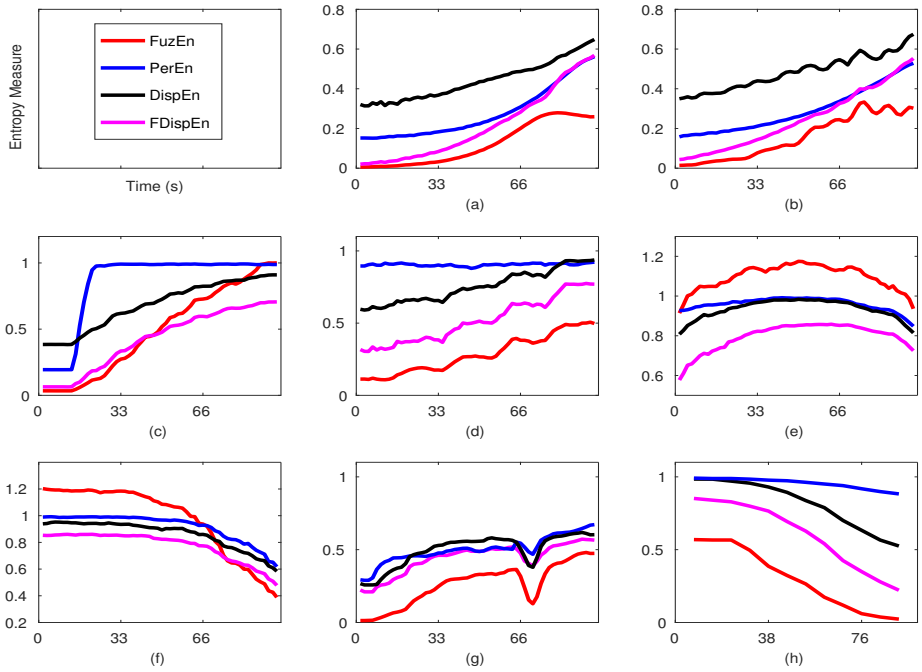


Figure 3.6: Results of the tests performed to understand better FuzEn (red), PerEn (blue), DispEn (black), and FDispEn (magenta) using: (a) chirp signal with constant amplitude (see Figure 3.1(a)); (b) amplitude-modulated chirp signal (see Figure 3.1(b)); (c) periodic signal with increasing additive noise power (see Figure 3.1(c)); (d) signal including five segments of colored noise with increasing bandwidth (see Figure 3.1(d)); (e) AR(1) process with variable parameter (see Figure 3.1(e)); (f) MIX process evolving from randomness to periodic oscillations (see Figure 3.1(f)), (g) logistic map with parameter α changing from 3.5 to 3.99 (see Figure 3.1(g)), and (h) signal including 12 segments of different kinds of noise with spectral density $\frac{1}{f^{\Gamma}}$, which Γ changes from 0 (white noise) to 2 (brown noise) from the first to twelfth segment, respectively (see Figure 3.1(h)). The time axis in this Figure corresponds to that of Figure 3.1.

frequency.

To understand the relationship between the entropy methods and simultaneous frequency and amplitude change, the amplitude-modulated chirp signal is used. As can be seen in Figure 3.6(b), DispEn and FuzEn, unlike PerEn and FDispEn, can detect the simultaneous change in amplitude and frequency. In other words, frequency-based techniques, i.e., PerEn and FDispEn, cannot detect change in simultaneous frequency and amplitude. Considering the periodicity of the changes in the FuzEn and the periodicity of the amplitude variations of the signal, the FuzEn change is not very clear while observing the curve associated with the

DispEn approach.

To inspect how the entropy methods change with the level of noise affecting periodic signals, we use an amplitude-modulated periodic time series with additive WGN with diverse power depicted in Figure 3.1(c). As can be seen in Figure 3.6(c), the results obtained by FuzEn, DispEn, and FDispEn demonstrate the change in the amount of noise power, while the PerEn values are saturated from temporal window of 16 ($N \approx 4,000$ sample points).

To study the dependence between the entropy methods and noise bandwidth, a 100 s signal composed of five segments of colored noise with increasing bandwidth described in Subsection 3.1.1 is used. The FuzEn, DispEn, FDispEn, and PerEn values are depicted in Figure 3.6(d). The results show that the FuzEn, DispEn, and FDispEn values, unlike PerEn ones, rise when the signal bandwidth increases, showing an advantage of DispEn, FDispEn, and FuzEn over PerEn when dealing with signal bandwidth with increasing bandwidth.

To investigate the relationship between the univariate entropy techniques and the spectral content of colored noise, AR(1) described in Subsection 3.1.1 is used. The results are shown in Figure 3.6(e). In the center of the signal (*i.e.,* $\rho = 0$), the sequence corresponds to WGN. The entropy values obtained by all the approaches in the center of the AR(1) process are maximum.

To understand how DispEn and FDispEn, compared with PerEn and FuzEn, vary when a stochastic sequence progressively turns into a periodic deterministic time series, we use the MIX process described in Subsection 3.1.1. The results, depicted in Figure 3.6(f), show that all the entropy values decrease along the signal as the series evolves from randomness to orderliness.

The results obtained by FDispEn, DispEn, PerEn, and FuzEn for the logistic map with the parameter α changing linearly from 3.5 to 3.99 are shown in Figure 3.6(g). As expected, the entropy values, obtained by the entropy techniques generally increase along the signal, except for the downward spikes in the windows of periodic behavior (e.g., for $\alpha = 3.8$). This fact is in agreement with Figure 4.10 (page 87 in [179]) and the other previous studies [30, 31].

To inspect how DispEn, FDispEn, FuzEn, and PerEn change when a signal changes from white to brown noise along 12 equal segments with length 1250 sample points, we use the series which Γ for $S(f) = \frac{1}{f^{\Gamma}}$ noise increases from 0 to 2 (see Subsection 3.1.1). Since the

adjacent segments with different Γ values have dissimilar correlations, the entropy of each of 12 segments is calculated, instead of moving the window with 80% overlap. The results, depicted in Figure 3.6(h), show that all the entropy values decrease along the signal, in agreement with the fact that the irregularity of white noise is the most irregular noise, followed by pink and brown noise, respectively [21, 22, 180, 181].

To sum, PerEn and FDispEn were not able to distinguish simultaneous change in amplitude and frequency. DispEn and FuzEn detect the change in amplitude and frequency, noise power, and frequency band, degree of randomness and periodicity, and types of noise. Nevertheless, DispEn, unlike FuzEn, does not lead to unreliable values for short signals and is considerably faster than FuzEn, especially for long time series (please see Section 3.4). Therefore, DispEn may be the most consistent algorithm for characterization different types of time series.

3.3.2 Real Biomedical Datasets

We employ several entropy approaches to discriminate focal signals from non-focal EEGs and elderly from young subjects using their blood pressure recordings.

3.3.2.1 Dataset of Focal and Non-focal Brain Activity

For the focal and non-focal EEG signals, the mean and median of results obtained by DispEn, FDispEn, PerEn, SampEn, and FuzEn, are shown in Figure 3.7. The results for FuzEn, SampEn, DispEn, and FDispEn, unlike those for PerEn, show that non-focal signals (NFS) are more irregular than focal ones (FS). This fact is consistent with previous studies [162, 163]. It should be noted that the average entropy values over 2 channels for these bivariate EEG signals are reported for the univariate entropy methods.

For each technique, the non-parametric Mann–Whitney U-test was employed to assess the differences between results for focal and non-focal signals, as the entropy values for all the entropy approaches did not follow a normal distribution. The results are presented in Table 3.2. The p-values show that DispEn, FDispEn, SampEn, and FuzEn discriminate the focal EEGs from non-focal signals batter than PerEn.

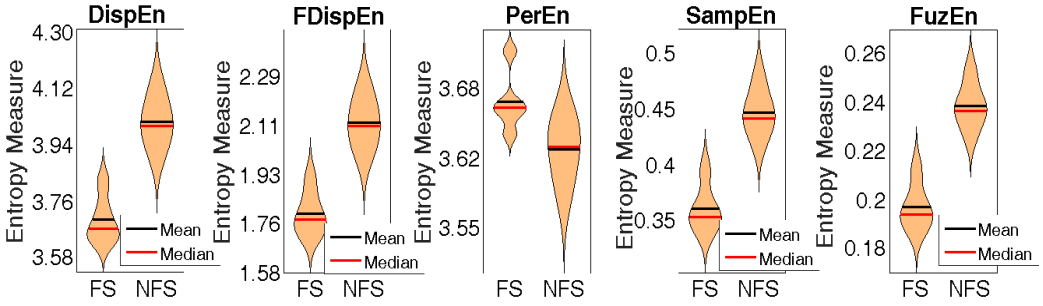


Figure 3.7: Mean and median of results obtained by the DispEn, FDispEn, PerEn, SampEn, and FuzEn computed from the focal (FS) and non-focal EEG signals (NFS).

Table 3.2: p-values obtained by the DispEn, FDispEn, PerEn, SampEn, and FuzEn for the focal and non-focal EEG signals (Mann-Whitney U-test).

DispEn	FDispEn	PerEn	SampEn	FuzEn
0.008	0.008	0.032	0.008	0.008

3.3.2.2 Fantasia Dataset

For the blood pressure recordings for young and old subjects in Fantasia database, the mean and median of results obtained by the univariate entropy methods are shown in Figure 3.8. The results for FuzEn, SampEn, DispEn, and FDispEn, unlike those for PerEn, show that young subjects' signals are more irregular than old ones. This fact is in agreement with [193].

For each method, the non-parametric Mann–Whitney *U*-test was employed to evaluate the differences between results for young and old subjects' times series. The results are depicted in Table 3.3. The *p*-values show that DispEn and FuzEn discriminate the young from old subjects' signals than PerEn, SampEn, and FDispEn.

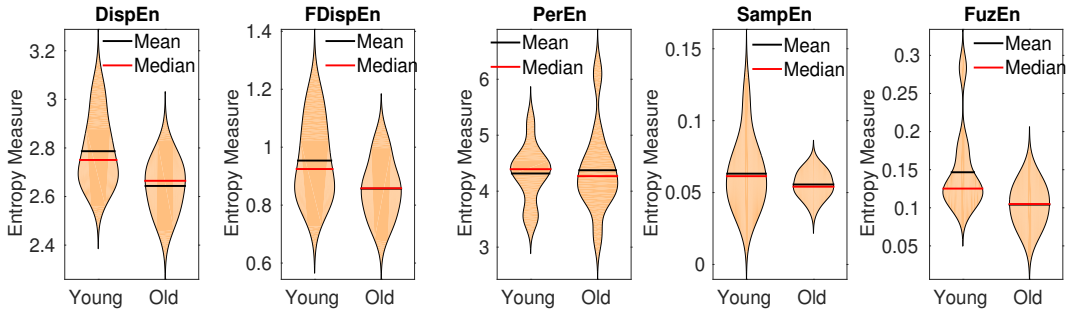


Figure 3.8: Mean and median of results obtained by the DispEn, FDispEn, PerEn, SampEn, and FuzEn of the blood pressure recordings for young and old subjects in Fantasia database.

Table 3.3: p-values obtained by the DispEn, FDispEn, PerEn, SampEn, and FuzEn of the blood pressure recordings for young and old subjects in Fantasia database (Mann-Whitney U-test).

DispEn	FDispEn	PerEn	SampEn	FuzEn
0.0452	0.1859	0.9097	0.7337	0.0452

Table 3.4: Computation time of DispEn, FDispEn, PerEn, SampEn, and FuzEn with $m = 2, 3$ and 4 for WGN with different lengths (300, 1000, 3,000, 10,000, 30,000, and 100,000 sample points).

Number of samples →	300	1,000	3,000	10,000	30,000	100,000
DispEn ($m = 2$)	0.0022 s	0.0022 s	0.0025 s	0.0057 s	0.0080 s	0.0225 s
DispEn ($m = 3$)	0.0028 s	0.0035 s	0.0076 s	0.0115 s	0.0284 s	0.0888 s
DispEn ($m = 4$)	0.0084 s	0.0094 s	0.0205 s	0.0505 s	0.1422 s	0.4752 s
FDispEn ($m = 2$)	0.0022 s	0.0025 s	0.0028 s	0.0034 s	0.0062 s	0.0175 s
FDispEn ($m = 3$)	0.0025 s	0.0031 s	0.0038 s	0.0062 s	0.0150 s	0.0490 s
FDispEn ($m = 4$)	0.0054 s	0.0064 s	0.0120 s	0.0284 s	0.0699 s	0.2535 s
PerEn ($m = 2$)	0.0028 s	0.0103 s	0.0323 s	0.1081 s	0.3148 s	1.1118 s
PerEn ($m = 3$)	0.0039 s	0.0134 s	0.0418 s	0.1323 s	0.4129 s	2.1395 s
PerEn ($m = 4$)	0.0103 s	0.0297 s	0.0869 s	0.3128 s	1.5200 s	3.7163 s
SampEn ($m = 2$)	0.0023 s	0.0208 s	0.1841 s	1.8478 s	16.8394 s	193.1970 s
SampEn ($m = 3$)	0.0022 s	0.0206 s	0.1808 s	1.8337 s	16.9200 s	189.4041 s
SampEn ($m = 4$)	0.0019 s	0.0193 s	0.1631 s	1.8322 s	16.5596 s	189.1037 s
FuzEn ($m = 2$)	0.0546 s	0.2784 s	1.5490 s	10.9053 s	63.5978 s	515.0032 s
FuzEn ($m = 3$)	0.0717 s	0.3161 s	1.7715 s	12.4459 s	77.3289 s	619.4129 s
FuzEn ($m = 4$)	0.055 s	0.3438 s	1.8905 s	13.9068 s	85.7283 s	726.0238 s

3.4 Computational Time of Univariate Entropy Methods

In order to assess the computational time of DispEn, FDispEn, PerEn, SampEn, and FuzEn, we use WGN times series with different lengths, logarithmically changing from 300 to 100,000 sample points. The results are depicted in Table 3.4. All the simulations in this Thesis have been carried out using a PC with Intel (R) Xeon (R) CPU, E5420, 2.5 GHz and 8-GB RAM by MATLAB R2015a. The embedding dimension values change from 2 to 4 for all the methods.

The results show that the computation times of SampEn with different *m* are very close, while for DispEn, FDispEn, PerEn, and FuzEn, the larger the *m* value, the higher the computation time. The SampEn is considerably faster than FuzEn for various lengths.

When dealing with short signals (300 and 1,000 sample points), the differences between the computation time values for SampEn, DispEn, FDispEn, and PerEn are negligible. However, for signals with length 3,000 sample points or longer, FDispEn is the fastest algorithm, followed by DispEn, PerEn, SampEn, and FuzEn, in that order. This is in agreement with the fact that the computation costs of DispEn-based methods, PerEn, SampEn, and FuzEn are

respectively O(N), O(N), O(N^2), and O(N^2) [18, 194]. It is worth noting that in the DispEn and FDispEn algorithms, it is needed to neither sort the amplitude values of each embedded vector (like PerEn) nor calculate every distance between any two composite delay vectors with embedding dimensions m and $m + 1$ (like SampEn and FuzEn). This makes DispEn and FDispEn noticeably faster than PerEn, SampEn, and FuzEn when dealing with long signals.

The DispEn, PerEn, SampEn, and FuzEn codes are available at [195], [196], [197], and [198], respectively. The SampEn code used in this Thesis was optimized, whereas FuzEn was implemented based on its straightforward algorithm without any optimization. Thus, this may be the reason for the considerable difference between the computational times for FuzEn and SampEn.

3.5 Summary

In this Chapter, our proposed AAPerEn, DispEn, and FDispEn to quantify the irregularity of physiological signals and their advantages and disadvantages were described. It was found that AAPerEn, unlike PerEn, deals with the equal amplitude values in each embedded vector and simultaneous change in amplitude and frequency of signals. DispEn and FDispEn, which are based on our introduced dispersion patterns and the Shannon's definition of entropy, address the problem of sensitivity of AAPerEn to noise and unreliable SampEn and FuzEn values for short signals.

The SampEn, FuzEn, and DispEn methods yield similar findings for synthetic and real signals although the latter has the following advantages: 1) DispEn led to more stable results; 2) it is considerably faster; and 3) it does not lead to undefined or unreliable results. In comparison with PerEn, FDispEn, and SampEn, FuzEn and DispEn better distinguished the elderly from young subjects' blood pressure recordings, and focal from non-focal EEGs, respectively, for Fantasia and focal and non-focal brain activity datasets.

FDispEn showed a behaviour between that for PerEn and DispEn. In fact, FDispEn and PerEn, unlike DispEn, cannot detect simultaneous change in frequency and amplitude. However, FDispEn and DispEn, unlike PerEn, are not very sensitive to noise and can also distinguish the change in signal bandwidth. It was also found that, for long signals, FDispEn is the fastest

Table 3.5: Ability for characterization simultaneous change in frequency and amplitude, characterization of short signals, sensitivity to noise, type of entropy, and computational cost for our proposed DispEn, FDispEn, and AAPerEn in comparison with the popular PerEn, SampEn, and FuzEn methods.

Characteristics	DispEn [18]	FDispEn [29]	AAPerEn [28]	PerEn [17]	SampEn [16]	FuzEn [103]
Simultaneous change in frequency and amplitude	yes	no	yes	no	yes	yes
Short signals	reliable	reliable	reliable	reliable	undefined	unreliable
Sensitivity to noise	no	no	yes	yes	no	no
Type of entropy	ShEn	ShEn	ShEn	ShEn	ConEn	ConEn
Computational cost	O(N)	O(N)	O(N)	O(N)	O(N^2)	O(N^2)

technique, followed by DispEn, PerEn, SampEn, and FuzEn, respectively. The results also illustrated that FDispEn, as a new frequency-based entropy technique, outperformed PerEn to discriminate various dynamics of synthetic signals and physiological recordings.

To summarize, the characteristics and limitations of our introduced DispEn, FDispEn, and AAPerEn, compared with SampEn, FuzEn, PerEn, are explained in Table 3.5.

Chapter 4

Univariate Multiscale Entropy Methods

Existing entropy approaches, such as sample entropy (SampEn) and permutation entropy (PerEn), are widely used to quantify the irregularity of signals at one temporal scale [18, 19]. They assess repetitive patterns and return maximum values for completely random processes [21, 22, 160]. However, SampEn and PerEn may fail to account for the multiple time scales inherent in biomedical recordings [21, 30]. To deal with the problem, multiscale SampEn (MSE) was proposed [39] and it has become a prevalent algorithm to quantify the complexity of univariate time series, especially physiological recordings [21, 106].

However, MSE is undefined for very short signals and slow for real-time applications as a result of using SampEn [25, 30]. To address these deficiencies, multiscale PerEn (MPE) was proposed [25]. To increase the stability of MPE-based profiles, especially at long temporal scale factors, the improved MPE (IMPE) is developed in this Thesis [169]. Although MPE and IMPE are able to deal with short signals and are considerably faster than MSE-based techniques, they do not fulfil the key hypotheses of the concept of complexity (see Section 2.5) [199].

To overcome the limitations of PerEn and the problem of undefined SampEn values for short univariate time series, we propose refined composite multiscale fuzzy entropy (MFE - RCMFE) [30]. Nevertheless, RCMFE may yield unreliable results for short signals and is also not quick enough for real-time applications.

Thus, we introduce multiscale dispersion entropy (DispEn - MDE) based on our developed DispEn to quantify the complexity of signals [31]. The refined composite MDE (RCMDE) to improve the stability of MDE for short or noisy signals is proposed as well. To evaluate the existing and developed univariate multiscale entropy methods, several synthetic and real signals

are used in this Chapter. The descriptions of these data are explained in the next Section.

4.1 Univariate Signals for Evaluation

In this section, the synthetic and real signals used in this study to evaluate the behaviour of the univariate multiscale entropy approaches are described.

4.1.1 Synthetic Signals

4.1.1.1 Univariate Multiscale Entropy Methods vs. Noise

As mentioned in Chapter 2, the entropy and multiscale entropy approaches respectively quantify the irregularity and complexity of signals [39]. White Gaussian noise (WGN) is more irregular than $1/f$ noise, although the latter is more complex because $1/f$ noise contains long-range correlations and its $1/f$ decay produces a fractal structure in time [21, 22, 161]. Therefore, WGN and $1/f$ noise are two important signals to evaluate the multiscale entropy techniques [21, 22, 30, 161, 199, 200].

4.1.1.2 Univariate Multiscale Entropy Methods vs. Chaotic Behavior

In order to investigate the change in the behavior of a non-linear system, the Lorenz attractor is used as:

$$\begin{aligned} \dot{x} &= \lambda(y - x), \\ \dot{y} &= x(\rho - z) - y, \\ \dot{z} &= xy - \beta z, \end{aligned} \tag{4.1}$$

where λ, β, and ρ denote the system parameters [179, 201]. We created 40 realizations of two Lorenz signals with lengths of 450 and 4,500 sample points and sampling frequency of 150 Hz. To have a nonlinear behavior, $\lambda = 10$, $\beta = \frac{8}{3}$, and $\rho = 28$ were set [179, 201]. The synthetic time series with length of 4,500 samples and its spectrogram are shown in Figure 4.1.

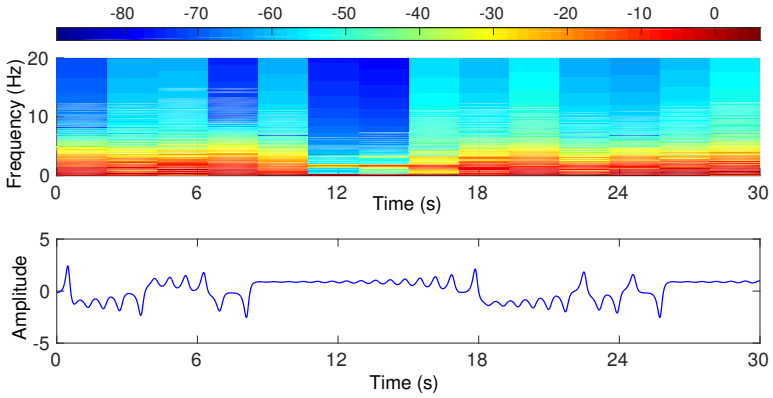


Figure 4.1: Spectrogram and time plot of the Lorenz signal ($\lambda = 10$, $\beta = \frac{8}{3}$, and $\rho = 28$). Red corresponds to high power, and blue corresponds to low power.

4.1.1.3 Univariate Multiscale Entropy Methods vs. Noise Power

To investigate the relationship between univariate multiscale entropy methods and the level of noise affecting periodic time series, the amplitude-modulated periodic signal with additive WGN with diverse power is used. The descriptions of the signal are found in Chapter 3 (Subsection 3.1.1).

4.1.2 Real Biomedical Datasets

Multiscale entropy techniques are broadly used to characterize physiological recordings [21, 30, 39, 106]. To this end, two non-invasive electroencephalograms (EEGs) [162] and blood pressure data [184] are used to distinguish different kinds of dynamics of time series. The descriptions of these real datasets are found in Chapter 3 (Subsection 3.1.2).

4.2 Multiscale Permutation Entropy-based Approaches

In this Section, MPE and IMPE are explained and then, their advantages and disadvantages are described.

4.2.1 Multiscale Permutation Entropy (MPE)

The MPE algorithm at scale factor τ includes the following two steps:

1. Univariate coarse-graining process: Assume the univariate time series $\mathbf{u} = \{u_1, u_2, \ldots, u_i, \ldots, u_L\}$. The coarse-grained signal $\mathbf{x}^{(\tau)}$, corresponding to the scale factor τ, is first constructed as follows [25, 39]:

$$x_j{}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} u_i, \quad 1 \leq j \leq \left\lfloor \frac{L}{\tau} \right\rfloor = N. \tag{4.2}$$

2. Calculation of PerEn: Given an embedding dimension *m*, for each scale factor τ, the PerEn of $\mathbf{x}^{(\tau)}$ is calculated [17, 25].

4.2.2 Improved Multiscale Permutation Entropy (IMPE)

The conventional univariate coarse-graining process [21] has two main drawbacks:

- This process is not symmetric. For example, as can be seen in Figure 2.4 (Chapter 2), at scale 3, we could rationally expect the metric to behave the same for u_3 and u_4, in comparison with u_2 and u_3. However, at scale 3, u_1, u_2, and u_3 are separated from u_4, u_5, and u_6.
- The second drawback is the relative variability of the MSE-based results for long temporal scales. The coarse-graining process causes the length of a signal decreases to $\left\lfloor \frac{L}{\tau_{max}} \right\rfloor$. Thus, when the scale factor is high, the number of samples in the coarse-grained sequence decreases. This may result in an unstable measure of entropy.

To deal with these shortcomings and increase the stability and reliability of MPE-based results, especially for short time series or at long temporal scale factors, we propose the IMPE technique [169] based on the idea originally developed in [202]. The IMPE of the univariate signal $\mathbf{u}$ with length L at scale τ is calculated in two main steps:

1. For scale factor τ, τ different time series ${}_\ell\mathbf{z}^{(\tau)}$ $(1 \leq \ell \leq \tau)$, corresponding to different starting points of the coarse-graining process, are created as follows:

$${}_\ell z_j{}^{(\tau)} = \frac{1}{\tau} \sum_{i=\ell+\tau(j-1)}^{\ell+\tau j-1} u_i, \quad 1 \leq j \leq \left\lfloor \frac{L}{\tau} \right\rfloor = N, \quad 1 \leq \ell \leq \tau. \tag{4.3}$$

The phenomenon of improved coarse-graining process (also called univariate refined composite coarse-graining process [30, 202]) is illustrated in Figure 4.2.

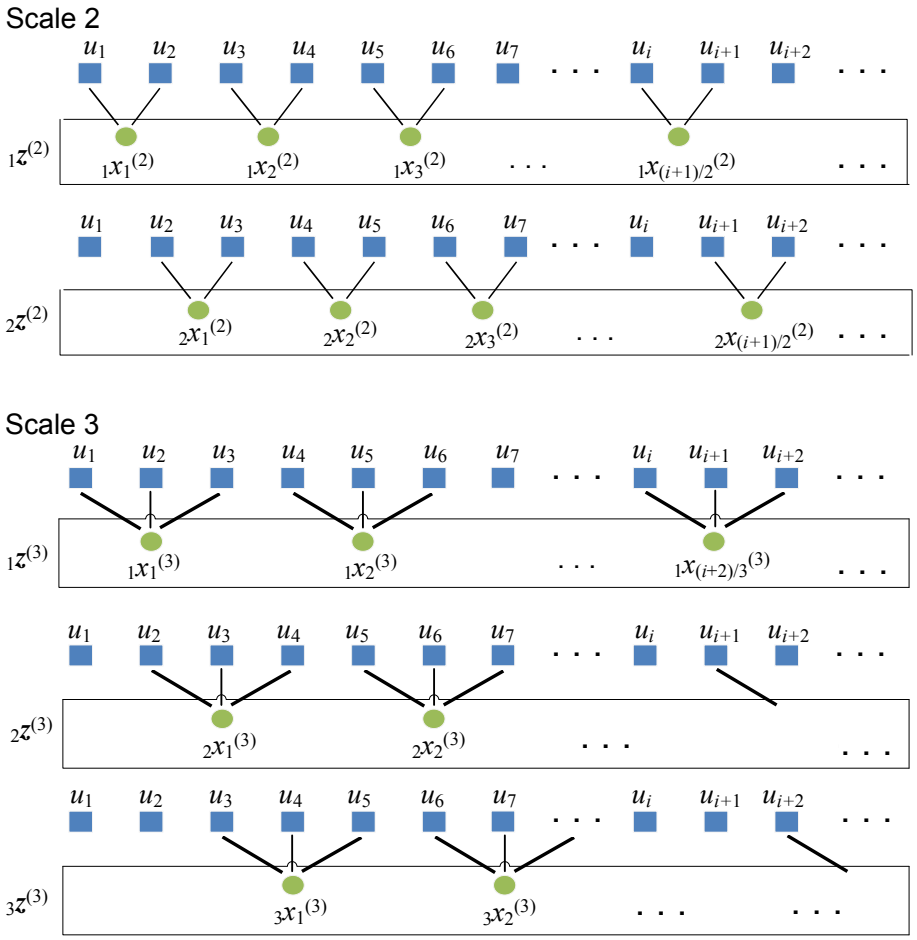


Figure 4.2: Demonstration of the univariate refined composite coarse-graining of a sequence for scale factor $\tau = 2$ and $\tau = 3$.

2. Given an embedding dimension m, the average of the PerEn values of τ different signals ${}_{\ell}\mathbf{z}^{(\tau)}$ $(1 \leq \ell \leq \tau)$ is defined as the IMPE value of the signal $\mathbf{u}$ at scale factor τ [169].

4.2.3 Parameters of MPE and IMPE

To work with reliable statistics to calculate PerEn, it is recommended that $(m+1)!$ is smaller than the length of the original signal [167, 187]. For MPE, since the coarse-graining process causes the length of a signal decreases to $\left\lfloor \frac{L}{\tau_{max}} \right\rfloor$, it is advisable $(m+1)! < \left\lfloor \frac{L}{\tau_{max}} \right\rfloor$ [115]. In IMPE, we consider τ coarse-grained time series with length $\left\lfloor \frac{L}{\tau_{max}} \right\rfloor$. Therefore, the total sample points calculated in IMPE is $\tau \times \left\lfloor \frac{L}{\tau_{max}} \right\rfloor \approx L$. Thus, IMPE follows $(m+1)! < L$, leading to more reliable results, especially for short signals [28].

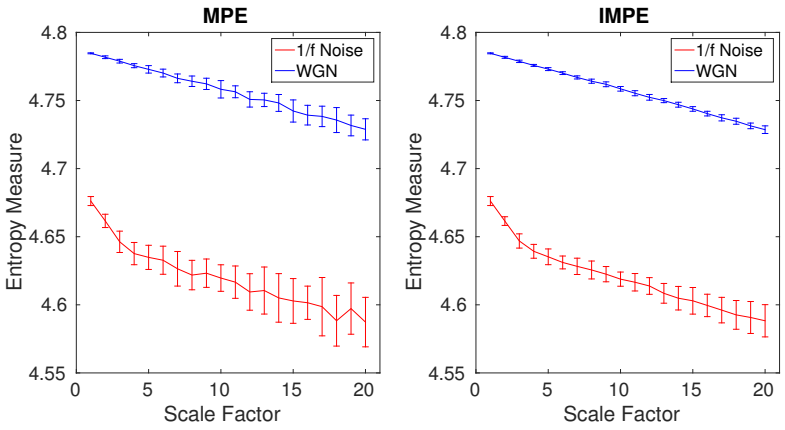


Figure 4.3: Mean value and SD of results obtained by the MPE and IMPE computed from 40 different realizations of $1/f$ *noise and WGN signals with length 20,000 samples. Red and blue demonstrate* $1/f$ *noise and WGN results, respectively.*

Table 4.1: CVs of MPE and IMPE values for $1/f$ *noise and WGN at scale 10.*

MPE of WGN	IMPE of WGN	MPE of $1/f$ noise	IMPE of $1/f$ noise
0.0013	0.0003	0.0021	0.0011

4.2.4 Results and Discussion

Figure 4.3 depicts the mean and standard deviation (SD) values of results obtained by MPE and IMPE for 40 different realizations of WGN and $1/f$ noise signals with length 20,000 samples. To compare the stability of MPE- and IMPE-based results, the coefficient of variation (CV) is used. Scale factor $\tau = 10$ as a trade-off between short and long scales is considered. The CV values, depicted in Table 4.1, show that IMPE-based CV values are noticeably smaller than those for MPE, illustrating an advantage of IMPE over MPE. Here, $m = 5$ is set for both the MPE and IMPE.

Nevertheless, the profiles illustrate that WGN is more complex than $1/f$ noise, in contradiction to the fact that, unlike WGN, $1/f$ noise contains complex structures across multiple scales, leading to higher complexity [21, 22, 161].

4.3 Univariate Refined Composite Multiscale Fuzzy Entropy (RCMFE)

As evidenced in the last Section, MPE and IMPE do not follow the key hypothesis of the concept of complexity for $1/f$ noise and WGN [31]. To overcome this problem, and increase the stability of MSE values for short signals (e.g., 100 sample points for embedding dimension 2 - see 4.3), refined composite MSE (RCMSE) was developed [202]. Although RCMSE better distinguishes the dynamics of short time series [202], it still leads to undefined or unreliable results for short signals [30]. As fuzzy entropy (FuzEn) addresses the problem of undefined SampEn values and increases the stability of SampEn for short time series [30, 103], RCMFE on the basis of RCMSE is proposed in this Thesis [30]. The algorithm of RCMFE at scale factor τ is defined as follows:

1. Assume we have a signal $\mathbf{u}$ with length L. In RCMFE, like IMPE [169] and RCMSE [202], for scale factor τ, τ different signals ${}_{\ell}\mathbf{z}^{(\tau)}$ are created according to Equation (4.3). Then, for each of the τ different series ${}_{\ell}\mathbf{z}^{(\tau)}$, all template vectors ${}_{\ell}\mathbf{z}_{\Lambda}^{m,d}$ $(\Lambda = 1, 2, \ldots, N-(m-1)\,d)$ are created as follows [16]:

$$ {}_{\ell}\mathbf{z}_{\Lambda}^{m,d} = \left\{ {}_{\ell}z_{\Lambda}, {}_{\ell}z_{\Lambda+d}, \ldots, {}_{\ell}z_{\Lambda+(m-1)d} \right\}, \tag{4.4} $$

   where *d* and *m* are the time delay and embedding dimension, respectively.

