DEPARTAMENTO DE INVESTIGACIÓN Y DOCTORADO

ITBA
Instituto Tecnológico de Buenos Aires

MODELADO ESTADÍSTICO Y CUANTIFICACIÓN DE SEÑALES DE EEG:

Aplicación a la caracterización y detección de inicio en crisis epilépticas

AUTOR: Mag. Ing. Antonio Quintero-Rincón

DIRECTOR: Dr. Marcelo Risk
CO-DIRECTOR: Dr. Hadj Batatia

TESIS PRESENTADA PARA OPTAR AL TÍTULO DE
DOCTOR EN INGENIERÍA

Jurado

Dr. Carlos H. Muravchik | Dr. Pablo I. Fierens | Dr. Alejandro Blenkmann

CIUDAD AUTÓNOMA DE BUENOS AIRES
Abril 2019

Mag. Ing. Antonio QUINTERO-RINCÓN: MODELADO ESTADÍSTICO Y CUANTIFICACIÓN DE SEÑALES DE EEG:. *Tesis presentada como requisito parcial para acceder al grado de* ***DOCTOR EN INGENIERÍA*** *del Instituto Tecnológico de Buenos Aires.*

DEPARTAMENTO DE INVESTIGACIÓN Y DOCTORADO

ITBA
Instituto Tecnológico de Buenos Aires

STATISTICAL MODELING AND QUANTIFICATION OF EEG SIGNALS:

Application to the characterization and onset detection in epileptic seizures

AUTHOR: Mag. Ing. Antonio Quintero-Rincón

ADVISOR: Dr. Marcelo Risk
CO-ADVISOR: Dr. Hadj Batatia

A THESIS SUBMITTED AS A REQUIREMENT FOR THE DEGREE OF
DOCTOR EN INGENIERÍA

Thesis committee

Dr. Carlos H. Muravchik | Dr. Pablo I. Fierens | Dr. Alejandro Blenkmann

CIUDAD AUTÓNOMA DE BUENOS AIRES
April 2019

Mag. Ing. Antonio QUINTERO-RINCÓN: STATISTICAL MODELING AND QUANTIFICATION OF EEG SIGNALS:. *A thesis submitted in partial fulfillment of the requirements for the degree of* ***DOCTOR EN INGENIERÍA*** *of Instituto Tecnológico de Buenos Aires.*

This work is dedicated to my family, my wife *Angye Carolina Garzón Narváez*, my father *Ernesto Quintero* and my mother *Maria Luisa Rincón*!

"It is not whether to be or not to be but to know how to be!"

..::..

Acknowledgements

Thanks to: *Hadj Batatia* from University of Toulouse, *Marcelo Pereyra* from Heriot-Watt University, *Jorge Prendes* from Mathworks and *Roxana Saint-Nom* my friends and partners for all the good times we shared; *Carlos D'Giano, Gabriela Ugarnes* and *Valeria Muro* from FLENI for all teachings; *Jean Yves Tourneret* from University of Toulouse and *Jose Carlos Bermudes* from Federal University of Santa Catarina for their recommendations and *Marcelo Risk* from ITBA for all the support throughout the process.

Resumen

Identificar la actividad cerebral epiléptica utilizando señales de electroencefalografía (EEG) en tiempo real es un problema difícil. Los métodos modernos de detección basados en técnicas avanzadas de aprendizaje automático, son efectivos pero requieren grandes conjuntos de datos de entrenamiento y son difíciles de implementar en sistemas de monitoreo en tiempo real, debido a su costo computacional relativamente alto. Esta tesis se centra en dos problemas centrales vinculados a la caracterización de las crisis epilépticas con señales de EEG. El primero se relaciona con la detección de inicio y el otro se refiere al reconocimiento de patrones epileptiformes. Usando el nuevo método de caracterización presentado en el capítulo 2, ambos problemas pueden implementarse en tiempo real y lograr un alto rendimiento de detección. En general, esta tesis permitió aportar cinco nuevas contribuciones para tratar los problemas desafiantes de la epilepsia. Estas contribuciones se resumen a continuación y se relacionan con sus correspondientes referencias a nuestras publicaciones.

En el capítulo 2, se presenta la principal contribución de esta tesis: un nuevo método de caracterización de las crisis epilépticas en señales de EEG, basado en modelos estadísticos. El enfoque propuesto tiene varias ventajas interesantes. En primer lugar, permite la detección del inicio de la crisis en los diferentes ritmos cerebrales de forma independiente. En segundo lugar, el modelo propuesto se basa únicamente en 2 parámetros, lo que hace que su cálculo sea factible en tiempo real. En tercer lugar, permite el desarrollo de métodos de clasificación automática, los cuales pueden ser entrenados con conjuntos de datos razonablemente pequeños. Estas propiedades se demuestran a través de los métodos desarrollados en los capítulos que componen esta tesis. El método de caracterización propuesto se desarrolla en tres etapas. Primero, las señales de EEG se separan en cinco ritmos cerebrales diferentes mediante el uso de un banco de filtros usando la transformada *wavelet*. Cada señal de cada ritmo cerebral, se representa mediante un modelo estadístico Gaussiano generalizado, que mapea los datos del EEG en un espacio de baja dimensión de dos parámetros: escala y forma [1]. Finalmente, las pruebas estadísticas están diseñadas para demostrar que estos parámetros caracterizan correctamente las crisis epilépticas. Además, en el capítulo 2, se desarrolla nuestra segunda contribución, un desarrollo analítico de la divergencia de Kullback-Leibler para medir la discrepancia entre las distribuciones estadísticas de crisis y no-crisis en las señales epilépticas [2]. El Capítulo 3, presenta la tercera contribución que consiste en un nuevo algoritmo para detectar en señales EEG, el inicio de las crisis epilépticas junto con la estimación de su propagación. Se muestra que el parámetro de *escala* está estrechamente

relacionado con la variabilidad de la actividad cerebral y, por lo tanto, es un buen descriptor para realizar la detección y seguimiento de las crisis [3]. Basado en el modelo estadístico desarrollado en el capítulo 2, el capítulo 4 presenta la cuarta contribución, la cual consiste en desarrollar y comparar cuatro enfoques de clacificación con alta sensibilidad y especificidad. Este enfoque está basado en un modelo para detectar si las señales exhiben actividad cerebral normal o anormal en cada ritmo cerebral. Primero, se propone un clasificador de análisis discriminante lineal que se basa en estadística univariada de los datos del EEG [4, 5]. En segundo lugar, se generaliza este enfoque desarrollando un clasificador Bayesiano multivariado [6]. En tercer lugar, un clasificador de ensamble de conjuntos basado en la entropía de la distribución Gaussiana generalizada [7]. Finalmente, se desarrolla un método de clasificación de regresión logística basado en el mayor exponente de Lyapunov del análisis de componentes independientes (ICA) de cada ritmo cerebral [8]. Este último método se usa como referencia, ya que proporciona un excelente rendimiento pero con costos computacionales significativos. La quinta contribución, en el capítulo 5, está relacionada con en el reconocimiento de patrones epileptiformes con el propósito de detectar descargas espiga-onda (SWD) en señales EEG de larga-duración. Se desarrollan y se comparan tres métodos: un clasificador de vecinos más cercanos (kNN) basado en las distribuciones Gaussiana generalizada y t-location-scale [9, 10], y un clasificador de árbol de decisión basado en la correlación cruzada [11]. Se demostró que estos métodos logran una precisión muy alta en la detección del patrón SWD.

Abstract

Identifying epileptic brain activity using electroencephalography signals (EEG) in real-time is a difficult problem. Modern detection methods based on advanced machine learning techniques are effective but require large training datasets, and are difficult to implement in real-time monitoring systems because of their relatively high computational cost. This thesis focuses on two central problems linked to the characterization of epileptic seizures with EEG signals. The first one is related to onset detection and the other one is about epileptiform pattern recognition. Using the new characterization method presented in chapter 2, both can be implemented in real-time and achieve a high detection performance. In general, this thesis brings five new contributions to deal with challenging epilepsy problems. These contributions are summarized next, with references to our related publications.

Chapter 2 presents the main contribution of this thesis: a new statistical model-based characterization method of epileptic seizures in EEG signals. The proposed approach has several interesting advantages. First, it allows the detection of seizure onset in the different brain rhythms, independently. Second, the proposed model relies on 2 parameters only, making its computation feasible in real-time. Third, it allows developing automatic classification methods trainable with reasonably small datasets. These properties are demonstrated through the methods developed in the subsequent chapters of this thesis. The proposed characterization method proceeds through three stages. First, EEG signals are separated into five different brain rhythms by using a wavelet filter bank. Each brain rhythm signal is then represented using a generalized Gaussian statistical model that maps the EEG data to a low-dimensional space of two parameters: scale and shape [1]. Finally, statistical tests are designed to show that these parameters characterize correctly epileptic seizures. In addition, chapter 2 develops our second contribution which is an analytical development of Kullback-Leibler divergence to measure the discrepancy between the statistical distributions of seizure and non-seizure epileptic signals [2], confirming the ability of our model to characterize seizures. Chapter 3 presents the third contribution consisting in a new algorithm for epilepsy seizure onset detection and spread estimation from EEG signals, where the *scale* parameter is shown to be closely related to the variability of the brain activity and makes, therefore, a good descriptor for performing seizure onset detection and tracking [3]. Based on the statistical model developed in chapter 2, chapter 4 presents our fourth contribution. This consists of the development and comparison of four model-based classification approaches to detect whether the signals exhibit normal or abnormal brain activity in each brain rhythm, with high sensitivity and specificity. First, we propose a linear discriminant analysis classifier which relies on univariate statistics of the EEG

data [4, 5]. Second, we generalize this approach by developing a multivariate Bayesian classifier [6]. Third, an ensemble bagging classifier based on the entropy of the generalized Gaussian distribution [7]. Finally, we develop a logistic regression classifier method based on the largest Lyapunov exponent from the independent component analysis of each brain rhythm [8]. The latest method is used as a reference since it provides excellent performance but with significant computational costs. The fifth contribution, in chapter 5, is related to epileptiform pattern recognition with the purpose of detecting spike-and-wave discharges (SWD) in EEG long-time signals. Three methods are developed and compared: k-nearest neighbors classifier based on the generalized Gaussian and t-location-scale distributions [9, 10], and a decision tree method based on the cross-correlation [11]. We show that these methods achieve very high accuracy in detecting SWD pattern.

Contents

List of Figures xv

List of Tables xxi

Symbols and Abbreviations xxv

1 EEG and Epilepsy 1

1.1 Introduction . . . 1
1.2 Electroencephalography . . . 1
1.3 Brain rhythms . . . 3
1.4 Epilepsy . . . 4
1.5 Seizure classification . . . 6
1.6 Seizure onset detection . . . 9
1.7 Conclusions . . . 12

2 Statistical-Model-based EEG signal characterization 13

2.1 Introduction . . . 13
2.2 Data set . . . 13
2.3 Data organization . . . 14
2.4 Wavelets and Filter banks . . . 15
2.4.1 Multilevel 1D wavelet decomposition . . . 16
2.4.2 Multilevel 2D wavelet decomposition . . . 16
2.4.3 Extraction of brain rhythms . . . 16
2.5 Statistical models . . . 18
2.5.1 Generalized Gaussian distribution . . . 19
2.5.2 Logistic distribution . . . 19
2.5.3 t-location-scale distribution . . . 20
2.5.4 Alpha-stable Distribution . . . 20
2.5.5 Cauchy distribution . . . 21
2.6 Goodness-of-fit test . . . 21
2.6.1 Distribution-fitting . . . 21
2.6.2 Kolmogorov-Smirnov (KS) . . . 24

2.6.3 Cramer-von Mises criterion (CvM) . . . 24
2.6.4 Goodness-of-fit test results . . . 25
2.7 Generalized Gaussian best distribution-fitting . . . 26
2.8 Characterization with the generalized Gaussian distribution parameters . . . 27
2.8.1 Estimation of σ and τ . . . 27
2.8.2 Model based characterization . . . 28
2.9 Kullback-Leibler divergence (KLD) . . . 31
2.9.1 A visual EEG epilepsy detection experience . . . 32
2.10 Conclusions . . . 34

3 Seizure onset detection and temporal spread estimation 35
3.1 Introduction . . . 35
3.2 Scale parameter sigma (σ) . . . 35
3.3 Algorithm . . . 38
3.4 Experimental Results . . . 39
3.4.1 General Spread . . . 39
3.4.2 Delta Band Spread . . . 39
3.4.3 Theta Band Spread . . . 39
3.4.4 Alpha Band Spread . . . 40
3.4.5 Beta Band Spread . . . 41
3.4.6 Gamma Band Spread . . . 41
3.5 Validation . . . 41
3.6 Conclusions . . . 48

4 Model-based classification for seizure onset detection 49
4.1 Introduction . . . 49
4.2 Seizure onset detection (SOD) classifiers . . . 50
4.3 Cross-validation and leave-one-out cross-validation . . . 53
4.4 Methodology . . . 53
4.4.1 Problem statement . . . 53
4.4.2 Proposed method . . . 54
4.4.3 Spectral decomposition by wavelet filter bank . . . 54
4.4.4 Statistical model of the spectral components . . . 54
4.5 Linear discriminant classifier . . . 55
4.5.1 Classifier parameters . . . 55
4.5.2 SOD by linear discriminant analysis classification . . . 56
4.5.3 Experimental results . . . 57
4.5.4 Pearson's product moment correlation . . . 59
4.6 Multivariate Bayesian Classifier . . . 60
4.6.1 Classifier parameters . . . 60
4.6.2 SOD by Bayesian Classifier . . . 61

4.6.3 Experimental results . . . 62
4.7 Ensemble Bagging Classifier . . . 69
4.7.1 Classifier Parameters . . . 69
4.7.2 SOD by ensemble of bagged decision trees classification . . . 70
4.7.3 Experimental results . . . 71
4.8 Logistic Regression Classifier . . . 74
4.8.1 Classifier parameters . . . 75
4.8.2 Largest Lyapunov exponent (LLE) . . . 77
4.8.3 λ-Scaling . . . 78
4.8.4 SOD by logistic regression classification . . . 78
4.8.5 Experimental results . . . 79
4.9 Conclusions . . . 81

5 Spike-and-wave epileptiform pattern recognition 83
5.1 Introduction . . . 83
5.2 Spike-and-Wave discharge (SWD) . . . 83
5.3 Morlet Wavelet . . . 85
5.4 Database . . . 86
5.5 Spike-and-wave detection using the generalized Gaussian distribution . . . 87
5.5.1 Methodology . . . 87
5.5.2 Spike-and-wave detection by *k*-nearest neighbors classifier . . . 88
5.5.3 Experimental Results . . . 89
5.6 Spike-and-wave detection using t-location-scale distribution . . . 90
5.6.1 Methodology . . . 91
5.6.2 Experimental Results . . . 92
5.7 Spike-and-wave detection using cross-correlation and decision-trees . . . 95
5.7.1 Methodology . . . 96
5.7.2 Experimental results . . . 98
5.8 Conclusions . . . 99

6 Conclusions, contributions, advantages, limitations and research perspectives 101
6.1 Conclusions and contributions . . . 101
6.2 Advantages and limitations . . . 105
6.3 Research perspectives . . . 106

A Estimation and regularization of source localization using EEG 107
A.1 Dipole assumption . . . 107
A.2 Lead-Field . . . 108
A.3 Head models . . . 109
A.3.1 Homogeneous sphere head volume conductor model . . . 110

A.3.2 Boundary element model . . . 111
A.3.3 Finite element model . . . 111
A.3.4 Dipole fitting model . . . 112
A.4 Inverse problem . . . 112
A.5 Source localization approaches in EEG inverse problem . . . 113
A.5.1 Equivalent current dipole approach . . . 114
A.5.2 Linear current distributed approach . . . 114
A.6 Current distributed-source estimation . . . 115
A.6.1 Minimum norm estimation . . . 115
A.6.2 Weighted minimum-norm . . . 116
A.6.3 LORETA . . . 117
A.6.4 FOCUSS . . . 118
A.6.5 sLORETA . . . 120
A.6.6 LPISS . . . 121
A.6.7 Bayesian approach . . . 122
A.6.8 Spatial estimation . . . 123
A.6.9 Spatiotemporal estimation . . . 124
A.7 Shrinking methods . . . 125
A.7.1 Shrinking LORETA-FOCUSS . . . 125
A.7.2 Standardized shrinking LORETA-FOCUSS (SSLOFO) . . . 126
A.7.3 Adaptive standardized LORETA-FOCUSS (ALF) . . . 126
A.8 Regularization . . . 127
A.8.1 ℓ_1 and ℓ_2 Regularization . . . 128
A.8.2 $1 < p < 2$ regularization . . . 132

Bibliography **162**

List of Figures

1.1 10-20 system: electrode placement method, the electrodes positions have a nomenclature according to the different functional regions of the cortex (temporal lobe, parietal lobe, occipital lobe, frontal lobe), which are responsible for motor control, cognitive and memory functions. A = Ear lobe, C = central, P = parietal, F = frontal, Fp = frontal polar, O = occipital. Both parts of the figure show the standard bipolar electrodes method between the front (F) and back (B) of the head and between the left (L) and right (R) hemispheres respectively 3

1.2 Examples of different artifacts: bruxism, tongue, arms and legs movement, exhibit a similar alteration in all channels while swallowing and cell phone have a high pronounced peak; lamp artifact present spikes that can be confused with epilepsy abnormal waveform by an inexpert eye. The examples shown correspond to one-second segments in duration. 4

1.3 Different interictal abnormal activity: a) spike waveforms using a monopolar configuration whose reference is the average of all channels (Avg). b) sharp-wave using bipolar configuration, note the phase reversal in FP1-FT7, FT7-F7 channels vs. FT9-T3, T3-T5 channels, this indicating that channel F7 is a seizure candidate; c) spike-and-wave using monopolar configuration whose reference are A1 or A2 channels, the spike-and-wave waveform is present in almost all channels. 7

1.4 Different ictal abnormal activity: polyspike, polyspike wave and polyspike, respectively. 8

2.1 db4 wavelet waveform, we can see the similarity with a seizure waveforms in Figures 1.3 and 1.4. 17

2.2 Example of good fit of generalized Gaussian distribution (GGD) statistical model for the Delta Band; we can observe the different data distribution-fitting and how the GGD is the best data-fit among all distributions considered. 22

2.3 Example of good fit of generalized Gaussian distribution (GGD) statistical model for the Theta Band; we can observe the different data distribution-fitting and how the GGD is the best data-fit among all distributions considered. 22

2.4 Example of good fit of generalized Gaussian distribution (GGD) statistical model for the Alpha Band; we can observe the different data distribution-fitting and how the GGD is the best data-fit among all distributions considered. 23

2.5 Example of good fit of generalized Gaussian distribution (GGD) statistical model for the Beta Band; we can observe the different data distribution-fitting and how the GGD is the best data-fit among all distributions considered. 23
2.6 Example of good fit of generalized Gaussian distribution (GGD) statistical model for the Gamma Band; we can observe the different data distribution-fitting and how the GGD is the best data-fit among all distributions considered. 24
2.7 KLD between the PDFs of the sliding window and the PDF of the *seizure* onset of the epileptic signal, showing a clear discrepancy between *seizure* or *non-seizure*. In this example, the *seizure* onset begins at minute 2, and its duration is 40 seconds. 33
2.8 KLD between adjacent PDFs coupled with a 7-order one-dimensional median filter, showing clearly the discrepancy given by the highest peak which emerges from the background of EEG. In this example, the *seizure* onset begins at minute 2, and its duration is 40 seconds. 34

3.1 Areas of the brain of the Cerebral Cortex (Lateral view) can be correlated with the electrodes positions, which have a nomenclature according to the lobules, namely (T) temporal lobe, (P) parietal lobe, (O) occipital lobe and (F) frontal lobe. . . . 38
3.2 General Spread: EEG electrodes array, suggests that the seizure starts in channel T8 and then the spread across the channels F8-T8-P8, F4-C4-P4. Amplitude (y-axis) in mV and time (x-axis) in min. 40
3.3 Delta Band Spread: EEG electrodes array, suggests delta activity spread across the right hemisphere and isolated activity in the left hemisphere. Amplitude (y-axis) in mV and time (x-axis) in min. 42
3.4 Theta Band Spread: EEG electrodes array, suggests theta activity spread across both hemispheres on temporal area of the cerebral cortex. Amplitude (y-axis) in mV and time (x-axis) in min. 43
3.5 Alpha Bands Spread: EEG electrodes array, suggests alpha activity spread across both hemispheres of the cerebral cortex, predominating the right side. Amplitude (y-axis) in mV and time (x-axis) in min. 44
3.6 Beta Band Spread: EEG electrodes array, suggests beta activity spread across both hemispheres of the cerebral cortex. Amplitude (y-axis) in mV and time (x-axis) in min. 45
3.7 Gamma Band Spread: EEG electrodes array, suggests gamma activity spread across both hemispheres of the cerebral cortex. Amplitude (y-axis) in mV and time (x-axis) in min. 46
3.8 For illustration, the figure shows the EEG of a patient (left) with his corresponding MRI (right top). A reconstruction of the brain is shown with the corresponding seizure spread (right bottom). On the right of each image, there is a color map which indicates the intensity of each activation source. 48

4.1 Algorithm used in linear classifier. 55

4.2 Scatter plots for the statistical parameters σ and τ for *seizure* signals (red crosses) and *non-seizure* signals (blue circles) for each spectral band, showing the good discrimination properties of the proposed representation. 57

4.3 Algorithm used in Bayesian classifier. The coefficients vector $\boldsymbol{C}_j^{(i)}$ and the vector of parameters estimated $\boldsymbol{\theta}_j^{(i)}$, from different brain rhythms are composed by the details H, V, and D, which refer to horizontal, vertical and diagonal respectively, see Section 2.4. 61