2. The distance between each of ${}_{\ell}\mathbf{z}_{\Lambda}^{m,d}$ and ${}_{\ell}\mathbf{z}_{a}^{m,d}$ is defined as ${}_{\ell}\Delta_{\Lambda,a} = \text{ChebDist}[{}_{\ell}\mathbf{z}_{\Lambda}^{m,d} - {}_{\ell}\mathbf{z}0(\Lambda), \mathbf{z}_{a}^{m,d} - {}_{\ell}\mathbf{z}0(a)], \Lambda \neq a$, where ChebDist denotes the Chebyshev distance, and ${}_{\ell}\mathbf{z}0(\Lambda)$ is the average of $\{{}_{\ell}z_{\Lambda}, {}_{\ell}z_{\Lambda+d}, \ldots, {}_{\ell}z_{\Lambda+(m-1)d}\}$ to remove the baseline [103]. Given a FuzEn power n_f and tolerance *r*, the similarity degree is calculated through a fuzzy function as $\exp\left(-\frac{{}_{\ell}\Delta_{\Lambda,a}^{n_f}}{r}\right)$. The function ${}_{\ell}\psi_{\tau}^{m,d}(n_f, r)$ is then calculated as follows:

$$ {}_{\ell}\psi_{\tau}^{m,d}(n_f, r) = \frac{1}{(N-md)(N-md-1)} \sum_{\Lambda=1}^{N-md} \sum_{a=1, a\neq\Lambda}^{N-md} \exp\left(-\frac{{}_{\ell}\Delta_{\Lambda,a}^{n_f}}{r}\right). \tag{4.5} $$

Afterwards, ${}_{\ell}\psi_{\tau}^{m+1,d}(n_f, r)$ is computed following the same procedure for embedding dimension m. Next, the average of ${}_{\ell}\psi_{\tau}^{m,d}(n_f, r)$ and ${}_{\ell}\psi_{\tau}^{m+1,d}(n_f, r)$ $(1 \leq \ell \leq \tau)$ are computed and shown as $\overline{\psi}_{\tau}^{m,d}(n_f, r)$ and $\overline{\psi}_{\tau}^{m+1,d}(n_f, r)$, respectively.

Finally, the RCMFE at scale τ is calculated as follows:

$$RCMFE(\mathbf{u}, m, r, d, \tau) = -\ln\left(\frac{\overline{\psi}_{\tau}^{m+1,d}(n_f, r)}{\overline{\psi}_{\tau}^{m,d}(n_f, r)}\right). \tag{4.6}$$

4.3.1 Parameters of MSE, MFE, RCMSE, and RCMFE

According to the suggestions for FuzEn and SampEn in Chapter 2, for all the MSE- and MFE-based methods, we set $d = 1$, $m = 2$, and $r = 0.15$ of the SD of the original signal [21, 30]. Moreover, $n_f = 2$ is set for MFE and RCMFE for all the stimulations in this Thesis [30, 103].

4.3.2 Sensitivity of MSE, RCMSE, MFE, and RCMFE to Signal Length

To evaluate the sensitivity of MSE, RCMSE, MFE, and RCMFE to the signal length, WGN and $1/f$ noise signals as functions of the number of sample points (*L*) are considered. Figures 4.4, 4.5, 4.6, and 4.7 respectively depict the MSE, RCMSE, MFE, and RCMFE values for the signal length 100, 300, 1,000, 3,000, 10,000, and 30,000 computed from 40 different realizations of WGN and $1/f$ noise.

For WGN, the entropy values of all multiscale approaches decrease monotonically with scale factor τ. However, for $1/f$ noise, the entropy values become approximately constant over larger-scale factors. These facts are in agreement with the fact that, unlike WGN, $1/f$ noise has structure across all scale factors [21, 22]. Note that each error bar of each scale factor τ depicts the SD of the results. The profiles suggest that the greater the value of *L*, the more robust the multiscale entropy estimations.

It has been advised that the number of sample points is at least 10^m, or preferably at least 20^m, to robustly estimate SampEn in time series [16, 20]. Because the coarse-graining step reduces the length of times series by the scale factor τ, and here we have $\tau_{max} = 10$ and $m = 2$, the original signal should have at least 1,000 samples. As mentioned in Chapter 2, in

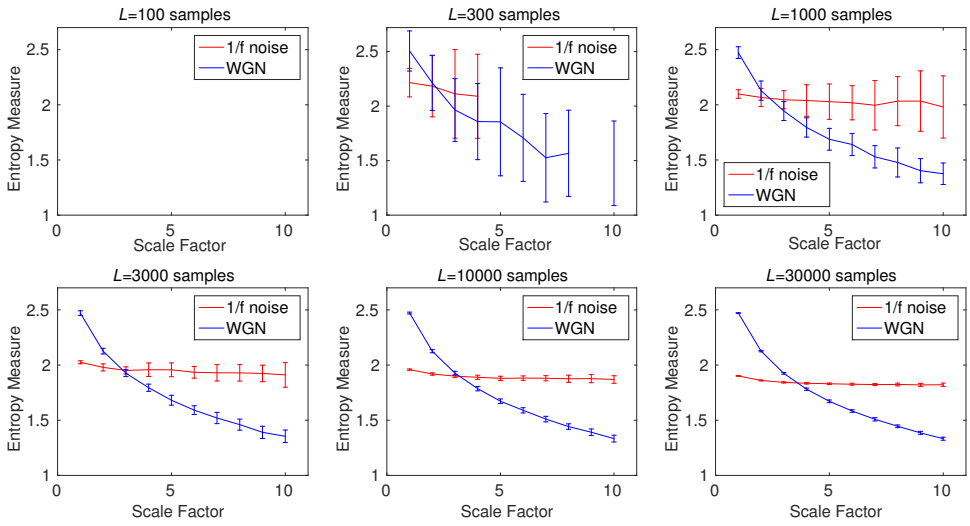


Figure 4.4: MSE as a function of data length L, 100, 300, 1,000, 3,000, 10,000, and 30,000 sample points computed from 40 different WGN and $1/f$ noise signals. The entropy values are undefined for noise signals with the length of 100 and 300 at all and large-scale factors, respectively. Red and blue demonstrate $1/f$ noise and WGN results, respectively.

SampEn, the number of matches whose differences are smaller than a predefined threshold r is counted. When the signal length is too small, this number may be 0, leading to undefined values. Accordingly, as can be seen in Figure 4.4, the SampEn values are undefined for the signals with the length of 100 and 300 samples at all and large-scale factors, respectively.

For RCMSE at scale factor τ, although the length of the signal decreases τ times, we take into account τ different coarse-grained signals [30, 202]. Therefore, in refined composite-based algorithms, we have τ times larger number of instances in comparison with their corresponding basic versions, leading to more reliable results, especially for short signals. This fact can be seen in Figure 4.5 in comparison with Figure 4.4. Although RCMSE outperforms MSE in terms of reliability for short signals, RCMSE values for $L = 100$ and $L = 300$ are still undefined at some scale factors.

However, the FuzEn-based algorithms do not count matches, yet consider all possible range of distances between any two composite vectors. Therefore, MFE and RCMFE avoid resulting in undefined entropy values in such situations. The results obtained by the RCMFE (Figure 4.7) have considerably smaller SD values, especially for short signals, than those obtained by MFE (Figure 4.6). Nevertheless, SampEn and FuzEn, as the second step of the algorithms of MSE

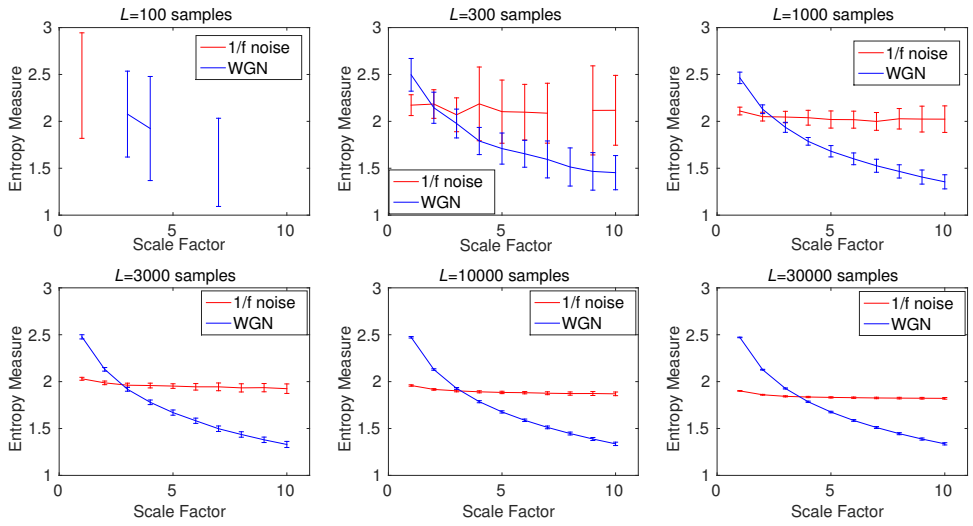


Figure 4.5: RCMSE as a function of data length L, 100, 300, 1,000, 3,000, 10,000, and 30,000 sample points computed from 40 different WGN and $1/f$ noise signals. The entropy values are undefined for noise signals with the length of 100 and 300 at most and large-scale factors, respectively. Red and blue demonstrate $1/f$ noise and WGN results, respectively.

and MFE respectively, have a computational cost of O(N^2) [18, 199, 203].

4.4 Multiscale Dispersion Entropy-based Approaches

To address the aforementioned limitations of SampEn, PerEn, and FuzEn, MDE and RCMDE are introduced in this Thesis.

4.4.1 Multiscale Dispersion Entropy (MDE)

MDE includes to main steps: 1) coarse-graining process and 2) calculation of DispEn at each scale factor [31]. However, MDE is more than the combination of the coarse-graining [21] with dispersion entropy (DispEn). Instead, crucially, the mapping based on the normal cumulative distribution function (NCDF) used in the calculation of DispEn [18] for the first temporal scale is maintained across all scales. In fact, in MDE and RCMDE, μ and σ of NCDF are respectively set at the average and SD of the original signal and they remain constant for all scale factors. This approach is similar to keeping r constant fixed (usually 0.15 of the SD of the original signal) in the MSE-based algorithms [21]. To increase the stability of MDE-based results, RCMDE is proposed in this Thesis as well.

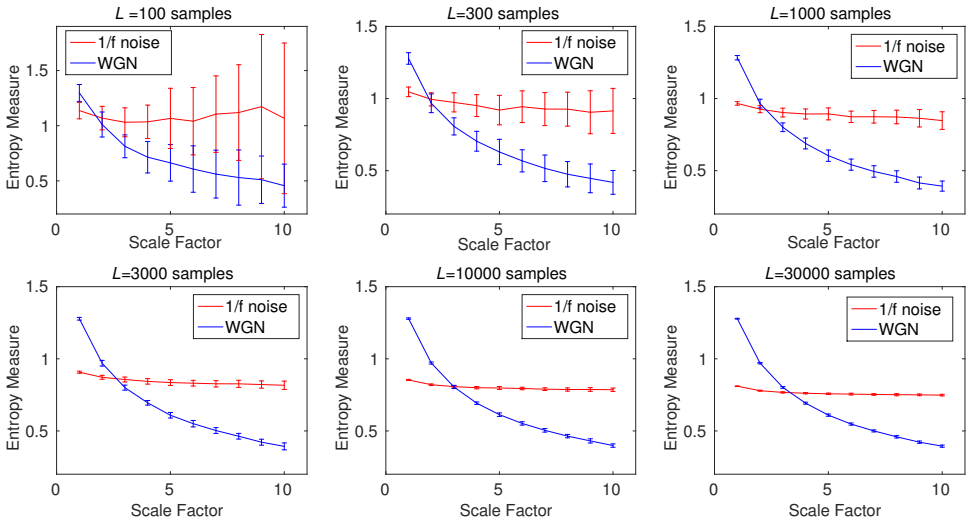


Figure 4.6: MFE as a function of data length L, 100, 300, 1,000, 3,000, 10,000, and 30,000 sample points computed from 40 different WGN and $1/f$ noise signals. Red and blue demonstrate $1/f$ noise and WGN results, respectively.

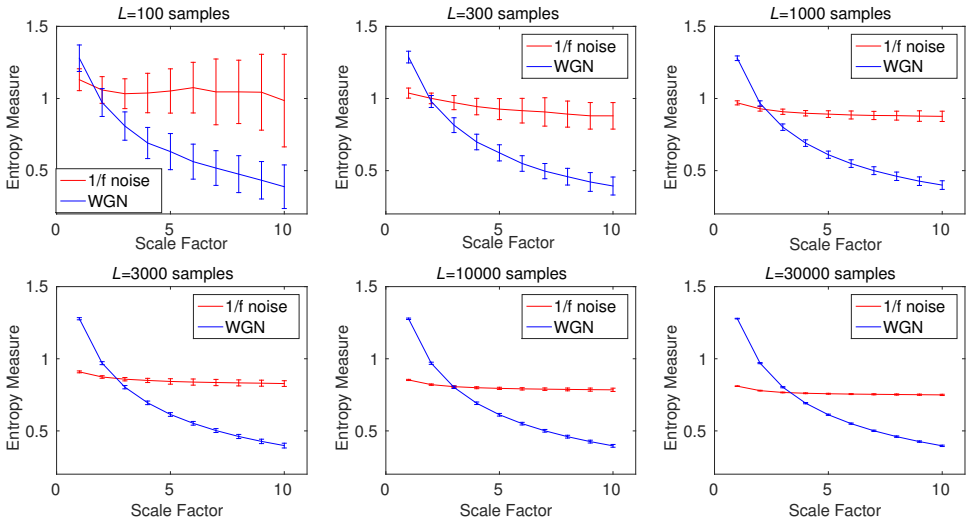


Figure 4.7: RCMFE as a function of data length L, 100, 300, 1,000, 3,000, 10,000, and 30,000 sample points computed from 40 different WGN and $1/f$ noise signals. Red and blue demonstrate $1/f$ noise and WGN results, respectively.

4.4.2 Refined Composite Dispersion Entropy (RCMDE)

The RCMDE of **u** with length *N* at scale factor τ is calculated as follows:

1. In RCMDE, like RCMFE and IMPE, for scale factor τ, τ different time series ${}_{\ell}\mathbf{z}^{(\tau)}$ $(1 \leq \ell \leq \tau)$ are created according to Equation (4.3).

2. For each ${}_{\ell}\mathbf{z}^{(\tau)}$ with length *N*, $\{{}_{\ell}z_1^{(\tau)}, \ldots, {}_{\ell}z_N^{(\tau)}\}$ are mapped to c classes with integer indices from 1 to c. To this end, the NCDF is first used to overcome the problem of assigning the majority of ${}_{\ell}z_j^{(\tau)}$ $(j = 1, \ldots, N)$ to only few classes in case maximum or minimum values are noticeable larger or smaller than the mean/median value of the signal. The NCDF maps ${}_{\ell}z_j^{(\tau)}$ into ${}_{\ell}y_j^{(\tau)}$ from 0 to 1 as follows:

$$ {}_{\ell}y_j^{(\tau)} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{{}_{\ell}z_j^{(\tau)}} e^{\frac{-(t-\mu)^2}{2\sigma^2}} \, \mathrm{d}t \tag{4.7} $$

   where σ and μ are the SD and mean of time series **u**, respectively.

   Then, we use a linear algorithm to assign each ${}_{\ell}y_j^{(\tau)}$ to an integer from 1 to c. To do so, for each member of the mapped signal, we use ${}_{\ell}\omega_j^{(\tau)} = \text{round}(c \cdot {}_{\ell}y_j^{(\tau)} + 0.5)$, where ${}_{\ell}\omega_j^{(\tau)}$ denotes the j^{th} member of the classified time series and rounding to the constant integer [18,31]. It is worth noting that the other non-linear mapping techniques described in Chapter 3 (Subsection 3.2.7) can be used in this step.

   Time series ${}_{\ell}\mathbf{w}_{\Lambda}^{m,\tau}$ are made with embedding dimension m and time delay d according to ${}_{\ell}\mathbf{w}_{\Lambda}^{m,\tau} = \{{}_{\ell}w_{\Lambda}^{(\tau)}, {}_{\ell}w_{\Lambda+d}^{(\tau)}, \ldots, {}_{\ell}w_{\Lambda+(m-1)d}^{(\tau)}\}$, $\Lambda = 1, \ldots, N-(m-1)d$ [17, 18].

   Each time series ${}_{\ell}\mathbf{w}_{\Lambda}^{m,\tau}$ is mapped to a dispersion pattern $\pi_{v_0 v_1 \ldots v_{m-1}}$, where ${}_{\ell}w_{\Lambda}^{(\tau)} = v_0$, ${}_{\ell}w_{\Lambda+d}^{(\tau)} = v_1$,..., ${}_{\ell}w_{\Lambda+(m-1)d}^{(\tau)} = v_{m-1}$. The number of possible dispersion patterns assigned to each vector ${}_{\ell}\mathbf{w}_{\Lambda}^{m,\tau}$ is equal to c^m since ${}_{\ell}\mathbf{w}_{\Lambda}^{m,\tau}$ has m elements and each can be one of the integers from 1 to c [18].

   For each of c^m potential dispersion patterns $\pi_{v_0 \ldots v_{m-1}}$, relative frequency is obtained as follows:

$$ {}_{\ell}Pr^{(\tau)}(\pi_{v_0 \ldots v_{m-1}}) = \frac{\#\{\Lambda \,|\Lambda \leq N-(m-1)d, \mathbf{w}_{\Lambda}^{m,\tau} \text{ has type } \pi_{v_0 \ldots v_{m-1}}\}}{N-(m-1)d} \tag{4.8} $$

where # means cardinality. In fact, ${}_{\ell}Pr^{(\tau)}(\pi_{v_0\ldots v_{m-1}})$ shows the number of dispersion patterns of $\pi_{v_0\ldots v_{m-1}}$ that is assigned to $\mathbf{w}_{\Lambda}^{m,\tau}$, divided by the total number of embedded signals with embedding dimension m.

Finally, RCMDE at scale factor τ is defined as the Shannon entropy value of the averages of the relative frequency values ${}_{\ell}Pr^{(\tau)}(\pi_{v_0\ldots v_{m-1}})$ $(1 \leq \ell \leq \tau)$. In fact, given the embedding dimension *d*, number of classes *c*, and time delay *d*, RCMDE at scale factor τ is defined as follows:

$$RCMDE(\mathbf{u}, m, c, d, \tau) = -\sum_{\pi=1}^{c^m} \bar{P}r(\pi_{v_0\ldots v_{m-1}}) \cdot \ln\left(\bar{P}r(\pi_{v_0\ldots v_{m-1}})\right), \tag{4.9}$$

where ln is the natural logarithm and $\bar{P}r(\pi_{v_0\ldots v_{m-1}}) = \frac{1}{\tau}\sum_{1}^{\tau} {}_{\ell}Pr^{(\tau)}(\pi_{v_0\ldots v_{m-1}})$ with the relative frequency of the dispersion pattern π in the series ${}_{\ell}\mathbf{z}^{(\tau)}$.

4.4.3 Parameters of MDE and RCMDE

There are four parameters for MDE, including the embedding dimension *m*, number of classes *c*, time delay *d*, and maximum scale factor τ_{max}. In practice, it is recommended to use $d = 1$, because aliasing may occur for $d > 1$ [18]. Clearly, we need $c > 1$ in order to avoid the trivial case of having only one dispersion pattern. For MDE and RCMDE, here, we use $c = 6$ for all signals according to [18], although the range $2 < c < 9$ leads to similar findings. For more information about *c*, *m*, and *d*, please refer to [18, 29].

To work with reliable statistics to calculate DispEn, it was suggested that the number of potential dispersion patterns is smaller than the length of the signal ($c^m < L$) [18]. For MDE, since the coarse-graining process causes the length of a signal decreases to $\left\lfloor \frac{L}{\tau_{max}} \right\rfloor$, it is recommended $c^m < \left\lfloor \frac{L}{\tau_{max}} \right\rfloor$. In RCDME, we consider τ coarse-grained time series with length $\left\lfloor \frac{L}{\tau_{max}} \right\rfloor$. Therefore, the total sample points calculated in RCMDE is $\tau \times \left\lfloor \frac{L}{\tau_{max}} \right\rfloor \approx L$. Thus, RCMDE follows $c^m < L$, leading to more reliable results, especially for short signals.

4.4.4 Simulation Results for Noise Signals

The MDE and RCMDE are used to distinguish the dynamics of $1/f$ noise and WGN with different lengths. The results obtained by MDE and RCMDE are respectively depicted in

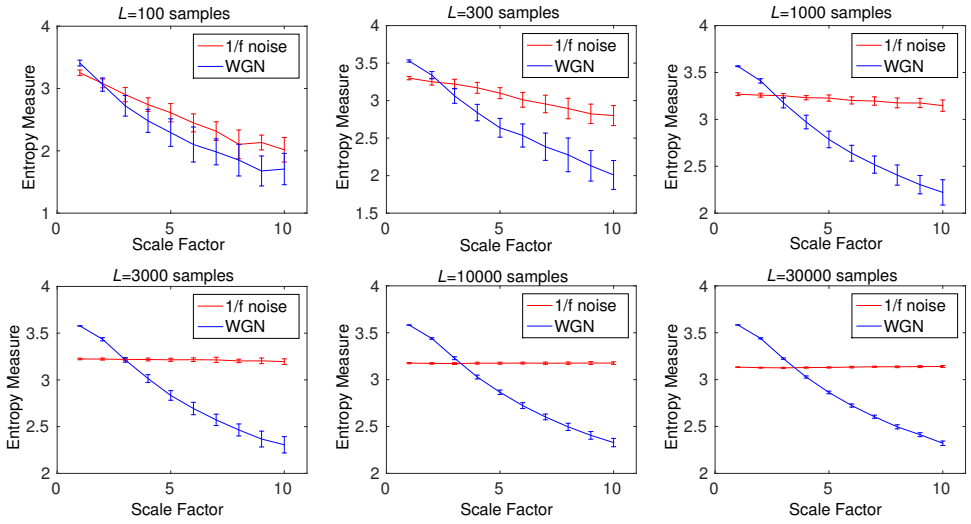


Figure 4.8: MDE as a function of data length L, 100, 300, 1,000, 3,000, 10,000, and 30,000 sample points computed from 40 different WGN and $1/f$ noise signals. Red and blue demonstrate $1/f$ noise and WGN results, respectively.

Figures 4.8 and 4.9. They are all consistent with the fact that $1/f$ noise is more complex though less irregular than WGN [21, 22, 30]. The results show that MDE and RCMDE, unlike MSE and RCMSE, do not yield undefined values as expected theoretically. Here, *m* is equal to 2.

4.5 Performance Results

In this Section, the ability of the proposed and existing univariate multiscale entropy approaches is investigated by the use of several synthetic and real univariate time series. MSE and MFE are based on conditional entropy [21, 103], while MDE is based on Shannon's definition of entropy [31]. Nevertheless, the comparison of MDE with MSE and MFE is meaningful as the latter two are the most common entropy approaches that follow the Costa's definition of complexity [39].

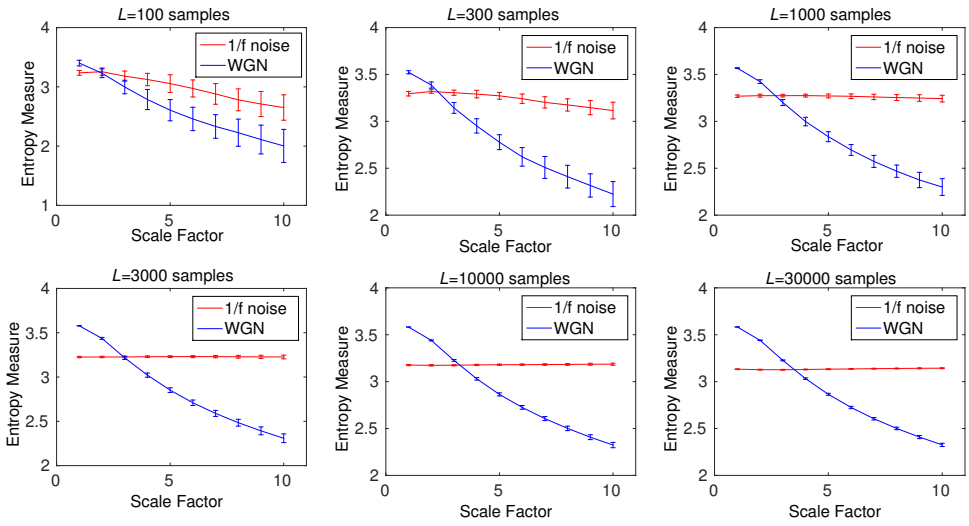


Figure 4.9: RCMDE as a function of data length L, 100, 300, 1,000, 3,000, 10,000, and 30,000 computed from 40 different WGN and $1/f$ noise signals. Red and blue demonstrate $1/f$ noise and WGN results, respectively.

4.5.1 Synthetic Signals

4.5.1.1 Multiscale Entropy Methods vs. Noise Signals

To compare the results obtained by MSE, MFE, MDE, RCMSE, RCMFE, and RCMDE, the CV values for both the WGN and $1/f$ noise with different lengths are depicted in Table 4.2. The refined composite technique makes MSE, MFE, and MDE more stable for short and long signals. The MFE and RCMFE methods are respectively more stable than MSE and RCMSE for short signals (100 and 300 sample points), while the CV values for MSE and RCMSE are smaller for long series (10,000 and 30,000 sample points). Overall, MDE and RCMDE lead to the smallest CV values for both the short and long time series.

4.5.1.2 Multiscale Entropy Methods vs. Noise Power

The multiscale methods are also applied to a periodic signal with additive noise using a moving window of 1500 samples (10 s) with 80% overlap. Here, for MDE and RCMDE, τ_{max} and m respectively were 15 and 2. Figure 4.10 shows the MSE-, MFE-, MDE-, RCMSE-, RCMFE-, and RCMDE-based profiles. As expected, the entropy values for all the methods increase along the signal. It is worth mentioning that the coarse-graining process at scale factor τ can

Table 4.2: CVs of MSE, RCMSE, MFE, RCMFE, MDE, and RCMDE values for $1/f$ noise and WGN with different lengths at scale five.

Number of samples →	100	300	1,000	3,000	10,000	30,000
MSE of WGN	undefined	0.2666	0.0586	0.0264	0.0123	0.0072
RCMSE of WGN	undefined	0.0970	0.0359	0.0167	0.0070	0.0048
MFE of WGN	0.2497	0.1265	0.0649	0.0312	0.0196	0.0127
RCMFE of WGN	0.1980	0.0896	0.0410	0.0210	0.0149	0.0071
MDE of WGN	0.0964	0.0474	0.0316	0.0182	0.0080	0.0049
RCMDE of WGN	0.0688	0.0286	0.0187	0.0093	0.0057	0.0037
MSE of $1/f$ noise	undefined	undefined	0.0789	0.0319	0.0102	0.0038
RCMSE of $1/f$ noise	undefined	0.1600	0.0452	0.0125	0.0067	0.0036
MFE of $1/f$ noise	0.2560	0.1108	0.0462	0.0238	0.0125	0.0065
RCMFE of $1/f$ noise	0.1458	0.0787	0.0260	0.0204	0.0095	0.0040
MDE of $1/f$ noise	0.0564	0.0235	0.0102	0.0050	0.0033	0.0019
RCMDE of $1/f$ noise	0.0488	0.0111	0.0063	0.0031	0.0021	0.0013

be considered as a low-pass filter with cut-off frequency $\frac{f_s}{2\tau}$ [204]. Thus, the entropy values decrease while increasing the scale factor.

To sum up, the results show that all the methods lead to the similar results, although the RCMDE, RCMFE, and RCMSE results are slightly more stable than, respectively, MDE, MFE, and MSE ones, as evidenced by the smoother nature of variations in Figure 4.10. Hence, when a high level of noise is present, the refined composite technique makes the multiscale approaches more stable.

4.5.1.3 Multiscale Entropy Methods vs. Chaotic Behavior

To understand the effect of refined composite technique on the output of a nonlinear system without noise, we use the Lorenz signal with lengths 450 and 4,500 sample points. The results obtained by MSE, MFE, MDE, RCMSE, RCMFE, and RCMDE are depicted in Figure 4.11. The increase in entropy values along the temporal scale may be explained by the two following reasons: 1) the coarse-graining process is considered with cut-off frequency $\frac{150}{20} = 7.5$ at scale 10 and this is too large to filter the Lorenz series with the frequency components between around 0 and 4 Hz (see Figure 4.1); and 2) increasing the time scale (*i.e.,* averaging consecutive data points) is an effective decorrelation of a data with a finite correlation time [201]. The results are in agreement with [30, 201]

To investigate the effect of the refined composite technique on the stability of results, the CVs for the multiscale approaches at scale 5 are calculated. The smallest CVs are obtained by MDE and RCMDE approaches. The results also suggest that the refined composite does not

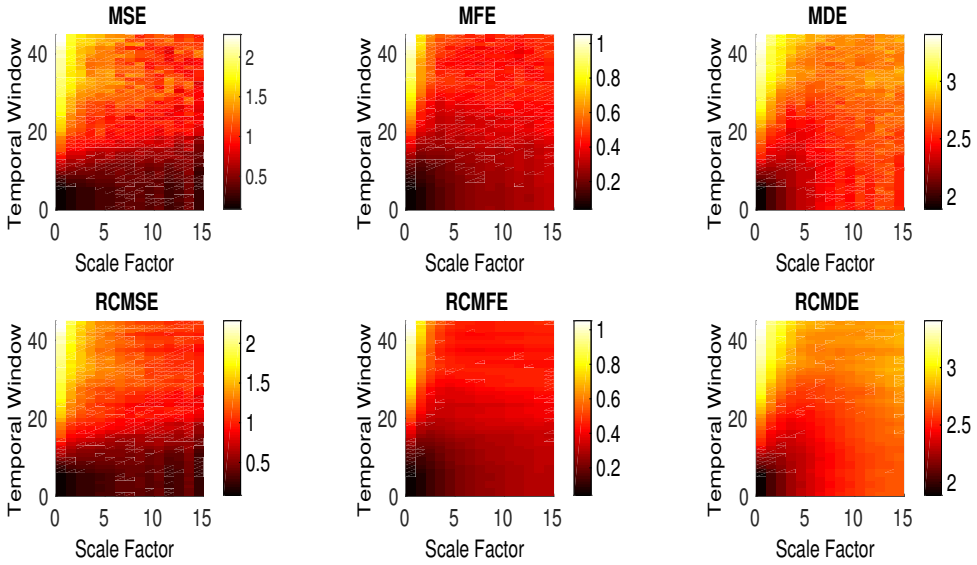


Figure 4.10: Results of the test performed to understand better the behavior of the MSE, MFE, MDE, RCMSE, RCMFE, and RCMDE for the periodic signal with additive noise (see Figure 3.1(c) in Chapter 3).

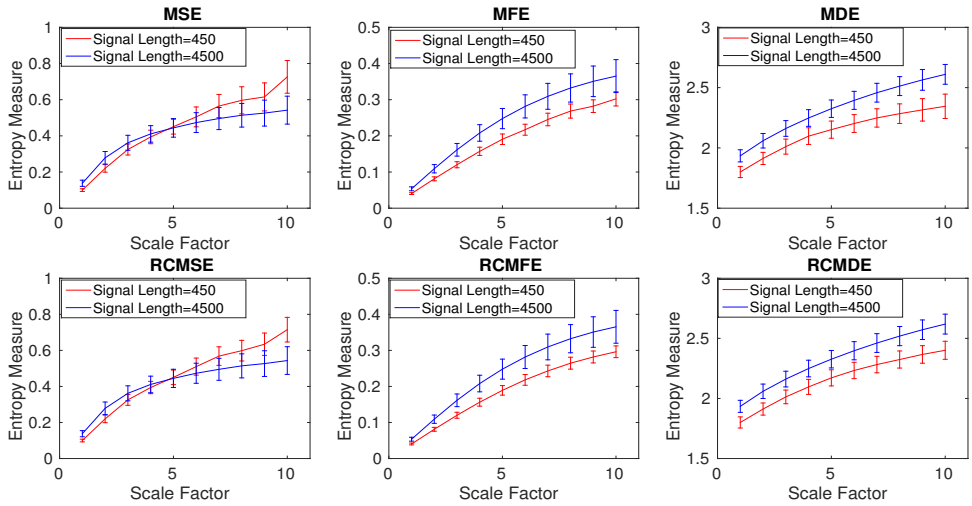


Figure 4.11: Mean and SD of the results obtained by the MSE, MFE, MDE, RCMSE, RCMFE, and RCMDE for the Lorenz series with lengths 450 and 4,500 sample points (see Figure 4.1) .

Table 4.3: CVs of MSE, RCMSE, MFE, RCMFE, MDE, and RCMDE values for the 40 different realizations of the Lorenz signals (see Figure 4.1) with length 450 and 4,500 samples at scale five.

Signal length	MSE	MFE	MDE	RCMSE	RCMFE	RCMDE
450 sample points	0.1000	0.0759	0.0898	0.0700	0.0331	0.0309
4,500 sample points	0.1156	0.1128	0.0310	0.1134	0.1119	0.0312

improve the stability of profiles for the signal with length 4,500 samples (long signals). For the Lorenz series with length 450 sample points, RCMSE, RCMFE, and RCMDE lead to smaller CV values in comparison with MSE, MFE, and MDE, in that order, showing the importance of the refined composite method to characterize small time series.

Overall, the results shown in Tables 4.2 and 4.3 suggest that when dealing with short or noisy signals, the refined composite technique makes results more stable, otherwise there is not a considerable difference between the results obtained by MSE, MFE, and MDE and RCMSE, RCMFE, and RCMDE, respectively.

4.5.2 Real Biomedical Datasets

4.5.2.1 Dataset of Focal and Non-focal Brain Activity

For the focal and non-focal EEG dataset (see Subsection 3.1.2), the results obtained by MSE, MFE, MDE, RCMSE, RCMFE, and RCMDE, depicted in Figure 4.12, show that the non-focal signals are more complex than the focal ones. This fact is in agreement with previous studies [162, 163]. The results demonstrate that all the techniques lead to the similar findings, albeit the MDE-based methods are significantly faster than MSE-based ones, as illustrated later. Note that, for MDE and RCMDE, τ_{max} and m respectively were 30 and 3. It also should be mentioned that the average entropy values over 2 channels for these bivariate EEG signals are reported for the univariate complexity techniques.

To compare the results, the CV values obtained by all the univariate multiscale techniques are calculated at scale factor 15. These are shown in Table 4.4. The CV values for MDE, RCMDE, MFE, and RCMFE illustrate that the refined composite approach does not increase the stability of the MDE and MFE profiles, while RCMSE-based CVs are slightly smaller than those for MSE. Overall, the smallest CV values are achieved by MDE and RCMDE.

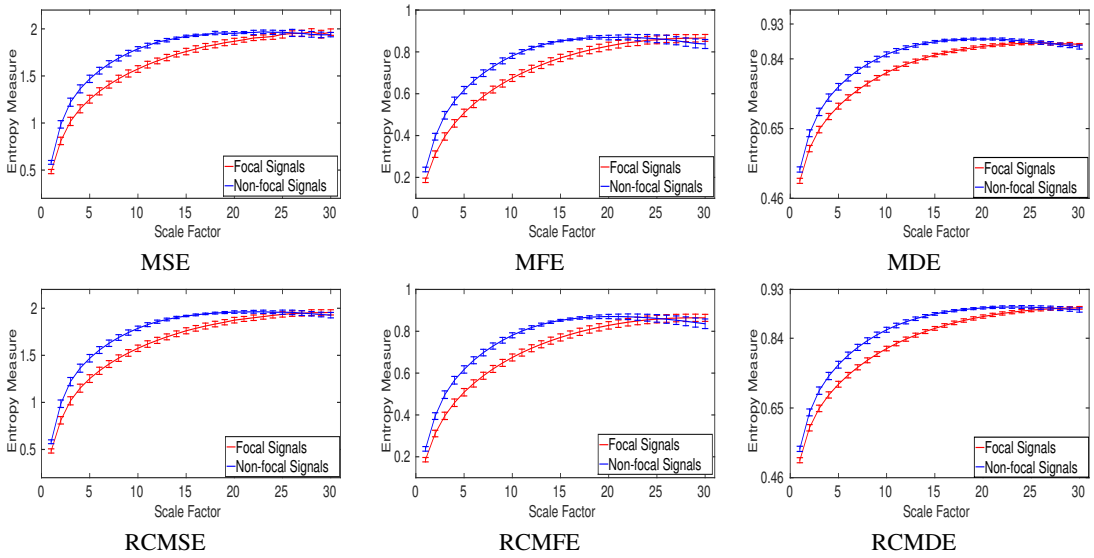


Figure 4.12: Mean value and SD of results obtained by the MSE, MFE, MDE, RCMSE, RCMFE, and RCMDE computed from the focal and non-focal EEGs.

Table 4.4: CVs of MSE, RCMSE, MFE, RCMFE, MDE, and RCMDE values for the focal and non-focal EEGs at scale 15.

Signals	MSE	MFE	MDE	RCMSE	RCMFE	RCMDE
Focal EEGs	0.0194	0.0211	0.0053	0.0183	0.0206	0.0053
Non-focal EEGs	0.0063	0.0144	0.0038	0.0052	0.0141	0.0038

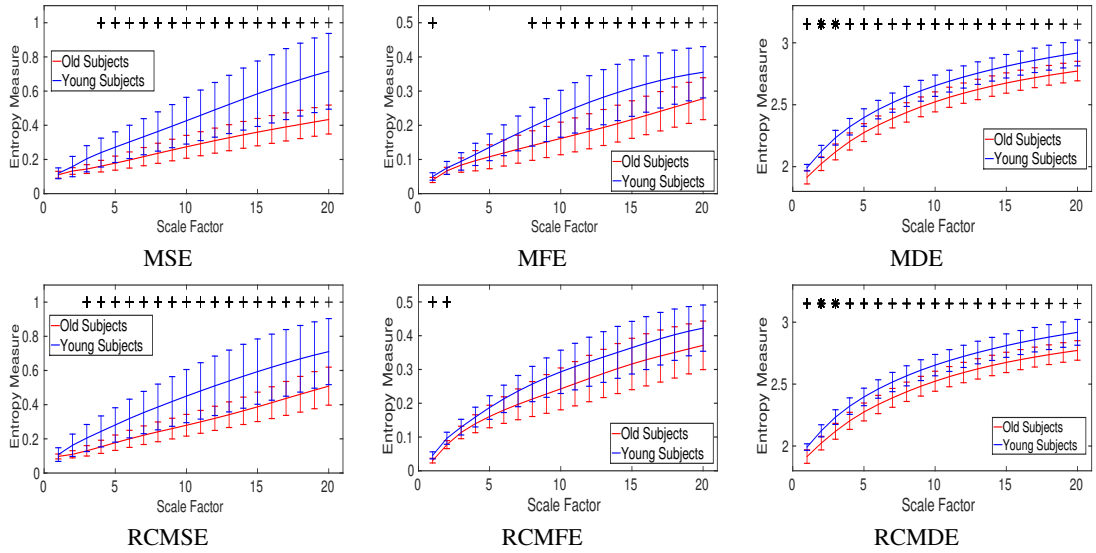


*Figure 4.13: Mean and SD of results obtained by the MSE, MFE, MDE, RCMSE, RCMFE, and RCMDE of the blood pressure recordings for young and old subjects in Fantasia database. The scale factors with p-values between 0.001 and 0.05, and smaller than 0.001 are respectively shown with + and *.*

4.5.2.2 Fantasia Dataset of Blood Pressure Recordings

In Figure 4.13, the mean and SD of the MSE, MFE, MDE, RCMSE, RCMFE, and RCMDE values computed from young and old subjects' blood pressure recordings in the Fantasia database are illustrated. For each scale factor, the average of entropy values for elderly subjects is smaller than that for young ones, in agreement with those obtained by the other entropy-based method [193] and the fact that recordings from healthy young subjects correspond to more complex states because of their ability to adapt to adverse conditions, whereas aged individuals' signals present complexity loss (see Chapter 2.5) [21–23].

The non-parametric Mann-Whitney U-test was employed to assess the differences between results for young versus old people, as the entropy values at each scale factor did not follow a normal distribution. The scales with the p-values between 0.001 and 0.05 (significant), and less than 0.001 (very significant) are shown with + and *, respectively. The results show that the MDE and RCMDE lead to the (very) significant differences for elderly and young subjects at all scale factors. However, the MSE- and MFE-based results do not show a significant difference at several temporal scales. The smallest p-values, illustrated in Table 4.5, demonstrate that MDE and RCMDE discriminate the young from old subjects better than MSE, RCMSE, MFE,

Table 4.5: Smallest p-values obtained by the MSE, MFE, MDE, RCMSE, RCMFE, and RCMDE of the blood pressure recordings for young and old subjects in Fantasia database (Mann-Whitney U-test).

MSE	MFE	MDE	RCMSE	RCMFE	RCMDE
0.0017	0.0073	0.0008	0.0058	0.0073	0.0008

and RCMFE. Moreover, the refined composite method does not improve the performance of MSE, MFE, and MDE.

4.6 Computational Time of Univariate Multiscale Entropy Methods

To evaluate the computation time of the proposed and existing multiscale entropy methods, we use WGN signals with different lengths, changing from 300 to 100,000 sample points. The results are shown in Table 4.6. The MDE (and RCMDE), MSE (and RCMSE), and MFE (and RCMFE) codes are respectively available at [195], [197], and [198].

For all the methods, the refined composite technique considerably increases the computational time. The running times for the MSE (or RCMSE) with m =2 and 3 are close since the value of *m* does not change the computational time of SampEn noticeably (see Subsection 3.4 in Chapter 3). In contrast, the larger the value of *m*, the high computational time of MDE and MFE. The MFE-based approaches are noticeably slower than MSE- and MDE-based ones. Note that the MSE code used in this Thesis was optimized, while MFE was implemented based on its straight-forward algorithm without any optimization. Therefore, this might be, at least partially, the reason for the difference in computational time.

For the WGN with 300 sample points, there is not a big difference between the computational times of MDE and MSE. However, for the WGN with 1,000 sample points or longer, MDE and RCMDE are noticeably faster than MSE and RCMSE, respectively. This computational advantage of MDE and RCMDE increases notably with the signal length. It is in agreement with the fact that the computational cost of SampEn (like FuzEn) and DispEn are O(N^2) [199] and O(N) [18], respectively.

Table 4.6: Computational time of MSE, MFE, MDE, RCMSE, MFE, and RCMDE for WGN signals with different lengths, changing from 300 to 100,000 sample points.

Number of samples →	300	1,000	3,000	10,000	30,000	100,000
MSE ($m = 2$)	0.021 s	0.072 s	0.371 s	3.181 s	26.641 s	293.371 s
MSE ($m = 3$)	0.017 s	0.067 s	0.366 s	3.199 s	26.637 s	290.151 s
RCMSE ($m = 2$)	0.062 s	0.248 s	1.076 s	7.495 s	56.775 s	589.912 s
RCMSE ($m = 3$)	0.061 s	0.238 s	1.045 s	7.364 s	55.950 s	597.718 s
MFE ($m = 2$)	0.161 s	0.601 s	2.806 s	17.995 s	102.810 s	754.163 s
MFE ($m = 3$)	0.168 s	0.716 s	3.149 s	21.799 s	116.134 s	865.114 s
RCMFE ($m = 2$)	0.612 s	1.928 s	7.693 s	43.539 s	252.410 s	1716.231 s
RCMFE ($m = 3$)	0.621 s	2.001 s	7.693 s	49.428 s	294.812 s	1932.197 s
MDE ($m = 2$)	0.016 s	0.043 s	0.119 s	0.381 s	1.248 s	3.850 s
MDE ($m = 3$)	0.026 s	0.049 s	0.125 s	0.443 s	1.171 s	3.901 s
RCMDE ($m = 2$)	0.070 s	0.198 s	0.555 s	1.805 s	5.429 s	17.568 s
RCMDE ($m = 3$)	0.109 s	0.247 s	0.638 s	1.886 s	5.634 s	18.586 s

4.7 Summary

In this Chapter, the existing and developed univariate multiscale entropy methods and their benefits and shortcomings were inspected. We first proposed IMPE to increase the stability of MPE-based results. However, both MPE and IMPE do not fulfil the main hypotheses of complexity. To address this deficiency, RCMFE was developed as well. RCMFE does not lead to undefined values, and, compared with MFE, results in more stable profiles for short or noisy time series. Nevertheless, RCMFE may not be stable enough for some short signals and its computation is not fast enough for some real-time applications.

To address the deficiencies of RCMSE, RCMFE, and IMPE at the same time, MDE and RCMDE were introduced. The ability of MSE, RCMSE, MFE, RCMFE, MDE, and RCMDE was investigated by the use of several relevant synthetic signals and two datasets of physiological signals. It was found that when dealing with short or noisy signals, the refined composite makes profiles more stable, otherwise there is not a noticeable difference between the results for MSE, MFE, and MDE and those for RCMSE, RCMFE, and RCMDE, respectively. The results also showed similar behavior in terms of complexity profiles for MFE or RCMFE, MSE or RCMSE, and MDE or RCMDE, although MDE and RCMDE were noticeably faster, especially for long signals. For short signals, MDE and RCMDE, unlike MSE and RCMSE, did not lead to undefined values. In comparison with MSE- and MFE- based approaches, MDE and RCMDE led to smaller CVs for synthetic and real signals. These benefits of MDE and RCMDE show their advantages over the state-of-the-art univariate

multiscale entropy approaches.

On the whole, thanks to their ability to distinguish dynamics of time series with low computation time, we expect that MDE and RCMDE play a prominent role in the evaluation of complexity in univariate signals.

Chapter 5

Multivariate Multiscale Entropy Methods

Univariate multiscale entropy-based methods, though widespread, are not able to reveal the patterns across channels of a multivariate signal. For such time series, evaluation of cross-statistical properties between multiple channels is important for a better understanding of the underlying signal-generating system [23, 32, 110]. To deal with this challenge, multivariate multiscale sample entropy (MSE - mvMSE), was proposed to take into account both the time and spatial domain at the same time [111].

However, mvMSE values are undefined for short signals. To address this limitation, multivariate multiscale permutation entropy (MPE - mvMPE) was developed [25]. Although mvMPE has been used in a sizeable number of applications to distinguish the dynamics of multivariate signals, it does not consider the spatial domain of multi-channel signals and does not fulfil the key hypotheses of the concept of complexity [32]. To address these problems, multivariate multiscale fuzzy entropy (MFE - mvMFE) was developed [205]. However, it is not fast enough for real-time applications.

To decrease the computation time of the original mvMFE method while maintaining its advantages, we propose a new fuzzy membership function for mvMFE [32]. To increase the stability of the mvMFE-based results, the refined composite mvMFE (RCmvMFE) is developed as well [32]. Nevertheless, RCmvMFE has three drawbacks: 1) RCmvMFE values are still unreliable for very short signals; 2) it is not fast enough for real-time applications; and 3) the computation of RCmvMFE for signals with a large number of channels requires the storage of a huge number of elements.

To deal with these problems and improve the stability of RCmvMFE, we introduce multivariate multiscale dispersion entropy (MDE - mvMDE), as an extension of our recently developed

MDE [30], to quantify the complexity of multivariate time series [33].

5.1 Multivariate Signals for Evaluation

In this Section, the descriptions of correlated and uncorrelated multi-channel noise signals, bivariate autoregressive (BAR) process, and real time series are given.

5.1.1 Multivariate Noise Time Series

The irregularity of multivariate $1/f$ noise is lower than multivariate white Gaussian noise (WGN), whereas the complexity of the former is higher than the latter, since $1/f$ noise contains long-range correlations and its $1/f$ decay produces a fractal structure in time [22, 23, 32, 110]. Thus, multi-channel $1/f$ noise and WGN signals have been commonly used to assess the multivariate multiscale entropy techniques [23, 32, 206]. For more information, please refer to [21–23, 32].

5.1.1.1 Uncorrelated Multivariate Noise Signals

To understand the behaviour of the multivariate multiscale methods on uncorrelated WGN and $1/f$ noise, we first generated a trivariate time series, where originally all three data channels were realizations of mutually independent $1/f$ noise. Then, we gradually decreased the number of data channels representing $1/f$ noise (from 3 to 0) and at the same time, increased the number of variates representing independent WGN (from 0 to 3) [206]. The number of channels was always three.

5.1.1.2 Correlated Multivariate Noise Signals

To create correlated bivariate noise time series, we first generated a bivariate uncorrelated random time series $\mathbf{H}$. Afterwards, $\mathbf{H}$ was multiplied with the standard deviation (hereafter, σ) and then, the value of the mean (hereafter, μ) was added. Next, $\mathbf{H}$ was multiplied by the upper triangular matrix $\mathbf{L}$ obtained from the Cholesky decomposition of a defined correlation matrix $\mathbf{R}$ (which is positive and symmetric) to set the cross-correlation. Here, we set $\mathbf{R}$ as

follows [23, 32]:

$$\mathbf{R} = \begin{bmatrix} 1 & 0.95 \\ 0.95 & 1 \end{bmatrix} \tag{5.1}$$

An in-depth study on the effect of correlated versus uncorrelated $1/f$ noise and WGN on multiscale entropy approaches can be found in [23].

5.1.1.3 Bivariate Autoregressive (BAR) Process

Based on the fact that the higher the order of an autoregressive (AR) process, the more complex it is [23], we evaluate the multivariate multiscale entropy methods on a BAR process describing the evolution of two variables as a linear function of their past values according to:

$$\mathbf{y}_n = \mathbf{e}_n + \sum_{\gamma=1}^{\gamma_{\max}} \mathbf{y}_{n-\gamma}\mathbf{A}_\gamma, \tag{5.2}$$

where $\mathbf{y}_n = \{y_n(1), y_n(2)\}$ is the n^{th} sample of a bidimensional time series, $\gamma_{\max}$ is the maximum lag in the BAR model, $\mathbf{A}_\gamma$ denotes the 2×2 matrix of parameters corresponding to lag order γ, and $\mathbf{e}_n$ is the 2×1 vector of error terms assumed to be WGN [207]. In practice, the stability of an empirical BAR process can be evaluated by the eigenvalues of the coefficient matrix $\mathbf{A}_\gamma$. Thus, for all the BAR processes used in this Chapter, we considered the fact that if the moduli of the eigenvalues of $\mathbf{A}_\gamma$ are smaller than 1, the BAR process is stable [208].

5.1.2 Real Biomedical Datasets

5.1.2.1 Dataset of Stride Internal Fluctuations

A decrease in irregularity for activity fluctuations with aging and Alzheimer's disease (AD) was shown [209]. To this end, and to compare mvMDE, mvMFE, and mvMSE for short multi-channel signals, stride interval recordings are used [110, 210]. The time series were recorded from ten young, healthy men. Mean age was 21.7 years, changing from 18 to 29 years. Height and weight were 1.77 $\pm$ 0.08 meters (mean $\pm$ standard deviation (SD)) and 71.8

$\pm$ 10.7 kg (mean $\pm$ SD), respectively. All ten participants provided informed written consent walking at slow, normal, and fast pace and also walking a metronome set to each subject's mean stride interval. The lengths of the recordings are around Three walking paces were considered as different variables from the same system. In this way, we expect to be able to discriminate between the metronomically-paced and self-spaced walking. For further information, please refer to [210].

5.1.2.2 Dataset of Focal and Non-focal Brain Activity

This bivariate electroencephalogram (EEG) dataset was described in Subsection 3.1.2 (Chapter 3).

5.2 Multivariate Multiscale Permutation Entropy (mvMPE)

The algorithm of mvMPE at scale factor τ includes two main steps [25]:

1. Multivariate coarse-graining process: Given a *p*-channel time series $\mathbf{U} = \{u_{k,i}\}_{k=1,2,\dots,p}^{i=1,2,\dots,L}$ of length L, for each channel, the original signal is first divided into non-overlapping segments of length τ. Next, the average of each segment is calculated to derive the coarse-grained signal ${x_{k,j}}^{(\tau)}$ as follows [23]:

$$ {x_{k,j}}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} u_{k,i},\ 1 \leq j \leq \left\lfloor \frac{L}{\tau} \right\rfloor = N\ ,\ 1 \leq k \leq p, \tag{5.3} $$

   where N is the length of the coarse-grained signal $\mathbf{X} = \{x_{k,j}\}_{k=1,2,\dots,p}^{j=1,2,\dots,N}$.

2. Calculation of multivariate permutation entropy (mvPE): For each series $\mathbf{x}_k$ and for each time Λ, we embed the signal $\mathbf{x}_k$ in an *m*-dimensional space to obtain the reconstruction vectors $\mathbf{x}_{k,\Lambda}^{m,d} = \{x_{k,\Lambda}, x_{k,\Lambda+d}, \dots, x_{k,\Lambda+(m-2)d}, x_{k,\Lambda+(m-1)d}\}$ for $\Lambda = 1, 2, \dots, N - (m-1)d$, where *m* and *d* denote the embedding dimension and time delay, respectively. To calculate the mvPE, the elements of $\mathbf{x}_{k,\Lambda}^{m,d}$ are associated with numbers from 0 to $m-1$ and arranged in increasing order as follows:

$$ \{x_{k,\Lambda+(\aleph_1-1)d}, x_{k,\Lambda+(\aleph_2-1)l}, \dots, x_{k,\Lambda+(\aleph_{m-1}-1)d}, x_{k,\Lambda+(\aleph_m-1)d}\}, \tag{5.4} $$

where $\aleph_*$ is the (time) index of the element in the reconstruction vector [25]. There are $m!$ potential ordinal patterns or symbol sequences η_t ($1 \leq t \leq m!$), termed "motifs". Then, the occurrence of each order pattern η_t in channel *k*, denoted as $f^{(k)}(\eta_t)$, is counted. For each η_t, $Pr^{(k)}(\eta_t)$ represents the relative frequency in channel *k* as follows:

$$Pr^{(k)}(\eta_t) = \frac{f^{(k)}(\eta_t)}{p(N-(m-1)d)}. \tag{5.5}$$

The differences between Equation (5.5) and its corresponding Equation in the original definition of permutation entropy (see Equation (3.3)) in Chapter 3) is that the relative frequency in mvPE is divided by the number of channels *p* so that $\sum_{k=1}^{p}\sum_{t=1}^{m!} Pr^{(k)}(\eta_t) = 1$ holds [25].

The marginal relative frequencies demonstrating the distribution of the motifs are defined as follows:

$$Pr(\eta_t) = \sum_{k=1}^{p} Pr^{(k)}(\eta_t). \tag{5.6}$$

Consequently, mvPE at each scale τ is defined as [25]:

$$\text{mvPE}(\mathbf{X}, m, d) = -\sum_{t=1}^{m!} Pr(\eta_t) \cdot \ln(Pr(\eta_t)), \tag{5.7}$$

where $\ln$ denotes the natural logarithm. When all marginal relative frequencies have equal probabilities, the largest value of mvPE is obtained, which has a value of $\ln(m!)$. In contrast, if there is only one $Pr(\eta_t)$ different from zero, which demonstrates a completely regular signal in every channel, the smallest value of mvPE is achieved 0 [25].

5.2.1 Parameters of mvMPE

Choosing an acceptable embedding dimension *m* in mvMPE is challenging. As mentioned in Chapter 4, to work with reliable statistics when calculating MPE, it is highly recommended $(m+1)! < \frac{L}{\tau_{\max}}$ [115]. Accordingly, because the number of samples increases to pL for the multi-channel $\mathbf{U}$, it is recommended $(m+1)! < \frac{pL}{\tau_{\max}}$ [115]. In addition, when *m* is large, the computation time of mvMPE increases. While *m* is high, the number of potential permutation

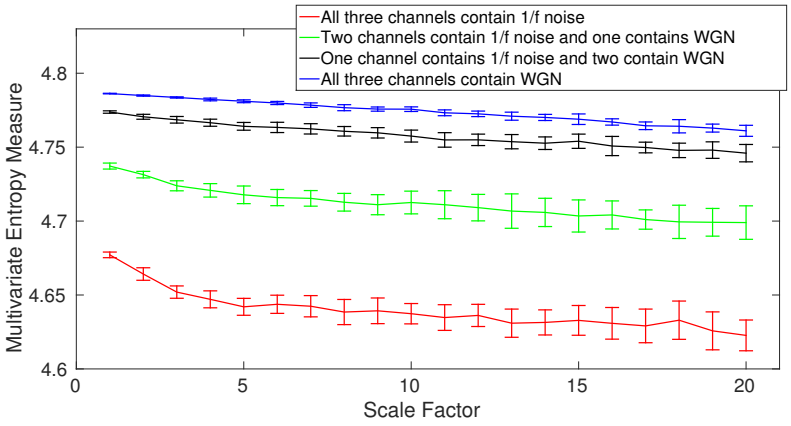


Figure 5.1: Mean value and SD of the results using mvMPE computed from 40 different uncorrelated trivariate WGN and $1/f$ noise time series with length 15,000.

patterns is large, leading to more reliable results. Overall, we should make a trade-off between the aforementioned cases.

5.2.2 Results and Discussion

The mvMPE method is used for 40 independent realizations of uncorrelated trivariate WGN and $1/f$ noise with length 15,000 sample points, described in Section 5.1. The mean value and SD of the results for mvMPE are depicted in Figure 5.1. It is found that the mvMPE values of trivariate WGN signals are higher than those of the other trivariate time series at all scale factors, in contradiction to the fact that multi-channel WGN signals are less complex than multi-channel $1/f$ noise [23, 32, 110]. Furthermore, as can be seen from its algorithm, the mvMPE technique does not consider the spatial domain of multivariate signals.

5.3 Multivariate Multiscale Fuzzy Entropy (mvMFE)

To address the aforementioned shortcomings of mvMPE and the problem of undefined mvMSE values, the mvMFE techniques with different fuzzy membership functions are proposed [32, 205]. In fact, the mvMFE approaches deal with the spatial and time domains simultaneously. The mvMFE algorithms include two steps:

1. Multivariate coarse-graining process: Given a p-channel time series $\mathbf{U} = \{u_{k,i}\}_{k=1,2,\dots,p}^{i=1,2,\dots,L}$ of length L, the multivariate coarse-grained signal $\mathbf{X} = \{x_{k,j}^{(\tau)}\}$ $(1 \leq j \leq N)$ is calculated according to Equation (5.3).

2. Calculation of multivariate fuzzy entropy (mvFE) at each scale factor: Based on the Taken's embedding theorem [114], the multivariate embedded vectors are initially generated as:

$$\begin{aligned} X_{\mathbf{m}}(b) = [&x_{1,b}, x_{1,b+d_1}, \ldots, x_{1,b+(m_1-1)d_1}, \\ &x_{2,b}, x_{2,b+d_2}, \ldots, x_{2,b+(m_2-1)d_2}, \ldots, \\ &x_{p,b}, x_{p,b+d_p}, \ldots, x_{p,b+(m_p-1)d_p}], \end{aligned} \tag{5.8}$$

where $\mathbf{m} = [m_1, m_2, \ldots, m_p]$ and $\mathbf{d} = [d_1, d_2, \ldots, d_p]$ are the embedding dimension and time delay vectors, respectively. Note that the length of $X_{\mathbf{m}}(b)$ is $\sum_{k=1}^{p} m_k$. For simplicity, $d_k = d$ and $m_k = m$ are set in the literature [23, 32, 110, 111, 205]. That is, all the embedding dimension values and all the delay values are equal.

For the *p*-variate time series $\{\mathbf{x}_k\}_{k=1}^{p}$, the mvFE algorithm, as a natural extension of the standard univariate fuzzy entropy, includes the following steps [32, 205]:

1. For each scale factor τ, form multivariate embedded vectors $X_{\mathbf{m}}(b) \in R^m$, where $b = 1, 2, \ldots, N - n$ and $n = \max\{\mathbf{m}\} \times \max\{\mathbf{d}\}$ [205]. Herein, $n = m \times d$.

2. Calculate the Chebyshev distance (ChebDist) between any two composite delay vectors $X_{\mathbf{m}}(b)$ and $X_{\mathbf{m}}(\beta)$ as the maximum norm.

3. One of the most important shortcomings of the multivariate sample entropy (mvSE) is that the method ignores every Chebyshev distance between two composite delay vectors $X_{\mathbf{m}}(b)$ and $X_{\mathbf{m}}(\beta)$ that is larger than a defined threshold *r* [32]. To alleviate this problem, a fuzzy membership function $\theta(d, r)$ was proposed as follows [205]:

$$\theta(\Delta_{b,\beta}, r) = \begin{cases} 1, & \Delta_{b,\beta} \leq r \\ \exp\left(-\ln(2)\left(\frac{r-\Delta_{b,\beta}}{r}\right)^2\right), & \Delta_{b,\beta} > r \end{cases} \tag{5.9}$$

where $\Delta_{b,\beta} = \text{ChebDist}[X_{\mathbf{m}}(b), X_{\mathbf{m}}(\beta)]$. Although the problem of undefined mvSE values is solved by using this function, the mvFE method considerably slower than mvSE, especially when the number of channels or sample points of every channel, or the value of embedding dimension *m* is large. To tackle this deficiency, we propose to

use the following well-known fuzzy membership function [32]:

$$\theta(\Delta_{b,\beta}, r) = \exp\left(-(\Delta_{b,\beta})^{n_f}/r\right), \tag{5.10}$$

where n_f shows the fuzzy power and is usually equal to 2 [32]. Next, define a global quantity $\phi^m(n_f, r)$, as the average membership grade as:

$$\phi^m(n_f, r) = \frac{1}{(N-n)} \sum_{b=1}^{N-n} \frac{\sum\limits_{\beta=1, b\neq\beta}^{N-n} \theta(\Delta_{b,\beta}, r)}{N-n-1} \tag{5.11}$$

4. Extend the dimensionality of the multivariate delay vector in Equation (5.8) from m to $(m+1)$. This can be done in *p* different ways, as from $[m_1, m_2, \ldots, m_h, \ldots, m_p]$ to $[m_1, m_2, \ldots, m_{h+1}, \ldots, m_p](h = 1, \ldots, p)$. In the process, the dimension of the other variables are unchanged [32].

5. Calculate $\phi^{(m+1)}(n_f, r)$, where denotes the average over all $\phi^{(m_h+1)}(n_f, r)$ values in an $(m+1)$-dimensional space.

6. Finally, mvFE is defined as [32]:

$$\mathrm{mvFE}(\mathbf{X}, \mathbf{m}, r, n_f, \mathbf{d}) = -\ln\left(\frac{\phi^{(m+1)}(n_f, r)}{\phi^m(n_f, r)}\right). \tag{5.12}$$

Since multivariate time series may have different amplitude ranges, the distances calculated from embedded vectors obtained with Takens embedding theorem may be dominated by components of the vectors coming from the time series with the largest amplitudes. Thus, we scale all of the data channels to the same amplitude range and normalise each data channel to unit SD so that the total variation becomes equal to the number of channels or variables [23, 32, 111].

5.3.1 Parameters of the mvMSE and mvMFE methods

In this Chapter, d_k, m_k, and r for the mvMSE and mvMFE were respectively set as 1, 2, and 0.15 of the SD of the original time series following recommendations in [23, 32]. The maximum scale factor for mvMSE and mvMFE also follows [23, 32]. In the algorithm of

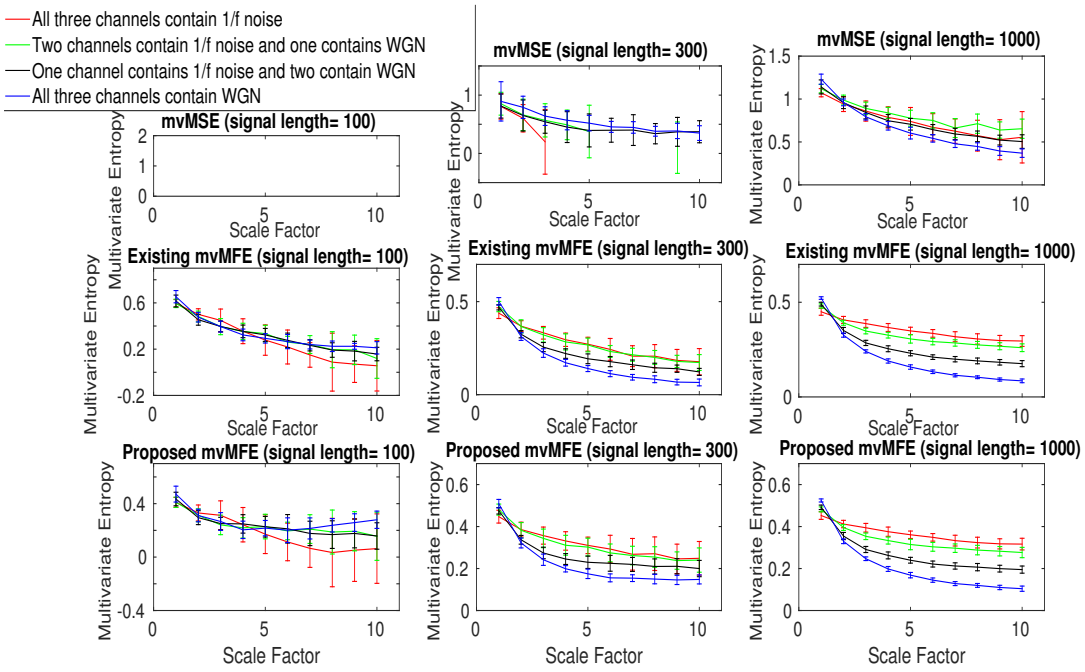


Figure 5.2: Mean value and SD of the results using the mvMSE, existing mvMFE, and proposed mvMFE methods computed from 40 different uncorrelated trivariate WGN and $1/f$ noise time series with lengths 100, 300, and 1,000 sample points. The mvMSE values are undefined for noise signals with the length of 100 and 300 at all and high scale factors, respectively.

mvSE and mvFE, at least $\binom{Np}{2} + Np(pm+1)$ elements are stored.