4.4 Scatter plots for the generalized Gaussian parameters σ and τ for seizure events (red stars) and non-seizure events (blue squares for before the seizure and black diamonds for after seizure) observed through the *Delta* frequency band, showing the good linear discrimination power of the proposed approach. 64

4.5 Scatter plots for the generalized Gaussian parameters σ and τ for seizure events (red stars) and non-seizure events (blue squares for before the seizure and black diamonds for after seizure) observed through the *Theta* frequency band, showing the good linear discrimination power of the proposed approach. 65

4.6 Scatter plots for the generalized Gaussian parameters σ and τ for seizure events (red stars) and non-seizure events (blue squares for before the seizure and black diamonds for after seizure) observed through the *Alpha* frequency band, showing the good linear discrimination power of the proposed approach. 66

4.7 Scatter plots for the generalized Gaussian parameters σ and τ for seizure events (red stars) and non-seizure events (blue squares for before the seizure and black diamonds for after seizure) observed through the *Beta* frequency band, showing the good linear discrimination power of the proposed approach. 67

4.8 Scatter plots for the generalized Gaussian parameters σ and τ for seizure events (red stars) and non-seizure events (blue squares for before the seizure and black diamonds for after seizure) observed through the *Gamma* frequency band, showing the good linear discrimination power of the proposed approach. 68

4.9 Algorithm used in bagging ensemble classifier. 71

4.10 Scatter plots from vector $\boldsymbol{p} = [\sigma, \tau, \varepsilon]$ observed through all brain rhythms using 105 events: 35 *seizures* (yellow dots) and 70 *non-seizures* (blue dots). We can see how the *seizure* event concentrates on high values of σ and ϵ. 72

4.11 Box plots of Shannon entropy observed through all brain rhythms using 105 events (35 *seizures* and 70 *non-seizures*). The maximum an minimum values for each box together with the quartiles can help to classify based on a thresholding approach. 73

4.12 Algorithm used in logistic regression classifier. 75

4.13 Example for seven electrodes, namely $\boldsymbol{X}_1, ..., \boldsymbol{X}_7$ and four sources $\boldsymbol{S}_t$ namely $\boldsymbol{s}_1, .., \boldsymbol{s}_4$, representation of assumption that the source signals arrive at the electrodes at the same time instantaneously. 76

4.14 Scatter plot for six largest Lyapunov exponents (LLE) for *seizure* (middle) and *non-seizure* before an after events observed through the different rhythms bands. Lowest valued LLE is in the *seizure* events for delta, theta, and alpha bands, while highest valued LLE are in the *seizure* events for beta and gamma bands. In this example, the start and end of the seizure in this EEG signal were labeled by the neurologist using two lines. The first line divides the EEG signal before the seizure and the second line after the seizure. 80

5.1 EEG signals extraction. A physician may determine the patient treatment according to the EEG waveforms; while in a biomedical signal approach the most relevant features employed to classify epileptogenic abnormality can be categorized in terms of spectral properties, signal morphology, and statistical measures. 84

5.2 Spike-and-wave and Morlet wavelet respectively, Note that the scales are different, but for illustration, we can see the symmetric and regular morphology in both signals. 85

5.3 Example of 6 channels of one monopolar EEG raw, we can see different SWD in all channels. 85

5.4 Electrodes position used with this database. 86

5.5 Scatter plots of the off-line training classification in database signals, for σ, ν and $\widetilde{x}$ parameters for spike-and-waves events (SWD = class 1 = red dots) and non-spike-and-waves events (non-SWD = class 0 = blue dots), showing the data dispersion of the proposed approach. In (a) Scale parameter (σ) vs variance (ν). For class 1 or SWD, we can see the direct relationship between the variance and sigma, both grow proportionally, while for class 0 or non-SWD both sigma and variance remain in a range of values. (b) Scale parameter (σ) vs median ($\widetilde{x}$). As sigma grows, median increases and decreases for both SWD and non-SWD, but is larger for SWD. (c) variance (ν) vs median ($\widetilde{x}$). As variance grows, median increases and decreases for SWD, while for non-SWD, remains in a small range. 90

5.6 Scatter plots of the off-line training classification in 192 dataset signals, for the t-location-scale parameters μ, σ and ν for spike-and-waves events (blue dots) and non-spike-and-waves events (red dots), showing the data dispersion of the proposed approach. In a) and c) spike-and-waves tend to have a higher scale σ, in b) non-spikes-and-waves tend to have a location μ concentrated between 0 and 100. . . 93

5.7 Scatter plot in off-line training classification in 192 dataset signals for the t-location-scale parameters μ, σ and ν, we can see the correct discrimination between two groups whose size is the same (96 spike-and-waves and 96 non-spikes-and-waves), label 1 for spike-and-wave and label 0 for non-spike-and-wave. 93

5.8 Scatter plots in on-line classification in 46 test signals, for the t-location-scale parameters μ, σ and ν for spike-and-waves events (blue dots) and non-spike-and-waves events (red dots), showing the data dispersion of the proposed approach. In a) spike-and-waves tend towards the center down, in b) the trend is not very clear, although there is a great concentration of spike-and-waves in the center down near zero, and in c) spike-and-waves tend to be located towards the right and near zero. 94

5.9 Scatter plot in on-line classification in 46 test signals for the t-location-scale parameters μ, σ and ν, we can see the correct discrimination between two groups whose size is different (spike-and-waves labeled by an expert neurologist and non-spikes-and-waves), label 1 for spike-and-wave and label 0 for non-spike-and-wave. 94

5.10 A decision tree example. Consider the decision problem as to whether or not to go ahead with a cross-correlation similarity. If we go ahead with the similarity and meets the threshold (0.6), then we have a spike-and-wave candidate; on the other hand, if we don't go ahead with the similarity (0.4) then the threshold is not met and therefore we don't have a spike-and-wave candidate. Note that, this tree has only two regions given by the similarity threshold for SWD or non-SWD. 95

5.11 Scatter plot example between all 23 channels (x axis) and the total coincidences (y axis) into spike-and-wave signals from database, the line in 40 is the threshold used. 98

5.12 Receiver operating characteristic curve (ROC) in 874 predictors 99

A.1 The magnetic field $\boldsymbol{B}$ at each electrode is computed with respect to the dipole(s) moment(s) and the distance between the center of the dipole volume and the electrode. 108

A.2 Forward and Inverse Problems 111

List of Tables

1.1 Brain rhythms 5

1.2 Abnormal activity and associated EEG morphologic waveforms 8

1.3 Classification of simple partial seizures 9

1.4 Classification of generalized seizures 10

2.1 Length of the 18 seizures used in this study and the corresponding number of overlapping 1-second segments using a rectangular sliding window of 2 seconds. An offset has been used for each epoch to avoid leading and trailing signals that were noisy. Consequently, the number of windows is irregular between epochs. . . 14

2.2 Frequencies of the different scales of the multilevel 2D wavelet decomposition. . . 18

2.3 Statistical parameters of each distribution under consideration. 18

2.4 Means of the Kolmogorov-Smirnov (KS) and the Cramer-von-Mises criterion (CvM) scores obtained for GGD pdfs estimated with all EEG segments of 54 events used, 18 *seizures* and 36 *non-seizures*. The GGD shows the lowest scores with respect to the other distributions considered. 25

2.5 Standard deviations of the Kolmogorov-Smirnov (KS) and the Cramer-von-Mises criterion (CvM) scores obtained for GGD pdfs estimated with all EEG segments of 54 events used, 18 *seizures* and 36 *non-seizures*. The GGD shows the lowest scores with respect to the other distributions considered. 26

2.6 Decision rules to asses the statistical significance of the difference of means of the GGD parameters for seizure and non-seizure signals. 30

2.7 Results of the t-tests to assess the ability of σ and τ to discriminate separately *seizure* and *non-seizure* EEG. The H_0^σ hypothesis is rejected for all rhythms, with highly statistical significant ($p < 0.01$). These scores are supported by very high Bayes factors. Contrarily, H_0^τ is accepted for all rhythms, except Delta band. The associated p-values are largely greater than 0.1 with Bayes factors lower than the evidence threshold. We conclude that the scale parameter σ is a marker to discriminate *seizure* and *non-seizure* events with a high statistical significance. The shape parameter τ is not a marker to discriminate *seizure* and *non-seizure* events. 31

3.1 Minimum value, maximum value and mean of σ parameter for all signals in 36 non-seizure events and 18 seizure events. The mean value allows using a thresholding approach with the scale parameter. 36

3.2 Minimum value, maximum value and mean of σ parameter for each brain rhythm in 36 non-seizure events and 36 seizure events. The mean value allows using a thresholding approach with the scale parameter. 36

3.3 Onset delay (in seconds) by frequency bands, the symbol (-) means that there is no clear difference at the beginning of the onset. In this example, according to the data, channels (21) FT9-FT10, (23) T8-P8, (13) FP2-F8, (14) F8-T8 and (15) T8-P8 are the best candidates by brain area for SOD because it has minor delays; in this case T8 is common to all channels in this brain area, therefore is the best option for being the onset and other channels form the spread. 37

4.1 Classification methods developed independently for each brain rhythm. The scale (σ) and shape (τ) correspond to the generalized Gaussian distribution parameters, ν is the variance, ϵ is the entropy from the generalized Gaussian, λ are the largest Lyapunov exponents from the Analysis of Independent Components (ICA), and ς are the λ-scaling between their $\pm$ standard deviation. 49

4.2 State-of-the-art methods to perform seizure detection automatically in EEG signals, summarized in terms of the classification techniques and features used and their reported performance on a test dataset. The performance metrics are the True Positives Rate or Sensitivity (TPR), the True Negative Rate or Specificity (TNR). 52

4.3 Seizure detection performance by using linear discriminant analysis classification for each brain rhythm and for 39 events (13 *seizure* and 26 *non-seizure*) of the Children's Hospital Boston database, in terms of: TPR = True Positives Rate or Sensitivity; TNR = True Negative Rate or Specificity; FPR = False positive Rate; ACC = Accuracy; and [$\pm$ standard deviation]. 59

4.4 Average latency between seizure onset and detection (in seconds), for the proposed method on each spectral band, and for the state-of-the-art methods. [12, 13, 14]. 59

4.5 Pearson's product moment correlation coefficient comparison between ω_s for *seizure* and ω_{ns} for *non-seizure* events, over the proposed epilepsy classification model for each brain rhythm; where $r = 1$ is total positive correlation, $r = 0$ is no correlation, $r = -1$ is total negative correlation and $IC95\%$ is the 95 percent confidence interval. The model presents high correlation for all brain rhythms, except for gamma band in *non-seizure* events. This suggests that Pearson's product moment correlation coefficient can be used to estimate changes between *seizure* and *non-seizure* events. 60

4.6 Seizure detection performance for each independent contribution between before the seizure and the seizure by using multivariate Bayesian classification for each brain rhythm and for 36 events (18 *seizure* and 18 *non-seizure*) of the Children's Hospital Boston database, in terms of: TPR = True Positives Rate or Sensitivity; TNR = True Negative Rate or Specificity; FPR = False positive Rate; and ACC = Accuracy [± standard deviation]. δ: Delta band, θH: Theta horizontal band, θV: Theta vertical band, θD: Theta diagonal band, αH: Alpha horizontal band, αV: Alpha vertical band, αD: Alpha diagonal band, βH: Beta horizontal band, βV: Beta vertical band, βD: Beta diagonal band, γH: Gamma horizontal band, γV: Gamma vertical band, γD: Gamma diagonal band. To simplify the visual interpretation we highlight in red the metric that achieves the highest sensitivity, specificity, and overall accuracy for each frequency band 63

4.7 Seizure detection performance for each independent contribution between the seizure and after the seizure by using multivariate Bayesian classification for each brain rhythm and for 36 events (18 *seizure* and 18 *non-seizure*) of the Children's Hospital Boston database, in terms of: TPR = True Positives Rate or Sensitivity; TNR = True Negative Rate or Specificity; FPR = False positive Rate; and ACC = Accuracy [± standard deviation]. δ: Delta band, θH: Theta horizontal band, θV: Theta vertical band, θD: Theta diagonal band, αH: Alpha horizontal band, αV: Alpha vertical band, αD: Alpha diagonal band, βH: Beta horizontal band, βV: Beta vertical band, βD: Beta diagonal band, γH: Gamma horizontal band, γV: Gamma vertical band, γD: Gamma diagonal band. To simplify the visual interpretation we highlight in red the metric that achieves the highest sensitivity, specificity, and overall accuracy for each frequency band 63

4.8 Seizure detection performance by using multivariate Bayesian classification for each brain rhythm and for 36 events (18 *seizure* and 18 *non-seizure*) of the Children's Hospital Boston database, in terms of: TPR = True Positives Rate or Sensitivity; TNR = True Negative Rate or Specificity; FPR = False positive Rate; and ACC = Accuracy [± standard deviation]. 64

4.9 Average latency between seizure onset and detection (in seconds), for the proposed method on each spectral band, and for the state-of-the-art methods. [12, 13, 14]. 64

4.10 Comparison between *means*, *standard deviations* of the entropy and 95% confidence interval (*IC95%*) of seizure and non-seizure, using 105 events (35 seizures and 70 non-seizures) for each brain rhythm. We can see how one can set a threshold for detecting the seizure 73

4.11 Seizure detection performance by using ensemble bagged classification for each brain rhythm and for 105 events (35 *seizure* and 70 *non-seizure*) of the Children's Hospital Boston database, in terms of: TPR = True Positives Rate or Sensitivity; TNR = True Negative Rate or Specificity; FPR = False positive Rate; and ACC = Accuracy [± standard deviation]. 73

4.12 Average latency between seizure onset and detection (in seconds), for the proposed method on each spectral band, and for the state-of-the-art methods. [12, 13, 14]. 73
4.13 Minimum, maximum and standard deviation from all LLE for each brain rhythm. . 81

5.1 Methods, features, classifiers, signals for training and testing considered in this chapter. The generalized Gaussian distribution uses the feature vector $[\sigma, \nu, \widetilde{x}]$, where σ is the scale parameter of the generalized Gaussian distribution, ν is the variance parameter from the generalized Gaussian distribution parameters, and $\widetilde{x}$ is the median parameter from the wavelet Morlet coefficients $\boldsymbol{C}^{(i)}$; the t-location-scale distribution uses the feature vector $[\mu, \sigma, \nu]$, which corresponds to the parameters of this distribution, namely location (μ), scale (σ) and shape (ν); and the cross-correlation measure uses the similarity feature vector r. 87
5.2 Range of values for sigma (σ), variance (ν) and median ($\widetilde{x}$) parameters for class 0 or non-spike-and-wave and for class 1 or spike-and-wave. 90

Symbols and Abbreviations

Letters

Symbol	Description
f	Frequency

Greek letters

Symbol	Description	Unidades
δ	Delta Band $f \leq 4$ Hz	
θ	Theta Band $4 \leq f \leq 7$ Hz	
α	Alpha Band $8 \leq f \leq 12$ Hz	
β	Beta Band $13 \leq f \leq 29$ Hz	
γ	Gamma Band $30 \leq f$ Hz	
σ	Scale parameter from GGD	
τ	Shape parameter from GGD	

Abbreviations

Abbreviation	Description
ACC	**Acc**uracy
ALF	**A**daptive Standardized **L**ORETA-**F**OCUSS
Bagging	**B**ootstrap **Agg**regat**ing**
BF	**B**ayes **F**actor
BF_t	**B**ayes **F**actor **t**hreshold
BEM	**B**oundary **E**lement **M**ethod
CCD	**C**ortical **C**urrent **D**ensity
CDR	**C**urrent **D**ensity **R**econstruction
CvM	**C**ramer-von **M**ises criterion

Abbreviation	Description
DFM	**D**ipole **F**itting **M**ethod
KLD	**K**ullback-**L**eibler **D**ivergence
KS	**K**olmogorov-**S**mirnov
ECD	**E**quivalent **C**urrent **D**ipole
ECDF	**E**mpirical **C**umulative **D**istribution **F**unction
ECoG	**E**lectro**Co**rtico**G**raphy
EEG	**E**lectro**E**cephalo**G**raphy
ERP	**E**vent-**R**elated **P**otencials
FEM	**F**inite **E**lement **M**ethod
FLENI	**F**undación **L**ucha contra las **E**nfermedades **N**eurológicas **I**nfantiles
fMRI	**f**unctional **M**agnetic **R**esonance **I**mage
FNR	**F**alse **N**egative **R**ate
FOCUSS	**Foc**al **U**nderdetermined **S**ystem **S**olver
FPR	**F**alse **P**ositive **R**ate
GGD	**G**eneralized **G**aussian **D**istribution
IC	**I**nterval **C**onfidence
ICA	**I**ndependent **C**omponent **A**nalysis
iEEG	**i**ntracranial **E**lectro**E**cephalo**G**raphy
IT	**I**nformation **T**echnology
ITBA	**I**nstituto **T**ecnológico de **B**uenos **A**ires
k-NN	**k**-**N**earest **N**eighbor
ILAE	**I**nternational **L**eague **A**gaints **E**pilepsy
LLE	**L**argest **L**yapunov **E**xponents
LS	**L**east-**S**quare
LORETA	**L**ow-**R**esolution **E**lectromagnetic **T**omography
LPISS	ℓ_p-norm **I**terative **S**parse **S**olution
p	**p**-value
MCMC	**M**arkov **C**hain **M**onte **C**arlo
MEG	**M**agneto **E**cephalo**G**raphy
MLE	**M**aximum **L**ikelihood **E**stimator
MNE	**M**inimun **N**orm **E**stimation
MNI	**M**ontreal **N**eurological **I**nstitute
MNLS	**M**inimun-**N**orm **L**east **S**quares
MRI	**M**agnetic **R**esonance **I**mage
N_E	**N**umber of **E**lectrodes
N_V	**N**umber of **V**oxels
PDF	**P**robability **D**ensity **F**unction
sEEG	**S**tereo**E**lectro**E**ncephalo**G**raphy

Abbreviation	Description
sLORETA	**S**tandardized **L**ow-**R**esolution **E**lectromagnetic **T**omography
SCA	**S**parse **C**omponent **A**nalysis
SNR	**S**ignal-to-**N**oise **R**atio
SOD	**S**eizure **O**nset **D**etection
SSLOFO	**S**tandardized **S**hrinking **LO**RETA-**FO**CUSS
SVD	**S**ingular **V**alue **D**ecomposition
SWD	**S**pike-and-**W**ave **D**ischarge
T_t	**T**-test threshold
T_ν	**T**-test score
TNR	**T**rue **N**egative **R**ate
TPR	**T**rue **P**ositive **R**ate
TPS	**T**hin **P**late **S**pline
WMN	**W**eighted **M**inimun-**N**orm

Chapter 1
EEG and Epilepsy

1.1 Introduction

Electroencephalography (EEG) is a non-invasive and widely available biomedical modality that is used to diagnose epilepsy and plan treatment. Neurologists trained in EEG are able to determine a correct diagnostics of epilepsy. They identify visually its onset and presence through the analysis of characteristic waveforms, known as spikes, associated with epileptic seizures, which include: mode of onset and termination, clinical manifestations, and abnormally enhanced synchrony. A spike is characterized by short bursts of high amplitude, synchronized and multi-phasic activity, in which polarity changes occur in different times, which manifest themselves at or around the epileptic focus and stand out from the background EEG. The spikes are associated with seizures through different brain areas and to the relationship with the brain rhythms. These waveforms are usually depicted according to their morphology (e.g. amplitude, duration, sharpness, and emergence from the background) to provide more insight into the epilepsy phenomena behind EEG measurements. This first chapter discusses the medical aspects of seizures, classification and onset detection in epilepsy, as well as brain rhythms.

1.2 Electroencephalography

The German psychiatrist Hans Berger made the first recording of the electric field of the human brain in 1924 [15]. This recording or *electroencephalogram* can measure brain activity in two ways: noninvasively through the scalp with an amplitude of approximately 100 μV or invasively on the surface of the brain with an amplitude of 1 to 2 mV. EEG is the result of the sum of the action potentials derived from the mixture of streams generated by extracellular populations of neurons. Therefore, EEG depends basically on the cytoarchitecture of neuronal populations, their connectivity and the geometry of their extracellular fields. The main physical sources of the potential scalp are the pyramidal cells of the cortical layers III and V [16]. This modality, improved over a century still remains the most widely used method in neurological and psychological laboratories.

EEG is used primarily for three types of studies [17]:

1. Brain activity.
2. Event-Related Potentials (ERP), which are EEG components that emerge in response to a stimulus (e.g. electric, auditory, visual). Such signals are usually below the noise level, therefore not easily distinguished, and require the use of stimuli and signal time averaging to improve the SNR.
3. Bioelectric events produced by single neurons. The behavior of a single neuron can be examined using microelectrodes that traverse the specific cells of interest. The study of a single cell or cell networks allows the construction of models that reflect the actual properties of the tissue.

The standardized international 10-20 system is generally used to record the EEG activity. This system has 21 electrodes located symmetrically on the surface of the scalp; these positions are computed as percentages of standard distances, the resulting records are comparable between different patients, see Figure 1.1. EEG electrode positions are determined as follows: the reference points are the nasion, which is the delve at the top of the nose, at the level of the eyes; and the inion, which is the bony lump at the base of the skull on the midline at the back of the head. From these points and once the central point (Cz) is localized, the skull perimeters are measured in the transverse and median planes. Electrode locations are determined by dividing these perimeters into 10% and 20% intervals, see Figure 1.1 and Figure 3.1. Additionally, the EEG measurement provides temporal and spatial information about the synchronous firing of many neurons inside the brain with a dominant frequency according to the brain rhythms (see section 1.3), namely delta (δ) with (f $\leq$ 4 Hz), theta (θ) with (4 Hz $\leq$ f $\leq$ 7 Hz), alpha (α) with (8 Hz $\leq$ f $\leq$ 12 Hz), beta (β) with (13 Hz $\leq$ f $\leq$ 29 Hz) and gamma (γ) with (30 Hz $\leq$ f), see Table 1.1 for more details. EEG measurement can use an *unipolar* electrodes configuration, where the potential of each electrode is compared either to a neutral electrode or to the average of all electrodes; or *bipolar* electrodes configuration, where the potential difference between a pair of electrodes spatially close is measured [17].