5.3.2 Results and Discussion

The mvMSE and existing and proposed mvMFE methods are used for 40 independent realizations of uncorrelated trivariate WGN and $1/f$ noise with lengths 100, 300, and 1,000 sample points, described in Section 5.1. The results obtained by these approaches are shown in Figure 5.2. The mvMSE values are undefined for trivariate time series with the length of 100 and 300 at all and high scale factors, respectively. In contrast, the mvMFE-based values are defined for different lengths and scale factors. The existing and proposed mvMFE lead to similar profiles, although the computation time for the existing mvMFE is at least 2 times higher than that for the proposed mvMFE.

5.4 Refined Composite Multivariate Multiscale Fuzzy Entropy (RCmvMFE)

Like the univariate coarse-graining process, the multivariate coarse-graining technique [23, 205] has two main limitations:

- This process is not symmetric. According to Figure 2.5 in Chapter 2, for instance at scale 3, we could rationally expect the measure to behave the same for $u_{k,3}$ and $u_{k,4}$, in comparison with $u_{k,2}$ and $u_{k,3}$. However, at scale 3, $u_{k,1}$, $u_{k,2}$, and $u_{k,3}$ are separated from $u_{k,4}$, $u_{k,5}$ and $u_{k,3}$.
- When the coarse graining process is computed, the number of samples of the resulting coarse-grained sequence is $\left\lfloor \frac{L}{\tau_{\max}} \right\rfloor$. When $\tau_{\max}$ is high, for each channel, the number of time sample points in the coarse-grained sequence decreases. This may yield unstable or undefined entropy values.

To tackle these shortcomings, extensions of refined composite MSE [202] and MFE [30] to their multichannel cases, i.e, RCmvMFE and RCmvMFE are developed in this Thesis [32]. The algorithm of RCmvMFE is explained below.

1. Refined composite multivariate coarse-graining process: The first step of RCmvMFE is generating τ multivariate coarse-grained time series ${}_{\ell}\mathbf{Z}^{(\tau)} = \{{}_{\ell}x_{k,j}^{(\tau)}\}$ $(1 \leq \ell \leq \tau)$, where

$$ {}_{\ell}x_{k,j}{}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+\ell}^{j\tau+\ell-1} u_{k,i}, \quad 1 \leq j \leq \left\lfloor \frac{L}{\tau} \right\rfloor = N, \quad 1 \leq k \leq p, \quad 1 \leq \ell \leq \tau \tag{5.13} $$

2. Calculation of the mvFE of each ${}_{\ell}\mathbf{Z}^{(\tau)}$: For each scale factor τ, we have τ different multivariate signals ${}_{\ell}\mathbf{Z}^{(\tau)}$ for $1 \leq \ell \leq \tau$. For each ${}_{\ell}\mathbf{Z}^{(\tau)}$, ${}_{\ell}\phi^{m}(n_f, r)$ and ${}_{\ell}\phi^{m+1}(n_f, r)$, where $\ell = 1, \ldots, \tau$, are separately calculated according to Equation (5.11). Next, the average of ${}_{\ell}\phi^{m}(n_f, r)$ and ${}_{\ell}\phi^{m+1}(n_f, r)$ are separately calculated and shown as $\overline{\phi}^{m}(n_f, r)$ and $\overline{\phi}^{m+1}(n_f, r)$. Finally, the RCmvMFE is computed as follows:

$$ \mathrm{RCmvMFE}(\mathbf{X}, \mathbf{m}, r, n_f, \mathbf{d}, \tau) = -\ln \left(\frac{\overline{\phi}^{m+1}(n_f, r)}{\overline{\phi}^{m}(n_f, r)} \right). \tag{5.14} $$

The algorithm of RCmvMSE is similar to that of RCmvMFE, except the Heaviside function is used for RCmvMSE instead of the fuzzy membership function employed in RCmvMFE [32]. As we illustrated in [32], the results for RCmvMFE are more stable than those obtained by mvMFE for noisy and short time series. However, the computation times for RCmvMFE and RCmvMSE are markedly higher than those for mvMFE and mvMSE, respectively, and the multivariate refined composite method cannot noticeably improve the stability of mvMSE or mvMFE for long real signals [32, 206, 211]. Therefore, the simulation results based on RCmvMFE and RCmvMSE are not shown in this Chapter. Of note is that the codes of our developed mvMFE, RCmvMFE, and RCmvMSE are publicly available at [212].

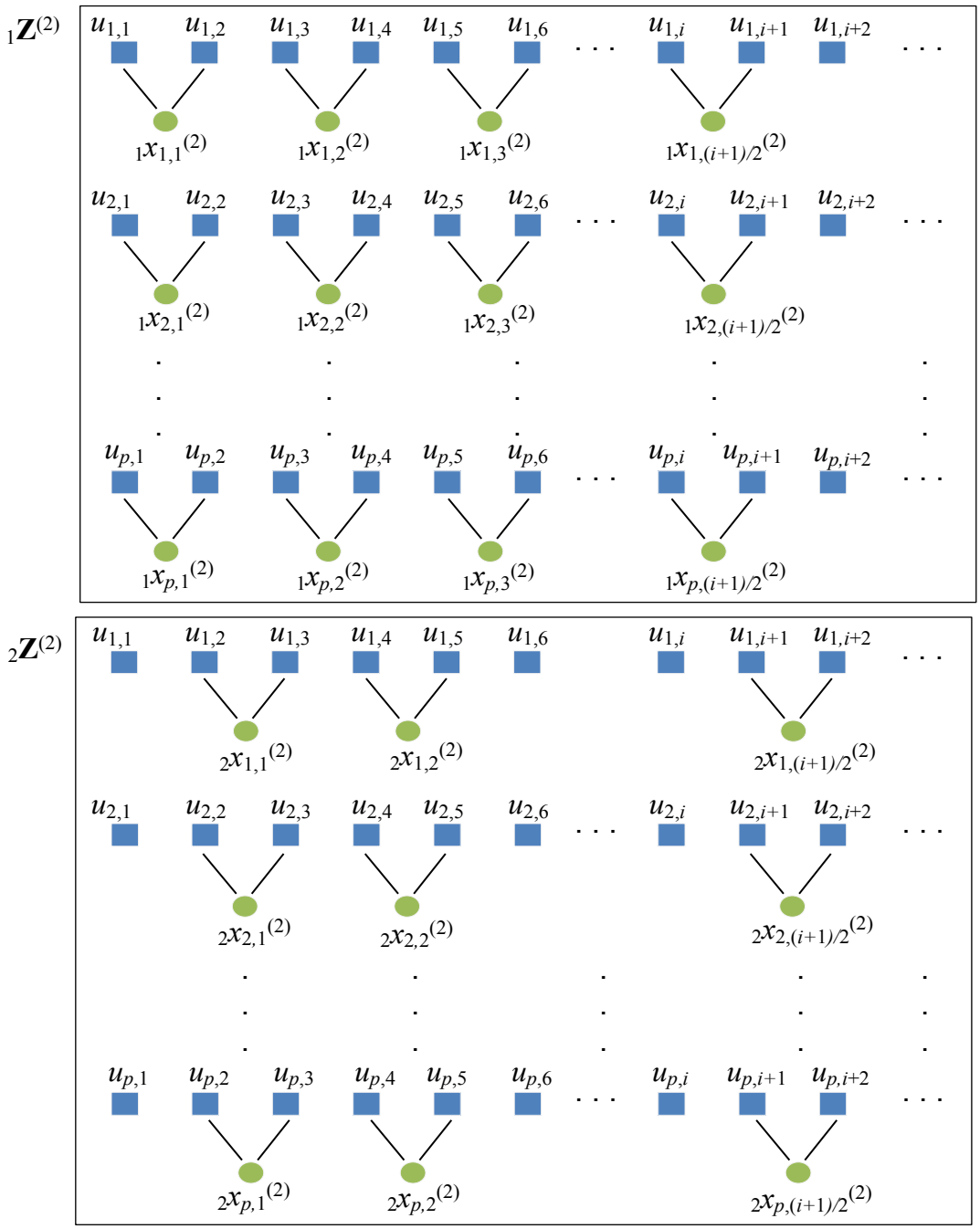
5.5 Multivariate Multiscale Dispersion Entropy (mvMDE)

In this Section, we propose and explore three different alternative implementations of mvMDE until we arrive at a fourth and preferred one. All the mvMDE implementations include two main steps: 1) coarse-graining process for multivariate time series; and 2) multivariate dispersion entropy (mvDE), as an extension of our recently developed DispEn [18]. It is worth noting that for all the mvMDE algorithms, the mapping based on the normal cumulative distribution function (NCDF) used in the calculation of mvDE for the first temporal scale factor is maintained fixed across all scales. In fact, in the mvMDE, μ and σ of the NCDF are respectively set at the average and SD of the original time series and they remain constant for all temporal scale factors. This fact is similar to r in the mvMSE and mvMFE, setting at a certain percentage of the SD of the original signal and remaining constant for all scales [23, 32].

5.5.1 Coarse-graining Process for Multivariate Signals

Multivariate coarse-graining process: Given a *p*-channel time series $\mathbf{U} = \{u_{k,i}\}_{k=1,2,\dots,p}^{i=1,2,\dots,L}$ of length L, for each channel, the multivariate coarse-grained signal $\mathbf{X} = x_{k,j}^{(\tau)}$ $(1 \leq j \leq N)$ is calculated according to Equation (5.3). The second step of mvMDE is calculating the mvDE of each coarse-grained signal.

Scale 2


$_1\mathbf{Z}^{(2)}$
$u_{1,1}$ $u_{1,2}$ $u_{1,3}$ $u_{1,4}$ $u_{1,5}$ $u_{1,6}$ . . . $u_{1,i}$ $u_{1,i+1}$ $u_{1,i+2}$ . . .
$_1x_{1,1}^{(2)}$ $_1x_{1,2}^{(2)}$ $_1x_{1,3}^{(2)}$ $_1x_{1,(i+1)/2}^{(2)}$
$u_{2,1}$ $u_{2,2}$ $u_{2,3}$ $u_{2,4}$ $u_{2,5}$ $u_{2,6}$ . . . $u_{2,i}$ $u_{2,i+1}$ $u_{2,i+2}$ . . .
$_1x_{2,1}^{(2)}$ $_1x_{2,2}^{(2)}$ $_1x_{2,3}^{(2)}$ $_1x_{2,(i+1)/2}^{(2)}$
$u_{p,1}$ $u_{p,2}$ $u_{p,3}$ $u_{p,4}$ $u_{p,5}$ $u_{p,6}$ . . . $u_{p,i}$ $u_{p,i+1}$ $u_{p,i+2}$ . . .
$_1x_{p,1}^{(2)}$ $_1x_{p,2}^{(2)}$ $_1x_{p,3}^{(2)}$ $_1x_{p,(i+1)/2}^{(2)}$
$_2\mathbf{Z}^{(2)}$
$u_{1,1}$ $u_{1,2}$ $u_{1,3}$ $u_{1,4}$ $u_{1,5}$ $u_{1,6}$ $u_{1,i}$ $u_{1,i+1}$ $u_{1,i+2}$ . . .
$_2x_{1,1}^{(2)}$ $_2x_{1,2}^{(2)}$ $_2x_{1,(i+1)/2}^{(2)}$
$u_{2,1}$ $u_{2,2}$ $u_{2,3}$ $u_{2,4}$ $u_{2,5}$ $u_{2,6}$ . . . $u_{2,i}$ $u_{2,i+1}$ $u_{2,i+2}$ . . .
$_2x_{2,1}^{(2)}$ $_2x_{2,2}^{(2)}$ $_2x_{2,(i+1)/2}^{(2)}$
$u_{p,1}$ $u_{p,2}$ $u_{p,3}$ $u_{p,4}$ $u_{p,5}$ $u_{p,6}$ . . . $u_{p,i}$ $u_{p,i+1}$ $u_{p,i+2}$ . . .
$_2x_{p,1}^{(2)}$ $_2x_{p,2}^{(2)}$ $_2x_{p,(i+1)/2}^{(2)}$

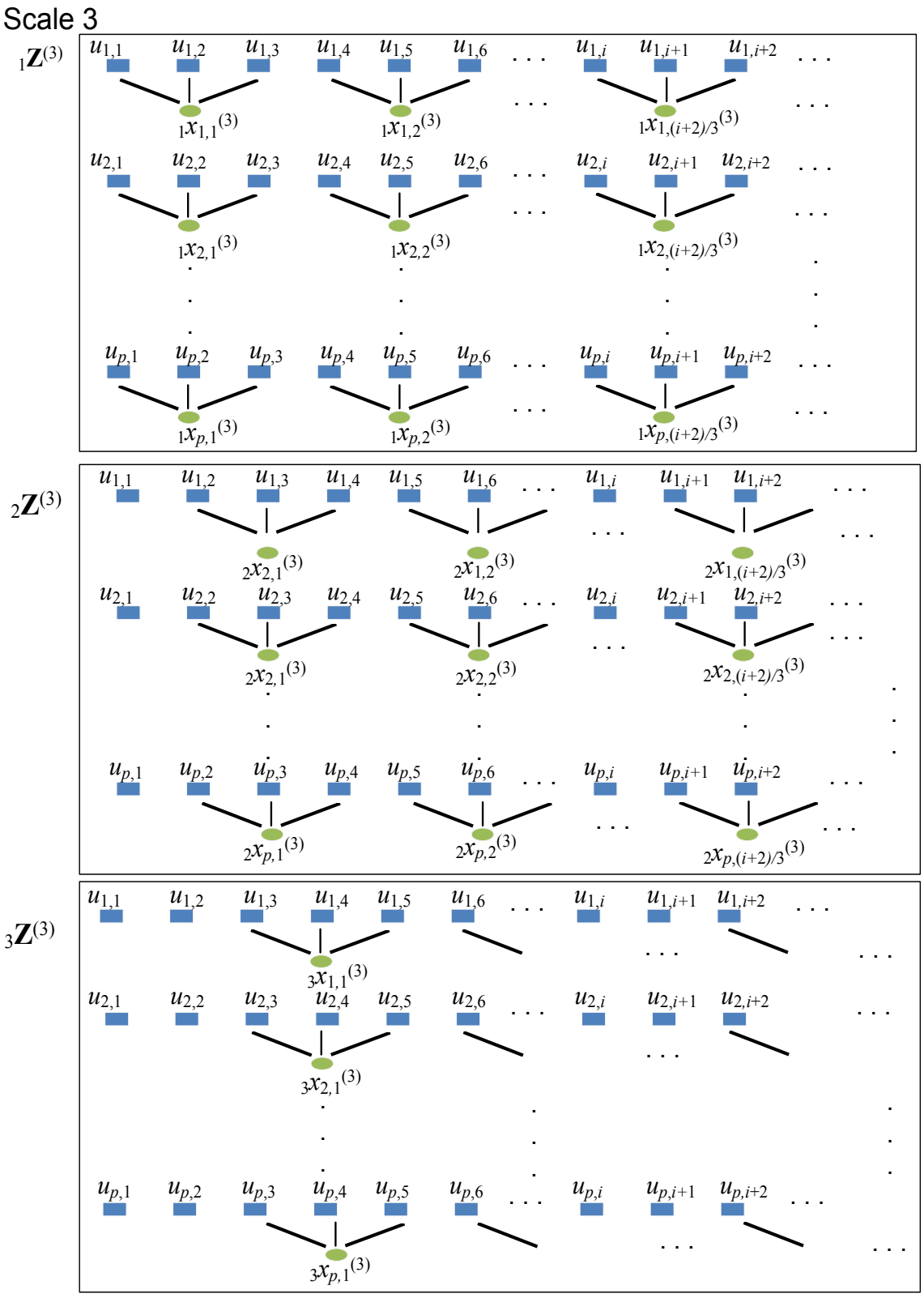


Figure 5.3: Demonstration of the refined composite multivariate coarse-graining of a multivariate sequence for scale factor $\tau = 2$ and $\tau = 3$.

5.5.2 Background Information for mvDE

We build four diverse alternative implementations of mvDE (mvDE_I to mvDE_III and mvDE) until we arrive at a preferred (or optimal) one, i.e., mvDE. However, we here present all the simpler alternatives (mvDE_I to mvDE_III), since they can still be useful in some settings and allow for clearer comparisons with other current approaches.

5.5.2.1 mvDE_I

The mvDE_I of the multi-channel coarse-grained time series $\mathbf{X} = \{x_{k,j}\}_{k=1,2,\dots,p}^{j=1,2,\dots,N}$, which is based on the mvMPE algorithm [25], is calculated as follows:

a) First, $\mathbf{X} = \{x_{k,j}\}_{k=1,2,\dots,p}^{j=1,2,\dots,N}$ are mapped to c classes with integer indices from 1 to c. Because the amplitude values of each of series $\mathbf{x}_k$ ($k = 1, 2, \dots, p$) may be dominated by the components of vectors coming from the time series with the largest amplitudes, we scale every data channel to the same amplitude range. To this end and to overcome the problem of assigning the majority of $x_{k,j}$ to only few classes when maximum or minimum values are noticeable larger or smaller than the mean/median value of the signal, the NCDF of each of $\mathbf{x}_k$ is first calculated. In fact, the NCDF maps $\mathbf{X}$ into $\mathbf{Y} = \{y_{k,j}\}_{k=1,2,\dots,p}^{j=1,2,\dots,N}$ from 0 to 1 as follows:

$$y_{k,j} = \frac{1}{\sigma_k\sqrt{2\pi}} \int_{-\infty}^{x_{k,j}} e^{\frac{-(t-\mu_k)^2}{2\sigma_k^2}} \, \mathrm{d}t \tag{5.15}$$

where σ_k and μ_k are the SD and mean of time series $\mathbf{x}_k$, respectively. Then, we use a linear algorithm to assign each $y_{k,j}$ to an integer from 1 to c. To do so, for each member of the mapped signal, we use $w_{k,j}^c = \text{round}(c \cdot y_{k,j} + 0.5)$, where $w_{k,j}^c$ denotes the j^{th} member of the classified signal in the k^{th} channel and rounding involves either increasing or decreasing a number to the next digit.

b) Time series $\mathbf{w}_{k,j}^{m,c}$ are next made with embedding dimension m and time delay d according to [16–18]:

$$\mathbf{w}_{k,\Lambda}^{m,c} = \{w_{k,\Lambda}^c, w_{k,\Lambda+d}^c, + \cdots + w_{k,\Lambda+(m-1)d}^c\}, \Lambda = 1, 2, \dots, N - (m-1)d. \tag{5.16}$$

Each time series $\mathbf{w}_{k,\Lambda}^{m,c}$ is mapped to a dispersion pattern $\pi_{v_0 v_1 \ldots v_{m-1}}$, where $w_{k,\Lambda}^{c} = v_0$, $w_{k,\Lambda+d}^{c} = v_1$,..., $w_{k,\Lambda+(m-1)d}^{c} = v_{m-1}$. The number of possible dispersion patterns that can be assigned to each time series $\mathbf{w}_{k,\Lambda}^{m,c}$ is equal to c^m because $\mathbf{w}_{k,\Lambda}^{m,c}$ has m elements and each of them can be one of the integers from 1 to c [18].

c) For each channel $1 \leq k \leq p$ and for each of c^m potential dispersion patterns $\pi_{v_0 \ldots v_{m-1}}$, relative frequency is obtained as follows:

$$Pr(\pi_{v_0 \ldots v_{m-1}}) = \frac{\#\{(k,\Lambda) \,\big|\, 1 \leq k \leq p, 1 \leq \Lambda \leq N-(m-1)d, \mathbf{w}_{k,\Lambda}^{m,c} \text{ has type } \pi_{v_0 \ldots v_{m-1}}\}}{(N-(m-1)d)p} \tag{5.17}$$

where $\#$ means cardinality. In fact, $Pr(\pi_{v_0 \ldots v_{m-1}})$ shows the number of dispersion patterns of $\pi_{v_0 \ldots v_{m-1}}$ that is assigned to $\mathbf{w}_{k,\Lambda}^{m,c}$, divided by the total number of embedded signals with embedding dimension m multiplied by the number of channels.

d) Finally, based on the Shannon's definition of entropy, the mvDE$_{\mathrm{I}}$ is calculated as follows:

$$\mathrm{mvDE_I}(\mathbf{X}, \mathrm{m}, \mathrm{c}, \mathrm{d}) = -\sum_{\pi=1}^{\mathrm{c^m}} \mathrm{Pr}(\pi_{\mathrm{v_0 \ldots v_{m-1}}}) \cdot \ln\left(\mathrm{Pr}(\pi_{\mathrm{v_0 \ldots v_{m-1}}})\right) \tag{5.18}$$

In case all possible dispersion patterns have equal probability value, the highest value of mvDE$_{\mathrm{I}}$ is obtained, which has a value of $ln(c^m)$. In contrast, if there is only one $Pr(\pi_{v_0 \ldots v_{m-1}})$ different from zero, which demonstrates a completely regular/certain signal, the smallest value of mvDE$_{\mathrm{I}}$ is obtained. In the algorithm of mvDE$_{\mathrm{I}}$, we compare Np dispersion patterns of a *p*-channel signal with c^m potential patterns. Thus, at least $c^m + Np$ elements are stored.

To work with reliable statistics to calculate MDE, it was recommended $c^m < \left\lfloor \frac{L}{\tau_{max}} \right\rfloor$ [31]. Since mvDE$_{\mathrm{I}}$ counts the dispersion patterns for every channel of a multivariate time series, it is suggested $c^m < \left\lfloor \frac{pL}{\tau_{max}} \right\rfloor$. mvDE$_{\mathrm{I}}$ extracts the dispersion patterns from each of channels regardless of their cross-channel information. Thus, mvDE$_{\mathrm{I}}$ works appropriately when the spatial components of a multivariate signal are statistically independent. However, the mvDE$_{\mathrm{I}}$ algorithm, like mvPE [25], does not consider the spatial domain of time series. To overcome this problem, we propose mvDE$_{\mathrm{II}}$ based on the Taken's theorem [32, 114].

5.5.2.2 mvDE$_{\text{II}}$

The algorithm of mvDE$_{\text{II}}$ is as follows:

a) First, like mvDE$_{\text{I}}$, $\mathbf{X} = \{x_{k,i}\}_{k=1,2,\dots,p}^{i=1,2,\dots,N}$ are mapped to $\mathbf{W} = \{w_{k,j}\}_{k=1,2,\dots,p}^{j=1,2,\dots,N}$ based on the NCDF.

b) To take into account both the spatial and time domains, multi-channel embedded vectors are generated according to the multivariate embedding theory [114]. The multivariate embedded reconstruction of **W** is defined as:

$$\begin{aligned} W_{\mathbf{m}}(\Lambda) = [&w_{1,\Lambda}, w_{1,\Lambda+d_1}, \dots, w_{1,\Lambda+(m_1-1)d_1}, \\ &w_{2,\Lambda}, w_{2,\Lambda+d_2}, \dots, w_{2,\Lambda+(m_2-1)d_2}, \dots, \\ &w_{p,\Lambda}, w_{p,\Lambda+d_p}, \dots, w_{p,\Lambda+(m_p-1)d_p}] \end{aligned} \tag{5.19}$$

where $\mathbf{m} = [m_1, m_2, \dots, m_p]$ and $\mathbf{d} = [d_1, d_2, \dots, d_p]$ denote the embedding dimension and the time lag vectors, respectively. Note that the length of $W_{\mathbf{m}}(\Lambda)$ is $\sum_{k=1}^{p} m_k$. For simplicity, we assume $d_k = d$ and $m_k = m$, that is, all the embedding dimension values and all the delay values are equal. Thus, the length of $W_{\mathbf{m}}(\Lambda)$ is mp.

c) Each series $W_{\mathbf{m}}(\Lambda)$ is mapped to a dispersion pattern $\pi_{v_0 v_1 \dots v_{mp-1}}$, where $w^c_{1,\Lambda} = v_0$, $w^c_{1,\Lambda+d} = v_1, \dots, w_{p,\Lambda+(m-1)d} = v_{mp-1}$. The number of possible dispersion patterns that can be assigned to each time series $W_{\mathbf{m}}(\Lambda)$ is equal to c^{mp}, as $W_{\mathbf{m}}(\Lambda)$ has mp elements and each of them can be one of the integers from 1 to c.

d) For each of c^{mp} potential dispersion patterns $\pi_{v_0 \dots v_{mp-1}}$, relative frequency is obtained based on the dispersion entropy algorithm [18] as follows:

$$Pr(\pi_{v_0 \dots v_{mp-1}}) = \frac{\#\{\Lambda \,\big|\, 1 \le \Lambda \le N - (m-1)d, W_{\mathbf{m}}(\Lambda) \text{ has type } \pi_{v_0 \dots v_{mp-1}}\}}{N - (m-1)d} \tag{5.20}$$

e) Finally, based on the Shannon's definition of entropy, the mvDE$_{\text{II}}$ is calculated as follows:

$$\mathrm{mvDE_{II}}(\mathbf{X}, \mathrm{m}, \mathrm{c}, \mathrm{d}) = -\sum_{\pi=1}^{\mathrm{c^{mp}}} \mathrm{Pr}(\pi_{\mathrm{v_0 \dots v_{mp-1}}}) \cdot \ln\left(\mathrm{Pr}(\pi_{\mathrm{v_0 \dots v_{mp-1}}})\right) \tag{5.21}$$

In the algorithm of mvDE$_{\text{II}}$, at least $c^{mp} + Np$ elements are stored. Thus, when p is large, the algorithm needs huge space of memory to store elements. To work with reliable statistics to calculate mvMDE$_{\text{II}}$, it is recommended $c^{mp} < \left\lfloor \frac{L}{\tau_{max}} \right\rfloor$. Thus, although mvDE$_{\text{II}}$ deals with both the spatial and time domains, the length of a signal and its number of channels should be very large and small, respectively, to reliably calculate mvDE$_{\text{II}}$ values. This leads us to consider mvDE$_{\text{III}}$.

5.5.2.3 mvDE$_{\text{III}}$

The algorithm of mvDE$_{\text{III}}$ is as follows:

a) First, like the mvDE$_{\text{I}}$ and mvDE$_{\text{II}}$ approaches, $\mathbf{X} = \{x_{k,j}\}_{k=1,2,\dots,p}^{i=j,2,\dots,N}$ are mapped to $\mathbf{W} = \{w_{k,j}\}_{k=1,2,\dots,p}^{j=1,2,\dots,N}$.

b) Multivariate embedded vectors $\mathbf{W}_{k,\mathbf{m}}(\Lambda)$ are generated according to the Taken's embedding theorem [114] with p embedding dimension vectors $\mathbf{m}_k = [1, 1, \dots, m_k, \dots, 1, 1]$ ($k = 1, \dots, p$) with length $m + p - 1$, where m_k denotes the k^{th} element of $\mathbf{m}$. For simplicity, we assume $m_k = m$ and $d_k = d$. Therefore, the length of $W_{\mathbf{m}}(\Lambda)$ is equal to $m + p - 1$.

c) Each series $\mathbf{W}_{k,\mathbf{m}}(\Lambda)$ is mapped to a dispersion pattern $\pi_{v_0 v_1 \dots v_{m+p-2}}$. The number of possible dispersion patterns that can be assigned to each time series $\mathbf{W}_{k,\mathbf{m}}(\Lambda)$ is equal to c^{m+p-1}, since the vector $\mathbf{W}_{k,\mathbf{m}}(\Lambda)$ has $m + p - 1$ elements and each of them can be one of the integers from 1 to c [18]. As we count the number of patterns for each of p different $\mathbf{m}_k$, we have considerably larger number of dispersion patterns in comparison with mvDE$_{\text{II}}$, leading to more reliable results for a signal with a small number of sample points, as shown later.

d) For each channel $1 \leq k \leq p$ and for each of c^{m+p-1} potential dispersion patterns $\pi_{v_0 \dots v_{m+p-2}}$, relative frequency is obtained as follows:

$$Pr(\pi_{v_0 \dots v_{m+p-2}}) = \frac{\#\{(k, \Lambda) \,|\, 1 \leq k \leq p, 1 \leq \Lambda \leq N - (m-1)d, \mathbf{W}_{k,\mathbf{m}}(\Lambda) \text{ has type } \pi_{v_0 \dots v_{m+p-2}}\}}{(N - (m-1)d)p} \tag{5.22}$$

e) Finally, based on the Shannon's definition of entropy, the mvDE$_{\text{II}}$ is calculated as follows:

$$\text{mvDE}_{\text{III}}(\mathbf{X}, \text{m}, \text{c}, \text{d}) = -\sum_{\pi=1}^{\text{c}^{\text{m+p-1}}} \Pr(\pi_{\text{v}_0 \ldots \text{v}_{\text{m+p-2}}}) \cdot \ln\big(\Pr(\pi_{\text{v}_0 \ldots \text{v}_{\text{m+p-2}}})\big) \tag{5.23}$$

mvDE$_{\text{III}}$ assumes embedding dimension 1 for all signals except one, which might limit the potential to explore the dynamics. Moreover, in the algorithm of mvDE$_{\text{III}}$, at least $c^{m+p-1}+Np$ elements are stored. Although this number is noticeably smaller than that for mvDE$_{\text{II}}$, the algorithm still needs to have large memory space for a signal with a large number of channels. To work with reliable statistics to calculate mvMDE$_{\text{III}}$, it is recommended $c^{m+p-1} < \left\lfloor \frac{pL}{\tau_{max}} \right\rfloor$. Therefore, albeit mvDE$_{\text{III}}$ takes into account both the spatial and time domains and needs to smaller number of sample points in comparison with mvDE$_{\text{II}}$, there is a need to have a large enough number of samples and small number of channels. To alleviate these deficiencies, we propose mvDE.

5.5.3 Multivariate Dispersion Entropy (mvDE)

The mvDE algorithm is as follows:

a) First, like mvDE$_{\text{I}}$ to mvDE$_{\text{III}}$, the multivariate signal $\mathbf{X} = \{x_{k,j}\}_{k=1,2,\ldots,p}^{j=1,2,\ldots,N}$ is mapped to c classes with integer indices from 1 to c.

b) Like mvDE$_{\text{II}}$, to consider both the spatial and time domains, multivariate embedded vectors $W_{\mathbf{m}}(\Lambda), 1 \leq \Lambda \leq N-(m-1)d$ are created based on the Taken's embedding theorem (see Equation (5.19)) [114]. For simplicity, we assume $d_k = d$ and $m_k = m$, leading to $W_{\mathbf{m}}(\Lambda)$ with length mp.

c) For every $W_{\mathbf{m}}(\Lambda)$, all combinations of the mp elements in $W_{\mathbf{m}}(\Lambda)$ taken *m* at a time, termed $\psi_q(\Lambda)$ $(q = 1, \ldots \binom{mp}{m})$, are created. The number of the combinations is equal to $\binom{mp}{m}$. Therefore, for all channels, we have $(N-(m-1)d)\binom{mp}{m}$ dispersion patterns.

d) For each $1 \leq q \leq \binom{mp}{m}$ and for each of c^m potential dispersion patterns $\pi_{v_0 \ldots v_{m-1}}$, relative frequency is obtained as follows:

$$Pr(\pi_{v_0 \ldots v_{m-1}}) = \frac{\#\{\Lambda \,|1 \leq \Lambda \leq N-(m-1)d, \psi_q(\Lambda) \text{ has type } \pi_{v_0 \ldots v_{m-1}}\}}{(N-(m-1)d)\binom{mp}{m}} \tag{5.24}$$

e) Finally, based on the Shannon's definition of entropy, the mvDE is calculated as follows:

$$\mathrm{mvDE}(\mathbf{X}, \mathrm{m}, \mathrm{c}, \mathrm{d}) = -\sum_{\pi=1}^{\mathrm{c^m}} \Pr(\pi_{\mathrm{v_0 \ldots v_{m-1}}}) \cdot \ln\left(\Pr(\pi_{\mathrm{v_0 \ldots v_{m-1}}})\right) \tag{5.25}$$

In fact, mvDE explores all combinations of patterns of length *m* within an *mp*-dimensional embedding vector. In the mvDE algorithm, at least $c^m + Np$ elements are stored. This number is noticeably smaller than those for $\mathrm{mvDE_I}$ to $\mathrm{mvDE_{III}}$, leading to more stable results for short signals with a large number of channels. As the number of patterns obtained by the mvDE method is $(N-(m-1)d)\binom{mp}{m}$, it is suggested $c^m < \left\lfloor \frac{L\binom{mp}{m}}{\tau_{max}} \right\rfloor$ to work with reliable statistics. It is worth noting that the key contribution of this Chapter is developing the mvDE method.

5.5.4 Parameters of the mvMDE Methods

In addition to the maximum scale factor τ_{max} described before, there are three other parameters for the mvMDE methods, including the embedding dimension vector **m**, the number of classes c, and the time delay vector **d**. It is better to set $d_k > 1$ for oversampled time series. However, some information with regard to the frequency of signals may be ignored for $d_k > 1$. Therefore, like previous studies about multivariate entropy methods [23,110], we set $d_k = 1$ for simplicity. We need $1 < c$ to keep away the trivial case of having only one dispersion pattern. For simplicity, we use $c = 5$ and $m_k = 2$ for all signals used in this Chapter, although the range $2 < c < 9$ leads to similar findings. For more information about c, m_k, and d_k, please refer to [18]. Overall, the characteristics and limitations of the mvSE, mvFE, and mvDE algorithms for a *p*-channel signal with length *N* are summarized in Table 5.1.