In EEG, an artifact is defined as an electrical potential that has originated outside of the brain; there are two basic artifact types, 1) physiological artifacts generated from the electrical activity associated with the normal functioning of the body of the patient (e.g. movement and blinking of the eyes, respiration, chewing, bruxism, swallowing, tongue movement, skin potentials, body tremor, cardiac activity, muscle activity, sweat glands, pulse in the tissues, and artificial cardiac pacemaker); 2) Non-physiological artifacts generated by electromagnetic fields outside the body or by technical problems (e.g. poor signal characteristics given by bad signal recording, line frequency 50/60 Hz, electrodes, the different types of medical equipment, cell phones, lights, and the environmental movement). Figure 1.2 shows some examples of artifacts, for more details see [18]. Noises are as important as artifacts. Acquiring EEG signal properly means mainly safety, biosignal measurement with higher signal to noise ratio (SNR) and no data loss, as much as possible. The system electronics include the circuitry and printed circuit board design, the

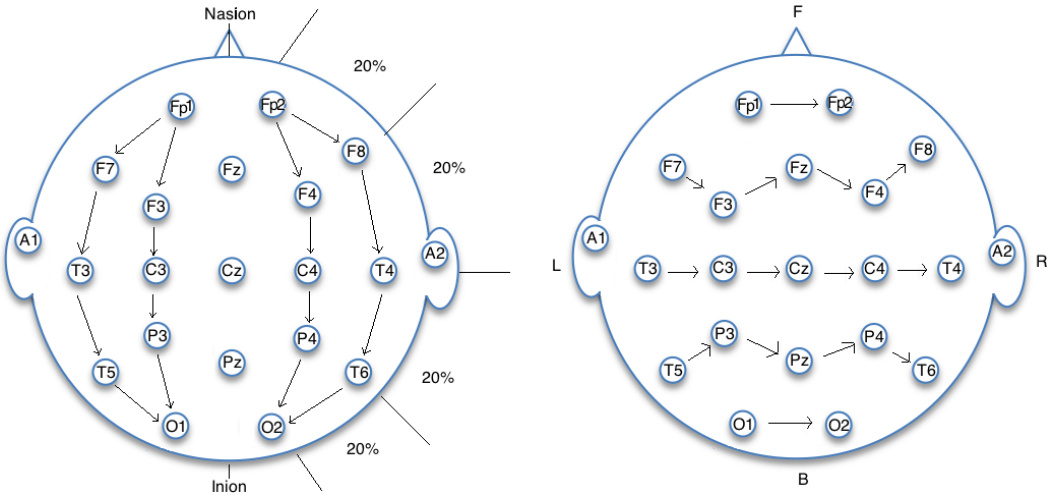


FIGURE 1.1: 10-20 system: electrode placement method, the electrodes positions have a nomenclature according to the different functional regions of the cortex (temporal lobe, parietal lobe, occipital lobe, frontal lobe), which are responsible for motor control, cognitive and memory functions. A = Ear lobe, C = central, P = parietal, F = frontal, Fp = frontal polar, O = occipital. Both parts of the figure show the standard bipolar electrodes method between the front (F) and back (B) of the head and between the left (L) and right (R) hemispheres respectively

filtering stages, electronic amplifier's noise control, correct signal conversion, data storing, contact resistance skin-electrodes, and background noise [19, 20, 21].

1.3 Brain rhythms

EEG is the predominant modality to study abnormal cerebral activity due to its low cost, small space requirements, very-high time resolution, medium space resolution and its tolerance to subject movement [22, 23]. EEG enables us to glimpse the generalized activity of the cerebral cortex. Brain activity produces a range of electrical or brain rhythms, which closely correlate with particular states of behavior or pathology. They help diagnose certain neurological conditions, especially the seizures of epilepsy. Brain rhythms play an important role in spike timing and brain communication. Different brain regions, see Figure (3.1), produce distinctly different brain rhythm frequencies that are thought to reflect unique forms of processing important for the localization, parceling, and routing of information within and between regions [24].

The amplitude of the EEG signal strongly depends on how synchronous is the activity of the underlying neurons. When a group of cells is excited simultaneously, the tiny signals sum up to generate one larger surface signal. However, when each cell receives the same amount of excitation but the excitations are spread out in time, the summed signals are meager and irregular and can be a pathological discharge pattern generated in the basal ganglia [24]. Notice that in this case, the number of activated cells and the total amount of excitation may not have changed, only the timing of the activity. If a synchronous excitation of this group of cells is repeated, again and again, the resulting EEG will consist of large-brain rhythmic that represent the normal activity of the brain [25], Table 1.1 summarizes the brain rhythms.

A recent study that compares the spectral power in the different brain rhythms across 10 mental health disorders such as depression, bipolar disorder, addiction, autism, ADHD, anxiety, panic disorder, obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD) and schizophrenia; suggest that it is necessary to have caution with any interpretation of results from studies that consider only one disorder in isolation [26]. Extrapolating this to epilepsy disease is important, due to the considerable variability in the studies, reports based on a subset of studies or the highly inconsistent between experts makes it difficult to normalize all the differences found in the researches.

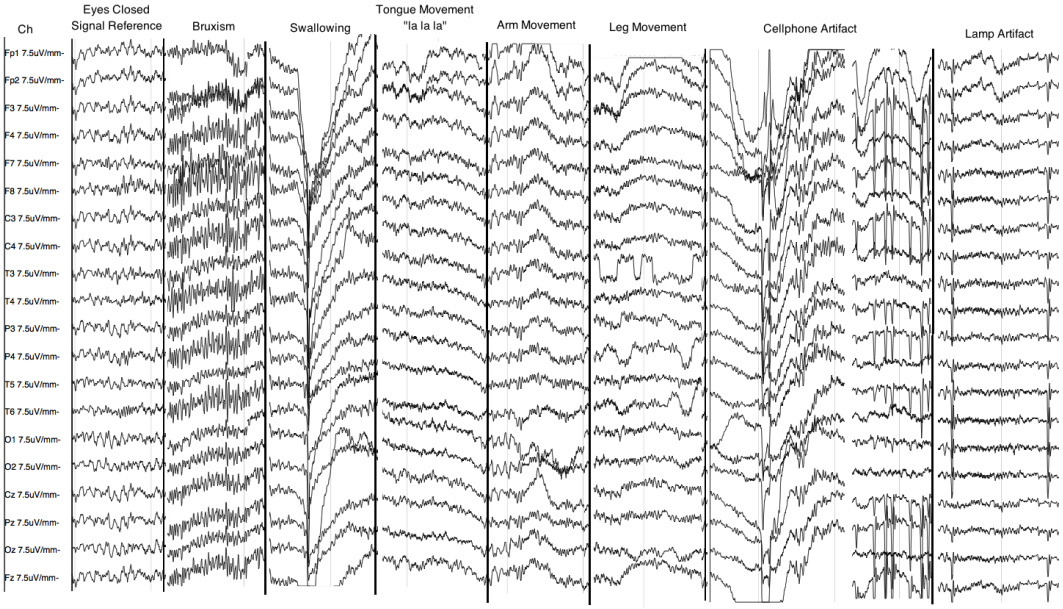


FIGURE 1.2: Examples of different artifacts: bruxism, tongue, arms and legs movement, exhibit a similar alteration in all channels while swallowing and cell phone have a high pronounced peak; lamp artifact present spikes that can be confused with epilepsy abnormal waveform by an inexpert eye. The examples shown correspond to one-second segments in duration.

1.4 Epilepsy

The term epilepsy derives from the Greek term *epilambanein* which means to seize, and it denotes the predisposition to have recurrent, unprovoked seizures. Seizures can be symptomatic; that is, result from specific precipitants such as fever, strokes, metabolic disturbances (e.g. hypoglycemia, drug abuse/withdrawal), trauma, infections in the central nervous system, and acute head injury. In epilepsy, however, seizures are unprovoked and expected to be recurrent [27]. Appropriate diagnosis and treatment of epilepsy is a main public health issue. According to the World Health Organization [28], there are more than 50 million people worldwide that suffer from some form of epilepsy, nearly 80% of them are in developing regions, where it is believed that 3 out of 4 people with these conditions do not get appropriate diagnostic and treatment.

Brain Rhythm	Frequency in Hz	Amplitude in μV	Region	Cognitive activity	Epileptic Clinical Association
Delta (δ)	< 4	20-200	Frontal, Temporal, Occipital	Deep sleep, waking state, normal in infants, sleeping adults.	Intermittent or non-rhythmic slow wave. Newborn seizures. Delta brush: beta-delta complexes and ripples of prematurity. Semirhythmic hallmarks of slow wave sleep. Sharply-contoured slow waves. Hypersynchrony. Intermittent rhythmic activity. Focal spiking. Chaotic bursts.
Theta (θ)	4-7	20-100	Temporal, Occipital	It is more common in children and young adults than in older adults, locomotion, sensory information, consciousness slips towards drowsiness, unconscious material, creative inspiration, deep meditation, maturational and emotional studies, sleeping adults, drowsiness, spatial memory processes.	Newborn seizures. Triphasic waves. Burst with a morphology very similar to ictal patterns. Rhythmic vertex. Semirhythmic hallmarks of the onset of drowsiness. Sleep-related hypersynchronies. Sharply-contoured slow waves. Sharp temporal discharges. *Theta pointu alternant*: Neonatal alternating sharp theta. Abnormal in the adult during wakefulness.
Alpha (α)	8-12	20-60	Occipital	When there is no attention, mental fatigue, cognitive disorders, awake but relaxed, attenuation as an indicator of visual activity during dreaming, semantic memory processes, to any type of task, during visually presented stimulations.	A slow decrease in frequency with an increase in amplitude.. Loss of reactivity to eye-opening or to mental alerting. Desynchronization when moving a body part. Intrude into a deep sleep or attention dramatically. An absence of the posterior rhythm.
Beta (β)	13-29	2-30	Frontal, Central, Parietal	Active thinking, active attention, focus on the outside world or solving concrete problems, is found in normal adults, panic state, rises immediately after the task, sensory-motor area, drowsiness, light sleep, REM sleep, a relatively sudden, diffuse increase in activity can mark onset of early drowsiness	Increase or decrease in waves activity. Triphasic waves. A smaller magnitude and delayed in motor movements. High voltage or plentiful activity. Asymmetry.
Gamma (γ)	$30 <$	5-10	Frontal, Central	Childhood, memory tasks, awakening, REM sleep, working, right and left index finger movement, right toes and the rather broad and bilateral area for tongue movement.	Highest levels of cerebral blood flow. Asynchrony bursts

TABLE 1.1: Brain rhythms

The International League Against Epilepsy (ILAE) [29] defines *"epileptic seizure as a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain"*. The elements defining an epileptic seizure include its mode of onset and termination, its clinical manifestations, and its abnormal enhanced synchrony [30]. Physical manifestations of epilepsy result mainly from the synchronous and excessive discharge of electricity by a group of neurons behaving abnormally in the cerebral cortex. Epileptic seizures usually have a sudden onset, they spread within seconds and, in most cases, are brief. The precise manifestation of a seizure depends on the location in the brain where it originates (onset detection) and on how far and fast it spreads. The correct identification of this location and spread information is key to proper treatment. Therefore, an epilepsy syndrome consists of a combination of clinical, seizure and EEG characteristics that make up a distinct entity. The diagnosis of epilepsy has implications for outcome and management, however, diagnosis of a particular syndrome does not imply a single cause, it has multiple etiologies [27].

The electroencephalogram (EEG) is the premier diagnostic tool for epilepsy and provides a key element for the classification and detection of epileptic seizures. The information about the morphology and dynamics of EEG signals can be used to accurately identify seizure onset and

quantify the severity and dynamical progression of seizure activity. A neurologist can discriminate normal from abnormal signals. A normal signal includes the general signal in the cortex, the typical brain rhythms and the varying degrees of thalamocortical interdependence while an abnormal signal includes burst suppression and seizures.

Epileptic attacks have two clinical manifestations of abnormal activity. *Ictal* or activity recorded during an epileptic seizure, and *Interictal* or abnormal signals recorded between epileptic seizures. Where the impaired consciousness plays an important role, which is defined as the inability to respond normally to exogenous stimuli by virtue of altered awareness and/or responsiveness [31]. In epilepsy context, normal activity includes physiological artifacts and different sleep potentials e.g. vertex waves, *K*-complexes, positive occipital sharp transients of sleep and benign waveform transients of sleep; while abnormal activity refers to interictal waveform potential, according to morphological characteristics. Interictal discharges are good epilepsy indicators and depending on EEG recordings duration and the inclusion of different states of vigilance; they can be shown in up to 90% of patients [32], see Table 1.2.

More common morphological waveforms are spike-wave and sharp-wave, see Figure 1.3; a spike-wave may last 20-70 ms and a sharp-wave may last 70-200 ms although not as sharply contoured as a spike; spike-wave complex is a spike followed by a slow-wave. If they occur at rates below 3 Hz then they are called spike-and-slow-wave complexes. Polyspikes have multiple spike complexes where several spikes occur in sequence, e.g. polyspike-and-slow-wave is a polyspike followed by a slow-wave, see Figure 1.4.

Background activity describes the context in which the spike occurs; it is used to normalize the spike parameters to account for varying electrical output from different patients and determine whether the spike is more than a random variation of the underlying rhythmic activity [33].

1.5 Seizure classification

Seizure classification is composed of four families: 1) Partial or Focal Seizures, which originate from a localized cortical area and represent 60% of epilepsy cases; 2) Generalized Seizures, which are characterized by initial synchronous discharges over both hemispheres and represent 40 % of cases (see Tables 1.3 and 1.4 for more details); 3) Unclassified Epileptic Seizures; and 4) Addendum, prolonged or repetitive seizures [31, 34]; see Table 1.1 for a relationship with the different brain rhythms. For a better compression of Table 1.4, the following definitions are introduced according to [35]

Absence seizures : They cause lapses in awareness, sometimes with staring. They are a type of generalized onset seizures, meaning they begin in both sides of the brain at the same time. An older term is petit mal seizures. They begin and end abruptly, lasting only a few seconds.

Atypical absences : They are a type of absence seizure that is atypical. This means it's different, unusual, or not typical compared to typical absence seizures, which were previously called petit mal seizures. They are a type of generalized onset seizure. When a single atypical absence seizure ends, the person usually is awake and continues doing whatever they were

doing before the seizure. No first aid is needed during a single seizure. Sometimes a person may have more than one atypical absence at a time or have groups of seizures.

Clonic seizures : "Clonus" means fast stiffening and relaxing of a muscle that happens repeatedly. In other words, it is repeated jerking. The movements cannot be stopped by restraining or repositioning the arms or legs. Clonic seizures are rare and most commonly occur in babies. Most often, clonic movements are seen as part of a tonic-clonic seizure.

Tonic seizures : Muscle "tone" is the muscle's normal tension at rest. In a tonic seizure, the tone is greatly increased: the body, arms, or legs become suddenly stiff or tense.

Myoclonic seizures : are brief shock-like jerks of a muscle or group of muscles. "Myo" means muscle. They occur in a variety of epilepsy syndromes that have different characteristics. During a myoclonic seizure, the person is usually awake and able to think clearly.

Tonic-clonic seizure : They usually begins on both sides of the brain, but can start in one side and spread to the whole brain. A person loses consciousness, muscles stiffen, and jerking movements are seen. These types of seizures usually last 1 to 3 minutes and take longer for a person to recover.

Atonic seizure : In this seizure, a person suddenly loses muscle tone. Their head or body may go limp and they may fall. They are also known as drop attacks.

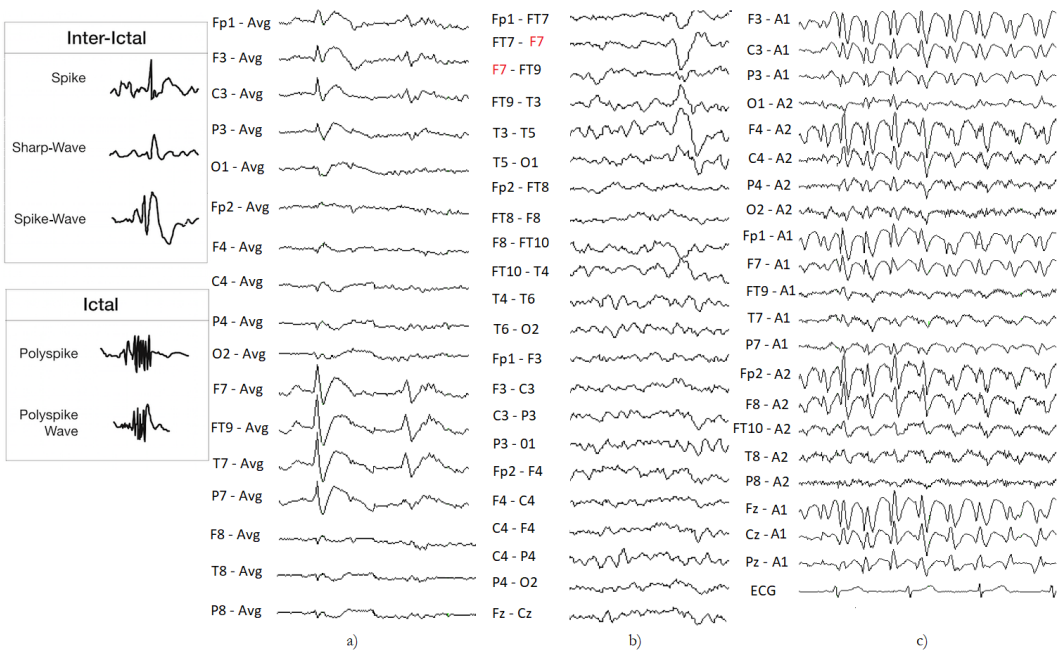


FIGURE 1.3: Different interictal abnormal activity: a) spike waveforms using a monopolar configuration whose reference is the average of all channels (Avg). b) sharp-wave using bipolar configuration, note the phase reversal in FP1-FT7, FT7-F7 channels vs. FT9-T3, T3-T5 channels, this indicating that channel F7 is a seizure candidate; c) spike-and-wave using monopolar configuration whose reference are A1 or A2 channels, the spike-and-wave waveform is present in almost all channels.

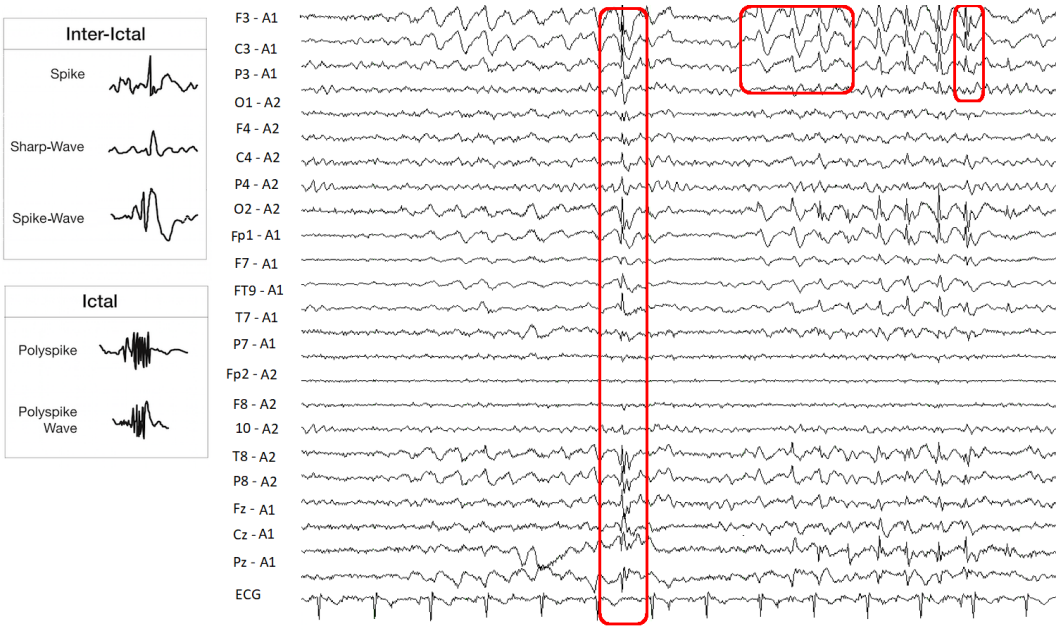


FIGURE 1.4: Different ictal abnormal activity: polyspike, polyspike wave and polyspike, respectively.

Clinical	Recorded	Onset	EEG waveform	Behavioral Disturbance
Ictal.	During an epileptic seizure.	Match.	Spikes, spike trains, isolated spikes, sharp-waves, spike-wave complexes, sharp-wave complexes	a) Bilateral involvement with impaired consciousness and abnormal EEG. b) Unilateral involvement with clear consciousness where the EEG can be normal.
Interictal.	Between epileptic seizure.	It can match or not.	Polyspike and polyspike waves.	a) Inadequate psychological adaptation; neurological, cognitive and intellectual deficits. b) Aberrant personality traits. Affective disorders. c) Psychoses. d) Memory. e) Depression.

TABLE 1.2: Abnormal activity and associated EEG morphologic waveforms

1.6 Seizure onset detection

Ictal discharges are clinical signs used to detect the onset seizure from the epileptogenic zone in the brain cortex. Seizure Onset Detection (SOD) helps physicians to improve therapy with drug treatment, diagnostic and alert procedures; while in biomedical technology, the goal is intended to recognize the start of a seizure, with the shortest possible delay and with the highest possible accuracy. SOD recordings may be: intracranial or extracranial. In intracranial recordings, for seizure identification and retrospective analysis of seizures, often in the context of presurgical evaluations, it can be relatively straightforward to detect onset with reasonable sensitivity and specificity since events often last over a minute [36]. In extracranial recordings to distinguish between primary and secondary irritative areas may be difficult. The primary irritative area is the ictal zone when the focal seizure starting and the second irritative area is related to his spread. This is because an ictal discharge can spread very fast to the normal anatomical connections between cortical areas, through the commissural fibers or via subcortical structures. Finally, this may lead to the widespread or bilateral occurrence of interictal discharges [37].