5.6 Performance Results and Discussion

In this Section, the ability of the proposed mvMFE and mvMDE methods, compared with mvMSE, is investigated by the use of several synthetic and real multi-channel signals. mvMSE and mvMFE are based on conditional entropy [23, 32, 110], whereas mvMDE is based on the Shannon entropy applied to dispersion patterns. This means that the methods work on different principles. However, the comparison of mvMDE with mvMSE and mvMFE is meaningful because the latter two are the most common multivariate entropy techniques and MDE has

Table 5.1: Ability to consider the spatial domain and characterization of short signals, minimum number of elements to be stored, and minimum number of samples needed for each of the mvSE, mvFE, and mvDE algorithms for a p-channel signal with length N.

Methods	Spatial domain	Short signals	Minimum number of elements stored	Minimum number of sample points
mvSE [111]	yes	undefined	$\binom{Np}{2} + Np(pm+1)$	$10^m < N$
mvFE [32]	yes	unreliable	$\binom{Np}{2} + Np(pm+1)$	$10^m < N$
mvPE [25] and mvWPE [213]	no	reliable	$m! + Np$	$m! < N$
mvDE_{I}	no	reliable	$c^m + Np$	$\frac{c^m}{p} < N$
mvDE_{II}	yes	unreliable	$c^{mp} + Np$	$c^{mp} < N$
mvDE_{III}	yes	unreliable	$c^{m+p-1} + Np$	$\frac{c^{m+p-1}}{p} < N$
mvDE	yes	reliable	$c^m + Np$	$\frac{c^m}{\binom{mp}{m}} < N$

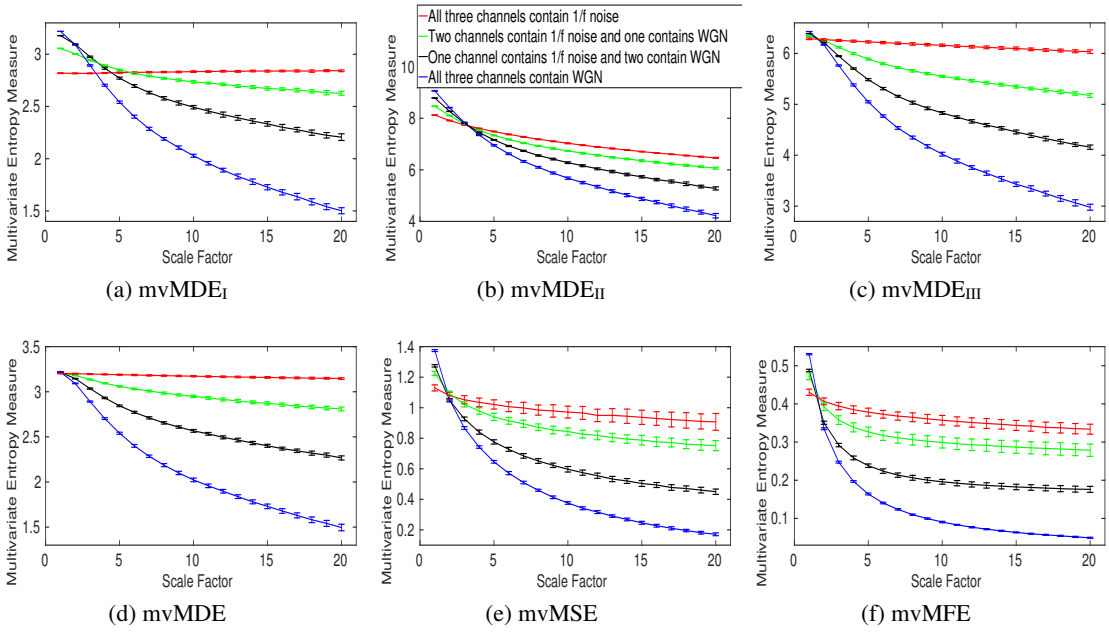
been shown to have similar behaviour to MSE when analysing real and synthetic signals [31].

5.6.1 Synthetic signals

5.6.1.1 Uncorrelated Multivariate Noise Signals

We first apply the proposed and existing methods to 40 independent realizations of uncorrelated trivariate WGN and $1/f$ noise, described in Section 5.1. The number of sample points for each of the $1/f$ noise and WGN signals were 15,000 sample points. The average and SD of the results for mvMDE_{I}, mvMDE_{II}, $\text{mvMDE}_{\text{III}}$, mvMDE, mvMSE, and mvMFE are depicted in Figure 5.4(a) to 5.4(f), respectively. Using all the existing and proposed methods, the entropy values of trivariate WGN signals are higher than those of the other trivariate time series at low scale factors. However, the entropy values for the coarse-grained trivariate $1/f$ noise signals stay almost constant or decrease slowly along the temporal scale factor, while the entropy values for the coarse-grained WGN signal monotonically decreases with the increase of scale factors. For WGN, no new structures are revealed at higher temporal scales. This demonstrates that a multivariate WGN time series has information only at small temporal scale factors. In contrast, for trivariate $1/f$ noise signals, the mean value of the fluctuations inside each signal does not converge to a constant value [21, 23, 32].

For all the methods, the higher the number of variates representing $1/f$ noise, the higher complexity the trivariate signal, in agreement with the fact that multivariate $1/f$ noise is structurally more complex than multivariate WGN [22, 23, 32].



(a) mvMDE$_{\text{I}}$ (b) mvMDE$_{\text{II}}$ (c) mvMDE$_{\text{III}}$

(d) mvMDE (e) mvMSE (f) mvMFE

Figure 5.4: Mean value and SD of the results using (a) mvMDE$_{I}$, (b) mvMDE$_{II}$, (c) mvMDE$_{III}$, (d) mvMDE, (e) mvMSE, and (f) mvMFE computed from 40 different uncorrelated trivariate WGN and $1/f$ noise time series with length 15,000 sample points.

To compare the results obtained by the mvMDE, mvMSE, and mvMFE methods, we used the coefficient of variation (CV). We investigate the results obtained by the uncorrelated noise signals at scale factor 10, as a trade-off between short and long scale factors. As can be seen in Table 5.2, the smallest CV values for the uncorrelated trivariate $1/f$ noise, uncorrelated combination of bivariate $1/f$ noise and univariate WGN, uncorrelated combination of bivariate WGN and univariate $1/f$ noise, and trivariate WGN are achieved by mvMDE, mvMDE$_{\text{II}}$, mvMDE$_{\text{II}}$, and mvMDE$_{\text{I}}$, respectively. Overall, the smallest CV values for trivariate $1/f$ noise and WGN profiles are reached by the mvMDE methods, showing the superiority of the mvMDE methods over mvMSE and mvMFE in terms of stability of results.

To assess the ability of the mvMDE methods to characterize short uncorrelated multi-channel signals in comparison with mvMFE and mvMSE, we use trivariate $1/f$ and WGN noise with length of 300 sample points. The results for the mvMDE, mvMSE, and mvMFE approaches at temporal scales 1 to 20 are depicted in Figure 5.5(a) to 5.5(f), respectively. As can be seen in Figure 5.5, the mvMDE$_{\text{I}}$ and mvMDE methods better discriminate different dynamics of the noise signals. However, the mvMSE values are undefined at higher scale factors. Although the mvMFE- and mvMDE$_{\text{II}}$-based values are defined at all scale factors, they cannot

Table 5.2: CV values of the proposed and existing multivariate multiscale entropy-based analyses at scale factor ten for the uncorrelated trivariate $1/f$ noise and WGN.

Time series	mvMDE_{I}	mvMDE_{II}	$\text{mvMDE}_{\text{III}}$
All three channels contain $1/f$ noise	0.0028	0.0025	0.0037
Two channels contain $1/f$ noise and one contains WGN	0.0042	0.0032	0.0036
One channel contains $1/f$ noise and two contain WGN	0.0066	0.0052	0.0058
All three channels contain WGN	0.0072	0.0080	0.0092

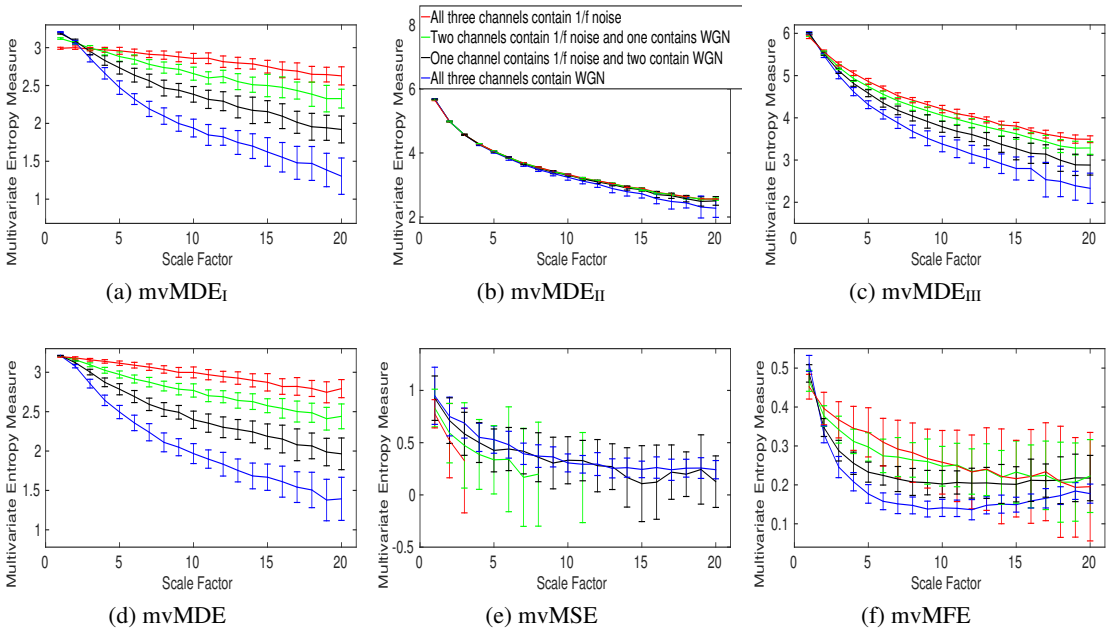
Time series	mvMDE	mvMSE	mvMFE
All three channels contain $1/f$ noise	0.0022	0.0405	0.0355
Two channels contain $1/f$ noise and one contains WGN	0.0044	0.0283	0.0274
One channel contains $1/f$ noise and two contain WGN	0.0061	0.0305	0.0292
All three channels contain WGN	0.0101	0.0232	0.0211

distinguish the dynamics of different noise signals. The profiles obtained by $\text{mvMDE}_{\text{III}}$ are more distinguishable than mvMDE_{II}, as mentioned that $\text{mvMDE}_{\text{III}}$ needs a smaller number of sample points. Nevertheless, the profiles obtained by $\text{mvMDE}_{\text{III}}$ have overlaps at all the scale factors. Overall, the results show the superiority of mvMDE_{I} and mvMDE over mvMDE_{II}, $\text{mvMDE}_{\text{III}}$, mvMSE, and mvMFE for short uncorrelated signals.

5.6.1.2 Correlated Multivariate Noise Signals

Univariate multiscale entropy approaches only consider every data channel separately and fail to take into account the cross-channel information of multivariate time series [23]. To assess the ability of the existing and proposed multivariate entropy methods to reveal the dynamics across the channels, we created 40 independent realizations of different combinations of bivariate $1/f$ noise and WGN time series with length 20,000 samples (according to [23,32]; see Section 5.1), making the channels correlated. Figure 5.6(a) to 5.6(f) respectively show the results obtained using the mvMDE_{I}, mvMDE_{II}, $\text{mvMDE}_{\text{III}}$, mvMDE, mvMSE, and mvMFE to model both the within- and cross-channel properties in multivariate signals.

As can be found from the algorithm of mvMDE_{I}, it cannot discriminate the correlated from uncorrelated WGN or $1/f$ noise. This fact is revealed in Figure 5.6(a). Therefore, mvMDE_{I} should only be used when the spatial components of a multi-channel time series are statistically independent. Multivariate multiscale entropy-based methods at scale factor 1 show the irregularity of multi-channel signals [23]. The mvMDE_{II}, $\text{mvMDE}_{\text{III}}$, and mvMDE values at scale 1 show that the uncorrelated WGN is the most irregular and unpredictable time series in agreement with [21], while the most irregular signal using mvMFE and mvMSE is



(a) mvMDE$_{\text{I}}$ (b) mvMDE$_{\text{II}}$ (c) mvMDE$_{\text{III}}$

(d) mvMDE (e) mvMSE (f) mvMFE

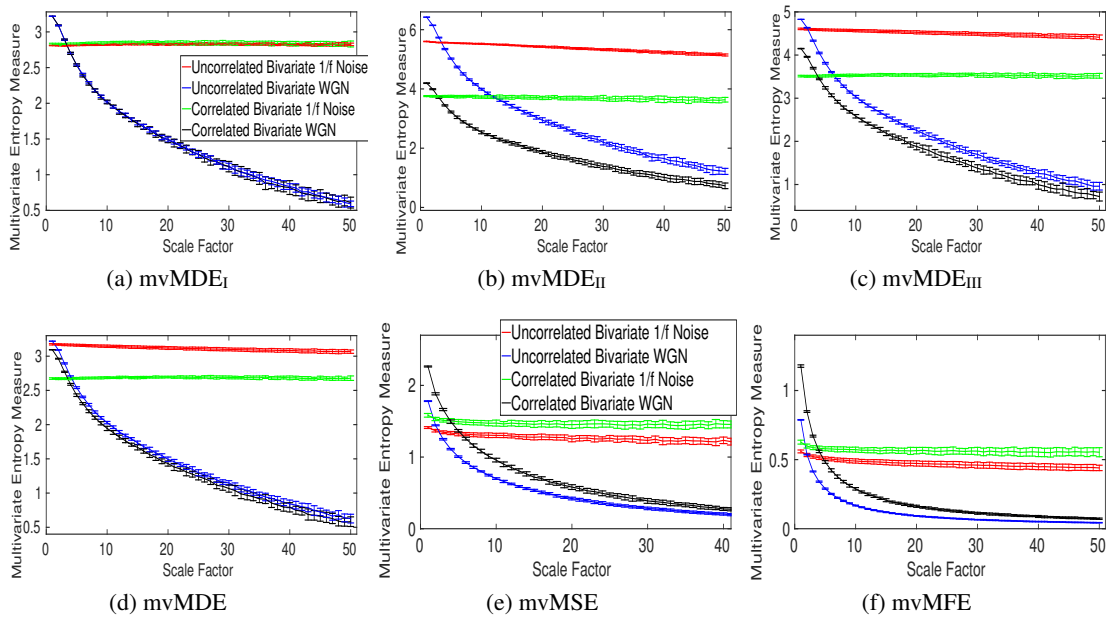
Figure 5.5: Mean value and SD of the results obtained by (a) mvMDE$_{I}$, (b) mvMDE$_{II}$, (c) mvMDE$_{III}$, (d) mvMDE, (e) mvMSE, and (f) mvMFE computed from 40 different uncorrelated trivariate WGN and $1/f$ noise time series with length 300 sample points.

the correlated WGN [23, 32], in contrast with the fact that correlated multi-channel WGN signals are more predictable and regular than uncorrelated WGN ones [21, 31].

The correlated bivariate $1/f$ noise is the most complex signal using the mvMDE$_{\text{II}}$, mvMDE$_{\text{III}}$, and mvMDE. The second most complex signal is the uncorrelated bivariate $1/f$ noise, as can be seen in Figure 5.6. The decreases of the uncorrelated bivariate WGN noise profiles using mvMDE$_{\text{II}}$, mvMDE$_{\text{III}}$, and mvMDE are the largest, evidencing the fact that the uncorrelated WGN is the least complex time series. These facts are also in agreement with the previous studies [22, 23, 32]. Therefore, as desired, the mvMDE$_{\text{II}}$, mvMDE$_{\text{III}}$, and mvMDE deal with both the cross- and within-channel correlations.

5.6.1.3 Bivariate Autoregressive Process

The ability of the mvMDE method to characterize multivariate AR processes is further evaluated using BAR(1), BAR(3), and BAR(5) with $\mathbf{A}_{\gamma_1} = \begin{bmatrix} 0.05 & 0.05 \\ 0.05 & 0.05 \end{bmatrix}$, $\mathbf{A}_{\gamma_2} = \begin{bmatrix} 0.10 & 0.10 \\ 0.10 & 0.10 \end{bmatrix}$, and $\mathbf{A}_{\gamma_3} = \begin{bmatrix} 0.15 & 0.15 \\ 0.15 & 0.15 \end{bmatrix}$. The results obtained by mvMDE are shown



(a) mvMDE$_{\text{I}}$ (b) mvMDE$_{\text{II}}$ (c) mvMDE$_{\text{III}}$

(d) mvMDE (e) mvMSE (f) mvMFE

Figure 5.6: Mean value and SD of the results obtained by (a) mvMDE$_{I}$, (b) mvMDE$_{II}$, (c) mvMDE$_{III}$, (d) mvMDE, (e) mvMSE, and (f) mvMFE computed from 40 different correlated and uncorrelated bivariate WGN and $1/f$ noise time series with length 20,000 sample points.

in Figure 5.7. As expected, when the lag order increases, the complexity of the corresponding time series using the mvMDE approaches increases, in agreement with the fact that a larger lag order denotes a more complex time series [23]. As the elements of $\mathbf{A}_{\gamma_1}$ are smaller than those of $\mathbf{A}_{\gamma_2}$ and $\mathbf{A}_{\gamma_3}$, the behaviour of the profiles obtained by the mvMDE method is more similar to the results for WGN (see Figure 5.4). In fact, the smaller the elements of $\mathbf{A}_{\gamma}$, the less complex the BAR, leading to lower entropy values at higher scale factors.

In order to investigate the dependence of the mvMDE method on the sensitivity to changes in

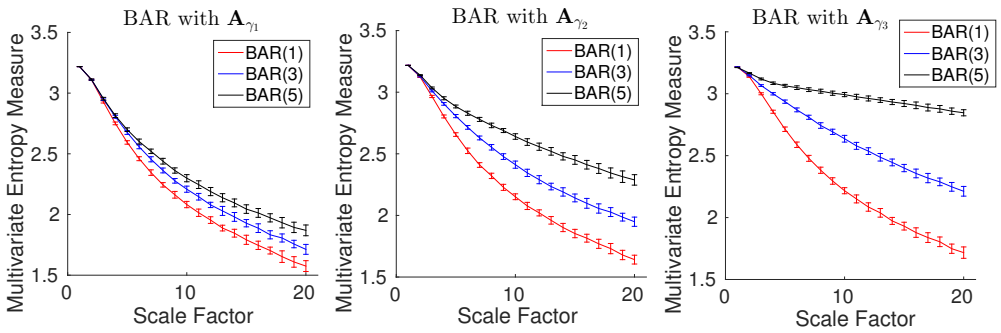


Figure 5.7: Mean and SD values of the results using mvMDE computed from 40 different BAR(1), BAR(3), and BAR(5) time series with $\mathbf{A}_{\gamma_1}$, $\mathbf{A}_{\gamma_2}$, and $\mathbf{A}_{\gamma_3}$.

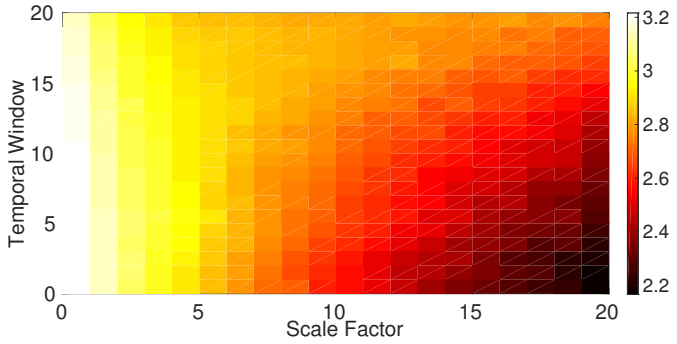
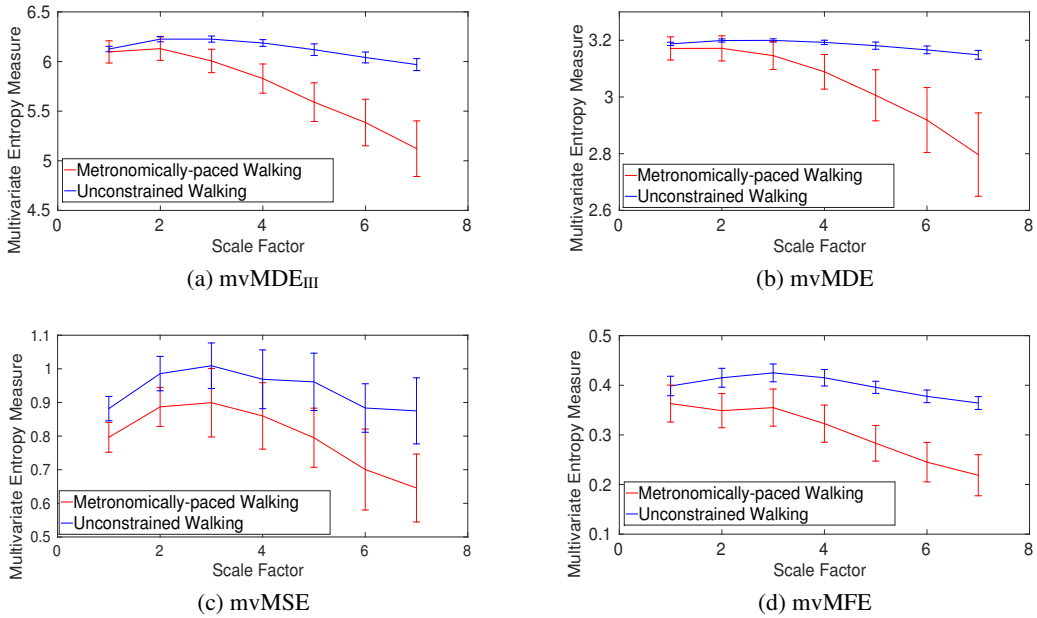


Figure 5.8: Results obtained by the mvMDE method using a bivariate window moving along the BAR(3) signal (temporal window), which the elements of anti-diagonal of the matrix $\boldsymbol{A}$ *linearly increase from 0 to 0.17, leading to more complex series.*

the signals, we generated BAR(3) with length of 10,000 sample points and sampling frequency of 150 Hz that $\mathbf{A}_\gamma$ linearly changes from $\begin{bmatrix} 0.17 & 0 \\ 0 & 0.17 \end{bmatrix}$ to $\begin{bmatrix} 0.17 & 0.17 \\ 0.17 & 0.17 \end{bmatrix}$. In fact, the elements of the diagonal of $\mathbf{A}$ are constant and those of anti-diagonal linearly increase from 0 to 0.17, leading to more complex series. We moved a bivariate window with length 2000 samples and 20% overlap along this BAR(3). The results, depicted in Figure 5.8 suggest that when the time window is occupied at the beginning of the BAR(3) ($\mathbf{A}_\gamma = \begin{bmatrix} 0.17 & 0 \\ 0 & 0.17 \end{bmatrix}$), the mvMDE values at higher scale factors are the smallest, showing the least complexity of BAR(3) in lower temporal windows, while their corresponding entropy values in the end of BAR(3) process ($\mathbf{A}_\gamma = \begin{bmatrix} 0.17 & 0.17 \\ 0.17 & 0.17 \end{bmatrix}$) are the largest.

5.6.2 Real biomedical datasets

In this Subsection, the existing and developed multivariate multiscale entropy approaches are used to detect various dynamics of multi-channel recordings of two physiological datasets. Of note is that we do not use the mvMDE_I for biomedical signals because it does not take into account the spatial domain of multi-channel recordings.



(a) mvMDE$_{\text{III}}$ (b) mvMDE

(c) mvMSE (d) mvMFE

Figure 5.9: Mean value and SD of the results using (a) mvMDE$_{III}$, (b) mvMDE, (c) mvMSE, and (d) mvMFE for the self-paced vs. metronomically-paced stride interval fluctuations.

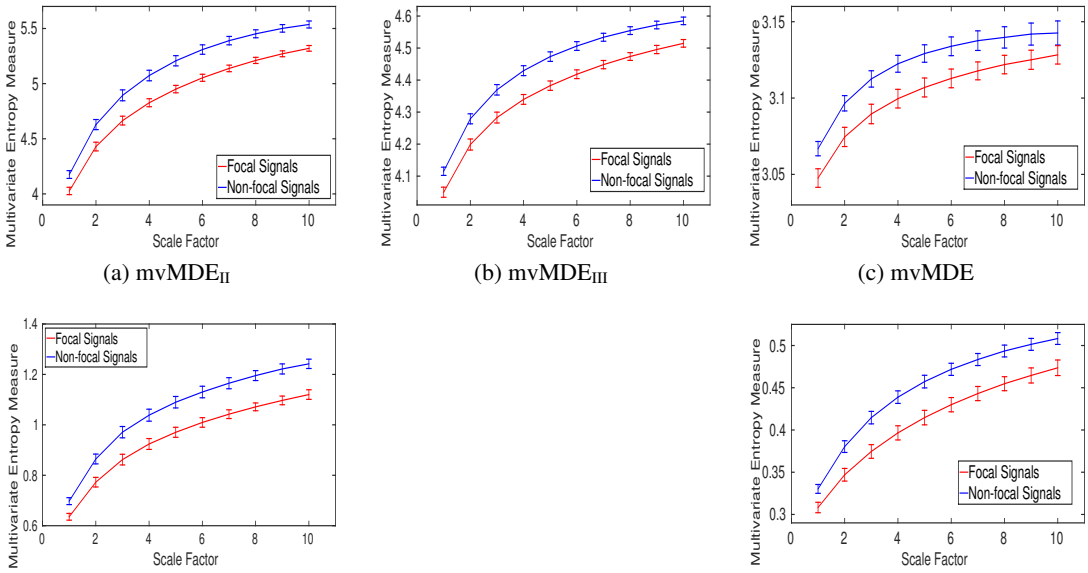
5.6.2.1 Dataset of Stride Internal Fluctuations

For the self-paced versus metronomically-paced stride interval fluctuations, the results obtained by the mvMDE$_{\text{III}}$, mvMDE, mvMSE, and mvMFE, respectively depicted in Figures 5.9(a), (b), (c), and (d), show that the self-paced unconstrained walk fluctuations have more complexity and greater long-range correlations than the metronomically-paced walk series, in agreement with those obtained by mvMSE, and multivariate empirical mode decomposition enhanced by mvSE [110]. We did not use mvMDE$_{\text{II}}$, as the signals do not follow the minimum number of samples required for mvMDE$_{\text{II}}$.

To compare the results, the CV values for both the metronomically- and self-paced walk at scale factor 4 are shown in Table 5.3. The CV values for our developed mvMFE are smaller than those for the mvMSE method. The CV values for the mvMDE$_{\text{III}}$- and mvMDE-based profiles are smaller than those for mvMFE, showing the superiority of the proposed methods over mvMFE and mvMSE in terms of the stability of results. The smallest CV values are achieved by mvMDE.

Table 5.3: CV values of the entropy results at scale factor 4 using mvMDE$_{III}$, mvMDE, mvSE, and mvMFE for the self-paced walk vs. metronomically-paced walk.

Signals	mvMSE	mvMFE	mvMDE$_{III}$	mvMDE
Self-paced walk	0.0901	0.040	0.005	0.002
Metronomically-paced walk	0.116	0.115	0.025	0.019



(a) mvMDE$_{II}$ (b) mvMDE$_{III}$ (c) mvMDE

(d) mvMSE (e) mvMFE

Figure 5.10: Mean value and SD of the results using (a) mvMDE$_{II}$, (b) mvMDE$_{III}$, (c) mvMDE, (d) mvMSE, and (e) mvMFE for the focal vs. non-focal time series.

5.6.2.2 Dataset of Focal and Non-focal Brain Activity

For the focal and non-focal EEG recordings, the results obtained by mvMDE$_{II}$, mvMDE$_{III}$, mvMDE, mvMSE, and mvMFE, depicted in Figure 5.10, show that the focal time series are less complex than the non-focal ones, in agreement with the previous studies [162, 163].

The CV values for the focal- and non-focal-based results at scale 6 are shown in Table 5.4. For non-focal EEGs, the CV values for mvMFE are slightly smaller than those for mvMSE. All the mvMDE-based CV values are smaller than those using mvMFE and mvMSE, showing more stability of the results obtained by the proposed mvMDE methods. Moreover, the CV values for mvMDE are smaller than those for mvMDE$_{III}$, and the latter ones are smaller than those for mvMDE$_{II}$, suggesting that the mvMDE leads to the most stable profiles.

Table 5.4: CV values of the entropy results at scale factor 6 using $mvMDE_{II}$, $mvMDE_{III}$, mvMDE, mvMSE, and mvMFE for focal vs. non-focal EEG recordings.

Signals	mvMSE	mvMFE	$mvMDE_{II}$	$mvMDE_{III}$	mvMDE
focal EEGs	0.019	0.019	0.006	0.003	0.002
Non-focal EEGs	0.021	0.015	0.008	0.003	0.002

5.7 Computational Time of Multivariate Multiscale Entropy Methods

To evaluate the computational time of mvMSE, mvMFE, $mvMDE_{I}$ to $mvDE_{III}$, and mvMDE, we use uncorrelated multivariate WGN time series with different lengths, changing from 100 to 10,000 sample points, and different number of channels, changing from 2 to 8. The results, depicted in Table 5.5, show that the computation times for mvMSE and mvMFE are close. The slowest algorithm is $mvMDE_{II}$, while the fastest ones are $mvMDE_{I}$ and mvMDE, in that order. For an 8-channel signal with 10,000 samples, using mvMSE, mvMFE, and $mvMDE_{II}$, the array exceeded the memory available. Overall, in terms of computation time and memory space, mvMDE outperforms all the existing and proposed methods taking into account both the time and spatial domains ($mvMDE_{I}$ does not consider the spatial domain). Note that the Matlab codes of mvMFE and mvMSE are available at [212].

5.8 Summary

To quantify the complexity of multivariate time series and to decrease the computation time of the existing mvMFE, we proposed mvMFE with the new fuzzy membership function. The results obtained by the existing and developed mvMFE were similar although the latter is at least 2 times faster in our considered implementations. The RCmvMFE and RCmvMSE approaches were also proposed as well. It was found that RCmvMSE and RCmvMFE, compared with respectively mvMSE and mvMFE, lead to more stable results when dealing with short or noisy data, whereas the results obtained by RCmvMFE (or RCmvMSE) and mvMFE (or mvMSE) are similar for long data. Nevertheless, the developed mvMFE and RCmvMFE are still not fast enough, especially for some real-time applications.

To decrease the running time of mvMFE and reliability of its results for short signals, we built

Table 5.5: Computational time of the mvMSE, mvMFE, and mvMDE algorithms with $\tau_{max} = 10$.

Number of channels and samples	mvMDE_{I}	mvMDE_{II}	$\text{mvMDE}_{\text{III}}$
2 channels and 1,000 samples	0.083 s	0.116 s	0.100 s
2 channels and 3,000 samples	0.240 s	0.3126 s	0.280 s
2 channels and 10,000 samples	0.736 s	1.010 s	0.919 s
5 channels and 1,000 samples	0.191 s	91.240 s	0.903 s
5 channels and 3,000 samples	0.568 s	169.275 s	2.209 s
5 channels and 10,000 samples	1.850 s	454.199 s	7.271 s
8 channels and 1,000 samples	0.298 s	out of memory	103.096 s
8 channels and 3,000 samples	0.820 s	out of memory	245.034 s
8 channels and 10,000 samples	2.687 s	out of memory	745.633 s

Number of channels and samples	mvMSE	mvMFE	mvMDE
2 channels and 1,000 samples	0.141 s	0.153 s	0.089 s
2 channels and 3,000 samples	0.598 s	0.723 s	0.265 s
2 channels and 10,000 samples	4.234 s	5.334 s	0.868 s
5 channels and 1,000 samples	0.544 s	0.636 s	0.229 s
5 channels and 3,000 samples	3.174 s	3.586 s	0.670 s
5 channels and 10,000 samples	28.229 s	31.242 s	2.312 s
8 channels and 1,000 samples	1.479 s	1.573 s	0.354 s
8 channels and 3,000 samples	9.421 s	9.972 s	1.028 s
8 channels and 10,000 samples	out of memory	out of memory	3.509 s

four diverse alternative implementations of mvMDE as further developments of our recently introduced MDE. These insights help towards a comprehensive understanding of four strategies to extend a univariate-based entropy method to its multivariate versions and therefore, provide invaluable information for future studies on multivariate time series. Although mvMDE was the best algorithm in terms of ability to discriminate dynamics of multivariate signals, computational time, and memory cost, the simpler alternatives (mvDE_{I} to mvDE_{III}) may still be useful in some settings.

We assessed their performance on the correlated and uncorrelated multivariate noise signals, the BAR time series, and two physiological datasets. The results showed the similar behavior of mvMSE-, mvMFE-, and mvMDE-based profiles. However, mvMDE had the following advantages over the existing methods: 1) it was noticeably faster than the existing methods; 2) mvMDE, in comparison with mvMSE and mvMFE, resulted in more stable profiles; 3) mvMDE better discriminated different kinds of biomedical signals; 4) for short multivariate time series, mvMDE, unlike mvMSE, did not result in undefined values; and 5) mvMDE, compared with mvMSE and mvMFE, needed to store a considerably smaller number of elements.

Overall, it is expected that the mvMDE approach plays a key role in the assessment of

complexity in multivariate physiological time series due to its great performance to distinguish different kinds of dynamics of multi-channel signals with low computation time.

Chapter 6

Illustration in Alzheimer's Disease

Discrimination of people with disease from healthy subjects on the basis of the analysis of their recorded time series is a long-lasting challenge in the physiological complexity literature [21,23,24,110,156]. Complexity techniques, such as multiscale entropy approaches, have been used to characterize electroencephalograms (EEGs) and magnetoencephalograms (MEGs) to distinguish Alzheimer's disease (AD) patients from controls [24, 25, 34, 109, 115, 156]. Nevertheless, there is room to develop new entropy-based algorithms to distinguish AD from healthy age-matched subjects with more significant differences. To this end, in this Chapter, the ability of existing and developed univariate and multivariate multiscale entropy methods to characterize two resting-state EEG and MEG datasets in AD is investigated. The univariate and multivariate approaches are also compared. Finally, we evaluate whether the results fulfil the hypotheses of the concept of complexity.

6.1 Resting-state Brain Activity Datasets

In this Section, two EEG and MEG datasets are briefly described.

6.1.1 Surface Electroencephalogram (EEG) Recordings

The 16-channel EEG dataset includes 11 AD patients (5 men; 6 women; age: 72.5 $\pm$ 8.3 years, all data given as mean $\pm$ SD) and 11 age-matched control subjects (7 men; 4 women; age: 72.8 $\pm$ 6.1 years) [140, 214]. To screen their cognitive status, a mini-mental state examination (MMSE) was done. The MMSE scores for AD patients and controls are 13.3 $\pm$ 5.6 and 30 $\pm$ 0, respectively. Table 6.1 shows the sociodemographic details of this dataset.

The subjects were recruited from the Alzheimer's Patients' Relatives Association of Valladolid

Table 6.1: AD patients' and controls' sociodemographic EEG data.

Control Subjects				Alzheimer's Disease Patients			
Identifier	Age (years)	Sex	MMSE	Identifier	Age (years)	Sex	MMSE
Con-1	72	Male	30/30	Alz-1	80	Female	7/30
Con-2	76	Male	30/30	Alz-2	69	Female	7/30
Con-3	70	Male	30/30	Alz-3	71	Female	7/30
Con-4	67	Female	30/30	Alz-4	74	Male	20/30
Con-5	76	Female	30/30	Alz-5	79	Female	10/30
Con-6	86	Male	30/30	Alz-6	72	Male	7/30
Con-7	79	Male	30/30	Alz-7	77	Male	14/30
Con-8	73	Male	30/30	Alz-8	79	Female	17/30
Con-9	69	Female	30/30	Alz-9	76	Male	23/30
Con-10	68	Male	30/30	Alz-10	71	Female	14/30
Con-11	65	Female	30/30	Alz-11	50	Male	18/30

(AFAVA), Spain [1]. The EEG signals were recorded with Oxford Instruments Profile Study Room 2.3.411 EEG equipment at the Hospital Clinico Universitario de Valladolid (Spain). The EEGs were recorded using the international 10-20 system, in an eyes closed and resting state. All 16 electrodes were referenced to the linked ear lobes of each individual. The signals were sampled at 256Hz and digitised with a 12-bit analog-to-digital converter.

Informed consent was obtained for all 22 subjects and the local ethics committee approved the study. Before band-pass filtering with cut-off frequencies 1 and 40 Hz and a Hamming window with order 200, the signals were visually examined by an expert physician to select 5 s epochs (1280 samples) with minimal artifacts for analysis. On average, 30.0 ± 12.5 epochs (mean$\pm$SD) were selected from each electrode and each subject. More details can be found in [140, 214].

6.1.2 Magnetoencephalogram (MEG) Recordings

Resting-state MEG time series were recorded with a 148-channel whole-head magnetometer (MAGNES 2500 WH, 4D Neuroimaging) in a magnetically shielded room at the Centre Dr. Perez-Modrego in Spain [2].

To screen the cognitive status, an MMSE was done. There were 36 AD patients (age = $74.06 \pm$

[1] We would like to thank Pedro Espino (Hospital Clinico San Carlos, Madrid, Spain) for his help in the recording and selection of EEG epochs. We also thank Daniel Abásolo for making the dataset available too.

[2] We would like to thank Alberto Fernández (Universidad Complutense de Madrid, Spain, and Universidad Politecnica de Madrid, Spain) for his help in the recording and selection of MEG epochs and making the dataset available to us.

6.95 years, all data given as mean $\pm$ SD, and MMSE score = 18.06 ± 3.36) and 26 controls (age = 71.77 ± 6.38 years, and MMSE score = 28.88 ± 1.18). The difference in age between two groups was not significant (p-value = 0.1911, Student's t-test) [215]. Table 6.2 demonstrates the sociodemographic details of this dataset.

The subjects lied on a hospital bed in a relaxed state with eyes closed. The subjects were asked to avoid moving head and eyes and sleeping. The distribution of MEG sensors is shown in Figure 6.1. For each participant, five minutes of MEG resting-state activity were recorded with sampling frequency (f_s) of 678.17 Hz. A hardware band-pass filter with cut-off frequencies at 0.1 and 200 Hz was then used. Afterwards, a notch filter at 50 Hz was employed to decrease the power supply interference. The recordings were next down-sampled by a factor of four. Therefore, the sampling frequency of the MEG signals are 169.5 Hz.

The signals were divided into 10 s segments (1695 samples) and visually inspected using an automated thresholding procedure to discard epochs noticeably contaminated with artifacts. All recordings were digitally band-pass filtered with a Hamming window FIR filter of order 200 and cut-off frequencies at 1 Hz and 40 Hz. The effect of cardiac artifact was reduced from the signals using a constrained blind source separation procedure [216]. For more information, please see [215]. Of note is that all 62 participants agreed to take part in the research, which was approved by the local ethics committee.

6.2 Application to Brain Activity in AD

In this Section, we compare the results obtained by the univariate and multivariate MSE, MFE, and MDE approaches to discriminate the controls from AD patients using the above-mentioned EEG and MEG recordings. Of note is that, as shown in Chapter 4 and the previous studies [32, 206, 211], the refined composite multiscale-based techniques do not improve the stability of results for real recordings with these characteristics, while these methods increase the computation time noticeably. Thus, the refined composite process is not considered in this Chapter.

Table 6.2: AD patients' and controls' sociodemographic MEG data.

Control Subjects				Alzheimer's Disease Patients			
Identifier	Age (years)	Sex	MMSE	Identifier	Age (years)	Sex	MMSE
Con-1	68	Female	30/30	Alz-1	71	Female	15/30
Con-2	61	Female	29/30	Alz-2	67	Male	12/30
Con-3	70	Female	30/30	Alz-3	56	Female	14/30
Con-4	64	Female	30/30	Alz-4	64	Female	15/30
Con-5	60	Male	30/30	Alz-5	59	Female	20/30
Con-6	63	Female	30/30	Alz-6	60	Male	16/30
Con-7	73	Male	29/30	Alz-7	72	Female	15/30
Con-8	69	Female	29/30	Alz-8	71	Female	15/30
Con-9	79	Male	29/30	Alz-9	75	Female	22/30
Con-10	79	Male	30/30	Alz-10	82	Female	21/30
Con-11	75	Female	29/30	Alz-11	72	Female	17/30
Con-12	67	Male	29/30	Alz-12	80	Male	24/30
Con-13	68	Female	29/30	Alz-13	83	Male	10/30
Con-14	84	Male	29/30	Alz-14	77	Female	21/30
Con-15	68	Female	27/30	Alz-15	82	Male	19/30
Con-16	73	Male	30/30	Alz-16	83	Female	20/30
Con-17	71	Female	29/30	Alz-17	73	Female	23/30
Con-18	74	Male	30/30	Alz-18	79	Male	19/30
Con-19	78	Male	27/30	Alz-19	83	Male	16/30
Con-20	76	Female	29/30	Alz-20	72	Female	23/30
Con-21	83	Female	26/30	Alz-21	69	Female	16/30
Con-22	68	Female	28/30	Alz-22	77	Male	21/30
Con-23	68	Female	30/30	Alz-23	74	Female	16/30
Con-24	72	Female	27/30	Alz-24	81	Female	21/30
Con-25	77	Female	29/30	Alz-25	81	Female	17/30
Con-26	78	Female	27/30	Alz-26	78	Female	15/30
			30/30	Alz-27	68	Male	21/30
			30/30	Alz-28	78	Female	15/30
			30/30	Alz-29	72	Female	22/30
			30/30	Alz-30	79	Male	15/30
			30/30	Alz-31	78	Female	18/30
			30/30	Alz-32	71	Male	20/30
			30/30	Alz-33	78	Male	18/30
			30/30	Alz-34	75	Female	16/30
			30/30	Alz-35	78	Female	21/30
			30/30	Alz-36	68	Female	21/30

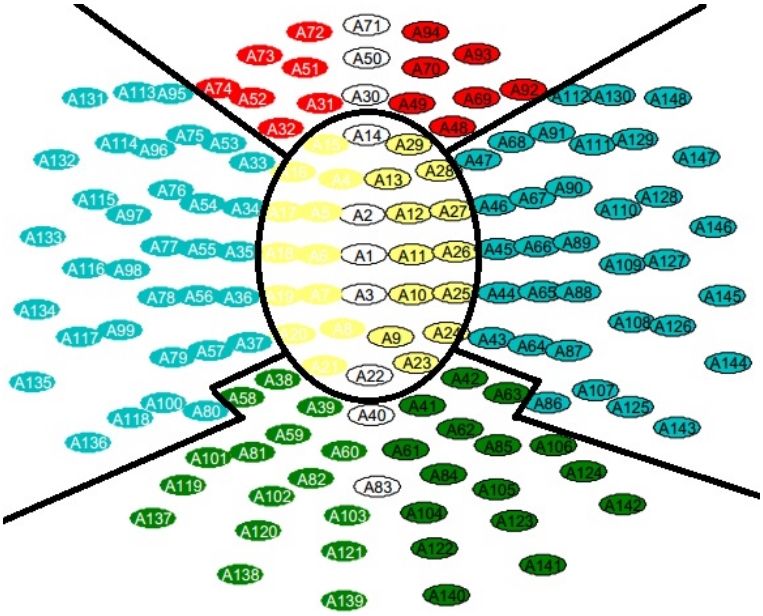


Figure 6.1: Distribution of the MEG electrodes into five regions: anterior (red), central (yellow), left (blue with white text) and right lateral (blue with black text), and posterior (green).

Table 6.3: Parameters values of MSE, MFE, MDE, mvMSE, mvMFE, and mvMDE (embedding dimension m, threshold r, number of classes c, and time delay d) for the resting-state EEG and MEG recordings.

m (all the methods)	r (MSE, MFE, mvMSE, and mvMFE)	c (MDE and mvMDE)	d (all the methods)
2	0.15×SD of a signal	6 (MDE) and 5 (mvMDE)	1

6.2.1 Multiscale Entropy-based Methods

6.2.1.1 Parameters of Multiscale Entropy-based Methods

The embedding dimension m is equal to 2 for all the univariate multiscale entropy techniques. The other values for the parameters of multiscale (sample) entropy (MSE), multiscale fuzzy entropy (MFE), and multiscale dispersion entropy (MDE) respectively follow [21], [30], and [31]. The parameters values of all the univariate multiscale entropy methods are shown in Table 6.3.

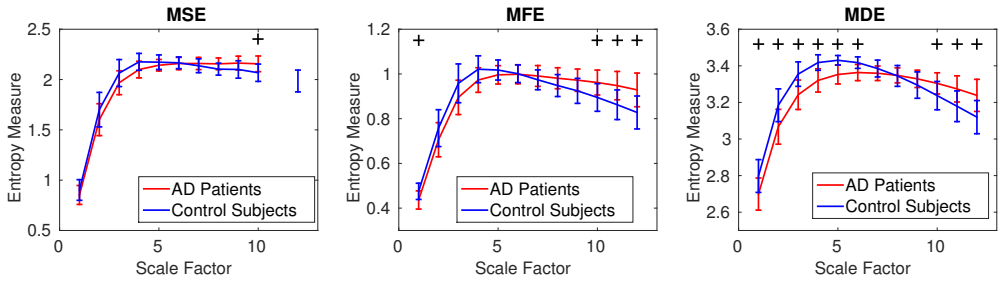


*Figure 6.2: Mean value and SD of the results obtained by (a) MSE, (b) MFE, and (c) MDE computed from 11 AD patients' EEGs versus 11 elderly age-matched controls' EEGs. Red and blue indicate AD patients and controls, respectively. The MSE values are undefined for AD patients' and controls' signals at scales 11 or 12. The scale factors with p-values between 0.001 and 0.05, and smaller than 0.001 are respectively shown with + and *.*

6.2.1.2 Surface Electroencephalogram (EEG) Recordings

MSE, MFE, and MDE are used to characterize the EEGs recorded from 11 patients with AD and 11 age-matched control subjects. The results are shown in Figure 6.2. The average of MSE, MFE, and MFE values for AD patients is smaller than those for controls at short-time scale factors, while the AD subjects' signals have larger entropy values at long-time scale factors. Herein, short-time (or low) scale factors mean the temporal scales that are smaller than or equal to the scale of crossing point of the curves for AD patients vs. controls. Long-time (or high) scale factors denote the temporal scales that are larger than the scale of crossing point of the curves for AD patients vs. controls. For example, short-time and long-time scale factors are 1-6 and 7-12, respectively, for MFE in Figure 6.2. All the results are in agreement with [24, 109, 156]. Of note is that the average of the entropy values for all the channels is reported for the univariate multiscale entropy methods herein.

The non-parametric Mann-Whitney U-test was employed to assess the differences between results for AD patients versus controls, as the MDE, MSE, and MFE values at each scale factor did not follow a normal distribution. The scales with the p-values between 0.001 and 0.05 (significant), and less than 0.001 (very significant) are shown with + and *, respectively, in this Chapter. As can be seen in Figure 6.2, the results show that MSE, MFE, and MDE respectively achieve to 1, 4, and 9 scale factors with significant differences. Furthermore, the MSE values, unlike the MDE and MFE ones, at scales 11 and 12 are undefined (see Figure 6.2), confirming the advantage of MDE and MFE over MSE for short signals.

Table 6.4: Smallest p-values obtained by MSE, MFE, and MDE for the resting-state EEG and MEG recordings (Mann-Whitney U-test).

Recordings	MSE	MFE	MDE
EEG	0.0356	0.0181	0.0010
MEG	0.00013	0.00013	0.00007

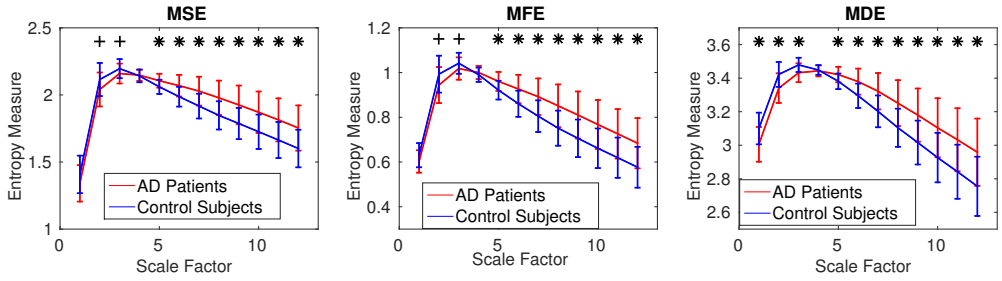


*Figure 6.3: Mean value and SD of the results obtained by (a) MSE, (b) MFE, and (c) MDE computed from 26 AD patients' MEGs versus 36 elderly age-matched controls' MEGs. Red and blue indicate AD patients and controls, respectively. The scale factors with p-values between 0.001 and 0.05, and smaller than 0.001 are respectively shown with + and *.*

The smallest p-values obtained by MSE, MFE, and MDE are represented in Table 6.4. Although the p-values for long scale factors are smaller than those for short scales, there is no optimal scale that always has the smallest p-values. This shows that the smallest p-value is achieved by MDE. On the whole, one finds the superiority of MDE over MSE and MFE for the discrimination of EEG background activity related to AD.

6.2.1.3 Magnetoencephalogram (MEG) Recordings

MSE, MFE, and MDE are also used to characterize the MEG recordings. The results, depicted in Figure 6.3, are consistent with Figure 6.2 and [24, 109, 156]. MSE, MFE, and MDE lead to the very significant differences at 8, 8, and 11 scale factors. Moreover, the smallest p-values, illustrated in Table 6.4, show that MDE is the best technique to discriminate the 26 AD patients' from 36 elderly age-matched controls' MEGs. Overall, the results for both datasets evidence the superiority of our introduced MDE over MSE and MFE to distinguish the controls from AD patients.

Table 6.5: Smallest p-values obtained by mvMSE, mvMFE, and mvMDE for the resting-state EEG recordings (Mann-Whitney U-test).

mvMSE	mvMFE	mvMDE
0.0071	0.2372	0.0086

6.2.2 Multivariate Multiscale Entropy-based Methods

As the resting-state EEGs and MEGs are multi-channel signals, in this Subsection, multivariate MSE (mvMSE), multivariate MFE (mvMFE), and multivariate MDE (mvMDE) are used to take into account both the time and spatial domains simultaneously.

6.2.2.1 Parameters of Multivariate Multiscale Entropy-based Methods

The embedding dimension m is equal to 2 for all the multivariate multiscale entropy approaches. The other values for the parameters of mvMSE, mvMFE, and mvMDE respectively follow [111], [32], and [33]. The parameters values of the techniques are shown in Table 6.3.

6.2.2.2 Surface Electroencephalogram (EEG) Recordings

The mvMSE, mvMFE, and mvMDE values for the 11 AD patients' and 11 controls' EEGs are depicted in Figure 6.4. For all the mvMSE and mvMFE methods, the controls' signals have more irregularity at short-time scales than the AD patients' recordings, whereas the latter ones are more irregular at long-time scales. The findings are in agreement with [34, 109, 144, 155, 156, 217].

The Mann-Whitney U-test is used to assess the differences between the results for AD patients versus controls. The smallest p-values are also illustrated in Table 6.5. The results show that mvMFE cannot detect the AD patients from controls with a significant difference. The p-values also suggest that a similar performance of mvMSE and mvMDE to distinguish AD patients from healthy subjects.

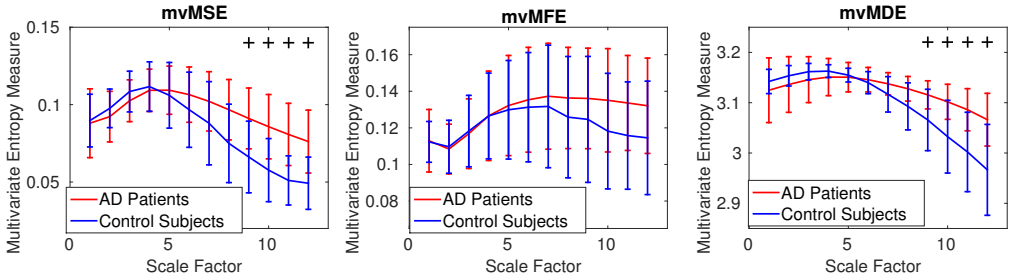


*Figure 6.4: Mean value and SD of the results obtained by (a) mvMSE, (b) mvMFE, and (c) mvMDE computed from 11 AD patients' EEGs versus 11 elderly age-matched controls' EEGs. Red and blue indicate AD patients and controls, respectively. The scale factors with p-values between 0.001 and 0.05, and smaller than 0.001 are respectively shown with + and *.*

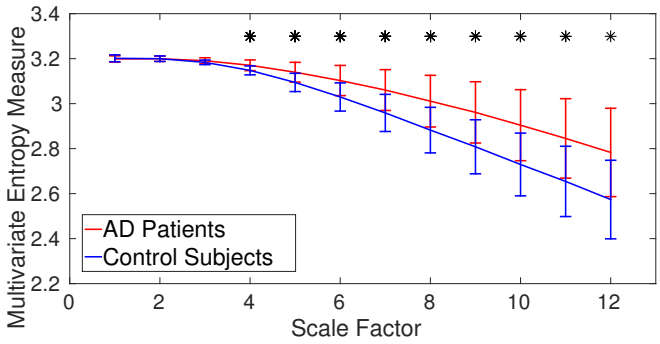


*Figure 6.5: Mean value and SD of the results obtained by mvMDE computed from 36 AD patients versus 26 elderly controls for all the 148 channels. Red and blue respectively indicate AD patients and controls. The scale factors with p-values between 0.001 and 0.05, and smaller than 0.001 are respectively shown with + and *.*

6.2.2.3 Magnetoencephalogram (MEG) Recordings

To assess the ability of mvMDE, in comparison with mvMFE and mvMSE, we applied the methods to the 148-channel MEG signals to discriminate AD patients from controls. Because mvMSE and mvMFE need to store a huge number of elements for a signal with a large number of channels, mvMSE and mvMFE were not able to simultaneously analyse all 148 time series (see Chapter 5 or [33]). However, the results using mvMDE are depicted in Figure 6.5. It represents an advantage of mvMDE over mvMSE and mvMFE for signals with a large number of channels. The profiles follow the aforementioned ones as well as previous studies [34, 109, 144, 155, 156]. The smallest p-value for the discrimination of AD subjects from controls is 0.0001.

To compare the mvMSE, mvMFE, and mvMDE techniques, we applied the methods to five

Table 6.6: Smallest p-values obtained by mvMSE, mvMFE, and mvMDE for the resting-state MEG recordings over five scalp regions (Mann-Whitney U-test).

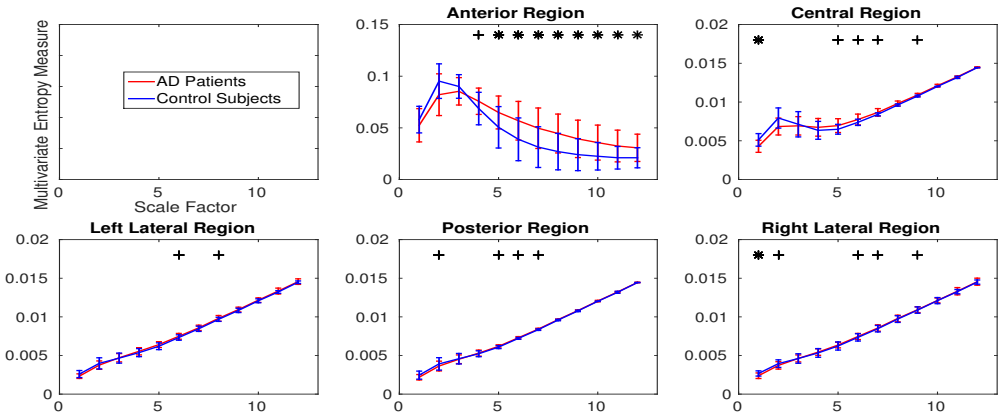
Region →	anterior	central	left lateral	posterior	right lateral
mvMSE	0.0001	0.0001	0.0040	0.0021	0.0001
mvMFE	0.0003	0.0022	0.0001	0.0009	0.0001
mvMDE	0.0001	0.0001	<0.0001	0.0001	<0.0001

main scalp regions (see Figure 6.1), namely, anterior (17 channels), right (34 channels) and left lateral (34 channels), central (29 channels), and posterior (34 channels) areas, leading to the smaller numbers of channels to noticeably decrease the number of elements stored by the use of the mvMFE and mvMSE algorithms.

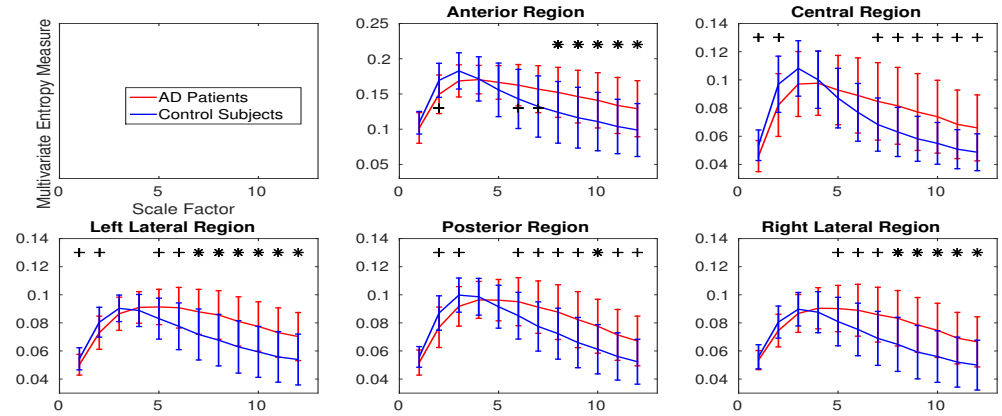
The average and SD of mvMSE, mvMFE, and mvMDE values for five regions are respectively shown in Figures 6.6(a), 6.6(b), and 6.6(c). As can be seen in Figures 6.5 and 6.6, the average mvMDE and mvMFE values for AD patients are smaller than those for controls at lower scale factors (short-time scale factors), while at higher scales, the AD subjects' recordings have larger entropy values (long-time scale factors), in agreement with [10, 34, 156].

The larger the number of channels, the smaller the mvMSE values (see Appendix B). Thus, when dealing with a multivariate signal with a large number of channels, the mvMSE values are close to 0. This might lead to less reliable results and cause the results for mvMSE are not reliable for central, left and right lateral, and posterior regions. It is worth noting that the mvMFE behaves similar to mvMSE when the number of channels rises, although mvMFE values do not decrease as much as mvMSE ones (please see Appendix B). Therefore, the mvMFE-based results are still reliable for the five different regions (see Figure 6.6(b)).

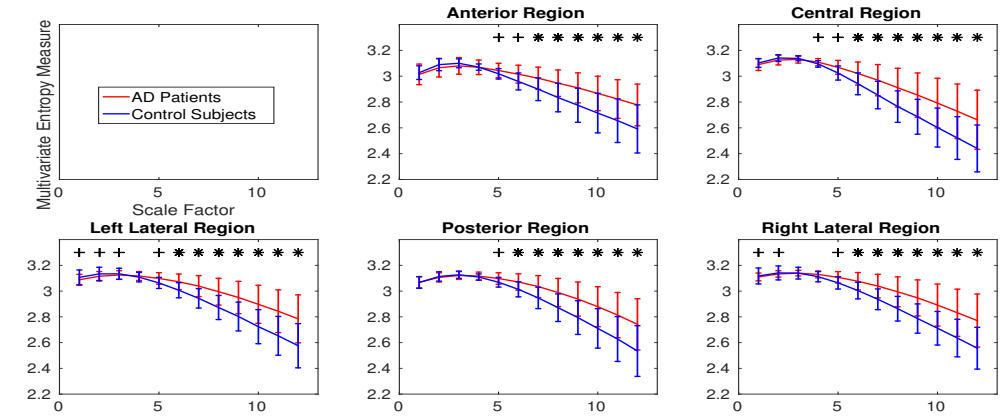
The Mann-Whitney U-test was used to assess the differences between the mvMSE, mvMDE and mvMFE profiles at each temporal scale for AD patients versus controls. The p-values show that the mvMDE, compared with the mvMFE, very significantly discriminated the controls from subjects with AD at a larger number of scale factors. Moreover, the smallest p-value was achieved by mvMDE (see Table 6.6) for each region, evidencing the superiority of mvMDE over mvMFE and mvMSE.



(a) mvMSE



(b) mvMFE



(c) mvMDE

*Figure 6.6: Mean value and SD of the results obtained by (a) mvMSE, (b) mvMFE, and (c) mvMDE computed from 36 AD patients' MEGs versus 26 elderly age-matched controls' MEGs over five scalp regions. Red and blue indicate AD patients and controls, respectively. The scale factors with p-values between 0.001 and 0.05, and smaller than 0.001 are respectively shown with + and *.*

6.2.3 Comparison between Univariate and Multivariate Multiscale Entropy Approaches

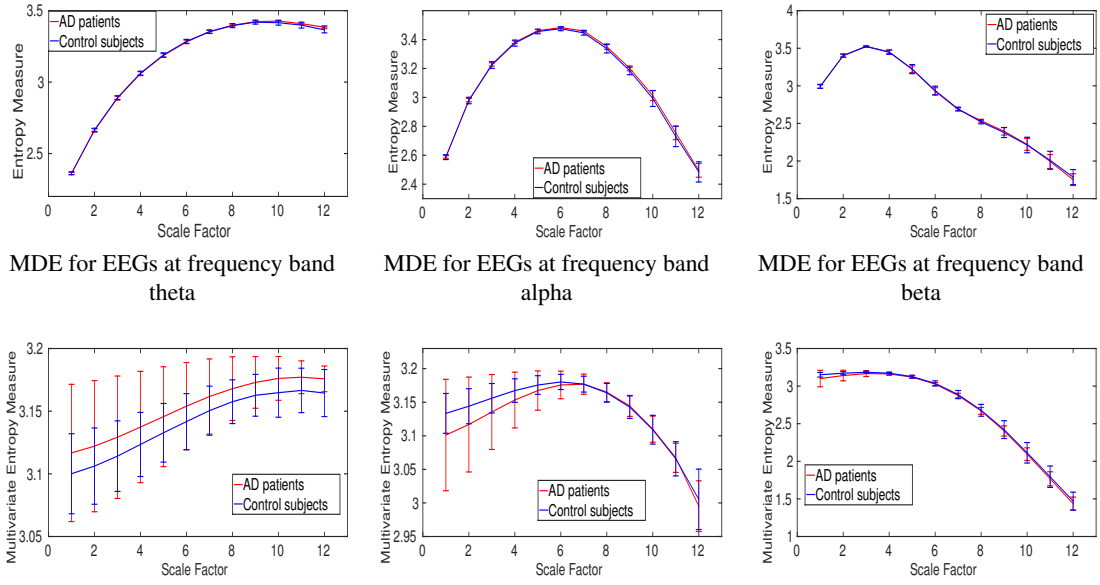
The multivariate entropy methods reveal the dynamics across the channels and consider the information in both the time and spatial domains, while the univariate entropy approaches only consider the time domain. Nevertheless, since the average over channels is reported for the MSE, MFE, and MDE methods, the profiles may become more stable (smaller coefficient of variation values). This may lead to lower *p*-values (see Tables 6.5 and 6.4).

6.2.4 Effect of EEG and MEG Frequency Bands on Univariate and Multivariate Multiscale Entropy Approaches

A potential strategy to increase the probability of an accurate AD diagnosis is to investigate specific frequency bands in EEGs (or MEGs), such as delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-40 Hz) [6]. AD affects these frequency bands in various ways: the power in delta, theta, and gamma bands increases, while the power in alpha and beta decreases in AD patients, compared with healthy age-matched controls [6, 119, 120, 124].

To this end, we investigated the changes in MDE and mvMDE for different frequency bands theta, alpha, and beta of the 11 AD patients' versus 11 controls' EEGs to understand the effect of AD and univariate and multivariate entropy approaches on each frequency range [34]. The results are shown in Figure 6.7. Note that delta and gamma, respectively, have too low and high frequency to be considered here based on the fact that these methods at scale factor τ can be considered as a low-pass filter with cut-off frequency $\frac{f_s}{2\tau}$.

The results do not show that controls' signals are less complex than AD patients' ones. This fact illustrates that complexity changes are best highlighted considering broadband activity. For more information, please refer to [34]. The results for the MEG dataset also suggested that the presence of broadband activity in MEGs is required for a comprehensive evaluation of complexity with univariate and multivariate multiscale entropy techniques. As the results for both the MEGs do not increase further knowledge to the Thesis, they are not reported.



MDE for EEGs at frequency band theta | MDE for EEGs at frequency band alpha | MDE for EEGs at frequency band beta

mvMDE for EEGs at frequency band theta | mvMDE for EEGs at frequency band alpha | mvMDE for EEGs at frequency band beta

Figure 6.7: Mean and SD of results obtained by the MDE and mvMDE methods for 11 AD patients' vs. 11 controls' EEGs at frequency bands theta (4-8 Hz), alpha (8-13 Hz), and beta (13-30 Hz).

6.3 Discussion about Multiscale-based Results and Hypothesis of Complexity

For both the MEG and EEG datasets, the results obtained by the univariate and multivariate multiscale entropy approaches show smaller entropy values at lower scale factors and larger entropies at higher temporal scales for AD patients, compared with healthy subjects. These profiles are consistent with the previous studies [34, 109, 144, 155, 156, 217]. The univariate and multivariate entropy values for AD patients are higher than those for controls at most of the scale factors. Therefore, AD subjects' recordings are not always less complex than controls' ones. This fact is in contradiction with the complexity hypothesis: age-matched healthy subjects' signals are always more complex than diseased individuals' (here AD patients') time series [21, 111]. Therefore, the hypothesis of complexity for the applications of AD may need to be revised based on the scale factors: AD patients' brain activity signals, compared with those for healthy controls, are less and more complex at short and long temporal scale factors, respectively.