Clinical	Impaired Consciousness	EEG Seizure	Onset	Interictal Expression
Simple partial seizure.	Not.	Local contralateral discharge starting over the corresponding area of cortical representation.	Lateral but not always recorded on the scalp.	Local contralateral discharge.
Complex partial seizure.	Yes.	Unilateral or frequently bilateral discharge, diffuse or focal in temporal.	Fronto-temporal.	Unilateral or bilateral, generally asynchronous focus.
Partial seizures evolving to secondarily generalized seizures.	Yes.	Above discharges become secondarily and rapidly generalized.	Focal, lateral.	Focal discharge.

TABLE 1.3: Classification of simple partial seizures

In a medical context, neurologists use EEG to determine the type of seizure the person may have had and if there are any detectable abnormalities in the person's brain rhythm waves. This analysis of waveform features permits to localize and quantify the epileptogenic zone. In the context of epilepsy surgery, the precise identification of the epileptogenic zone is crucial. The seizure onset zone is the area of the cortex from which clinical seizures generate. There is currently no diagnostic modality that can be used to directly measure the entire epileptogenic zone. The precise localization of the epileptiform discharges is essential to the localization of the epileptogenic zone,

Clinical	EEG Seizure	Abnormal Activity	EEG Waveforms	Interictal Expression
Absence seizures, It is the most typical.	Regular and symmetric between 2 Hz and 4 Hz.	Bilateral, regular and symmetric.	Spike-wave, slow-wave complexes, multiple spike-wave and slow-wave complexes.	In background activity usually normal, although paroxysmal activity may occur.
Atypical absence.	EEG more heterogeneous; may include fast activity or other paroxysmal activity.	Bilateral, irregular and asymmetric.	In background usually abnormal; paroxysmal activity.	Regular, spike-wave, irregular spike, slow-wave complexes, and slower generalized spike-wave.
Myoclonic seizures, Myoclonic jerks. (single or multiple).	Polyspike and wave, or sometimes spike and wave or sharp and slow waves.	Bilateral.	Polyspike, sharp-wave.	Same as ictal.
Clonic seizures.	Fast activity of 10 Hz or more.	Focal contralateral.	Slow-waves, occasional spikes and wave patterns.	Spike and spike-waves, polyspike and wave discharges.
Tonic seizures.	Low voltage fast activity or a fast rhythm of 9 Hz - 10 Hz or more decreasing in frequency and increasing in amplitude.	Sometimes asymmetric.	Sharp and slow-waves.	More or less rhythmic discharges of sharp and slow waves. The background is often abnormal for age.
Tonic-clonic seizures.	Rhythm at 10 Hz or more, decreasing in frequency and increasing in amplitude during tonic phase, interrupted by slow waves during clonic phase.	Bilateral.	Polyspike.	Polyspike and waves, spike and wave, or, sometimes, sharp waves and slow-wave discharges.
Atonic seizures.	Flattening or low-voltage fast activity.	Bilateral.	Polyspikes and slow-waves.	Polyspikes and slow-waves.

TABLE 1.4: Classification of generalized seizures

particularly in patients considered for resective epilepsy surgery [38]. The multifaceted aspects of these discharges can be explored in vivo by electroencephalographic recordings [39].

The seizure onset is assessed visually and defined as an unequivocal and sustained rhythmic change from the background activity in the EEG accompanied by subsequent clinically typical seizure activity, and clearly distinguished from background EEG and interictal activity. When this information cannot be properly identified from the scalp EEG the intracranial monitoring with grids or depth electrodes are indicated. The EEG seizure-onset patterns may exhibit:

1. A low-voltage fast activity that produces an attenuation of background activity.
2. Low-frequency high-amplitude periodic spikes, high-voltage spiking at 0.5-2 Hz.
3. Sharp activity, low to medium-voltage sharply-contoured rhythmic activity most commonly in the alpha/theta range.
4. Spike and wave activity occurring at a frequency of 2-4 Hz.

5. A burst of high amplitude polyspikes.

6. Burst suppression, brief bursts of medium- to high-voltage repetitive spikes alternating with brief periods of voltage attenuation

7. Delta brush, rhythmic delta waves at 1-2 Hz, with superimposed brief bursts of 20-30 Hz activity overriding each delta wave [40].

Seizure onset detection was first investigated in the seventies by Viglione et al. [41] and Liss et al. [42], and with later contributions by Ktonas et al. [43], Gotman et al. [44] and Iasemidis et al. [45, 46]. Moreover, different works studied linear and nonlinear prediction techniques to separate transients from background activity. For example, filter techniques [47], power spectrum techniques [48], cross-correlation techniques [49], principal or independent component analysis techniques [50, 51] have been investigated. Other examples include techniques based on wavelet representations [52], state space reconstruction [53], correlation measures [54], signal dimension [55], density and correlation integrals [56, 57], mutual prediction [58], Lyapunov exponents [59, 8], synchronization [60], similarity measures [61], recurrence quantification measures [62], and nonlinear predictability [63]. We refer the reader to [64] for a comprehensive treatment of measurement, models, detection and prediction techniques. Other important surveys of the literature in this topic can be found in [65, 52, 66, 67, 68, 33, 69, 70, 71, 72]. See Section 4.2 for an extension of this state-of-art.

Moreover, modern SOD methods can be grouped into the following categories: 1) *Template matching*: These are techniques based on finding events that match previously selected spikes; the detection is made whenever the cross correlation of the EEG with a template exceeds a threshold [73]; 2) *Parametric methods*: These techniques are based on traditional signal processing and consider that a seizure has occurred when the difference between the EEG signal and a predicted value (based on the assumption that the background is stationary) exceeds a threshold [74]; 3) *Mimetic methods*: these techniques seek to mimic the human expert (i.e. neurophysiologist) and operate by monitoring the value of parameters computed from each wave and applying thresholds [75, 76, 77]; 4) *Morphologic analysis*: these techniques are based on the characterization of the waveforms with respect to sharpness, amplitude, duration, convexity, frequency bands or time-frequency representations of spikes [75, 78]; 5) *Syntactic methods*: these techniques are based on the detection of the presence of structural features [79]; 6) *Neural networks*: this approach adopts a machine learning perspective to learn transients related to epileptic seizures [80, 81]; 7) *Expert systems*: this approach detects seizures by mimicking an expert's knowledge and reasoning process [82]; 8) *Data mining techniques*: this approach also adopts a machine learning perspective to train a classifier [71, 83, 84]; 9) *Clustering techniques*: detection is based on hierarchical agglomerative processes and self-organizing maps [47, 72]; 10) *Knowledge-based rules*: similar to expert systems, these techniques seek to incorporate knowledge from neurophysiologists through spatial and temporal rules [85, 67, 85].

1.7 Conclusions

This chapter presented the medical context of this thesis by defining electroencephalography, brain rhythms, epilepsy, and seizure activity. These topics are important for developing biomedical solutions, especially for detecting epileptic seizure activity in EEG signals in real-time. In order to quantify and characterize such brain disorders, this thesis adopts a statistical modeling and mathematics computation approach. The following chapter develops this approach and derives quantitative characterization indicators.

Chapter 2
Statistical-Model-based EEG signal characterization

2.1 Introduction

In chapter 1, EEG, brain rhythms, epilepsy, and seizure activity were explained in the medical context. In this chapter, a statistical model is established for EEG data, in order to characterize the epileptic seizure activity within each brain rhythm.

In the first part, the method of decomposing EEG data into different brain rhythms using wavelets filter banks is presented. Next, different statistical models are studied and compared in order to choose the best model that fits the brain rhythms. The generalized Gaussian distribution (GGD) is shown to give the best goodness-of-fit. This constitutes the main contribution of this thesis, which shows that the GGD statistical model represents correctly the epileptic seizure activity in EEG signals. This model permits the characterization and quantification of EEG signals using its scale and shape parameters. In addition, this proposed model-based characterization allows a considerable dimensional reduction and makes possible developing fast classification algorithms with low complexity. In the second part, the Kullback-Leibler divergence (KLD) is used to measure the discrepancy between probability density functions (PDF), in order to detect changes between *seizure* and *non-seizure*. Our second methodological contribution consists in the analytical development of the KLD between two generalized Gaussian distributions.

2.2 Data set

In order to establish the best statistical model for epileptic EEG, we studied the Children's Hospital Boston database [86], previously considered in [37]. This dataset which consists of 36 EEG recordings from pediatric subjects with intractable seizures. In this thesis, we used 54 events of 18 recordings from 9 different subjects. The events include 18 *seizures* and 36 *non-seizures* (18 before and 18 after the seizure).

The signals were acquired with a 23-channel array operating at a 256 Hz sampling rate. The neurologist annotated each signal to indicate the beginning and end of the seizure epochs, which we use as ground truth. Moreover, for each *seizure* epoch, the neurologist also selected two adjacent *non-seizure* signal segments of the same length to represent challenging or control of the

non-pathological brain activity. The length of all seizures used are summarized in Table 2.1.

No distinction was considered regarding the types of seizure onsets; the data contains focal, lateral, and generalized seizure onsets. Furthermore, the recordings were made in a routine clinical environment, therefore *non-seizure* activity and artifacts such as head/body movement, chewing, blinking, early stages of sleep, and electrode pops/movement are present in the data, see Figure 1.2 in Section 1.2.

				Processing duration in msec				
Epoch	Seizure	Duration	Segments	Delta	Theta	Alpha	Beta	Gamma
01	04	1m30sec	181	7	6	7	7	8
02	05	1m41sec	203	7	7	7	8	9
03	10	1m04sec	129	7	7	7	7	9
04	11	1m07sec	137	7	6	7	7	8
05	12	2m00sec	241	7	7	7	8	8
06	13	1m57sec	235	7	7	7	8	9
07	17	1m26sec	173	7	7	7	8	8
08	18	2m23sec	287	7	7	7	8	9
09	19	3m09sec	321	7	6	6	7	8
10	20	3m46sec	343	7	6	6	7	8
11	21	5m38sec	529	7	6	7	7	8
12	22	1m04sec	129	7	6	7	7	8
13	23	1m03sec	125	7	9	9	7	13
14	26	1m05sec	131	7	9	10	11	35
15	27	1m02sec	117	7	7	7	7	9
16	28	1m16sec	153	7	7	7	8	9
17	29	1m29sec	179	7	6	7	7	8
18	30	0m32sec	65	7	7	7	7	8

TABLE 2.1: Length of the 18 seizures used in this study and the corresponding number of overlapping 1-second segments using a rectangular sliding window of 2 seconds. An offset has been used for each epoch to avoid leading and trailing signals that were noisy. Consequently, the number of windows is irregular between epochs.

2.3 Data organization

Let $\boldsymbol{X} \in \mathbb{R}^{M \times N}$ denote the matrix gathering M EEG signals $\boldsymbol{x}_m \in \mathbb{R}^{1 \times N}$ measured simultaneously on different channels and at N discrete time instants. We use the representation [22],

$$\boldsymbol{X} = \boldsymbol{K}\, \boldsymbol{J} + \boldsymbol{\eta} \tag{2.1}$$

where $\boldsymbol{X}$ are the EEG signals, $\boldsymbol{J} \in \mathbb{R}^{n_s \times N}$ is a matrix representing the sources, $\boldsymbol{K} \in \mathbb{R}^{M \times n_s}$ is the so-called lead field or gain matrix, n_s is the number of sources, and $\boldsymbol{\eta}$ is additive noise.

The original signal $\boldsymbol{X}$ is split into a set of overlapping 1-second segments using a rectangular sliding window of 2 seconds so that

$$\begin{aligned} \boldsymbol{X}^{(i)} &= \boldsymbol{\Omega}^{(i)}\boldsymbol{X} \\ \boldsymbol{\Omega}^{(i)} &= \left[\mathbf{0}^{L\times iL}, \boldsymbol{I}^{L\times L}, \mathbf{0}^{L\times N-iL-L}\right] \end{aligned} \tag{2.2}$$

where $\mathbf{0}^{N\times M} \in \mathbb{R}^{N\times M}$ is the null matrix, $\boldsymbol{I}^{N\times N} \in \mathbb{R}^{N\times N}$ is the identity matrix and L is the number of measurement obtained in 2 seconds.

2.4 Wavelets and Filter banks

Wavelets are localized waves that, instead of oscillating forever, drop to zero. They come from the iteration of filters with scaling [87, 88]. They are obtained from a single prototype "mother" wavelet $\psi(t)$ by rescaling and shifting, i.e.,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2.3}$$

where a is the scaling parameter and b is the shifting parameter. The wavelet transform is given by

$$W_f(a,b) = \int_{-\infty}^{\infty} x(t)\psi_{a,b}(t)\mathrm{dt} \tag{2.4}$$

The discrete wavelet transform (DWT) transforms a discrete time signal to a discrete wavelet representation. It converts an input series $\boldsymbol{x} = [x_0, \ldots, x_{L-1}]^T$ of length L, into one high-pass ($\boldsymbol{h}$) wavelet coefficient series and one low-pass ($\boldsymbol{l}$) wavelet coefficient series, each one of length $\frac{L}{2}$, given by

$$h_j = \sum_{k=0}^{K-1} x_{2j-k}\ s_k, \qquad l_j = \sum_{k=0}^{K-1} x_{2j-k}\ t_k \qquad \forall\ 0 \le j < \frac{L}{2} \tag{2.5}$$

where $\boldsymbol{s} = [s_0, \ldots, s_{K-1}]^T$ and $\boldsymbol{t} = [t_0, \ldots, t_{K-1}]^T$ are called the wavelet filters. Recalling that $X^{(i)}$ represents a multichannel signal where each column contains a different channel and each row represents the temporal evolution of the EEG signal.
The discretized wavelet for the DWT takes the following form

$$W_d(a,b) = \sum_{-\infty}^{\infty} x(n)\frac{1}{\sqrt{a^j}}\psi_{a,b}\left(\frac{n-b}{a^j}\right) \tag{2.6}$$

In the discrete wavelet transform, the scale parameter is always discretized to integer powers of 2, 2^j with $j = 1, 2, 3, \cdots$, so that the number of voices per octave is always 1.

The CWT and DWT differ in how they discretize the scale parameter. The CWT typically uses exponential scales with a base smaller than 2 (e.g. $2^{1/12}$), while the DWT always uses exponential scales with the base equal to 2. The scales are powers of 2. Therefore, the physical interpretation of scales for both the CWT and DWT requires the inclusion of the signal's sampling interval if it is not equal to one [89].

2.4.1 Multilevel 1D wavelet decomposition

The multilevel 1D wavelet transform decomposes the matrix $\boldsymbol{X}^{(i)}$ using eq. (2.5) into two component matrices, namely $\boldsymbol{L}_j^{(i)}$, $\boldsymbol{H}_j^{(i)}$, where (L) corresponds to applying a low-pass frequency operation to the temporal component (rows) of $\boldsymbol{X}^{(i)}$ and (H) refers to the high-pass filter applied to the channel component (columns) of $\boldsymbol{X}^{(i)}$; each one according the scale j. The lowest frequency sub-band $\mathbf{L}_j^{(i)}$ is the approximation (A) coefficients and $\mathbf{H}_j^{(i)}$ is the detail (D) coefficients of the original signal $\boldsymbol{X}^{(i)}$. This process is repeated recursively replacing the input signal $X^{(i)}$ with the last approximation series $\boldsymbol{L}_j^{(i)}$ until the desired number of scales $j = [1, 2, \ldots, J]^T$ is obtained.

2.4.2 Multilevel 2D wavelet decomposition

The multilevel 2D wavelet transform decomposes the matrix $\boldsymbol{X}^{(i)}$ using eq. (2.5) into four component matrices, namely $\boldsymbol{LL}_j^{(i)}$, $\boldsymbol{LH}_j^{(i)}$, $\boldsymbol{HL}_j^{(i)}$ and $\boldsymbol{HH}_j^{(i)}$, where the first letter corresponds to applying a low-pass (L) or high-pass (H) frequency operation to the temporal component (rows) of $\boldsymbol{X}^{(i)}$ and the second letter refers to the filter applied to the channel component (columns) of $\boldsymbol{X}^{(i)}$, each one according to the scale j. The lowest frequency sub-band $\mathbf{LL}_j^{(i)}$ is the approximation coefficients of the original signal $\boldsymbol{X}^{(i)}$. The remaining three frequency sub-bands are the detail parts of the signal and give the vertical high ($\boldsymbol{LH}_j^{(i)}$), horizontal high ($\boldsymbol{HL}_j^{(i)}$) and diagonal high ($\boldsymbol{HH}_j^{(i)}$) coefficients. This process is repeated recursively replacing the input signal $X^{(i)}$ with the last approximation series $\boldsymbol{LL}_j^{(i)}$ until the desired number of scales $j = [1, 2, \ldots, J]^T$ is obtained.

2.4.3 Extraction of brain rhythms

The coefficients $\boldsymbol{X}^{(i)}$ associated with all wavelet scales for a 2-second segment with overlapping 1 second, are represented using a Daubechies wavelet filter bank with 6 scales of order 4 (Db4), see Figure 2.1, to obtain a time-frequency decomposition [90]. The purpose of this decomposition is to evaluate the energy distribution throughout the neurological frequency spectrum or brain rhythms, namely the *delta (δ), theta (θ), alpha (α), beta (β)* and *gamma (γ)* bands [91, 92]. Db4 offers the number of vanishing moments that allow representing the signal with sufficient smoothness. Performing wavelet decomposition fits naturally the dyadic structure of the neurological spectral bands, and provides a computationally efficient filtering algorithm that can be implemented straightforwardly on real-time signal processing hardware.

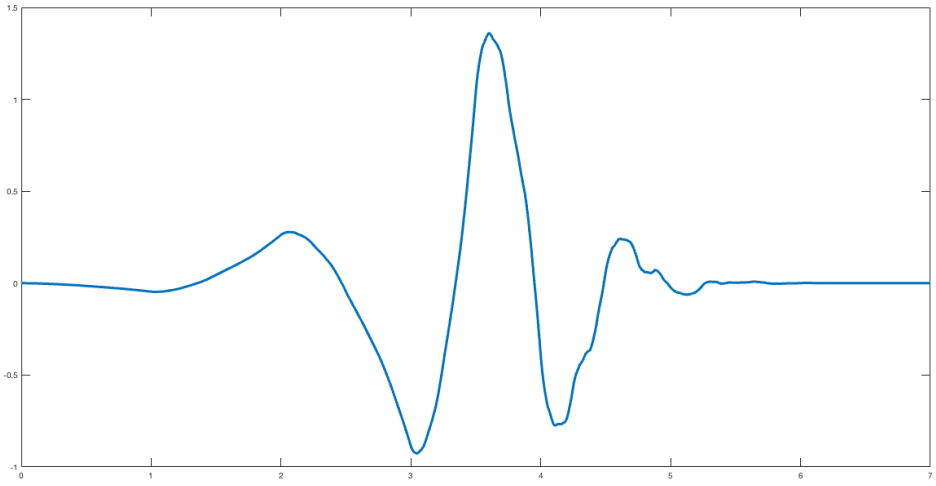


FIGURE 2.1: db4 wavelet waveform, we can see the similarity with a seizure waveforms in Figures 1.3 and 1.4.

Let $\boldsymbol{C}_j^{(i)}$ denote the coefficients corresponding to the different sub-bands, namely approximation (A) and detail (D) coefficients for the multilevel 1D wavelet decomposition (Section 2.4.1) or horizontal (H), vertical (V), diagonal (D) and approximation (A) for the multilevel 2D wavelet decomposition (Section 2.4.2).

Let $\boldsymbol{\theta}^{(i)}$ denote the vector of parameters estimated from the different brain rhythms ($\gamma, \beta, \alpha, \theta, \delta$, see Table 2.2):

$$\boldsymbol{\theta}^{(i)} = \left[\boldsymbol{\theta}_3^{(i)}, \boldsymbol{\theta}_4^{(i)}, \boldsymbol{\theta}_5^{(i)}, \boldsymbol{\theta}_6^{(i)}, \boldsymbol{\theta}_{L_6^{(i)}}^T\right]^T = \left[\boldsymbol{\theta}_\gamma^{(i)}, \boldsymbol{\theta}_\beta^{(i)}, \boldsymbol{\theta}_\alpha^{(i)}, \boldsymbol{\theta}_\theta^{(i)}, \boldsymbol{\theta}_\delta^{(i)}\right]^T. \tag{2.7}$$

For the multilevel 1D wavelet decomposition (Section 2.4.1) we obtain a 5-dimensional vector $\boldsymbol{\theta}^{(i)}$ and the components:

$$\boldsymbol{\theta}_j^{(i)} = \left[\boldsymbol{\theta}_{\boldsymbol{L}_j^{(i)}}^T, \boldsymbol{\theta}_{\boldsymbol{H}_j^{(i)}}^T\right]. \tag{2.8}$$

For the multilevel 2D wavelet decomposition (Section 2.4.2) we obtain a 13-dimensional vector $\boldsymbol{\theta}^{(i)}$ and the components:

$$\boldsymbol{\theta}_j^{(i)} = \left[\boldsymbol{\theta}_{\boldsymbol{LH}_j^{(i)}}^T, \boldsymbol{\theta}_{\boldsymbol{HL}_j^{(i)}}^T, \boldsymbol{\theta}_{\boldsymbol{HH}_j^{(i)}}^T\right]. \tag{2.9}$$

Table 2.2 presents frequencies corresponding to different levels of decomposition for the Daubechies wavelets of order 4 with a sampling frequency of 256 Hz, where A and D refer to Approximation and details respectively for Section 2.4.1 or where H, V and D refer to horizontal, vertical and diagonal details respectively for Section 2.4.2 and the number is the scale. The rest of approximations and details are discarded because they are outside of the brain rhythms. Note that in high-frequency oscillations (HFOs) the gamma band is extended to the 80–500 Hz frequency range, which provides

to the neurophysiologist new information about the extent of the epileptogenic tissue in addition to ictal and interictal lower frequency events[93].