As an alternative to consider complexity values at every scale factor, we can use the slope values as features of the complexity profiles to discriminate the controls' from AD subjects' signals [31, 144]. As can be seen in Figures 6.2, 6.3, 6.6, and 6.4, the curves increase until the scale factor of 4 or 5. Then, the slope decreases and the entropy values are nearly constant or decrease slightly. Therefore, we can divide each of the curves into two segments: (I) the first part corresponds to the steep increasing slope (low scale factors), and (II) the second one contains the scale factors in which the slope of the entropy values is smoother (high scale factors) [31, 144].

6.4 Summary

In this Chapter, we investigated the ability of our developed univariate and multivariate multiscale entropy methods, compared with the existing ones, to characterize AD using two EEG and MEG datasets.

For univariate multiscale entropy techniques, the results illustrated that MDE and MFE, unlike MSE, did not lead to undefined values. Moreover, based on the p-values, the differences between the MDE-based results for AD palatines versus controls were more significant than their corresponding MSE- and MFE-based results for both the EEG and MEG datasets. The number of scale factors with significant and very significant differences for MDE was larger than those for MSE and MFE for the EEG and MEG datasets, respectively. These advantages, besides its low computation time (see Chapter 4), make MDE to play a key role in complexity analyses for the characterisation of EEG and MEG signals in AD.

In the context of multivariate multiscale entropy approaches, for the EEG dataset, mvMDE and mvMSE outperformed mvMFE in terms of discrimination of AD patients' and controls' signals. For the MEGs, since mvMSE and mvMFE, unlike mvMDE, require to store a huge number of elements for a signal with a large number of channels, mvMSE and mvMFE were not able to analyse all 148 signals at the same time. Thus, we divided 148 channels into five main scalp regions to have a smaller number of channels. The results suggested that mvMDE led to the smallest p-values and that the number of temporal scales with very significant differences for mvMDE was larger than those for mvMSE and mvMFE for each region. These advantages and low computational time of mvMDE (see Chapter 5) evidence the superiority of mvMDE

over mvMSE and mvMFE.

The results obtained by the multiscale-based techniques showed that EEGs and MEGs recorded from age-matched controls are less complex than those for AD patients. Although the complexity profiles were in agreement with the previous studies, they were in contradiction with the hypothesis of complexity (healthy subjects' recordings are more complex than diseased ones). Thus, it may be needed to revise the hypothesis of complexity for the application of AD using EEGs and MEGs: healthy controls' brain activity signals, compared with those for AD patients, are more and less complex at low and high temporal scale factors, respectively.

Overall, the results support the relevance of multivariate and univariate multiscale analyses and superiority of MDE and mvMDE over their corresponding complexity approaches for the characterisation of EEG and MEG signals in AD.

Chapter 7

Summary, Conclusions, Limitations, and Future Research

7.1 Summary and Conclusions

Alzheimer's disease (AD) is a major neurodegenerative disease [2]. It is the most common form of dementia [2]. AD affects the interaction between neurons in the brain. Thus, it changes brain activity [4, 6]. Some of these changes may be recorded by electrophysiological techniques, including electroencephalograms (EEGs) and magnetoencephalograms (MEGs) [6, 8, 10].

Since EEGs and MEGs are considered as outputs of a nonlinear system, there has been an interest in nonlinear methods for the analysis of such signals [11–13]. Entropy is a powerful nonlinear metric to evaluate the irregularity of time series [16, 17, 103]. The purpose of this Thesis is to develop entropy-based metrics for characterization of physiological signals paying close attention to EEGs and MEGs in AD.

The simplest approach to alleviate noise in EEG and MEG recordings is epoch rejection, which divides a raw EEG or MEG signal into some epochs and discards those epochs highly contaminated by noise [218]. Thus, the time series are considerably shortened. Furthermore, many physiological recordings, such as EEGs and MEGs, are usually multi-channel [23, 110, 111]. Accordingly, recent trends in the field of entropy for the characterization of such time series have tried: 1) to enhance the stability and reliability of entropy-based results, especially for short signals; and 2) to extend the univariate entropy approaches to their multivariate cases to be able to reveal the patterns across channels.

Sample entropy (SampEn) [16] and its univariate and multivariate multiscale extensions, respectively named MSE [39] and mvMSE [23], are the most popular univariate and multivariate entropy approaches. However, these methods have two shortcomings: 1) they lead to undefined values for short signals; and 2) they are not fast enough for real-time applications [32].

To address the problem of undefined SampEn, MSE, and mvMSE values, respectively, fuzzy entropy (FuzEn) [103], multiscale FuzEn (MFE) [219], and multivariate MFE (mvMFE) [205] were proposed. However, the mvMFE algorithm was even slower than mvMSE. To this end, we developed the mvMFE with a new fuzzy membership function that decreases the computation time while keeping its benefits [32]. To increase the stability of MFE and mvMFE-based results for respectively univariate and multivariate signals, the refined composite MFE (RCMFE) and refined composite mvMFE (RCmvMFE) approaches were proposed in this Thesis as well [32]. However, the FuzEn-based methods are still slow for real time-applications [33].

Permutation entropy (PerEn) [17] and its univariate and multivariate multiscale extensions, respectively denoted as MPE [25] and mvMPE [25], are able to characterize short signals and are fast enough for real-time applications. However, these techniques suffer from the following deficiencies [31, 33, 167]:

1. They ignore the emergence of equal values.
2. When a time series is symbolized based on the permutation patterns, only the order of amplitude values is considered and some information with regard to the amplitudes may be discarded.
3. They are considerably sensitive to noise, since a small change in amplitude value may vary the order relations among amplitudes.
4. MPE and mvMPE do not follow the hypotheses of complexity.

Although our improved MPE [169] increased the stability of MPE-based results, and our developed amplitude-aware PerEn (AAPerEn) [28] addressed the first and second shortcomings, the third and forth ones are still present when dealing with these techniques. In sum, a whole new approach was needed.

To address the deficiencies of SampEn, FuzEn, and AAPerEn at the same time, dispersion

entropy (DispEn) [18,29] and frequency-based DispEn (FDispEn) [29] based on our introduced dispersion patterns and the Shannon's definition of entropy were developed in this Thesis. DispEn and FDispEn outperformed SampEn, PerEn, and FDispEn to discriminate various states of physiological signals. FDispEn, as a new frequency-based approach, had a better performance than PerEn to distinguish different kinds of dynamics of synthetic and real signals. Nevertheless, FDispEn, like PerEn, did not detect simultaneous change in amplitude and frequency. In terms of computation time, FDispEn was the fastest method for long time series, followed by DispEn, PerEn, SampEn, and FuzEn, in that order. It is in agreement with the fact that the computation costs of FDispEn, DispEn, PerEn, SampEn, and FuzEn are respectively O(N), O(N), O(N), O(N^2), and O(N^2) [18, 194, 199].

We also developed multiscale DispEn (MDE) [31] and multivariate MDE (mvMDE) [33] on the basis of DispEn to quantify the complexity of univariate and multivariate signals, respectively. MDE and mvMDE had the following advantages over the state-of-the-art univariate and multivariate multiscale methods [31, 33]:

1. They were noticeably faster.
2. They resulted in more stable profiles for synthetic and real signals.
3. They better discriminated different kinds of physiological time series.
4. They did not result in undefined values.
5. mvMDE, compared with mvMSE and mvMFE, required to store a considerably smaller number of elements.

To understand the ability of existing and developed univariate and multivariate multiscale techniques to characterize EEGs and MEGs in AD, two resting-state electrophysiological datasets were used: 1) 148-channel MEGs recorded from 62 subjects (36 AD patients versus 26 age-matched controls); and 2) 16-channel EEGs recorded from 22 subjects (11 AD patients versus 11 age-matched controls). The results obtained by both the MDE and mvMDE illustrated that the controls' EEG and MEG signals have more irregularity at short-time scales than the AD patients' recordings, whereas the latter ones are more irregular at long-time scales. The findings are in agreement with [34, 109, 144, 155, 156]. The univariate and multivariate entropy values for AD patients were higher than those for controls at the majority of scale factors. Thus, AD subjects' EEGs and MEGs are more complex than controls' ones.

This fact is in contradiction with the complexity hypothesis: age-matched control individuals' signals are more complex than diseased subjects' (here AD patients') time series [21, 111]. Therefore, it may be required to revise the hypothesis of complexity for the application of AD using EEGs and MEGs: healthy controls' brain activity signals, compared with those for AD patients, are more complex for short-time scale factors and less complex for long-time scale factors.

Based on Mann-Whitney U-test for AD patients versus controls, MDE, compared with MSE and MFE, led to more significant differences for both the EEG and MEG datasets (e.g., 0.0010, 0.0181, and 0.0356 for respectively MDE, MFE, and MSE using EEG database). The number of scale factors with significant and very significant differences for MDE was larger than those for MSE and MFE for both the EEG and MEG datasets (e.g., 11, 10, and 10 for respectively MDE, MFE, and MSE using EEG database), respectively. For multivariate multiscale entropy approaches, the results also suggested that the number of temporal scales with very significant differences for mvMDE was larger than those for mvMSE and mvMFE. These advantages and low computational time of MDE and mvMDE illustrate the superiority of these techniques over the state-of-the-art approaches for the discrimination of AD patients from age-matched controls' EEGs and MEGs.

On the whole, it is expected that our introduced MDE and mvMDE play a key role in the evaluation of complexity in physiological time series in general, and characterisation of EEGs and MEGs in AD particularly.

7.2 Limitations of the Thesis

In spite of the promising results for the AD application, the number of subjects for both the EEG and MEG datasets was relatively small. To ascertain the usefulness of these methods, these novel nonlinear signal processing approaches should be used on larger databases of AD patients and control subjects, potentially including subjects with mild cognitive impairment. Furthermore, the EEG and MEG datasets used in this Thesis are resting-state. In addition, to better assess the introduced methods to distinguish AD patients from controls, other kinds of EEG and MEG datasets with different characteristics (e.g., lengths and sampling frequency values) should be used.

Additionally, the correlation between mini-mental state examination (MMSE) and entropy-based values for AD patients' EEG signals has not be evaluated. Moreover, the detected complexity increase in the EEG and MEG might not be exclusive to AD and supplementary work should be carried out to analyse whether these EEG and MEG complexity changes also happen in other types of dementia, such as vascular dementia.

In this Thesis, the most commonly used coarse-graining process was employed [21, 23, 30, 32, 111]. The alternative coarse-graining processes based on multivariate empirical mode decomposition (MEMD) [110, 220, 221], and finite impulse response (FIR) filters [204] should be employed instead of the classical implementation of coarse-graining process used herein. Moreover, other filter banks may be used instead of the existing coarse-graining approaches. Finally, there is a need to compare these methods and to illustrate their advantages and disadvantages.

It is hypothesized that the profiles obtained by the univariate and multivariate multiscale entropy methods at short and long scale factors originate different pathophysiologic mechanisms toward regular or random process [156]. Nevertheless, the physiological nature of them needs to be further investigated. To this end, it is recommended to somehow create a model for normal EEGs and MEGs based on their dynamics in different frequency bands and then, change it based on the fact that the power in delta, theta, and gamma increases, whereas power in higher frequencies alpha and beta decreases in EEGS and MEGs recorded from AD patients [6, 119, 120, 124, 125]. Finally, the entropy-based values on the basis of these two models should be compared with the existing complexity-based profiles to understand if these results arise from the effects of AD on the different frequency-bands or not.

7.3 Future Research

The concept of DispEn can be extended on the basis of some concepts in information theory. For instance, based on conditional entropy [38] and DispEn, dispersion conditional entropy may be defined to quantify the average rate of creation of new information by DispEn with increasing the embedding dimension from m to $m + 1$. Cross-DispEn can also be defined to compare two different time series to assess their degree of asynchrony or dissimilarity based on cross-SampEn [16] and DispEn. The multivariate extension of cross-DispEn can be defined

to evaluate the degree of dissimilarity for two multi-channel signals as well.

A recent development in the field has tried to generalize multivariate and univariate multiscale algorithms to a family of statistics by using different moments (e.g., variance, skewness, and kurtosis) in the univariate and multivariate coarse-graining process [30, 32, 222, 223]. To this end, as pilot studies, we proposed RCMFE and RCmvMFE based on variance for extraction various dynamical properties (or features) of spread over multiple time scales [30, 32]. The results illustrated that the different moments for the multiscale technique lead to distinguishing different types of dynamics of a particular signal. Finally, it is recommended to compare these techniques in the context of signals processing and to investigate their interpretations.

The ability of our introduced FDispEn, DispEn, MDE, and mvMDE to distinguish different kinds of dynamics of other univariate and multivariate physiological and even non-physiological time series can be inspected. Surrogate signals can also be used by randomly shuffling the measured time series to find significantly different serial correlations in the data and the shuffles. Accordingly, we can reject or confirm the hypothesis of independence [224]. In fact, the correlations among the signal samples are destroyed in shuffles, while preserving statistical properties of the distributions (especially the first and second moments), and the complexity of the surrogates is lower than or equal to (in case the original series is completely random) that of the original signal [23, 224]. Finally, a comprehensive comparison among all employed nonlinear techniques is recommended to be done in the future.

The univariate entropy methods can also be extended to their two-dimensional cases to quantify the irregularity of textures or images [225, 226]. To this end, we have recently introduced two-dimensional distribution entropy (DistrEn_{2D}) - as an insensitive feature extraction method to rotation - [225]. The results indicated that DistrEn_{2D} can detect different amounts of white Gaussian and salt and pepper noise, and discriminate periodic from synthesized textures. The results also showed that DistrEn_{2D} distinguishes different kinds of textured surfaces. Nevertheless, this article was a pilot study and there is enough room to introduce novel two-dimensional entropy methods and find new applications for such techniques.

Appendix A: List of Publications

The outcomes of this thesis are described in the following papers:

Journals:

1. H. Azami, M. Rostaghi, D. Abásolo, and J. Escudero, "Refined composite multiscale dispersion entropy and its application to biomedical signals," *IEEE Transactions on Biomedical Engineering*, DOI: 10.1109/TBME.2017.2679136, 2017.
2. H. Azami, A. Fernandez, and J. Escudero, "Refined multiscale fuzzy entropy based on standard deviation for biomedical signal analysis," *Medical & Biological Engineering & Computing*, vol. 55, no. 11, pp. 2037–2052, 2017.
3. H. Azami and J. Escudero, "Refined composite multivariate generalized multiscale fuzzy entropy: A tool for complexity analysis of multichannel signals," *Physica A: Statistical Mechanics and its Applications*, vol. 465, pp. 261–276, 2017.
4. H. Azami, D. Abásolo, S. Simons, and J. Escudero, "Univariate and multivariate generalized multiscale entropy to characterise EEG signals in Alzheimer's disease," *Entropy*, vol. 19, no. 1, p. 31, 2017 (**Invited Feature Paper**).
5. H. Azami, J. Escudero, and A. Humeau-Heurtier, "Bidimensional distribution entropy to analyze the irregularity of small-sized textures," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1338–1342, 2017.
6. M. Rostaghi and H. Azami, "Dispersion entropy: A measure for time series analysis," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 610–614, 2016.
7. H. Azami and J. Escudero, "Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation," *Computer Methods and Programs in Biomedicine*, vol. 128, pp. 40–51, 2016.
8. H. Azami and J. Escudero, "Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings," *Biomedical Signal Processing and Control*, vol. 23, pp. 28–41, 2016.

Preprints:

1. H. Azami, A. Fernández, and J. Escudero, "Multivariate multiscale dispersion entropy of biomedical times series," *arXiv preprint:1704.03947*, 2017.
2. H. Azami and J. Escudero, "Amplitude- and frequency-based dispersion patterns and entropy," *arXiv preprint:1708.01066*, 2017.

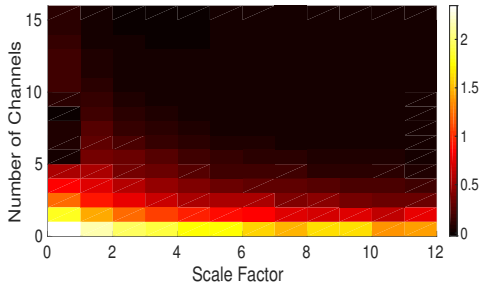
Conferences:

1. H. Azami, E. Kinney-lang, A. Ebied, A. Fernández and J. Escudero, "Multiscale dispersion entropy for the regional analysis of resting-state magnetoencephalogram complexity in Alzheimer's disease," *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 3182-3185.
2. H. Azami, M. Rostaghi, A. Fernández, and J. Escudero, "Dispersion entropy for the analysis of resting-state MEG regularity in Alzheimer's disease," *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 6417–6420.
3. H. Azami, K. Smith and J. Escudero, "MEMD-enhanced multivariate fuzzy entropy for the evaluation of complexity in biomedical signals," *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 3761-3764.
4. H. Azami, J. Escudero, and A. Fernández, "Refined composite multivariate multiscale entropy based on variance for analysis of resting-state magnetoencephalograms in Alzheimer's disease," *International Conference for Students on Applied Engineering (ICSAE)*, IEEE, 2016, pp. 413–418.
5. H. Azami, K. Smith, A. Fernández, and J. Escudero, "Evaluation of resting-state magnetoencephalogram complexity in Alzheimer's disease with multivariate multiscale permutation and sample entropies," *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 7422–7425.

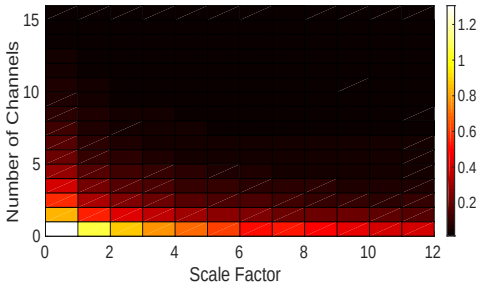
Appendix B: Effect of Number of Channels on Multivariate Sample and Fuzzy Entropy Approaches

The second step of the algorithms of multiscale sample entropy (SampEn - MSE) and multivariate MSE (mvMSE) are SampEn and multivariate sample entropy (mvSE), respectively [16, 23]. SampEn is based on the conditional probability that sequences close to each other for *m* consecutive data points will also be close to each other when one more point is added to each sequence [16]. Thus, the proportion of unseen, new samples over the number of samples included in the previous pattern for the embedding dimension $m = 2$ is 50%. However, in mvSE, multivariate embedded vectors are initially generated with the length of $m_1 + m_2 + \cdots + m_p$, where *p* denotes the number of channels of a time series [23]. For example, for a trivariate time series with the embedding dimension $m = [2, 2, 2]$, the length of embedded vectors is 6. Then, the conditional probability that sequences with the embedding dimension $m = [2, 2, 2]$ close to each other for six data points will also be close to each other for seven data points, associated with the embedding dimensions $[2, 2, 3]$, $[2, 3, 2]$, or $[3, 2, 2]$, is calculated. Note that the length of the newly embedded vectors is 7. Therefore, the proportion of unseen samples over the number of total samples in previous patterns for the embedding dimension $m = [2, 2, 2]$ is 16.66%. Likewise, for a four-channel time series with the embedding dimension $m = [2, 2, 2, 2]$, the proportion of unseen samples over the number of samples of previous patterns is 12.5%. Consequently, the proportion of new samples decreases proportionally to the number of channels, thus decreasing the likelihood of the longer new pattern not being a match with the shorter ones.

To investigate the change in multivariate entropy values as the number of channels increases, we used an uncorrelated multivariate white noise that the number of its channels changes from 1 to 16 and the length of each of them is 1280 samples (equal to the length of electroencephalograms in Chapter 6). Figure A2.1(a) shows how the number of channels



(a) mvMSE



(b) mvMFE

Figure A2.1: (a) mvMSE and (b) mvMFE values for the uncorrelated 1- to 16-channel white noise.

affects the mvMSE output values. It can be seen that the larger the number of channels, the smaller the mvMSE values, something that agrees with our results in Chapter 6.

Similarly, the multivariate multiscale fuzzy entropy (mvMFE) values for uncorrelated multivariate white noise that the number of its channels changes from 1 to 16 are shown in Figure A2.1(b). The results illustrate that the the larger the number of channels, the smaller the mvMFE, like mvMSE, values. Nevertheless, as can be seen in the ranges of Figures A2.1(a) and (b), the mvMFE values do not change as much as mvMSE ones.

Bibliography

[1] A. Wimo and M. Prince, *World Alzheimer report 2010: the global economic impact of dementia*. Alzheimer's Disease International, 2010.

[2] Alzheimer's Association, "2017 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 13, no. 4, pp. 325–373, 2017.

[3] H. O. Lords, "Ready for ageing?," *Report Retrieved January*, vol. 2, p. 2015, 2013.

[4] M. Prince, M. Knapp, M. Guerchet, P. McCrone, M. Prina, A. Comas-Herrera, R. Wittenberg, B. Adelaja, B. Hu, D. King, *et al.*, "Dementia uk: update," *Alzheimer's Society*, 2014.

[5] R. S. Wilson, E. Segawa, P. A. Boyle, S. E. Anagnos, L. P. Hizel, and D. A. Bennett, "The natural history of cognitive decline in Alzheimer's disease," *Psychology and Aging*, vol. 27, no. 4, p. 1008, 2012.

[6] J. Dauwels, F. Vialatte, and A. Cichocki, "Diagnosis of Alzheimer's disease from EEG signals: where are we standing?," *Current Alzheimer Research*, vol. 7, no. 6, pp. 487–505, 2010.

[7] Alzheimer's Association, "2016 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 12, no. 4, pp. 459–509, 2016.

[8] S. Bhat, U. R. Acharya, N. Dadmehr, and H. Adeli, "Clinical neurophysiological and automated EEG-based diagnosis of the Alzheimer's disease," *European Neurology*, vol. 74, no. 3-4, pp. 202–210, 2015.

[9] M. Donaghy, *Brain's diseases of the nervous system*. Oxford University Press Oxford, pp. 750–754, 2001.

[10] R. Hornero, D. Abásolo, J. Escudero, and C. Gómez, "Nonlinear analysis of electroencephalogram and magnetoencephalogram recordings in patients with Alzheimer's disease," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1887, pp. 317–336, 2009.

[11] A. Fornito, A. Zalesky, and M. Breakspear, "The connectomics of brain disorders," *Nature Reviews Neuroscience*, vol. 16, no. 3, p. 159, 2015.

[12] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, p. 061907, 2001.

[13] M. Breakspear, "Dynamic models of large-scale brain activity," *Nature Neuroscience*, vol. 20, no. 3, pp. 340–352, 2017.

[14] R. Hornero, J. Escudero, A. Fernández, J. Poza, and C. Gómez, "Spectral and nonlinear analyses of MEG background activity in patients with Alzheimer's disease," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 6, pp. 1658–1665, 2008.

[15] S. M. Pincus, "Approximate entropy as a measure of system complexity.," *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[16] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[17] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002.

[18] M. Rostaghi and H. Azami, "Dispersion entropy: A measure for time series analysis," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 610–614, 2016.

[19] C. E. Shannon, "Communication theory of secrecy systems," *Bell Labs Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.

[20] S. M. Pincus and A. L. Goldberger, "Physiological time-series analysis: what does regularity quantify?," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 266, no. 4, pp. H1643–H1656, 1994.

[21] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Physical Review E*, vol. 71, no. 2, p. 021906, 2005.

[22] H. C. Fogedby, "On the phase space approach to complexity," *Journal of Statistical Physics*, vol. 69, no. 1-2, pp. 411–425, 1992.

[23] M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy: A tool for complexity analysis of multichannel data," *Physical Review E*, vol. 84, no. 6, p. 061918, 2011.

[24] D. Labate, F. La Foresta, G. Morabito, I. Palamara, and F. C. Morabito, "Entropic measures of EEG complexity in Alzheimer's disease through a multivariate multiscale approach," *Sensors Journal, IEEE*, vol. 13, no. 9, pp. 3284–3292, 2013.

[25] F. C. Morabito, D. Labate, F. La Foresta, A. Bramanti, G. Morabito, and I. Palamara, "Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer's disease EEG," *Entropy*, vol. 14, no. 7, pp. 1186–1202, 2012.

[26] J. Jeong, "EEG dynamics in patients with Alzheimer's disease," *Clinical Neurophysiology*, vol. 115, no. 7, pp. 1490–1505, 2004.

[27] B. Jelles, J. Van Birgelen, J. Slaets, R. Hekster, E. Jonkman, and C. Stam, "Decrease of non-linear structure in the EEG of Alzheimer patients compared to healthy controls," *Clinical Neurophysiology*, vol. 110, no. 7, pp. 1159–1167, 1999.

[28] H. Azami and J. Escudero, "Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation," *Computer Methods and Programs in Biomedicine*, vol. 128, pp. 40–51, 2016.

[29] H. Azami and J. Escudero, "Amplitude- and frequency-based dispersion patterns and entropy," *arXiv preprint:1708.01066*, 2017.

[30] H. Azami, A. Fernández, and J. Escudero, "Refined multiscale fuzzy entropy based on standard deviation for biomedical signal analysis," *Medical & Biological Engineering & Computing*, vol. 55, no. 11, pp. 2037–2052, 2017.

[31] H. Azami, M. Rostaghi, D. Abasolo, and J. Escudero, "Refined composite multiscale dispersion entropy and its application to biomedical signals," *IEEE Transactions on Biomedical Engineering*, DOI: 10.1109/TBME.2017.2679136, 2017.

[32] H. Azami and J. Escudero, "Refined composite multivariate generalized multiscale fuzzy entropy: A tool for complexity analysis of multichannel signals," *Physica A: Statistical Mechanics and its Applications*, vol. 465, pp. 261–276, 2017.

[33] H. Azami, A. Fernández, and J. Escudero, "Multivariate multiscale dispersion entropy of biomedical times series," *arXiv preprint:1704.03947*, 2017.

[34] H. Azami, D. Abásolo, S. Simons, and J. Escudero, "Univariate and multivariate generalized multiscale entropy to characterise EEG signals in Alzheimer's disease," *Entropy*, vol. 19, no. 1, p. 31, 2017.

[35] C. J. Stam, "Nonlinear dynamical analysis of EEG and MEG: review of an emerging field," *Clinical Neurophysiology*, vol. 116, no. 10, pp. 2266–2301, 2005.

[36] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain," *Reviews of Modern Physics*, vol. 65, no. 2, p. 413, 1993.

[37] S. Sanei and J. A. Chambers, *EEG signal processing*. John Wiley & Sons, pp. 1-20, 2007.

[38] L. Faes, A. Porta, and G. Nollo, "Information decomposition in bivariate systems: theory and application to cardiorespiratory dynamics," *Entropy*, vol. 17, no. 1, pp. 277–303, 2015.

[39] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6, p. 068102, 2002.

[40] M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri, "The global prevalence of dementia: a systematic review and metaanalysis," *Alzheimer's & Dementia*, vol. 9, no. 1, pp. 63–75, 2013.

[41] N. C. Berchtold and C. W. Cotman, "Evolution in the conceptualization of dementia and Alzheimer's disease: Greco-roman period to the 1960s," *Neurobiology of Aging*, vol. 19, no. 3, pp. 173–189, 1998.

[42] R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack, J. Kaye, T. J. Montine, *et al.*, "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's & Dementia*, vol. 7, no. 3, pp. 280–292, 2011.

[43] M. Pietto, M. A. Parra, N. Trujillo, F. Flores, A. M. Garcia, J. Bustin, P. Richly, F. Manes, F. Lopera, A. Ibanez, *et al.*, "Behavioral and electrophysiological correlates of memory binding deficits in patients at different risk levels for Alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 53, no. 4, pp. 1325–1340, 2016.

[44] L. E. Hebert, J. Weuve, P. A. Scherr, and D. A. Evans, "Alzheimer disease in the united states (2010–2050) estimated using the 2010 census," *Neurology*, vol. 80, no. 19, pp. 1778–1783, 2013.

[45] R. C. Green, L. A. Cupples, R. Go, K. S. Benke, T. Edeki, P. A. Griffith, M. Williams, Y. Hipps, N. Graff-Radford, D. Bachman, *et al.*, "Risk of dementia among white and african american relatives of patients with Alzheimer's disease," *Jama*, vol. 287, no. 3, pp. 329–336, 2002.

[46] C. Qiu, M. Kivipelto, H. Agüero-Torres, B. Winblad, and L. Fratiglioni, "Risk and protective effects of the apoe gene towards Alzheimer's disease in the kungsholmen project: variation by age and sex," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 75, no. 6, pp. 828–833, 2004.

[47] C.-C. Liu, T. Kanekiyo, H. Xu, and G. Bu, "Apolipoprotein e and Alzheimer disease: risk, mechanisms and therapy," *Nature Reviews Neurology*, vol. 9, no. 2, pp. 106–118, 2013.

[48] M. Baumgart, H. M. Snyder, M. C. Carrillo, S. Fazio, H. Kim, and H. Johns, "Summary of the evidence on modifiable risk factors for cognitive decline and dementia: a population-based perspective," *Alzheimer's & Dementia*, vol. 11, no. 6, pp. 718–726, 2015.

[49] C. T. Liverman, K. Yaffe, D. G. Blazer, *et al.*, *Cognitive aging: Progress in understanding and opportunities for action*. National Academies Press, 2015.

[50] J. S. Saczynski, L. A. Pfeifer, K. Masaki, E. S. Korf, D. Laurin, L. White, and L. J. Launer, "The effect of social engagement on incident dementia: the honolulu-asia aging study," *American Journal of Epidemiology*, vol. 163, no. 5, pp. 433–440, 2006.

[51] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease report of the nincds-adrda work group under the auspices of department of health and human services task force on Alzheimer's disease," *Neurology*, vol. 34, no. 7, pp. 939–939, 1984.

[52] C. R. Jack, M. S. Albert, D. S. Knopman, G. M. McKhann, R. A. Sperling, M. C. Carrillo, B. Thies, and C. H. Phelps, "Introduction to the recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's & Dementia*, vol. 7, no. 3, pp. 257–262, 2011.

[53] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, *et al.*, "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's & Dementia*, vol. 7, no. 3, pp. 263–269, 2011.

[54] B. L. Wright, J. T. Lai, and A. J. Sinclair, "Cerebrospinal fluid and lumbar puncture: a practical review," *Journal of neurology*, vol. 259, no. 8, pp. 1530–1545, 2012.

[55] J. R. Steinerman and L. S. Honig, "Laboratory biomarkers in Alzheimer's disease," *Current Neurology and Neuroscience Reports*, vol. 7, no. 5, pp. 381–387, 2007.

[56] L. Minati, T. Edginton, M. Grazia Bruzzone, and G. Giaccone, "Reviews: current concepts in Alzheimer's disease: a multidisciplinary review," *American Journal of Alzheimer's Disease & Other Dementias*, vol. 24, no. 2, pp. 95–121, 2009.

[57] K. Blennow, B. Dubois, A. M. Fagan, P. Lewczuk, M. J. de Leon, and H. Hampel, "Clinical utility of cerebrospinal fluid biomarkers in the diagnosis of early Alzheimer's disease," *Alzheimer's & Dementia*, vol. 11, no. 1, pp. 58–69, 2015.

[58] G. Muehllehner and J. S. Karp, "Positron emission tomography," *Physics in Medicine and Biology*, vol. 51, no. 13, p. R117, 2006.

[59] K. Blennow, M. Leon, and H. Zetterberg, "Alzheimer's disease," *The Lancet*, vol. 368, no. 9533, pp. 387–403, 2006.

[60] B. Dubois, H. H. Feldman, C. Jacova, S. T. DeKosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, *et al.*, "Research criteria for the diagnosis of Alzheimer's disease: revising the nincds–adrda criteria," *The Lancet Neurology*, vol. 6, no. 8, pp. 734–746, 2007.

[61] C. Hooijer, C. Jonker, J. Posthuma, and S. L. Visser, "Reliability, validity and follow-up of the EEG in senile dementia: sequelae of sequential measurement," *Electroencephalography and Clinical Neurophysiology*, vol. 76, no. 5, pp. 400–412, 1990.

[62] C. Stam, "Use of magnetoencephalography (MEG) to study functional brain networks in neurodegenerative disorders," *Journal of the Neurological Sciences*, vol. 289, no. 1, pp. 128–134, 2010.

[63] E. Niedermeyer and F. L. da Silva, *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[64] R. Hari, S. Levänen, and T. Raij, "Timing of human cortical functions during cognition: role of MEG," *Trends in Cognitive Sciences*, vol. 4, no. 12, pp. 455–462, 2000.

[65] J. Zimmerman, P. Thiene, and J. Harding, "Design and operation of stable rf-biased superconducting point-contact quantum devices, and a note on the properties of perfectly clean metal contacts," *Journal of Applied Physics*, vol. 41, no. 4, pp. 1572–1580, 1970.

[66] D. Cohen, "Magnetoencephalography: detection of the brain's electrical activity with a superconducting magnetometer," *Science*, vol. 175, no. 4022, pp. 664–666, 1972.