Decomposed Signal	Frequency range (Hz)	Brain Rhythms
D3 - H3 - V3	32-64	Gamma (γ)
D4 - H4 - V4	16-32	Beta (β)
D5 - H5 - V5	8-16	Alpha (α)
D6 - H6 - V6	4-8	Theta (θ)
A6	0-4	Delta (δ)

TABLE 2.2: Frequencies of the different scales of the multilevel 2D wavelet decomposition.

2.5 Statistical models

In this thesis, we are interested in the characterization of the physical processes underlying the EEG signals. Deterministic physical models of such processes are difficult to establish as epileptic seizure have usually sudden onsets, spread in a matter of seconds, and are in most cases very brief. Also, they are contaminated with noise that intrinsically follows a stochastic process model. In addition, such physical model should also take into account the complexity related to the location of the seizure source and the temporal and spatial scope of its spread. Therefore, we resort to statistical modeling to capture the characteristics of the electrical processes underlying EEG signals. Our aim is to establish quantitative indicators based on a statistical model to characterize EEG signals. The objective is to develop a machine learning algorithm that uses such indicators as features to detect epileptic seizures.

A statistical model is a probability distribution constructed to enable inferences to be drawn through histogram or decisions made from data [94]. A histogram can be interpreted through a parametric statistical distribution, where the probability density function (PDF) is denoted

$$p\left(\boldsymbol{C}_j^{(i)}; \boldsymbol{\theta}_j^{(i)}\right) \tag{2.10}$$

where j refers to each brain rhythm, i is related to the window segment, $\boldsymbol{C}_j^{(i)}$ are the coefficients corresponding to the different brain rhythms, and $\boldsymbol{\theta}_j^{(i)}$ is a set of model parameters associated with the wavelet coefficients of each brain rhythm. Statistical parameters are a quantity that indexes a family of probability distributions [95], see Table 2.3 for some statistical parameters examples.

Parameter	Logistic	t-location-scale	Alpha-stable	Cauchy	GGD
Location	μ	μ	δ	x_0	μ
Scale	ζ	σ	γ	γ	σ
Shape		ν			τ
Skewness			τ		
Stability			σ		

TABLE 2.3: Statistical parameters of each distribution under consideration.

For the EEG signal $\boldsymbol{X}$ with long segment L from model $p\left(\boldsymbol{C}_j^{(i)}; \boldsymbol{\theta}_j^{(i)}\right)$, the maximum likelihood estimation consists in maximizing

$$\frac{1}{L} \log \left(c; \boldsymbol{\theta}_j^{(i)}\right) = \frac{1}{L} \sum_{i=1}^{L} \log\ p\left(c^{(i)}; \boldsymbol{\theta}_j^{(i)}\right) \tag{2.11}$$

giving the estimate

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{C}_j^{(i)}} = \underset{\left[\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}\right]^T}{\operatorname{argmax}} \log\ p\left(c; \boldsymbol{\theta}\right). \tag{2.12}$$

2.5.1 Generalized Gaussian distribution

The univariate generalized Gaussian distribution (GGD) is a flexible statistical model for one-dimensional signals [96] that has found numerous applications in science and engineering [97, 98, 99, 100]. Since the series has zero-mean because it was subtracted (detrending), then there aren't non-zero coefficients for the lower frequency resolution band [101], it can be safely assumed that they can be represented by a zero-mean distribution. Consequently, the distribution of the wavelet coefficients $\boldsymbol{C}_j^{(i)}$ can be represented by using a zero-mean GGD statistical model [102, 103] with probability density function (PDF) given by

$$f_{\mathrm{GGD}}(x; \sigma, \tau) = \frac{\tau}{2\sigma\Gamma(\tau^{-1})} \exp\left(-\left|\frac{x}{\sigma}\right|^{\tau}\right) \tag{2.13}$$

where $\sigma \in \mathbb{R}^+$ is a scale parameter and $\tau \in \mathbb{R}^+$ is a parameter that controls the shape of the density tail and $\Gamma(\cdot)$ is the Gamma function. Note that the GGD parametric distribution family includes many popular distributions that are commonly used in biomedical signal processing. For example, setting $\tau = 1$ leads to a Laplacian or double-exponential distribution, $\tau = 2$ leads to Gaussian or normal distribution, and $\tau \to \infty$ leads to a uniform distribution.
From eq. (2.13) and eq. (2.12), the statistical properties of the wavelet coefficients $\boldsymbol{C}_j^{(i)}$ can be summarized by parameter-vector $\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}$:

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{C}_j^{(i)}} = \left[\sigma_j^{(i)}, \tau_j^{(i)}\right]^T = \underset{[\sigma,\tau]^T}{\arg\max}\, f_{\mathrm{GGD}}(\boldsymbol{C}_j^{(i)}; \sigma, \tau). \tag{2.14}$$

2.5.2 Logistic distribution

The logistic distribution models a continuous random variable whose probability density is the logistic function. It is very popular in different areas, such as biology, epidemiology, sociology and energy [104, 105] . This parametric distribution has two parameters estimated by maximum

likelihood [106], and the probability density function (PDF)

$$f_{LD}(x;\mu,\zeta) = \frac{\exp\left(-\frac{x-\mu}{\zeta}\right)}{\zeta\left(1+\exp\left(-\frac{x-\mu}{\zeta}\right)^2\right)} \tag{2.15}$$

where $\mu \in \mathbb{R}$ is a location parameter and $\zeta > 0 \in \mathbb{R}$ is a scale parameter. From eq. (2.15) and eq. (2.12), the statistical properties of the wavelet coefficients $\boldsymbol{C}_j^{(i)}$ can be summarized by the parameter-vector $\boldsymbol{\theta}_{C_j^{(i)}}$

$$\hat{\boldsymbol{\theta}}_{C_j^{(i)}} = \left[\mu_j^{(i)}, \zeta_j^{(i)}\right]^T = \underset{[\mu,\zeta]^T}{\text{argmax}}\, f_{LD}(\boldsymbol{C}_j^{(i)};\mu,\zeta). \tag{2.16}$$

2.5.3 t-location-scale distribution

The t-location-scale distribution, is heavy-tailed and has been extensively used in many different areas [107, 108]. This parametric distribution has 3 parameters estimated by maximum likelihood [109], and the probability density function (PDF)

$$f_{TLSD}(x;\mu,\nu,\sigma) = \frac{\Gamma(\frac{\nu+1}{2})}{\sigma\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left[\frac{\nu+(\frac{x-\mu}{\sigma})^2}{\nu}\right]^{-\frac{\nu+1}{2}} \tag{2.17}$$

where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter and $\nu > 0$ is the shape parameter, see Section 5.6. From eq. (2.17) and eq. (2.12), the statistical properties of the wavelet coefficients $\boldsymbol{C}_j^{(i)}$ can be summarized by the parameter-vector $\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}$

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{C}_j^{(i)}} = \left[\mu_j^{(i)}, \nu_j^{(i)}\sigma_j^{(i)}\right]^T = \underset{[\mu,\nu,\sigma]^T}{\text{argmax}}\, f_{TLSD}(\boldsymbol{C}_j^{(i)};\mu,\nu,\sigma). \tag{2.18}$$

2.5.4 Alpha-stable Distribution

The alpha-stable distribution, is heavy-tailed and has found several applications in economics and physics as models of rare, but extreme events, such as earthquakes or stock market crashes [110]. In engineering and mathematics, it has also have a variety of applications [111, 112, 113, 114, 115]. This parametric distribution has 4 parameters estimated by maximum likelihood [116, 117, 118], and the probability density function (PDF)

$$f_{ASD}(x;\sigma,\tau,\gamma,\delta) \tag{2.19}$$

where $\sigma = 2$ is a Gaussian characteristic exponent parameter that describes the tail of the distribution, $\tau \in [-1,1]$ is a skewness parameter, with a right-skewed distribution for $\tau > 0$ and the left-skewed for $\tau < 0$; $\gamma > 0$ is a scale parameter and $\delta \in \mathbb{R}$ is the location parameter.

From eq. (2.19) and eq. (2.12), the statistical properties of the wavelet coefficients $\boldsymbol{C}_j^{(i)}$ can be summarized in a formal sense by the parameter-vector $\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}$

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{C}_j^{(i)}} = \left[\sigma_j^{(i)}, \tau_j^{(i)}, \gamma_j^{(i)}, \delta_j^{(i)}\right]^T = \underset{[\sigma,\tau,\gamma,\delta]^T}{\operatorname{argmax}} f_{ASD}(\boldsymbol{C}_j^{(i)}; \sigma, \tau, \gamma, \delta). \tag{2.20}$$

2.5.5 Cauchy distribution

The Cauchy distribution has no mean, variance or higher moments defined. Mode and media are both equal to x_0 [119, 120]. Some applications in engineering and mathematics can be found in [121, 122, 123]. This parametric distribution has two parameters estimated by maximum likelihood [124], and the probability density function (PDF)

$$f_{CD}(x, x_0, \gamma) = \frac{1}{\pi\gamma[1 + (\frac{x-x_0}{\gamma})^2]} \tag{2.21}$$

where $x_0 \in \mathbb{R}$ is a location parameter and $\gamma > 0 \in \mathbb{R}$ is a scale parameter. From eq. (2.21) and eq. (2.12), the statistical properties of the wavelet coefficients $\boldsymbol{C}_j^{(i)}$ can be summarized by the parameter-vector $\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}$

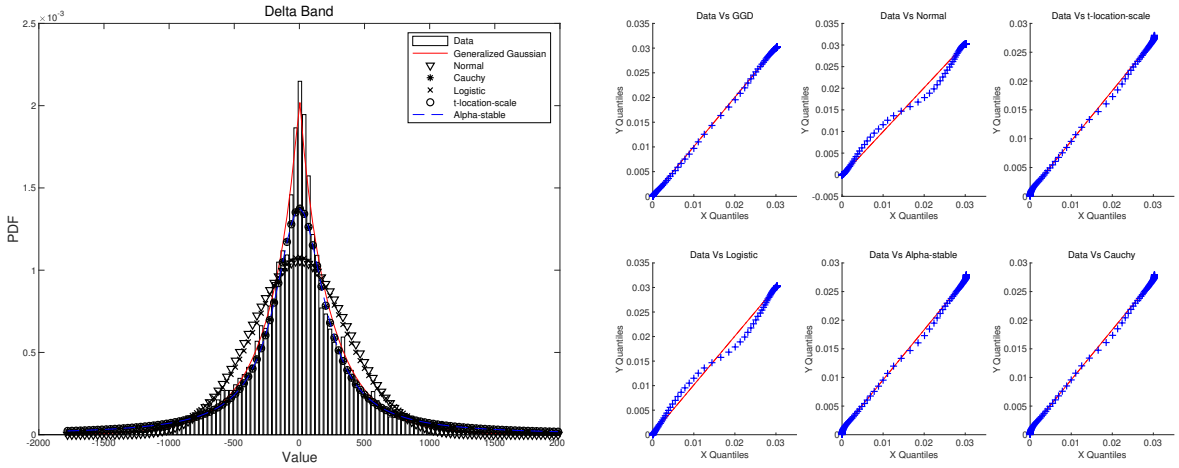
$$\hat{\boldsymbol{\theta}}_{\boldsymbol{C}_j^{(i)}} = \left[x_0{}_j^{(i)}, \gamma_j^{(i)}\right]^T = \underset{[x_0,\gamma]^T}{\operatorname{argmax}} f_{CD}(\boldsymbol{C}_j^{(i)}; x_0, \gamma). \tag{2.22}$$

2.6 Goodness-of-fit test

2.6.1 Distribution-fitting

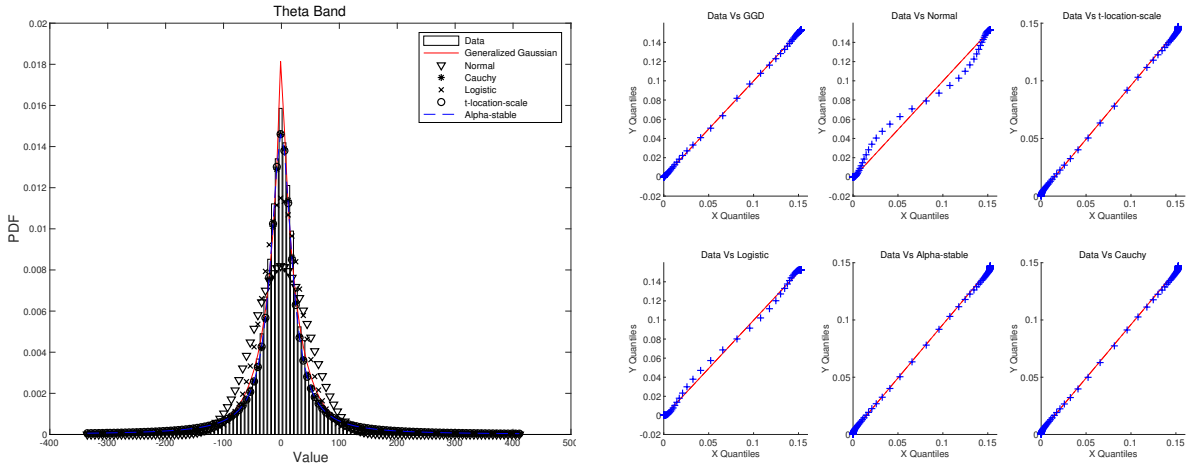
For each statistical distribution, the parameters $\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}$ (see Table 2.3) of the corresponding probability density function (PDF) were estimated in order to define which is the best model for our data set. This decision is determined visually through the histogram according to the presence or absence of symmetry of the data set with respect to the mean value.

Figures 2.2 to 2.6 depict the fit of the different statistical models to wavelet coefficients for each brain rhythm. This information is relevant to neurologists and allows discriminating clinical events of different nature, see Section 1.3. We notice that the generalized Gaussian distribution (GGD) gives the best fit among all the statistical distributions, for all brain rhythms by using all signals from the Children's Hospital Boston database. Both visual inspection of curve fitting and Q-Q plot analysis confirm this good-fit.



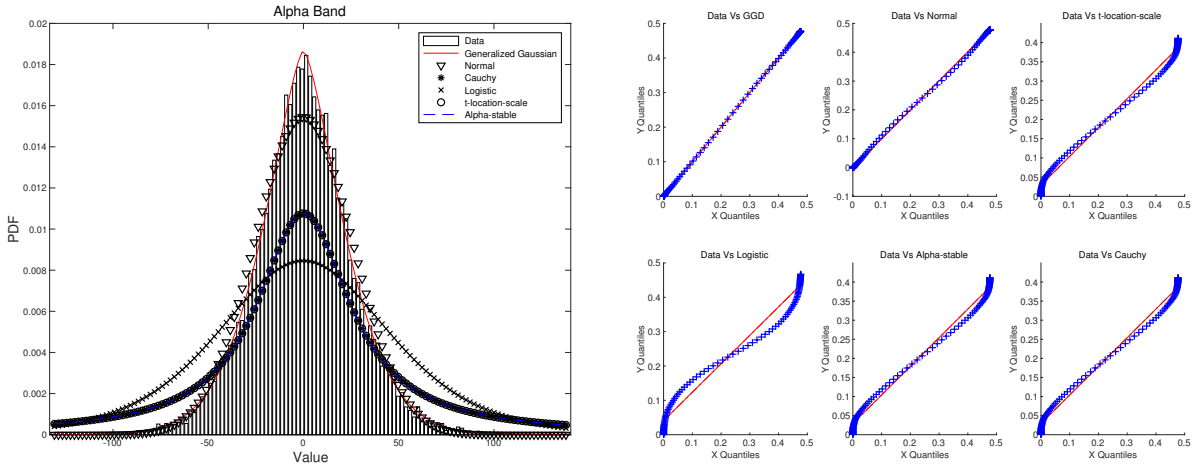
(a) Distributions-Fitting (b) Q-Q Plot

FIGURE 2.2: Example of good fit of generalized Gaussian distribution (GGD) statistical model for the Delta Band; we can observe the different data distribution-fitting and how the GGD is the best data-fit among all distributions considered.



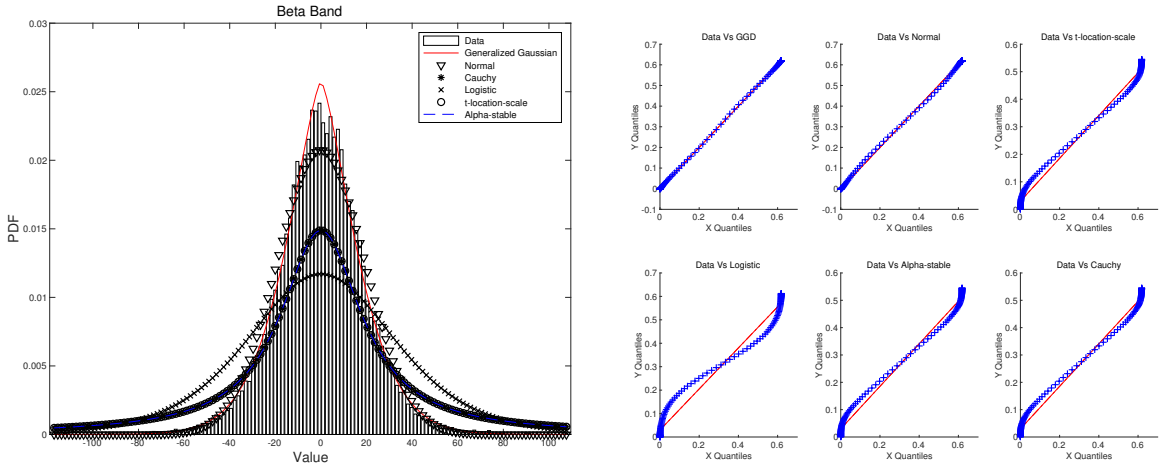
(a) Distributions-Fitting (b) Q-Q Plot

FIGURE 2.3: Example of good fit of generalized Gaussian distribution (GGD) statistical model for the Theta Band; we can observe the different data distribution-fitting and how the GGD is the best data-fit among all distributions considered.



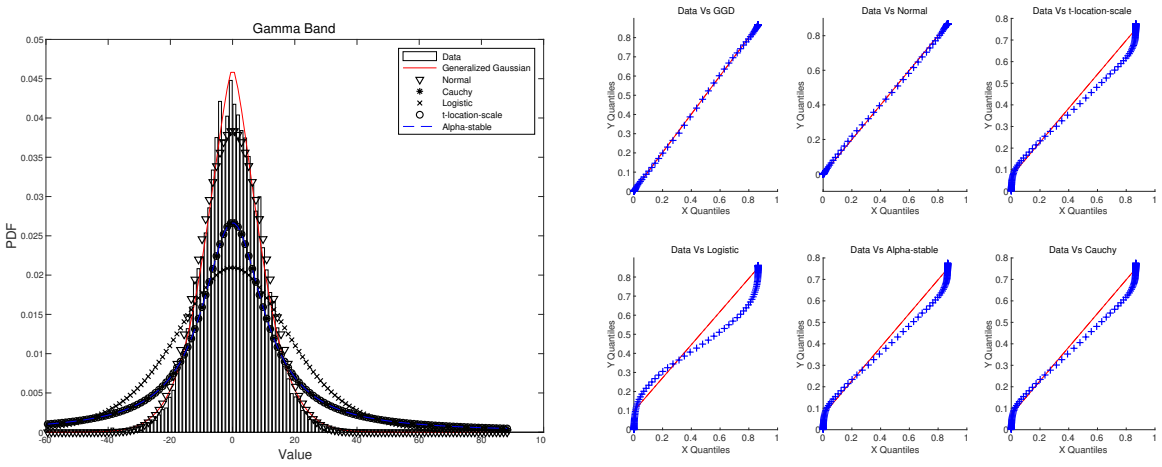
(a) Distributions-Fitting (b) Q-Q Plot

FIGURE 2.4: Example of good fit of generalized Gaussian distribution (GGD) statistical model for the Alpha Band; we can observe the different data distribution-fitting and how the GGD is the best data-fit among all distributions considered.



(a) Distributions-Fitting (b) Q-Q Plot

FIGURE 2.5: Example of good fit of generalized Gaussian distribution (GGD) statistical model for the Beta Band; we can observe the different data distribution-fitting and how the GGD is the best data-fit among all distributions considered.



(a) Distributions-Fitting (b) Q-Q Plot

FIGURE 2.6: Example of good fit of generalized Gaussian distribution (GGD) statistical model for the Gamma Band; we can observe the different data distribution-fitting and how the GGD is the best data-fit among all distributions considered.

2.6.2 Kolmogorov-Smirnov (KS)

Kolmogorov-Smirnov [125, 126] is a nonparametric test, used to decide if a sample comes from a population with a specific distribution or between two empirical (cumulative) distributions. It is defined by

$$\begin{aligned} H_0 &= \text{The data come from a specified distribution} \\ H_1 &= \text{The data don't come from a specified distribution} \\ KS &= \max_x \left| F^*_{\boldsymbol{X}^{(i)}} - F_{\boldsymbol{X}^{(i)}} \right| \end{aligned} \tag{2.23}$$

where F is the empirical distribution and F^* is the specified cumulative distribution. The hypothesis regarding the distributional form is rejected if the significant p-value is greater than 0.05.

2.6.3 Cramer-von Mises criterion (CvM)

Cramer-von Mises (CvM) is a criterion used for judging the goodness of fit of a cumulative distribution function compared to a given empirical distribution function or for comparing two empirical distributions [127, 128]. It is defined by

$$\begin{aligned} H_0 &= \text{The data come from a specified distribution} \\ H_1 &= \text{The data don't come from a specified distribution} \\ CvM &= \int_{\infty}^{+\infty} \left| F^*_{\boldsymbol{X}^{(i)}} - F_{\boldsymbol{X}^{(i)}} \right|^2 dF(\boldsymbol{X}^{(i)}) \end{aligned} \tag{2.24}$$

where F is the empirical distribution and F^* is the specified cumulative distribution. The hypothesis regarding the distributional form is rejected if the significant p-value is greater than 0.05.

2.6.4 Goodness-of-fit test results

For comparison, we computed the goodness-of-fit score for the following statistical models that are also commonly used to model wavelet coefficients: logistic, t-location-scale distribution, and alpha-stable. Moreover, we computed the scores for each spectral band and by separating the data into *seizures* and *non-seizure* groups for each brain rhythms. The resulting scores are summarized in Tables 2.4 and 2.5 below, which report respectively the mean and standard deviation of the Kolmogorov-Smirnov (KS) and the Cramer-von-Mises (CvM) scores. Observe that the generalized Gaussian distribution clearly provides the best model-fit-to-data.

KS	Means	GGD	Logistic	t-location-scale	Alpha-stable
delta	Non-Seizure	0.002	0.007	0.007	0.007
	Seizure	0.002	0.004	0.004	0.004
theta	Non-Seizure	0.008	0.037	0.042	0.042
	Seizure	0.005	0.018	0.021	0.021
alpha	Non-Seizure	0.005	0.045	0.051	0.051
	Seizure	0.003	0.021	0.024	0.024
beta	Non-Seizure	0.002	0.024	0.027	0.027
	Seizure	0.001	0.011	0.012	0.012
gamma	Non-Seizure	0.003	0.022	0.027	0.027
	Seizure	0.001	0.010	0.012	0.012
CvM	Means	GGD	Logistic	t-location-scale	Alpha-stable
delta	Non-Seizure	< 0.001	< 0.001	< 0.001	< 0.001
	Seizure	0.004	0.007	0.006	0.006
theta	Non-Seizure	< 0.001	0.013	0.016	0.016
	Seizure	0.001	0.006	0.008	0.008
alpha	Non-Seizure	< 0.001	0.021	0.027	0.027
	Seizure	< 0.001	0.005	0.006	0.006
beta	Non-Seizure	< 0.001	0.009	0.012	0.012
	Seizure	0.001	0.005	0.006	0.006
gamma	Non-Seizure	< 0.001	0.010	0.016	0.016
	Seizure	< 0.001	0.002	0.003	0.003

TABLE 2.4: Means of the Kolmogorov-Smirnov (KS) and the Cramer-von-Mises criterion (CvM) scores obtained for GGD pdfs estimated with all EEG segments of 54 events used, 18 *seizures* and 36 *non-seizures*. The GGD shows the lowest scores with respect to the other distributions considered.

KS	st. deviations	GGD	Logistic	t-location-scale	Alpha-stable
delta	Non-Seizure	0.001	< 0.001	< 0.001	< 0.001
	Seizure	0.024	0.032	0.031	0.031
theta	Non-Seizure	< 0.001	0.002	0.002	0.002
	Seizure	0.014	0.022	0.023	0.023
alpha	Non-Seizure	< 0.001	0.005	0.005	0.005
	Seizure	0.008	0.007	0.007	0.007
beta	Non-Seizure	< 0.001	0.004	0.004	0.004
	Seizure	0.008	0.014	0.016	0.016
gamma	Non-Seizure	0.001	0.004	0.005	0.005
	Seizure	0.002	0.003	0.003	0.003
CvM	st. deviations	GGD	Logistic	t-location-scale	Alpha-stable
delta	Non-Seizure	0.001	< 0.001	< 0.001	< 0.001
	Seizure	0.178	0.290	0.258	0.258
theta	Non-Seizure	0.001	0.004	0.003	0.003
	Seizure	0.013	0.064	0.078	0.078
alpha	Non-Seizure	< 0.001	0.007	0.008	0.008
	Seizure	0.005	0.016	0.015	0.015
beta	Non-Seizure	< 0.001	0.006	0.007	0.007
	Seizure	0.028	0.145	0.174	0.174
gamma	Non-Seizure	< 0.001	0.007	0.011	0.011
	Seizure	0.001	0.004	0.005	0.005

TABLE 2.5: Standard deviations of the Kolmogorov-Smirnov (KS) and the Cramer-von-Mises criterion (CvM) scores obtained for GGD pdfs estimated with all EEG segments of 54 events used, 18 *seizures* and 36 *non-seizures*. The GGD shows the lowest scores with respect to the other distributions considered.

2.7 Generalized Gaussian best distribution-fitting

In Section 2.6 different statistical models were compared for their ability to model wavelet coefficients, namely the generalized Gaussian, logistic, t-location-scale, alpha-stable and Cauchy. Based on the visual comparison, it was determined that the generalized Gaussian is the best distribution. Precisely, each scale of the wavelet decomposition is represented by the statistical parameters σ and τ of the generalized Gaussian distribution. The proposed parameter-vector $\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}$ is obtained by collecting the parameters associated with all wavelet scales for a 2-second segment with overlapping 1 second in all brain rhythms. This vector of parameters is adopted as a quantitative descriptor of the EEG signals. It acts as a strong reduction of the dimension of $\boldsymbol{X}$ (eq. (2.1)). Therefore an epileptic signal can be modeled by the probability density function (PDF) of the generalized Gaussian distribution represented by its parameter-vector given by

$$\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}} = \left[\sigma_j^{(i)}, \tau_j^{(i)}\right]^T = \underset{[\sigma,\tau]^T}{\operatorname{argmax}}\, f_{GGD}(\boldsymbol{C}_j^{(i)}; \sigma, \tau) = \left[\boldsymbol{\theta}_\gamma^{\boldsymbol{C}_j^{(i)}}, \boldsymbol{\theta}_\beta^{\boldsymbol{C}_j^{(i)}}, \boldsymbol{\theta}_\alpha^{\boldsymbol{C}_j^{(i)}}, \boldsymbol{\theta}_\theta^{\boldsymbol{C}_j^{(i)}}, \boldsymbol{\theta}_\delta^T\right]^T. \tag{2.25}$$

2.8 Characterization with the generalized Gaussian distribution parameters

2.8.1 Estimation of σ and τ

The electrical brain signal analysis is mostly qualitative according to the level of expertise of the physician. The development of new quantitative methods that can characterize the dynamical changes of the electrical activity is crucial for restricting the subjectivity in the study in epileptic seizures. In section (2.6), the goodness-of-fit test showed that the σ and τ parameters are good descriptors capable to quantify the variation of the EEG signal in both time and frequency. We adopt a pseudo-likelihood approach [129, 130, 131] and construct a log-likelihood function under the next assumption: by ignoring dependency, we get a realistic approximation of the distribution of the wavelet coefficients. However, this approximation significantly simplifies that estimation of the parameters of the model and generally produces accurate estimation results.

The log-likelihood of the sample c having independent component can be expressed as:

$$L(c; \sigma, \tau) = \log \prod_{i=1}^{L} f_{GGD}(c^{(i)}; \sigma, \tau) \tag{2.26}$$

where σ and τ are parameters to be estimated. Maximizing this log-likelihood requires solving the following equations:

$$\frac{\partial L(c; \sigma, \tau)}{\partial \sigma} = -\frac{L}{\sigma} + \sum_{i=1}^{L} \frac{\tau \left|c^{(i)}\right|^\tau \sigma^{-\tau}}{\sigma} = 0 \tag{2.27}$$

$$\frac{\partial L(c; \sigma, \tau)}{\partial \tau} = \frac{L}{\tau} + \frac{L\psi(\frac{1}{\tau})}{\tau^2} - \sum_{i=1}^{L} \left(\frac{\left|c^{(i)}\right|}{\sigma}\right)^\tau \log\left(\frac{\left|c^{(i)}\right|}{\sigma}\right) = 0 \tag{2.28}$$

where $\psi = \frac{\Gamma'(z)}{\Gamma(z)}$ is the digamma function. Fixing $\tau > 0$ in (2.27), the estimation of the scale σ parameter has a unique, real an positive solution:

$$\sigma = \left(\frac{\tau}{L} \sum_{i=1}^{L} \left|c^{(i)}\right|^\tau\right)^{\frac{1}{\tau}}. \tag{2.29}$$

Substituting this into (2.28), the estimation of the shape τ parameter is given by

$$g(\tau) = 1 + \frac{\psi(\frac{1}{\tau})}{\tau} - \frac{\sum_{i=1}^{L} \left|c^{(i)}\right|^\tau \log\left|c^{(i)}\right|}{\sum_{i=1}^{L} \left|c^{(i)}\right|^\tau} + \frac{\log\left(\frac{\tau}{L}\sum_{i=1}^{L} \left|c^{(i)}\right|^\tau\right)}{\tau} = 0 \tag{2.30}$$

Finally using a Newton-Raphson iterative procedure [102], we compute the new guess for the root of $g(\tau)$, τ_{k+1}, based on the previous one, τ_k, using

$$\tau_{k+1} = \tau_k - \frac{g(\tau_k)}{g'(\tau_k)}. \tag{2.31}$$

where

$$\begin{aligned} g'(\tau) = &-\frac{\psi(\frac{1}{\tau})}{\tau^2} - \frac{\psi'(\frac{1}{\tau})^2}{\tau^3} + \frac{1}{\tau^2} - \frac{\sum_{i=1}^{L} \left|c^{(i)}\right|^{\tau} \left(\log\left|c^{(i)}\right|\right)^2}{\sum_{i=1}^{L} \left|c^{(i)}\right|^{\tau}} \\ &+ \frac{\left(\sum_{i=1}^{L} \left|c^{(i)}\right|^{\tau} \log\left|c^{(i)}\right|\right)^2}{\left(\sum_{i=1}^{L} \left|c^{(i)}\right|^{\tau}\right)^2} + \frac{\sum_{i=1}^{L} \left|c^{(i)}\right|^{\tau} \log\left|c^{(i)}\right|}{\tau \sum_{i=1}^{L} \left|c^{(i)}\right|^{\tau}} \\ &- \frac{\log\left(\frac{\tau}{L} \sum_{i=1}^{L} \left|c^{(i)}\right|^{\tau}\right)}{\tau^2} \end{aligned} \tag{2.32}$$

where ψ' is the first poligamma or trigamma function. Note the fact that $g(\tau)$ and $g'(\tau)$ share many common terms which can be used for saving computation at each iteration step in (2.31). For a GGD, it can be shown that the ratio of mean absolute value of the standard deviation is a steadily increasing function of the τ:

$$F(\tau) = \frac{\Gamma(\frac{2}{\tau})}{\sqrt{\Gamma(\frac{1}{\tau})\Gamma(\frac{3}{\tau})}}. \tag{2.33}$$

The initial guess τ_0 of the maximum log-likelihood estimator is given by

$$\tau_0 = F^{-1}\left(\frac{\frac{1}{L}\sum_{i=1}^{L}|c^{(i)}|}{\sqrt{\frac{1}{L}\sum_{i=1}^{L}(c^{(i)})^2}}\right). \tag{2.34}$$

Given a value for τ, it is possible to estimate μ by finding the minimum of

$$\underset{[\mu]}{\operatorname{argmin}} \sum_{i=1}^{L} \left|\boldsymbol{C}^{(i)} - \mu\right|^{\tau}. \tag{2.35}$$

2.8.2 Model based characterization

In order to use the parameters σ and τ as features to classify *seizure* and *non-seizure* EEG segments, we first propose to assess their ability to separate such signals, in each brain rhythm, see section 2.4.3. We consider a dataset composed of n_1 *non-seizure* events and n_2 *seizure* events of EEG segments.

Let 2.36 and 2.37 the set of parameters estimated from *non-seizure* (N) and *seizure* (S) events respectively for a given brain rhythm j, which are the scale and shape parameters of the GGD associated with the wavelet coefficients from seizure events

$$\left(\boldsymbol{\sigma}_j^{(N)}, \boldsymbol{\tau}_j^{(N)}\right) = \left(\left\{\sigma_{1,j}^{(N)}, \ldots, \sigma_{n_1,j}^{(N)}\right\}, \left\{\tau_{1,j}^{(N)}, \ldots, \tau_{n_1,j}^{(N)}\right\}\right) \tag{2.36}$$

$$\left(\boldsymbol{\sigma}_j^{(S)}, \boldsymbol{\tau}_j^{(S)}\right) = \left(\left\{\sigma_{1,j}^{(S)}, \ldots, \sigma_{n_2,j}^{(S)}\right\}, \left\{\tau_{1,j}^{(S)}, \ldots, \tau_{n_2,j}^{(S)}\right\}\right). \tag{2.37}$$

It is assumed that these four parameters are independent and follow normal distributions. This assumption is because our parameters are estimated by using maximum likelihood estimation (MLE). MLE has two properties: consistency and asymptotic normality. That means that when we repeat MLE, the estimated values follow asymptotically a Gaussian distribution [132, 133], see for example the Gaussian fitting in the histograms from Figures 2.2 to 2.6.

$$\boldsymbol{\sigma}_j^{(N)} \sim \mathcal{N}\left(\mu_{\boldsymbol{\sigma}_j}^{(N)}, \sigma_{\boldsymbol{\sigma}_j}^{(N)}\right) \tag{2.38}$$

$$\boldsymbol{\sigma}_j^{(S)} \sim \mathcal{N}\left(\mu_{\boldsymbol{\sigma}_j}^{(S)}, \sigma_{\boldsymbol{\sigma}_j}^{(S)}\right) \tag{2.39}$$

$$\boldsymbol{\tau}_j^{(N)} \sim \mathcal{N}\left(\mu_{\boldsymbol{\tau}_j}^{(N)}, \sigma_{\boldsymbol{\tau}_j}^{(N)}\right) \tag{2.40}$$

$$\boldsymbol{\tau}_j^{(S)} \sim \mathcal{N}\left(\mu_{\boldsymbol{\tau}_j}^{(S)}, \sigma_{\boldsymbol{\tau}_j}^{(S)}\right). \tag{2.41}$$

A univariate T-test [134] was designed to compare the means $\mu_{\boldsymbol{\sigma}}^{(N)}$ and $\mu_{\boldsymbol{\sigma}}^{(S)}$

$$H_0^{(\sigma_j)} : \mu_{\boldsymbol{\sigma}_j}^{(N)} = \mu_{\boldsymbol{\sigma}_j}^{(S)} \tag{2.42}$$

$$H_1^{(\sigma_j)} : \mu_{\boldsymbol{\sigma}_j}^{(N)} \neq \mu_{\boldsymbol{\sigma}_j}^{(S)}. \tag{2.43}$$

The variances of the distributions (2.38)-(2.41) are not equal and unknown. Consequently, we designed the test as follows:
Let $\bar{\boldsymbol{\sigma}}_j^{(N)}$ and $\bar{\boldsymbol{\sigma}}_j^{(S)}$ denote the empirical conditional means of $\boldsymbol{\sigma}_j^{(N)}$ and $\boldsymbol{\sigma}_j^{(S)}$, and $D_{\boldsymbol{\sigma}_j} = \bar{\boldsymbol{\sigma}}_j^{(N)} - \bar{\boldsymbol{\sigma}}_j^{(S)}$ their difference. Denoting as $s^2_{\boldsymbol{\sigma}_j^{(N)}}$ and $s^2_{\boldsymbol{\sigma}_j^{(S)}}$ the unbiased estimators of the variances in each group of events, i.e. *seizure* and *non-seizure*, the standard deviation of $D_{\boldsymbol{\sigma}_j}$ can be estimated as

$$\hat{\sigma}_{D_{\sigma_j}} = \sqrt{\frac{s^2_{\boldsymbol{\sigma}_j^{(N)}}}{n_1} + \frac{s^2_{\boldsymbol{\sigma}_j^{(S)}}}{n_2}}. \tag{2.44}$$

The statistics of the T-test associated with (2.42) and (2.43) is then

$$T_\nu^{(\sigma_j)} = \frac{\bar{\boldsymbol{\sigma}}_j^{(N)} - \bar{\boldsymbol{\sigma}}_j^{(S)}}{\sqrt{\frac{s^2_{\boldsymbol{\sigma}_j^{(N)}}}{n_1} + \frac{s^2_{\boldsymbol{\sigma}_j^{(S)}}}{n_2}}}, \tag{2.45}$$

which is distributed according to a Student's t-distribution with ν degrees of freedom,

$$\nu = \frac{\left(\frac{s^2_{\sigma_j^{(N)}}}{n_1} + \frac{s^2_{\sigma_j^{(S)}}}{n_2}\right)^2}{\frac{s^4_{\sigma_j^{(N)}}}{n_1^2(n_1-1)} + \frac{s^4_{\sigma_j^{(S)}}}{n_2^2(n_2-1)}}. \tag{2.46}$$

The hypothesis $H_0^{(\sigma_j)}$ is rejected if $\left|T_\nu^{(\sigma_j)}\right| > T_t$ and we chose a probability of false alarm $t = 0.05$. To assess the statistical significance, the p-value of each test has been calculated. Table 2.6 shows the decision rules that were applied.

p-value	Observed difference
> 0.10	not significant
≤ 0.10	marginally significant
≤ 0.05	significant
≤ 0.01	highly significant

TABLE 2.6: Decision rules to asses the statistical significance of the difference of means of the GGD parameters for seizure and non-seizure signals.

A similar test has also been designed to compare $\mu_{\tau_j}^{(N)}$ and $\mu_{\tau_j}^{(S)}$ for each brain rhythm j. A bivariate T-test has also been designed for the pair (σ_j, τ_j). Its results were not significant, therefore it is not reported here.

In addition, to further support the statistical significance given by the p-value, we calculated the Bayes factor indicator following the method proposed by [135]. This method establishes a correspondence between frequency significance tests, such as the ones designed here, with Bayesian tests. As a result, it allows one to equate the size of the classical hypothesis tests with evidence thresholds in Bayesian tests. Following this work (and assuming equal variances), we calculated the Bayes factor (BF) that provides the same evidence as the p-values given by our tests

$$BF = \left(\frac{\nu + T_\nu^{(\sigma_j)}}{\nu + \left(T_\nu^{(\sigma_j)} - \sqrt{(\nu\gamma^*)}\right)^2}\right)^{(n_1+n_2)/2} \tag{2.47}$$

where the hypothesis H_0 is rejected when $t > \sqrt{\nu\gamma^*}$ with $\gamma^* = \gamma^{2/(n_1+n_2-1)} - 1$ and $\gamma = ((T_t^{(\sigma_j)})^2/\nu - 1)^{(n_1+n_2)/2}$.

These statistical tests were implemented with the parameters obtained from all events in all epochs according to section 2.2, with $n_1 = 36$ *non-seizure* events and $n_2 = 18$ *seizure* events. Table 2.7 shows the T-scores, and their associated p-values and Bayes factors. The corresponding thresholds are shown. We observe that the t-scores ($T_\nu^{\sigma_\cdot}$) are all greater than the threshold (T_t). The corresponding p-values (p) are all lower than 0.01. The equivalent Bayes factors (BF) are also all greater than the threshold (BF_t). The H_0^σ hypothesis is therefore rejected for all bands, with high statistical significance according to the decision rules presented in Table 2.6. The scale

parameter σ is a good marker to distinguish *seizure* and *non-seizure* EEG events. Contrarily, t-scores for τ are lower than the threshold, except for the Delta band. The associated p-values are higher than 0.1. The Bayes factors are lower than the thresholds. Consequently, H_0^{τ} hypothesis is accepted implying that *beta* band cannot discriminate *seizure* and *non-seizure* EEG events. Based on these results, it becomes credible to classify EEG into two classes *seizure* and *non-seizure* based on the scale parameter σ of the GGD associated with their wavelet coefficients in each brain rhythm.

	Delta Band		Theta Band		Alpha Band		Beta Band		Gamma Band	
GGD Parameter	$\boldsymbol{\sigma}$	$\boldsymbol{\tau}$	$\boldsymbol{\sigma}$	$\boldsymbol{\tau}$	$\boldsymbol{\sigma}$	$\boldsymbol{\tau}$	$\boldsymbol{\sigma}$	$\boldsymbol{\tau}$	$\boldsymbol{\sigma}$	$\boldsymbol{\tau}$
T_{ν}	6.15	3.19	5.86	0.17	6.47	0.50	7.08	0.48	6.40	0.91
T_t	2.09	2.01	2.09	2.03	2.09	2.01	2.07	2.03	2.08	2.01
p	$<$0.001	$<$0.001	$<$0.001	0.90	$<$0.001	0.62	$<$0.001	0.63	$<$0.001	0.37
BF	$>$1000	98.31	$>$1000	0.03	$>$1000	0.30	$>$1000	0.08	$>$1000	0.64
BF_t	3.63	2.12	3.70	2.45	3.79	2.09	3.39	2.46	3.51	2.17

TABLE 2.7: Results of the t-tests to assess the ability of σ and τ to discriminate separately *seizure* and *non-seizure* EEG. The H_0^{σ} hypothesis is rejected for all rhythms, with highly statistical significant ($p < 0.01$). These scores are supported by very high Bayes factors. Contrarily, H_0^{τ} is accepted for all rhythms, except Delta band. The associated p-values are largely greater than 0.1 with Bayes factors lower than the evidence threshold. We conclude that the scale parameter σ is a marker to discriminate *seizure* and *non-seizure* events with a high statistical significance. The shape parameter τ is not a marker to discriminate *seizure* and *non-seizure* events.

2.9 Kullback-Leibler divergence (KLD)

The Kullback-Leibler divergence (KLD) or relative entropy [136] is used to measure the discrepancy or similarity between probability density functions (PDF) [137, 138, 139, 140]. Specifically, we used KLD between the PDFs of the generalized Gaussian distribution for EEG signals in order to discriminate *seizure* from *non-seizure*. See [141, 142, 143] for some works on this topic in epilepsy and [144, 145, 146] for some applications in EEG Signals.