[67] J. Vrba and S. E. Robinson, "Signal processing in magnetoencephalography," *Methods*, vol. 25, no. 2, pp. 249–271, 2001.

[68] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[69] L. Venables and S. H. Fairclough, "The influence of performance feedback on goal-setting and mental effort regulation," *Motivation and Emotion*, vol. 33, no. 1, pp. 63–74, 2009.

[70] G. Pfurtscheller, D. Flotzinger, and C. Neuper, "Differentiation between finger, toe and tongue movement in man based on 40 hz EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 90, no. 6, pp. 456–460, 1994.

[71] J. Dauwels, F. Vialatte, T. Musha, and A. Cichocki, "A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG," *NeuroImage*, vol. 49, no. 1, pp. 668–693, 2010.

[72] K. Najarian and R. Splinter, *Biomedical signal and image processing*. CRC press, 2005.

[73] S. Rampp and H. Stefan, "On the opposition of EEG and MEG," 2007.

[74] K. Sternickel and A. I. Braginski, "Biomagnetism using squids: status and perspectives," *Superconductor Science and Technology*, vol. 19, no. 3, p. S160, 2006.

[75] C. T. Chen, *Linear system theory and design*. Oxford University Press, 1998.

[76] G. Boeing, "Visual analysis of nonlinear dynamical systems: Chaos, fractals, self-similarity and the limits of prediction," *Systems*, vol. 4, no. 4, p. 37, 2016.

[77] R. Sleimen-Malkoun, D. Perdikis, V. Müller, J.-L. Blanc, R. Huys, J.-J. Temprado, and V. K. Jirsa, "Brain dynamics of aging: multiscale variability of EEG signals at rest and during an auditory oddball task," *Eneuro*, vol. 2, no. 3, pp. ENEURO–0067, 2015.

[78] D. Hoyer, U. Leder, H. Hoyer, B. Pompe, M. Sommer, and U. Zwiener, "Mutual information and phase dependencies: measures of reduced nonlinear cardiorespiratory interactions after myocardial infarction," *Medical Engineering & Physics*, vol. 24, no. 1, pp. 33–43, 2002.

[79] M. Palacios, H. Friedrich, C. Götze, M. Vallverdú, A. B. de Luna, P. Caminal, and D. Hoyer, "Changes of autonomic information flow due to idiopathic dilated cardiomyopathy," *Physiological Measurement*, vol. 28, no. 6, p. 677, 2007.

[80] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt, "A comparison of waveform fractal dimension algorithms," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, vol. 48, no. 2, pp. 177–183, 2001.

[81] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Physica D: Nonlinear Phenomena*, vol. 16, no. 3, pp. 285–317, 1985.

[82] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 75–81, 1976.

[83] J. J. Heisz and A. R. McIntosh, "Applications of EEG neuroimaging data: Event-related potentials, spectral power, and multiscale entropy," *Journal of visualized experiments: JoVE*, no. 76, 2013.

[84] B. Mandelbrot, "Fractals: form, chance and dimension," *Fractals: form, Chance and Dimension., by Mandelbrot, BB. San Francisco (CA, USA): WH Freeman & Co., 16+ 365 p.*, 1979.

[85] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D: Nonlinear Phenomena*, vol. 9, no. 1-2, pp. 189–208, 1983.

[86] J. Guckenheimer, "Strange attractors in fluid dynamics," in *Dynamical System and Chaos*, pp. 149–156, Springer, 1983.

[87] M. J. Katz, "Fractals and the analysis of waveforms," *Computers in Biology and Medicine*, vol. 18, no. 3, pp. 145–156, 1988.

[88] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, 1988.

[89] A. Petrosian, "Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns," in *Computer-Based Medical Systems, Proceedings of the Eighth IEEE Symposium on*, pp. 212–217, IEEE, 1995.

[90] O. Faust and M. G. Bairy, "Nonlinear analysis of physiological signals: a review," *Journal of Mechanics in Medicine and Biology*, vol. 12, no. 04, p. 1240015, 2012.

[91] T. Tanaka, K. Aihara, and M. Taki, "Analysis of positive lyapunov exponents from random time series," *Physica D: Nonlinear Phenomena*, vol. 111, no. 1-4, pp. 42–50, 1998.

[92] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A practical method for calculating largest Lyapunov exponents from small data sets," *Physica D: Nonlinear Phenomena*, vol. 65, no. 1-2, pp. 117–134, 1993.

[93] Y. Bai, Z. Liang, and X. Li, "A permutation Lempel-Ziv complexity measure for EEG analysis," *Biomedical Signal Processing and Control*, vol. 19, pp. 102–114, 2015.

[94] W. Xiong, L. Faes, and P. C. Ivanov, "Entropy measures, entropy estimators, and their performance in quantifying complex dynamics: Effects of artifacts, nonstationarity, and long-range correlations," *Physical Review E*, vol. 95, no. 6, p. 062114, 2017.

[95] R. M. Gray, *Entropy and information theory*. Springer Science & Business Media, 2011.

[96] T. M. Cover and J. A. Thomas, "Entropy, relative entropy and mutual information," *Elements of Information Theory*, vol. 2, pp. 1–55, 1991.

[97] P. Grassberger and I. Procaccia, "Estimation of the Kolmogorov entropy from a chaotic signal," *Physical Review A*, vol. 28, no. 4, p. 2591, 1983.

[98] J.-P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Reviews of Modern Physics*, vol. 57, no. 3, p. 617, 1985.

[99] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz, "A regularity statistic for medical data analysis," *Journal of Clinical Monitoring and Computing*, vol. 7, no. 4, pp. 335–345, 1991.

[100] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. Bhat, and J. E. Koh, "Application of entropies for automated diagnosis of epilepsy using EEG signals: a review," *Knowledge-Based Systems*, vol. 88, pp. 85–96, 2015.

[101] J. M. Yentes, N. Hunt, K. K. Schmid, J. P. Kaipust, D. McGrath, and N. Stergiou, "The appropriate use of approximate entropy and sample entropy with short data sets," *Annals of Biomedical Engineering*, vol. 41, no. 2, pp. 349–365, 2013.

[102] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[103] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 15, no. 2, pp. 266–272, 2007.

[104] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 283, no. 3, pp. R789–R797, 2002.

[105] S. Lu, X. Chen, J. K. Kanters, I. C. Solomon, and K. H. Chon, "Automatic selection of the threshold value r for approximate entropy," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1966–1972, 2008.

[106] A. Humeau-Heurtier, "The multiscale entropy algorithm and its variants: A review," *Entropy*, vol. 17, no. 5, pp. 3110–3123, 2015.

[107] A. L. Goldberger and M. D. Costa, "Complexity based methods and systems for detecting depression," 2012. US Patent App. 14/233,999.

[108] C.-C. Chung, J.-H. Kang, R.-Y. Yuan, D. Wu, C.-C. Chen, N.-F. Chi, P.-C. Chen, and C.-J. Hu, "Multiscale entropy analysis of electroencephalography during sleep in patients with Parkinson disease," *Clinical EEG and Neuroscience*, vol. 44, no. 3, pp. 221–226, 2013.

[109] J. Escudero, E. Acar, A. Fernández, and R. Bro, "Multiscale entropy analysis of resting-state magnetoencephalogram with tensor factorisations in Alzheimer's disease," *Brain Research Bulletin*, vol. 119, pp. 136–144, 2015.

[110] M. Ahmed, N. Rehman, D. Looney, T. Rutkowski, and D. Mandic, "Dynamical complexity of human responses: a multivariate data-adaptive framework," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, vol. 60, no. 3, pp. 433–445, 2012.

[111] M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy analysis," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 91–94, 2012.

[112] S. Cerutti, D. Hoyer, and A. Voss, "Multiscale, multiorgan and multivariate complexity analyses of cardiovascular regulation," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1892, pp. 1337–1358, 2009.

[113] S. Cerutti, "Multivariate and multiscale analysis of biomedical signals: Towards a comprehensive approach to medical diagnosis," in *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, pp. 1–5, IEEE, 2012.

[114] L. Cao, A. Mees, and K. Judd, "Dynamics from multivariate time series," *Physica D: Nonlinear Phenomena*, vol. 121, no. 1, pp. 75–88, 1998.

[115] H. Azami, K. Smith, A. Fernandez, and J. Escudero, "Evaluation of resting-state magnetoencephalogram complexity in Alzheimer's disease with multivariate multiscale permutation and sample entropies," in *Engineering in Medicine and Biology Society, 37th Annual International Conference of the*, pp. 7422–7425, 2015.

[116] L. Zhao, S. Wei, H. Tang, and C. Liu, "Multivariable fuzzy measure entropy analysis for heart rate variability and heart sound amplitude variability," *Entropy*, vol. 18, no. 12, p. 430, 2016.

[117] S. Ramdani, V. Bonnet, G. Tallon, J. Lagarde, P. L. Bernard, and H. Blain, "Parameters selection for bivariate multiscale entropy analysis of postural fluctuations in fallers and non-fallers older adults," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 8, pp. 859–871, 2016.

[118] C. Herrmann and T. Demiralp, "Human EEG gamma oscillations in neuropsychiatric disorders," *Clinical Neurophysiology*, vol. 116, no. 12, pp. 2719–2733, 2005.

[119] B. Czigler, D. Csikós, Z. Hidasi, Z. A. Gaál, É. Csibri, É. Kiss, P. Salacz, and M. Molnár, "Quantitative EEG in early Alzheimer's disease patients—power spectrum and complexity features," *International Journal of Psychophysiology*, vol. 68, no. 1, pp. 75–80, 2008.

[120] K. Van der Hiele, A. Vein, R. Reijntjes, R. Westendorp, E. Bollen, M. Van Buchem, J. Van Dijk, and H. Middelkoop, "EEG correlates in the spectrum of cognitive decline," *Clinical Neurophysiology*, vol. 118, no. 9, pp. 1931–1939, 2007.

[121] C. Besthorn, R. Zerfass, C. Geiger-Kabisch, H. Sattel, S. Daniel, U. Schreiter-Gasser, and H. Förstl, "Discrimination of Alzheimer's disease and normal aging by EEG data," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 2, pp. 241–248, 1997.

[122] M. Baker, K. Akrofi, R. Schiffer, and M. W. O'Boyle, "EEG patterns in mild cognitive impairment (mci) patients," *The Open Neuroimaging Journal*, vol. 2, p. 52, 2008.

[123] V. Jelic, S. Johansson, O. Almkvist, M. Shigeta, P. Julin, A. Nordberg, B. Winblad, and L. Wahlund, "Quantitative electroencephalography in mild cognitive impairment: longitudinal changes and possible prediction of Alzheimer's disease," *Neurobiology of Aging*, vol. 21, no. 4, pp. 533–540, 2000.

[124] D. Moretti, C. Fracassi, M. Pievani, C. Geroldi, G. Binetti, O. Zanetti, K. Sosta, P. Rossini, and G. Frisoni, "Increase of theta/gamma ratio is associated with memory impairment," *Clinical Neurophysiology*, vol. 120, no. 2, pp. 295–303, 2009.

[125] J. Poza, R. Hornero, D. Abásolo, A. Fernández, and M. García, "Extraction of spectral based measures from MEG background oscillations in Alzheimer's disease," *Medical engineering & physics*, vol. 29, no. 10, pp. 1073–1083, 2007.

[126] M. Signorino, E. Pucci, N. Belardinelli, G. Nolfe, and F. Angeleri, "EEG spectral analysis in vascular and Alzheimer dementia," *Electroencephalography and Clinical Neurophysiology*, vol. 94, no. 5, pp. 313–325, 1995.

[127] H. Soininen, J. Partanen, V. Laulumaa, E.-L. Helkala, M. Laakso, and P. Riekkinen, "Longitudinal EEG spectral analysis in early stage of Alzheimer's disease," *Electroencephalography and Clinical Neurophysiology*, vol. 72, no. 4, pp. 290–297, 1989.

[128] D. V. Moretti, C. Babiloni, G. Binetti, E. Cassetta, G. Dal Forno, F. Ferreric, R. Ferri, B. Lanuzza, C. Miniussi, F. Nobili, *et al.*, "Individual analysis of EEG frequency and band power in mild Alzheimer's disease," *Clinical Neurophysiology*, vol. 115, no. 2, pp. 299–308, 2004.

[129] A. Fernández, R. Hornero, A. Mayo, J. Poza, P. Gil-Gregorio, and T. Ortiz, "MEG spectral profile in Alzheimer's disease and mild cognitive impairment," *Clinical Neurophysiology*, vol. 117, no. 2, pp. 306–314, 2006.

[130] D. Osipova, J. Ahveninen, O. Jensen, A. Ylikoski, and E. Pekkonen, "Altered generation of spontaneous oscillations in Alzheimer's disease," *Neuroimage*, vol. 27, no. 4, pp. 835–841, 2005.

[131] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Progress in Neurobiology*, vol. 77, no. 1, pp. 1–37, 2005.

[132] S. Brassen, D. F. Braus, W. Weber-Fahr, H. Tost, S. Moritz, and G. Adler, "Late-onset depression with mild cognitive deficits: electrophysiological evidences for a preclinical dementia syndrome," *Dementia and Geriatric Cognitive Disorders*, vol. 18, no. 3-4, pp. 271–277, 2004.

[133] B. Jelles, P. Scheltens, W. Van der Flier, E. Jonkman, F. L. da Silva, and C. Stam, "Global dynamical analysis of the EEG in Alzheimer's disease: frequency-specific changes of functional interactions," *Clinical Neurophysiology*, vol. 119, no. 4, pp. 837–841, 2008.

[134] C. Babiloni, R. Ferri, G. Binetti, F. Vecchio, G. B. Frisoni, B. Lanuzza, C. Miniussi, F. Nobili, G. Rodriguez, F. Rundo, *et al.*, "Directionality of EEG synchronization in Alzheimer's disease subjects," *Neurobiology of Aging*, vol. 30, no. 1, pp. 93–102, 2009.

[135] J. Jeong, J. C. Gore, and B. S. Peterson, "Mutual information analysis of the EEG in patients with Alzheimer's disease," *Clinical Neurophysiology*, vol. 112, no. 5, pp. 827–835, 2001.

[136] Y.-M. Park, H.-J. Che, C.-H. Im, H.-T. Jung, S.-M. Bae, and S.-H. Lee, "Decreased EEG synchronization and its correlation with symptom severity in Alzheimer's disease," *Neuroscience Research*, vol. 62, no. 2, pp. 112–117, 2008.

[137] C. J. Stam, A. M. v. C. van Walsum, Y. A. Pijnenburg, H. W. Berendse, J. C. de Munck, P. Scheltens, and B. W. van Dijk, "Generalized synchronization of MEG recordings in Alzheimer's disease: evidence for involvement of the gamma band," *Journal of Clinical Neurophysiology*, vol. 19, no. 6, pp. 562–574, 2002.

[138] H. Berendse, J. Verbunt, P. Scheltens, B. Van Dijk, and E. Jonkman, "Magnetoencephalographic analysis of cortical activity in Alzheimer's disease: a pilot study," *Clinical Neurophysiology*, vol. 111, no. 4, pp. 604–612, 2000.

[139] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources," *Human Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, 2007.

[140] D. Abásolo, J. Escudero, R. Hornero, C. Gómez, and P. Espino, "Approximate entropy and auto mutual information analysis of the electroencephalogram in Alzheimer's disease patients," *Medical & Biological Engineering & Computing*, vol. 46, no. 10, pp. 1019–1028, 2008.

[141] P. Zhao, P. Van-Eetvelt, C. Goh, N. Hudson, S. Wimalaratna, and E. Ifeachor, "Characterization of EEGs in Alzheimer's disease using information theoretic methods," in *Engineering in Medicine and Biology Society. 29th Annual International Conference of the IEEE*, pp. 5127–5131, 2007.

[142] C. Besthorn, H. Sattel, C. Geiger-Kabisch, R. Zerfass, and H. Förstl, "Parameters of EEG dimensional complexity in Alzheimer's disease," *Electroencephalography and Clinical Neurophysiology*, vol. 95, no. 2, pp. 84–89, 1995.

[143] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A spatio-temporal wavelet-chaos methodology for EEG-based diagnosis of Alzheimer's disease," *Neuroscience Letters*, vol. 444, no. 2, pp. 190–194, 2008.

[144] J. Escudero, D. Abásolo, R. Hornero, P. Espino, and M. López, "Analysis of electroencephalograms in Alzheimer's disease patients with multiscale entropy," *Physiological Measurement*, vol. 27, no. 11, p. 1091, 2006.

[145] D. Abásolo, R. Hornero, C. Gómez, M. García, and M. López, "Analysis of EEG background activity in Alzheimer's disease patients with Lempel-Ziv complexity and central tendency measure," *Medical Engineering & Physics*, vol. 28, no. 4, pp. 315–322, 2006.

[146] A. Babloyantz and A. Destexhe, "The Creutzfeld-Jakob disease in the hierarchy of chaotic attractors," in *From Chemical to Biological Organization*, pp. 307–316, Springer, 1988.

[147] J. Jeong, S. Y. Kim, and S.-H. Han, "Non-linear dynamical analysis of the EEG in Alzheimer's disease with optimal embedding dimension," *Electroencephalography and Clinical Neurophysiology*, vol. 106, no. 3, pp. 220–228, 1998.

[148] R. Hornero, P. Espino, A. Alonso, and M. Lopez, "Estimating complexity from EEG background activity of epileptic patients," *IEEE Engineering in Medicine and Biology Magazine*, vol. 18, no. 6, pp. 73–79, 1999.

[149] C. Stam, B. Jelles, H. Achtereekte, S. Rombouts, J. Slaets, and R. Keunen, "Investigation of EEG non-linearity in dementia and parkinson's disease," *Electroencephalography and Clinical Neurophysiology*, vol. 95, no. 5, pp. 309–317, 1995.

[150] J.-P. Eckmann and D. Ruelle, "Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems," *Physica D: Nonlinear Phenomena*, vol. 56, no. 2-3, pp. 185–187, 1992.

[151] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Physical Review Letters*, vol. 50, no. 5, p. 346, 1983.

[152] S. Simons and D. Abásolo, "Distance-based Lempel-Ziv complexity for the analysis of electroencephalograms in patients with Alzheimer's disease," *Entropy*, vol. 19, no. 3, p. 129, 2017.

[153] C. Gómez, R. Hornero, D. Abásolo, A. Fernández, and J. Escudero, "Analysis of MEG background activity in Alzheimer's disease using nonlinear methods and anfis," *Annals of Biomedical Engineering*, vol. 37, no. 3, pp. 586–594, 2009.

[154] A. L. Goldberger, C.-K. Peng, and L. A. Lipsitz, "What is physiologic complexity and how does it change with aging and disease?," *Neurobiology of Aging*, vol. 23, no. 1, pp. 23–26, 2002.

[155] T. Mizuno, T. Takahashi, R. Y. Cho, M. Kikuchi, T. Murata, K. Takahashi, and Y. Wada, "Assessment of EEG dynamical complexity in Alzheimer's disease using multiscale entropy," *Clinical Neurophysiology*, vol. 121, no. 9, pp. 1438–1446, 2010.

[156] A. C. Yang, S.-J. Wang, K.-L. Lai, C.-F. Tsai, C.-H. Yang, J.-P. Hwang, M.-T. Lo, N. E. Huang, C.-K. Peng, and J.-L. Fuh, "Cognitive and neuropsychiatric correlates of EEG dynamic complexity in patients with Alzheimer's disease," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 47, pp. 52–61, 2013.

[157] A. C. Yang and S.-J. Tsai, "Is mental illness complex? from behavior to brain," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 45, pp. 253–257, 2013.

[158] T. Takahashi, "Complexity of spontaneous brain activity in mental disorders," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 45, pp. 258–266, 2013.

[159] Y. Bar-Yam, *Dynamics of complex systems*, vol. 213. Addison-Wesley Reading, MA, 1997.

[160] Y.-C. Zhang, "Complexity and 1/f noise. a phase space approach," *Journal de Physique I*, vol. 1, no. 7, pp. 971–977, 1991.

[161] L. E. V. Silva, B. C. T. Cabella, U. P. da Costa Neves, and L. O. M. Junior, "Multiscale entropy-based methods for heart rate variability complexity analysis," *Physica A: Statistical Mechanics and its Applications*, vol. 422, pp. 143–152, 2015.

[162] R. G. Andrzejak, K. Schindler, and C. Rummel, "Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients," *Physical Review E*, vol. 86, no. 4, p. 046206, 2012.

[163] R. Sharma, R. B. Pachori, and U. R. Acharya, "Application of entropy measures on intrinsic mode functions for the automated identification of focal electroencephalogram signals," *Entropy*, vol. 17, no. 2, pp. 669–691, 2015.

[164] J. Hayano, F. Yamasaki, S. Sakata, A. Okada, S. Mukai, and T. Fujinami, "Spectral characteristics of ventricular response to atrial fibrillation," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 273, no. 6, pp. H2811–H2816, 1997.

[165] T. Kamizawa, T. Hara, and M. Ohya, "On relations among the entropic chaos degree, the Kolmogorov-Sinai entropy and the Lyapunov exponent," *Journal of Mathematical Physics*, vol. 55, no. 3, p. 032702, 2014.

[166] Y. B. Pesin, "Characteristic lyapunov exponents and smooth ergodic theory," *Russian Mathematical Surveys*, vol. 32, no. 4, pp. 55–114, 1977.

[167] M. Zanin, L. Zunino, O. A. Rosso, and D. Papo, "Permutation entropy and its main biomedical and econophysics applications: a review," *Entropy*, vol. 14, no. 8, pp. 1553–1577, 2012.

[168] M. Aboy, R. Hornero, D. Abásolo, and D. Álvarez, "Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 11, pp. 2282–2288, 2006.

[169] H. Azami and J. Escudero, "Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings," *Biomedical Signal Processing and Control*, vol. 23, pp. 28–41, 2016.

[170] J. Escudero, R. Hornero, and D. Abásolo, "Interpretation of the auto-mutual information rate of decrease in the context of biomedical signal analysis. application to electroencephalogram recordings," *Physiological Measurement*, vol. 30, no. 2, p. 187, 2009.

[171] G. D. Clifford, F. Azuaje, and P. McSharry, *Advanced Methods And Tools for ECG Data Analysis*. Norwood, MA, USA: Artech House, Inc., 2006.

[172] J. Lam, "Preserving useful info while reducing noise of physiological signals by using wavelet analysis," pp. 1–20, 2011.

[173] C. Houdré, D. M. Mason, P. Reynaud-Bouret, and J. Rosinski, *High Dimensional Probability VII*. Springer, 2016.

[174] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Transactions on Rehabilitation Engineering*, vol. 6, no. 3, pp. 316–325, 1998.

[175] M. Mendez, A. M. Bianchi, M. Matteucci, S. Cerutti, and T. Penzel, "Sleep apnea screening by autoregressive models from a single ECG lead," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 12, pp. 2838–2850, 2009.

[176] M. Ferrario, M. G. Signorini, G. Magenes, and S. Cerutti, "Comparison of entropy-based regularity estimators: application to the fetal heart rate signal for the identification of fetal distress," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 1, pp. 119–125, 2006.

[177] I. Mitov, "A method for assessment and processing of biomedical signals containing trend and periodic components," *Medical Engineering & Physics*, vol. 20, no. 9, pp. 660–668, 1998.

[178] D. E. Bahr and J. L. Reuss, "Method and apparatus for processing a physiological signal," Jan. 15 2002. US Patent 6,339,715.

[179] G. L. Baker and J. P. Gollub, *Chaotic dynamics: an introduction*. Cambridge University Press, 1996.

[180] L. Cohen, "The history of noise [on the 100th anniversary of its birth]," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 20–45, 2005.

[181] E. Sejdić and L. A. Lipsitz, "Necessity of noise in physiology and medicine," *Computer Methods and Programs in Biomedicine*, vol. 111, no. 2, pp. 459–470, 2013.

[182] M. S. Keshner, "1/f noise," *Proceedings of the IEEE*, vol. 70, no. 3, pp. 212–218, 1982.

[183] S. M. Pincus, "Assessing serial irregularity and its implications for health," *Annals of the New York Academy of Sciences*, vol. 954, no. 1, pp. 245–267, 2001.

[184] N. Iyengar, C. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 271, no. 4, pp. R1078–R1084, 1996.

[185] S. Zhao, D. Rangaprakash, A. Venkataraman, P. Liang, and G. Deshpande, "Investigating focal connectivity deficits in Alzheimer's disease using directional brain networks derived from resting-state fMRI," *Frontiers in Aging Neuroscience*, vol. 9, p. 211, 2017.

[186] M. C. Power, J. Weuve, J. J. Gagne, M. B. McQueen, A. Viswanathan, and D. Blacker, "The association between blood pressure and incident Alzheimer disease: a systematic review and meta-analysis," *Epidemiology*, vol. 22, no. 5, p. 646, 2011.

[187] J. Li, J. Yan, X. Liu, and G. Ouyang, "Using permutation entropy to measure the changes in EEG signals during absence seizures," *Entropy*, vol. 16, no. 6, pp. 3049–3061, 2014.

[188] B. Fadlallah, B. Chen, A. Keil, and J. Príncipe, "Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information," *Physical Review E*, vol. 87, no. 2, p. 022911, 2013.

[189] C. Bian, C. Qin, Q. D. Ma, and Q. Shen, "Modified permutation-entropy analysis of heartbeat dynamics," *Physical Review E*, vol. 85, no. 2, p. 021906, 2012.

[190] B. García-Martínez, A. Martínez-Rodrigo, A. Fernández-Caballero, J. Moncho-Bogani, J. M. Pastor, and R. Alcaraz, "Nonlinear symbolic assessment of electroencephalographic recordings for negative stress recognition," in *International Work-Conference on the Interplay Between Natural and Artificial Computation - Nonlinear Symbolic Assessment of Electroencephalographic Recordings for Negative Stress Recognition*, pp. 203–212, Springer, 2017.

[191] S. Tufféry, *Data mining and statistics for decision making*, vol. 2. Wiley Chichester, 2011.

[192] M. N. Gibbs and D. J. MacKay, "Variational gaussian process classifiers," *IEEE Transactions on Neural Networks*, vol. 11, no. 6, pp. 1458–1464, 2000.

[193] S. Nemati, B. A. Edwards, J. Lee, B. Pittman-Polletta, J. P. Butler, and A. Malhotra, "Respiration and heart rate complexity: effects of age and gender assessed by band-limited transfer entropy," *Respiratory Physiology & Neurobiology*, vol. 189, no. 1, pp. 27–33, 2013.

[194] Y. Jiang, D. Mao, and Y. Xu, "A fast algorithm for computing sample entropy," *Advances in Adaptive Data Analysis*, vol. 3, no. 01n02, pp. 167–186, 2011.

[195] http://dx.doi.org/10.7488/ds/1982.

[196] http://dx.doi.org/10.7488/ds/273.

[197] https://physionet.org/physiotools/matlab/wfdb-app-matlab.

[198] http://dx.doi.org/10.7488/ds/1477.

[199] S.-D. Wu, C.-W. Wu, and A. Humeau-Heurtier, "Refined scale-dependent permutation entropy to analyze systems complexity," *Physica A: Statistical Mechanics and its Applications*, vol. 450, pp. 454–461, 2016.

[200] A. Humeau-Heurtier, C.-W. Wu, S.-D. Wu, G. Mahé, and P. Abraham, "Refined multiscale hilbert–huang spectral entropy and its application to central and peripheral cardiovascular data," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 11, pp. 2405–2415, 2016.

[201] R. A. Thuraisingham and G. A. Gottwald, "On multiscale entropy analysis for physiological data," *Physica A: Statistical Mechanics and its Applications*, vol. 366, pp. 323–332, 2006.

[202] S.-D. Wu, C.-W. Wu, S.-G. Lin, K.-Y. Lee, and C.-K. Peng, "Analysis of complex time series using refined composite multiscale entropy," *Physics Letters A*, vol. 378, no. 20, pp. 1369–1374, 2014.

[203] A. Humeau-Heurtier, C.-W. Wu, and S.-D. Wu, "Refined composite multiscale permutation entropy to overcome multiscale permutation entropy length dependence," *Signal Processing Letters, IEEE*, vol. 22, no. 12, pp. 2364–2367, 2015.

[204] J. F. Valencia, A. Porta, M. Vallverdu, F. Claria, R. Baranowski, E. Orlowska-Baranowska, and P. Caminal, "Refined multiscale entropy: Application to 24-h holter recordings of heart period variability in healthy and aortic stenosis subjects," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 9, pp. 2202–2213, 2009.

[205] P. Li, L. Ji, C. Yan, K. Li, C. Liu, and C. Liu, "Coupling between short-term heart rate and diastolic period is reduced in heart failure patients as indicated by multivariate entropy analysis," in *Computing in Cardiology Conference (CinC), 2014*, pp. 97–100, IEEE, 2014.

[206] A. Humeau-Heurtier, "Multivariate generalized multiscale entropy analysis," *Entropy*, vol. 18, no. 11, p. 411, 2016.

[207] W. Penny and S. Roberts, "Bayesian multivariate autoregressive models with structured priors," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 149, no. 1, pp. 33–41, 2002.

[208] B. Pfaff *et al.*, "VAR, SVAR and SVEC models: Implementation within R package vars," *Journal of Statistical Software*, vol. 27, no. 4, pp. 1–32, 2008.

[209] K. Hu, E. J. Van Someren, S. A. Shea, and F. A. Scheer, "Reduction of scale invariance of activity fluctuations with aging and Alzheimer's disease: Involvement of the circadian pacemaker," *Proceedings of the National Academy of Sciences*, vol. 106, no. 8, pp. 2490–2494, 2009.

[210] J. M. Hausdorff, P. L. Purdon, C. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger, "Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations," *Journal of Applied Physiology*, vol. 80, no. 5, pp. 1448–1457, 1996.

[211] H. Azami, J. Escudero, and A. Fernández, "Refined composite multivariate multiscale entropy based on variance for analysis of resting-state magnetoencephalograms in Alzheimer's disease," in *Students on Applied Engineering (ICSAE), International Conference for*, pp. 413–418, IEEE, 2016.

[212] `http://dx.doi.org/10.7488/ds/1432`.

[213] Y. Yin and P. Shang, "Multivariate weighted multiscale permutation entropy for complex time series," *Nonlinear Dynamics*, pp. 1–16, 2017.

[214] D. Abásolo, R. Hornero, P. Espino, D. Alvarez, and J. Poza, "Entropy analysis of the EEG background activity in Alzheimer's disease patients," *Physiological Measurement*, vol. 27, no. 3, p. 241, 2006.

[215] J. Escudero, S. Sanei, D. Jarchi, D. Abásolo, and R. Hornero, "Regional coherence evaluation in mild cognitive impairment and Alzheimer's disease based on adaptively extracted magnetoencephalogram rhythms," *Physiological Measurement*, vol. 32, no. 8, p. 1163, 2011.

[216] J. Escudero, R. Hornero, D. Abásolo, and A. Fernández, "Quantitative evaluation of artifact removal in real magnetoencephalogram signals with blind source separation," *Annals of Biomedical Engineering*, vol. 39, no. 8, pp. 2274–2286, 2011.

[217] P.-H. Tsai, S.-C. Chang, F.-C. Liu, J. Tsao, Y.-H. Wang, and M.-T. Lo, "A novel application of multiscale entropy in electroencephalography to predict the efficacy of acetylcholinesterase inhibitor in Alzheimer's disease," *Computational and Mathematical Methods in Medicine*, vol. 2015, 2015.

[218] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey," *Clinical Neurophysiology*, vol. 118, no. 3, pp. 480–494, 2007.

[219] J. Zheng, J. Cheng, Y. Yang, and S. Luo, "A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination," *Mechanism and Machine Theory*, vol. 78, pp. 187–200, 2014.

[220] M. Hu and H. Liang, "Adaptive multiscale entropy analysis of multivariate neural data," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 1, pp. 12–15, 2012.

[221] H. Azami, K. Smith, and J. Escudero, "MEMD-enhanced multivariate fuzzy entropy for the evaluation of complexity in biomedical signals," in *Engineering in Medicine and Biology Society, IEEE 38th Annual International Conference of the*, pp. 3761–3764, 2016.

[222] M. D. Costa and A. L. Goldberger, "Generalized multiscale entropy analysis: application to quantifying the complex volatility of human heartbeat time series," *Entropy*, vol. 17, no. 3, pp. 1197–1203, 2015.

[223] M. Xu and P. Shang, "Analysis of financial time series using multiscale entropy based on skewness and kurtosis," *Physica A: Statistical Mechanics and its Applications*, vol. 490, pp. 1543–1550, 2018.

[224] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D: Nonlinear Phenomena*, vol. 142, no. 3, pp. 346–382, 2000.

[225] H. Azami, J. Escudero, and A. Humeau-Heurtier, "Bidimensional distribution entropy to analyze the irregularity of small-sized textures," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1338–1342, 2017.

[226] L. Silva, A. Senra Filho, V. Fazan, J. Felipe, and L. M. Junior, "Two-dimensional sample entropy: assessing image texture through irregularity," *Biomedical Physics & Engineering Express*, vol. 2, no. 4, p. 045002, 2016.