Let p and q two PDFs, then a Kullback-Leibler Divergence (KLD) is given by

$$D_{KL}(p||q) = \int_{-\infty}^{\infty} \log\left(\frac{p_x(x)}{q_x(x)}\right) p_x(x)dx \tag{2.48}$$

$$= -\int_{-\infty}^{\infty} \log(q_x(x))p_x(x)dx + \int_{-\infty}^{\infty} \log(p_x(x))p_x(x)dx \tag{2.49}$$

$$\tag{2.50}$$

Notice that in general $D_{KL}(p||q) \neq D_{KL}(q||p)$, and that $D_{KL}(p,q) = 0$ if and only if $p = q$ [138]. Rewriting (2.13), the probability density function of GGD is given by

$$p(x,\sigma,\tau) = \frac{1}{2\sigma\Gamma[1+\tau^{-1}]} \exp\left(-\left|\frac{x}{\sigma}\right|^{\tau}\right). \tag{2.51}$$

The divergence between two generalized Gaussian models with parameters (σ_1,τ_1,μ_1) and (σ_2,τ_2,μ_2) using eq. (2.51) subject to the constraint $\mu_1 = \mu_2 = 0$ (since our wavelet coefficients have zero-mean) given by

$$\begin{aligned} KLD_{pdf}(p||q) &= \int_{-\infty}^{\infty} p(x,\sigma_1,\tau_1)\log\left(\frac{p(x,\sigma_1,\tau_1)}{p(x,\sigma_2,\tau_2)}\right)dx \\ &= \int_{-\infty}^{\infty} p(x,\sigma_1,\tau_1)\left[\left|\frac{x}{\sigma_2}\right|^{\tau_2} - \left|\frac{x}{\sigma_1}\right|^{\tau_1} + \log\left(\frac{\sigma_2\Gamma(1+\frac{1}{\tau_2})}{\sigma_1\Gamma\left(1+\frac{1}{\tau_1}\right)}\right)\right]dx \\ &= \frac{-\Gamma\left(\frac{1}{\tau_1}\right) + \left(\frac{\sigma_1}{\sigma_2}\right)^{\tau_2}\tau_1\Gamma\left(\frac{1+\tau_2}{\tau_1}\right) + \tau_1\Gamma\left(\frac{1}{\tau_1}\right)\log\left(\frac{\sigma_2\Gamma\left(1+\frac{1}{\tau_2}\right)}{\sigma_1\Gamma\left(1+\frac{1}{\tau_1}\right)}\right)}{\tau_1\Gamma\left(\frac{1}{\tau_1}\right)} \\ &= -\frac{1}{\tau_1} + \frac{\left(\frac{\sigma_1}{\sigma_2}\right)^{\tau_2}\Gamma\left(\frac{1+\tau_2}{\tau_1}\right)}{\Gamma\left(\frac{1}{\tau_1}\right)} + \log\left(\frac{\sigma_2\Gamma\left(1+\frac{1}{\tau_2}\right)}{\sigma_1\Gamma(1+\frac{1}{\tau_1})}\right) \end{aligned} \tag{2.52}$$

In our study, we estimated the parameters (σ,τ) for each 1-second segment of the EEG signal. The estimated GGD distributions for all segments were compared using the KLD metric (eq. (2.52)) according to the following rules:

1. Between the PDFs $p^{(i)}$ of the sliding window and the PDF of the annotated seizure onset q

$$KLD_{pdf}(p^{(i)}||q^{onset}) = \mathbf{\Omega}^{(i)}KLD_{pdf}(p||q)$$

2. Between adjacent PDFs coupled with a 7-order one-dimensional median filter [147]

$$KLD_{pdf}(p^{(i)}||q^{(i+1)}) = \mathbf{\Omega}^{(i)}\mathbb{F}^{(i)}KLD_{pdf}(p||q)$$

with

$$\mathbb{F}^{(i)} = medianFilter(KLD_{pdf}(p^{(i)}||q^{(i)}))$$

See eq. (2.2) for more details about the sliding window.

2.9.1 A visual EEG epilepsy detection experience

A good performance of KLD method via visual inspection by an experienced neurologist from FLENI was obtained in 8 signals studied from the database described in Section 2.2. We use an example to show the proposed method. Figures 2.7 and 2.8 depicts the different brain rhythms: *delta*, *theta*, *alpha*, *beta* and *gamma*, where the *seizure* is 40 seconds of duration. We can see an increase in the activity between 2 minutes and 2.4 minutes in all brain rhythms. In Figure 2.7 one can notice that the signal shows a *seizure* onset starting at minute two. We can see clearly the discrepancy between *seizure* or *non-seizure* in epileptic signals; while in Figure 2.8 the *seizure*

onset is detected clearly given by the highest peak which emerges from the background of EEG showing a discrepancy between *seizure* or *non-seizure*. Once the *seizure* finished, there are several medical pathological factors that cause the signal to take time to stabilize. This explains why the *seizure* does not have an instantaneous change after 2 minutes 40 seconds. Nevertheless, the discrepancy is very clear after the seizure.

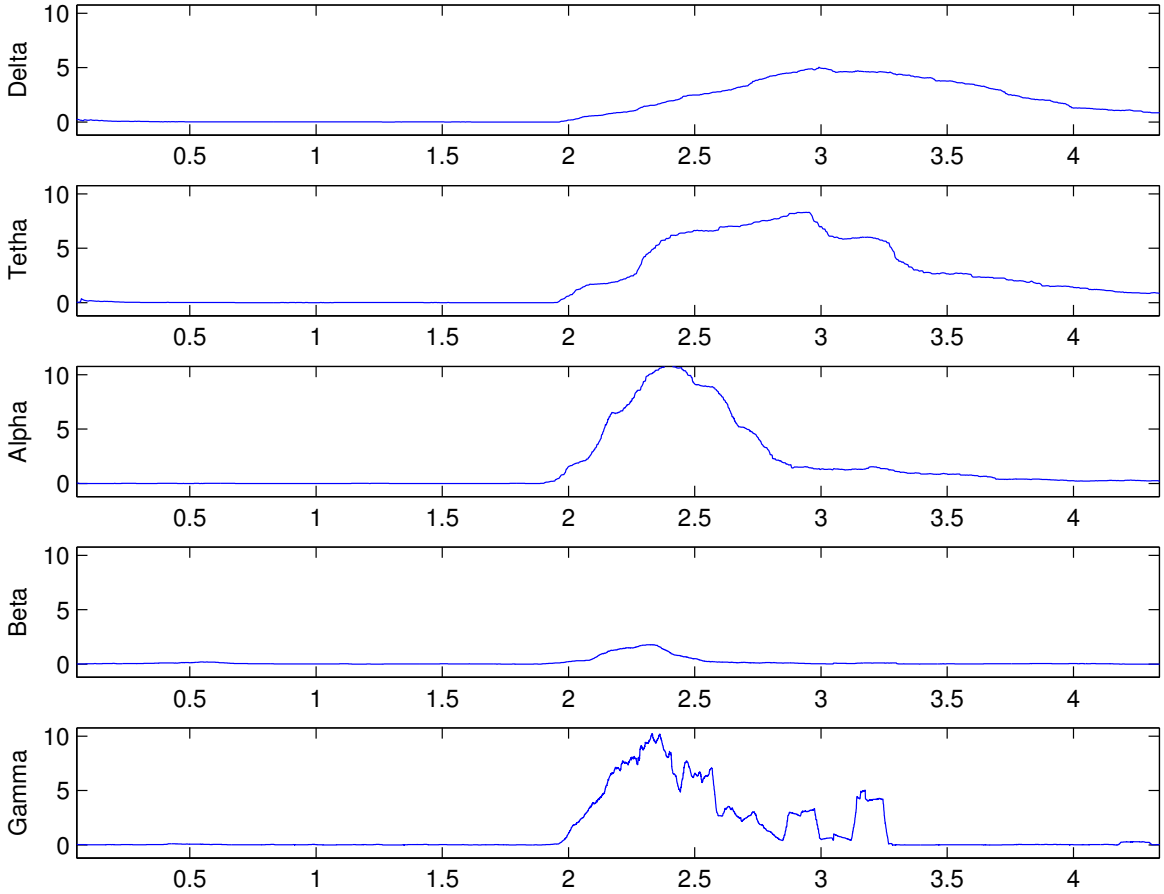


FIGURE 2.7: KLD between the PDFs of the sliding window and the PDF of the *seizure* onset of the epileptic signal, showing a clear discrepancy between *seizure* or *non-seizure*. In this example, the *seizure* onset begins at minute 2, and its duration is 40 seconds.

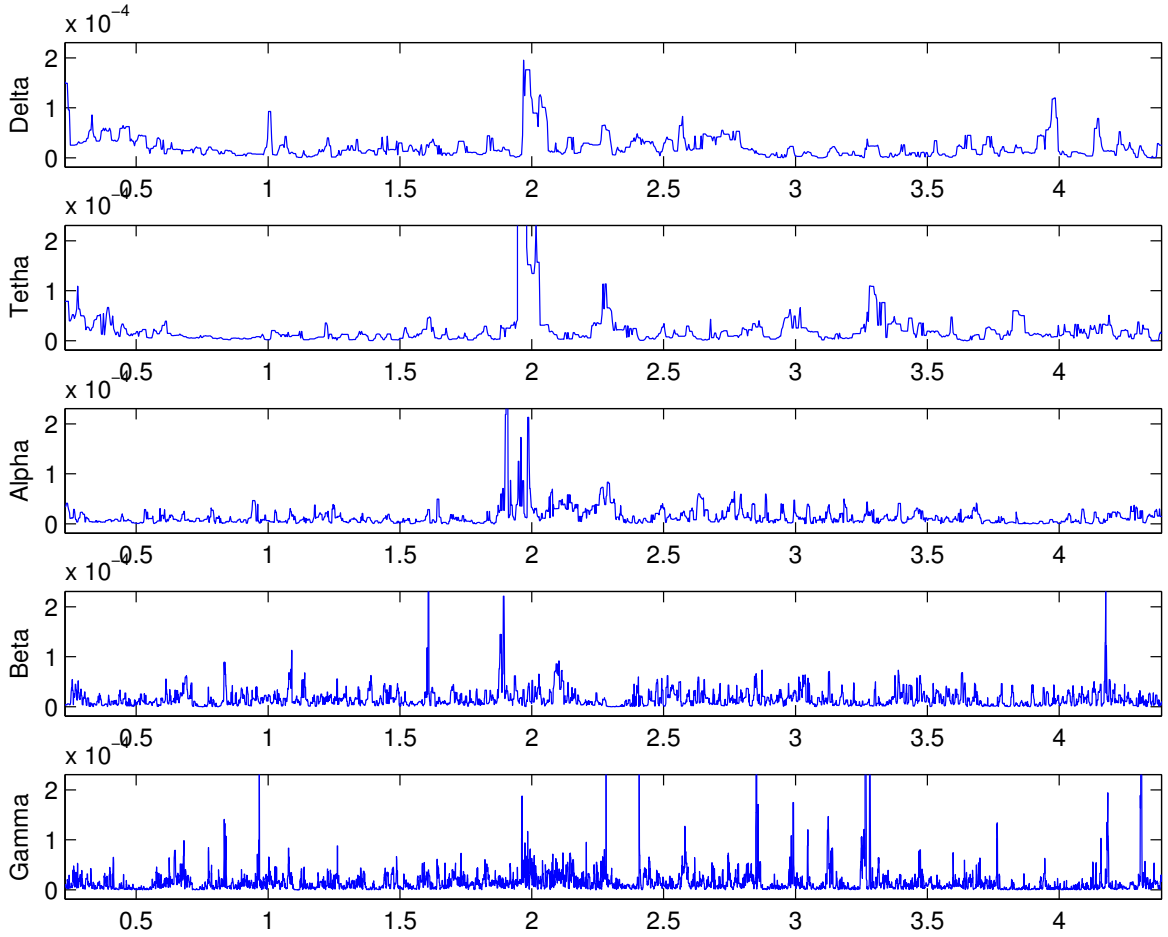


FIGURE 2.8: KLD between adjacent PDFs coupled with a 7-order one-dimensional median filter, showing clearly the discrepancy given by the highest peak which emerges from the background of EEG. In this example, the *seizure* onset begins at minute 2, and its duration is 40 seconds.

2.10 Conclusions

Seizure activity characterization requires an efficient and accurate statistical modeling. In this chapter, the general framework to achieve good epileptic seizure activity detection and quantification in EEG signals was presented. We have compared different statistical models and discussed the best distribution-fitting of the brain rhythms (or wavelet coefficients). The generalized Gaussian statistical model whose parameters can be found by maximum-likelihood estimation was the best distribution. Each brain rhythm is then mapped to a low-dimensional manifold by this model, which can be implemented in real time and makes possible developing classification algorithms with low complexity. Additionally, an analytical Kullback-Leibler divergence (KLD) was developed for the generalized Gaussian in order to detect epileptic seizures. A study that will be developed deeper in future work.

Chapter 3
Seizure onset detection and temporal spread estimation

3.1 Introduction

In chapter 2, we have established a generalized Gaussian statistical model for EEG data and derived the characterization of epileptic EEG signals using the parameters of this model. Precisely, we have shown that the scale and shape parameters of the Gaussian distribution are relevant features to classify EEG signals and detect epileptic seizures. This chapter starts from the fact that a seizure onset is a sudden change in the spectral energy distribution, which exhibit in a set of EEG channels. Such seizure progresses and spreads throughout the brain, while its characteristics evolve. We propose the idea that the scale parameter (which depends on the shape parameter) is characteristic of the variability of brain activity. Consequently, we develop a new algorithm that shows that the scale parameter is a descriptor that allows seizure onset detection and his spread across different brain rhythms in both focal and generalized seizures in epileptic EEG signals.

3.2 Scale parameter sigma (σ)

Based on the brain rhythm decomposition of EEG signals using multilevel 1D wavelets (see sections 2.4.1 and 1.3) and the associated GGD parameters, σ and τ (see eq. (2.13) and eq. (2.25)), a new algorithm for seizure onset detection and temporal spread estimation was developed. The purpose of this algorithm is to show that *the scale parameter σ (which depends on the shape parameter) is closely related to the variability of the brain activity and is, therefore, a good descriptor for performing seizure onset detection (SOD)*. Each signal in each channel were edited to have an epoch with the following characteristics: 2 minutes before the seizure, seizure at minute 2 and 2 minutes after the seizure. For each epoch, we know where the seizure begins (see section 2.2) and can calculate the time delay for the onset based on a thresholding approach (see section 1.6). Table 3.1 reports the minimum value, maximum value and mean of σ parameter for all EEG signals in 36 *non-seizure* events and 18 *seizure* events, which allow using a threshold approach [78]. Precisely, based on the σ mean for each brain rhythm, it is possible to differentiate between epileptic events, see Table 3.2.

σ	min	max	mean
Non-Seizure	1375.45	474942.97	69221.84
Seizure	4383.92	1476232.96	331300.79

TABLE 3.1: Minimum value, maximum value and mean of σ parameter for all signals in 36 non-seizure events and 18 seizure events. The mean value allows using a thresholding approach with the scale parameter.

σ	Non-Seizure			Seizure		
Bands	min	max	mean	min	max	mean
Delta	935.85	381863.65	58908.66	2579.55	1264885.38	279701.55
Theta	12.56	3774.65	359.08	10.26	11187.93	1834.65
Alpha	50.95	5216.49	544.67	92.77	17783.92	2890.21
Beta	198.17	42831.14	5253.32	583.46	100595.75	25139.34
Gamma	177.91	41257.05	4156.10	1117.88	81779.97	21735.05

TABLE 3.2: Minimum value, maximum value and mean of σ parameter for each brain rhythm in 36 non-seizure events and 36 seizure events. The mean value allows using a thresholding approach with the scale parameter.

The proposed algorithm requires solving basically three problems. The first problem is related to the seizure onset detection, where the scale parameter helps to detect a sudden change in the EEG signal by using the statistical-threshold crossing of the mean of σ, see Table 3.2. The second problem is the moment estimation in which this sudden change begins. This estimation is given by the time delay between the annotated seizure onset detection and the statistical-threshold crossing, given by the mean of σ. Finally, the combination of different time delays in each brain rhythm is given for each brain area according to the type of seizure, such as focal or generalized. Therefore, the onset time delay between the annotated seizure onset data and the detection was estimated using the mean of σ for *non-seizures* and *seizures* events. Once onset delay estimation has been performed for each brain rhythm and each channel, we collect this information off-line in a table, such as Table 3.3. This table allows identifying channels with low delay and thus understanding how the seizure originated and propagated temporarily. The most common electrode to all channels in the brain area is the best possible candidate.

For illustration, we will focus on one example with a generalized seizure. But in general, no distinction was considered regarding the types of seizure such as focal or generalized because the goal is to show the possibilities of the analysis of our algorithm through the scale parameter (which depends on the shape parameter). See the example in Table 3.3, where channels 13 (Fp2-F8), 14 (F8-T8), 15 (T8-P8), 21 (FT9-Ft10) and 23 (T8-P8) have a delay in all brain rhythms, therefore they are the candidate-electrodes. Note that these electrodes correspond to the right temporal brain area, see Figure 3.1. FT9 or FT10 have the least delay, but T8 is the most common electrode to all according to the clinical observations, therefore it is considered the best candidate. It is interesting to notice that for the observed activation region, the candidate channels involved belong all to the same region. This provides hints on the temporal spread. Figure 3.1 and Figures

3.5 to 3.7 show a correlation between the different brain areas and the electrode positions. Note that, despite the fact that channel 1 has a previous detection to any of the indicated channels, it was discarded by the neurologist because this brain area doesn't have a generalized seizure through all the brain bands. But it opens interesting research questions to explore deeper. They are related to the analysis of the frequency bands detected before the onset seizure, and which channels are active despite not participating directly in the detection of the seizure onset, see for example the negative value -0.099 in channel 1 for delta band in Table 3.3.

Channel Number	Channel Name	Delta	Theta	Alpha	Beta	Gamma
1	FP1-F7	-0.099	0.001	0.002	0.130	0.131
2	F7-T7	-	0.069	0.234	0.131	0.058
3	T7-P7	-	0.037	0.215	0.369	0.371
4	P7-O1	-	-	-	0.224	0.162
5	FP1-F3	-	0.037	0.011	0.133	0.133
6	F3-C3	-	-	-	0.221	0.132
7	C3-P3	-	-	-	0.378	0.140
8	P3-O1	-	-	-	0.584	0.552
9	FP2-F4	0.083	0.060	0.037	0.212	0.098
10	F4-C4	-	0.001	-	0.125	0.110
11	C4-P4	-	0.018	0.029	-	0.103
12	P4-O2	-	0.035	-	-	-
13	FP2-F8	0.105	0.012	0.025	0.021	0.101
14	F8-T8	0.073	0.087	0.134	0.097	0.100
15	T8-P8	0.081	0.018	0.013	0.099	0.103
16	P8-O2	-	0.036	0.016	0.258	0.086
17	FZ-CZ	-	-	-	0.034	0.141
18	CZ-PZ	-	-	-	-	-
19	P7-T7	0.086	0.085	0.040	2.360	0.350
20	T7-FT9	-	0.069	-	0.372	0.377
21	FT9-FT10	0.005	0.001	0.016	0.098	0.117
22	FT10-T8	0.075	0.066	0.010	0.102	0.101
23	T8-P8	0.065	0.032	0.015	0.098	0.117

TABLE 3.3: Onset delay (in seconds) by frequency bands, the symbol (-) means that there is no clear difference at the beginning of the onset. In this example, according to the data, channels (21) FT9-FT10, (23) T8-P8, (13) FP2-F8, (14) F8-T8 and (15) T8-P8 are the best candidates by brain area for SOD because it has minor delays; in this case T8 is common to all channels in this brain area, therefore is the best option for being the onset and other channels form the spread.

In the next sections, we are going to introduce the proposed algorithm (Section 3.3), which was applied to 8 real EEG signals from 8 patients suffering from epileptic seizures. By using one example with generalized seizure, the performance of the algorithm is demonstrated across all brain

rhythms, namely the delta, theta, alpha, beta, and gamma bands, (Section 3.4). Next, a validation stage is used by using the Brainstorm software with the purpose of validating the qualitative results made by visual inspection by the neurologist from FLENI (See Section 3.5).

3.3 Algorithm

The proposed algorithm can be summarized as follows:

Data: EEG raw

Result: Temporal spread estimation

for *each brain rhythm* **do**

1. Estimate the GGD parameters σ and τ, see eq. (2.25);
2. Use the scale parameter σ to calculate the time delays of each SOD for each channel and each brain rhythms;
3. Create a table with the information and calculate each delay for each channel;
4. Organize the table ascendantly;
5. Calculate the Seizure Onset Detection (SOD) candidate channels for the different brain areas namely Frontal, Parietal, Temporal and Occipital, see Figure 3.1;

The candidates for the onset are the channels that are common to all other channels by brain rhythm in a determined brain area; and the other channels are the possible temporal spread accompanying the seizures.

end

Algorithm 1: Temporal spread estimation algorithm

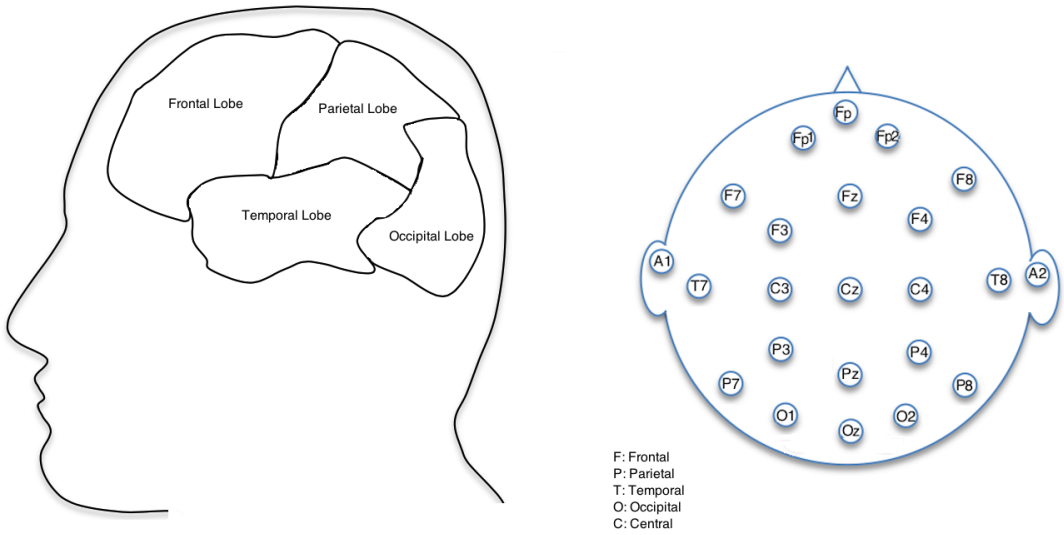


FIGURE 3.1: Areas of the brain of the Cerebral Cortex (Lateral view) can be correlated with the electrodes positions, which have a nomenclature according to the lobules, namely (T) temporal lobe, (P) parietal lobe, (O) occipital lobe and (F) frontal lobe.

Figure 3.1 shows the four lobe divisions of the cerebral cortex. These areas can be correlated with the electrode positions whose nomenclature is directly related, (F) of Frontal Lobule, (P) of Parietal Lobule, (T) of Temporal Lobule and (O) Occipital Lobule, see Figure 1.1.

3.4 Experimental Results

The performance of the proposed algorithm for seizure temporal spread estimation was assessed using 8 signals from the Children's Hospital Boston database described in Section 2.2, and compared with results obtained by qualitative visual inspection by an experimented neurologist from FLENI, who relied on EEG and MRI data, see Figure 3.8. In the next sections, we present the seizure detection at minute 2 and estimated spread across the cerebral cortex for each rhythm brain in one example with a generalized seizure.

3.4.1 General Spread

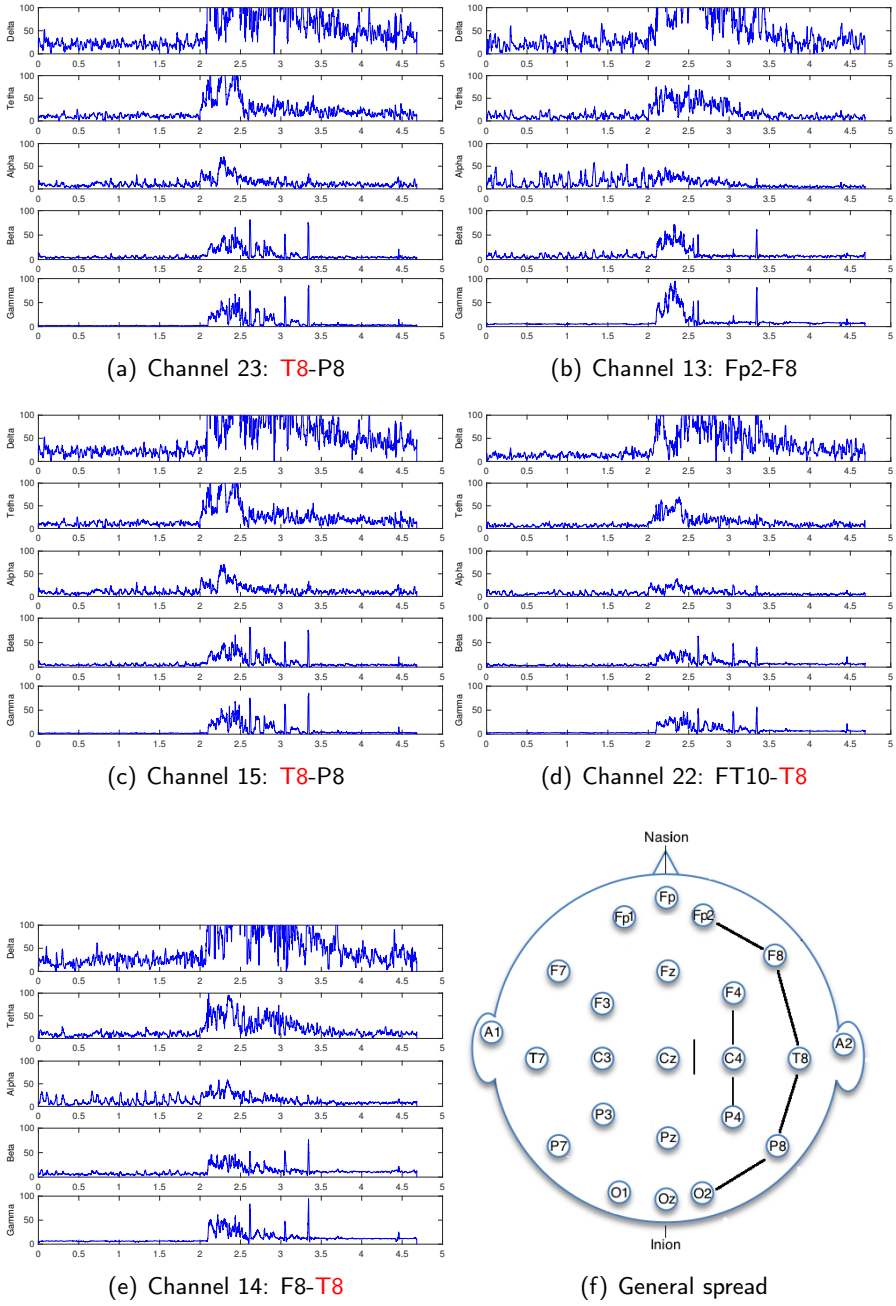
The channels with the smaller delay in the right temporal brain area are (21) FT9-FT10, (23) T8-P8, (13) FP2-F8, (15) T8-P8, (22) FT10-T8, (14) F8-T8, (9) FP2-FP4, (19) P7-T7. This distribution suggests that T8 can be the onset of the seizure. Figure 3.2 reports the spread across the cerebral cortex namely: (F) Frontal, (T) Temporal and (P) Parietal areas, with a small share in the (C) Central area. This result suggests a simple seizure across the right hemisphere of the cerebral cortex.

3.4.2 Delta Band Spread

The delta band sequence corresponding to low delays is: (10) F4-C4, (21) FT9-FT10, (11) C4-P4, (23) T8-P8, (14) F8-T8, (22) FT10-T8, (15) T8-P8, (9) FP2-F4, (19) P7-T7, (1) FP1-F7, (13) FP2-F8. Figure 3.3 reports the spread across the cerebral cortex namely: (F) Frontal and (T) Temporal and (C) Central areas. This result suggests a delta activity across the right hemisphere and isolated activity in the left hemisphere of the cerebral cortex.

3.4.3 Theta Band Spread

The theta band sequence corresponding to low delays is: (1) FP1-F7, (21) FT9-FT10, (13) FP2-F8, (15) T8-P8, (11) C4-P4, (23) T8-P8, (16) P8-O2, (3) T7-P7, (5) FP1-F3, (9) FP2-F4, (22) FT10-T8, (2) F7-T7, (20) T7-FT9, (19) P7-T7, (14) F8-T8. Figure 3.4 reports the spread across the cerebral cortex namely: (F) Frontal, (T) Temporal and (P) Parietal areas, with a small share in the (C) Central area. This result suggests theta activity on temporal area across both hemispheres of the cerebral cortex.



(a) Channel 23: T8-P8 (b) Channel 13: Fp2-F8

(c) Channel 15: T8-P8 (d) Channel 22: FT10-T8

(e) Channel 14: F8-T8 (f) General spread

FIGURE 3.2: General Spread: EEG electrodes array, suggests that the seizure starts in channel T8 and then the spread across the channels F8-T8-P8, F4-C4-P4. Amplitude (y-axis) in mV and time (x-axis) in min.

3.4.4 Alpha Band Spread

The alpha band sequence corresponding to low delays is: (1) FP1-F7, (22) FT10-T8, (5) FP1-F3, (15) T8-P8, (23) T8-P8, (16) P8-O2, (21) FT9-FT10, (13) FP2-F8, (12) P4-O2, (9) FP2-F4, (19) P7-T7, (14) F8-T8, (3) T7-P7, (2) F7-T7. Figure 3.5 reports the spread across the cerebral cortex namely: (F) Frontal, (T) Temporal and (P) Parietal areas. This result suggests alpha activity spread across both hemispheres of the cerebral cortex, predominating the right side.

3.4.5 Beta Band Spread

The beta band sequence corresponding to low delays is: (13) FP2-F8, (17) FZ-CZ, (14) F8-T8, (21) FT9-FT10, (23) T8-P8, (15) T8-P8, (22) FT10-T8, (10) F4-C4, (1) FP1-F7, (2) F7-T7, (5) FP1-F3, (9) FP2-F4, (6) F3-C3,(4) P7-O1, (16) P8-O2, (3) T7-P7, (20) T7-FT9, (7) C3-P3, (8) P3-O1, 19 P7-T7. Figure 3.6 reports the spread across the cerebral cortex namely: (F) Frontal, (T) Temporal and (P) Parietal areas. This result suggests beta activity spread across both hemispheres of the cerebral cortex.

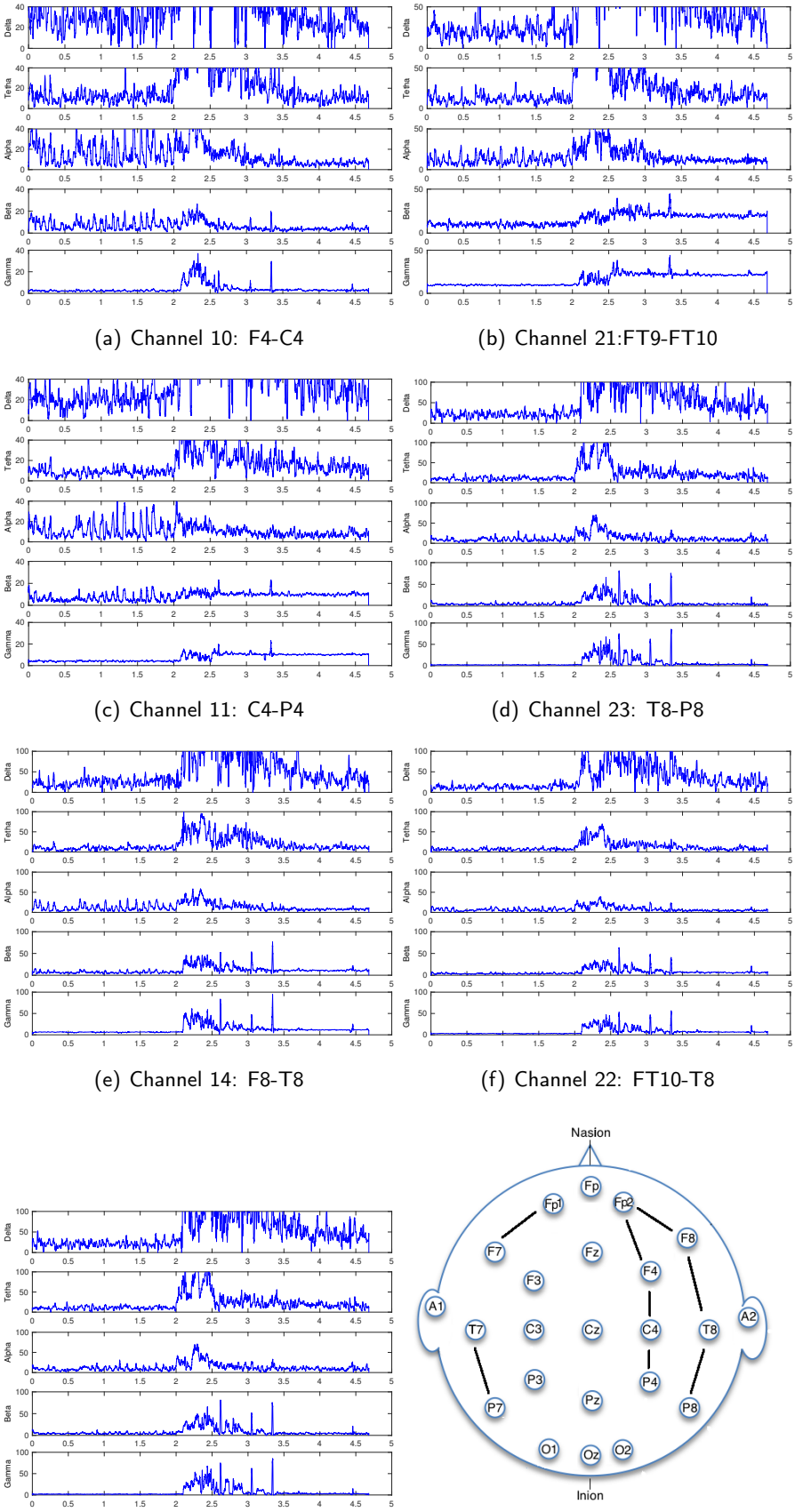
3.4.6 Gamma Band Spread

The delta gamma sequence corresponding to low delays is: (2) F7-T7, (16) P8-O2, (9) FP2-F4, (14) F8-T8, (13) FP2-F8, (22) FT10-T8, (11) C4-P4, (15) T8-P8, (10) F4-C4, (21) FT9-FT10, (23) T8-P8, (1) FP1-F7, (6) F3-C3, (5) FP1-F3, (7) C3-P3, (17) FZ-CZ, (4) P7-O1, (19) P7-T7, (3) T7-P7, 20 T7-FT9, (8) P3-O1. Figure (3.7) reports the spread across the cerebral cortex namely: (F) Frontal, (T) Temporal and (P) Parietal areas. This result suggests gamma activity spread across both hemispheres of the cerebral cortex.

3.5 Validation

In Section 3.4, we applied our algorithm described in Section 3.3 and we predicted the seizure onset with its temporal spread in each channel of each epileptic signal. For the validation process, we use Brainstorm software (this is a collaborative, open-source application dedicated to MEG/EEG/sEEG/ECoG data analysis, such as visualization, processing and advanced source modeling [148]) in two stages: MRI image and EEG raw signals. For each MRI of each subject, we used the default surfaces of Brainstorm, which allow calculating the coordinate system to create a realistic head model. The underlying method consists in exploiting anatomy and surfaces (information on head tissues and sensor characteristics) extracted from the image, see Section A.3.2 from appendix A. This MRI is coupled with each EEG signal for each subject in order to project the electrical activity in the image, see Figure 3.8. Each final image was validated by qualitative visual inspection by an experimented neurologist from FLENI relying on EEG and MRI data.

In other words, it is necessary to estimate the brain sources which produced the data, according to the head and information of the sensor array by solving the inverse problem. We used standardized low-resolution brain electromagnetic tomography (sLORETA) algorithm, see Section A.6.5, which assumes that all possible locations of the sources are simultaneous, see Section A.5.2 from appendix A.



(a) Channel 10: F4-C4

(b) Channel 21:FT9-FT10

(c) Channel 11: C4-P4

(d) Channel 23: T8-P8

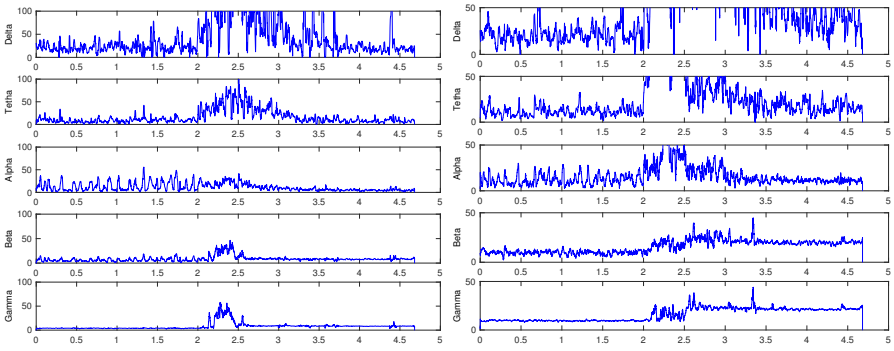
(e) Channel 14: F8-T8

(f) Channel 22: FT10-T8

(g) Channel 15: T8-P8

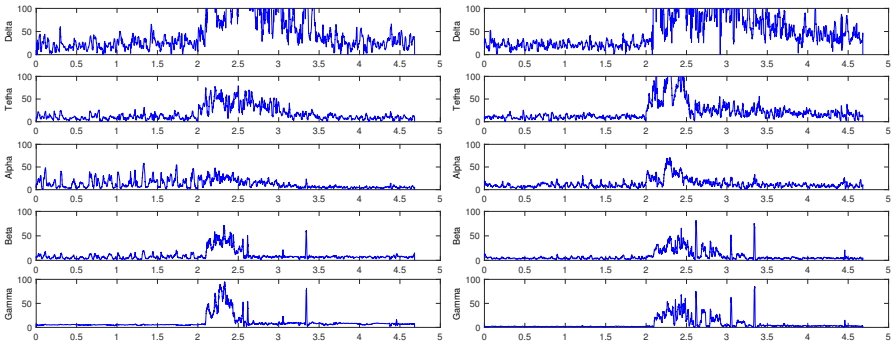
(h) Delta spread

FIGURE 3.3: Delta Band Spread: EEG electrodes array, suggests delta activity spread across the right hemisphere and isolated activity in the left hemisphere. Amplitude (y-axis) in mV and time (x-axis) in min.



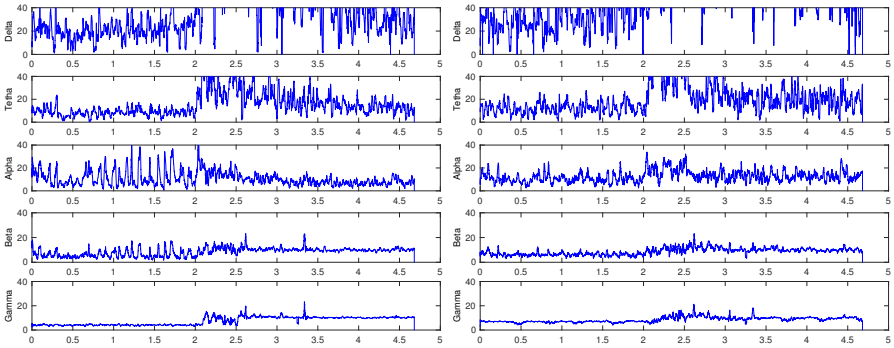
(a) Channel 1: Fp1-F7

(b) Channel 21:FT9-FT10



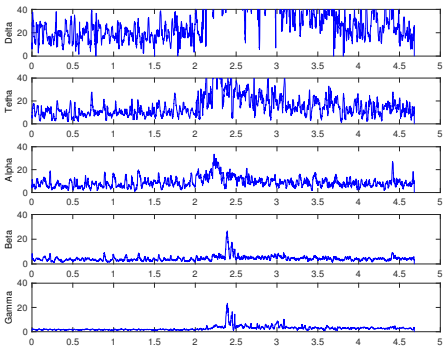
(c) Channel 13: C4-P4

(d) Channel 15: T8-P8

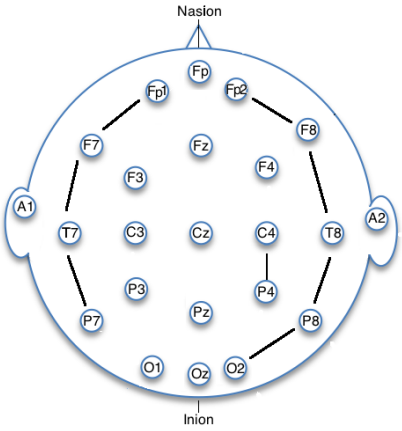


(e) Channel 11: C4-P4

(f) Channel 16: P8-O2

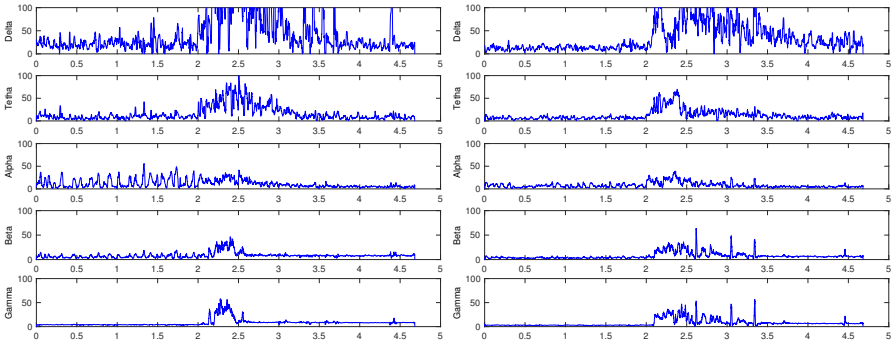


(g) Channel 3: T7-P7



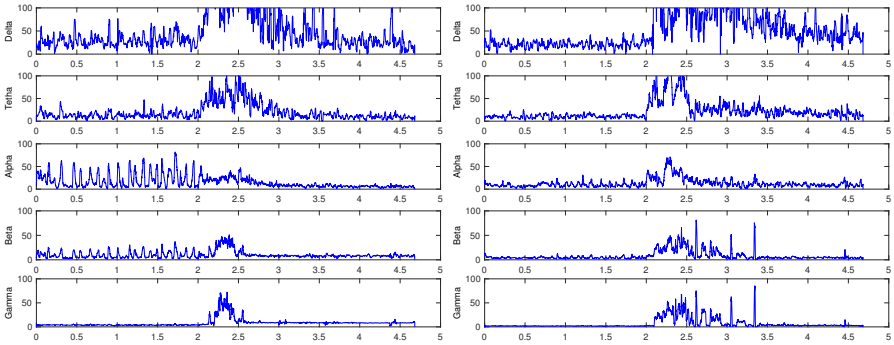
(h) Theta spread

FIGURE 3.4: Theta Band Spread: EEG electrodes array, suggests theta activity spread across both hemispheres on temporal area of the cerebral cortex. Amplitude (y-axis) in mV and time (x-axis) in min.



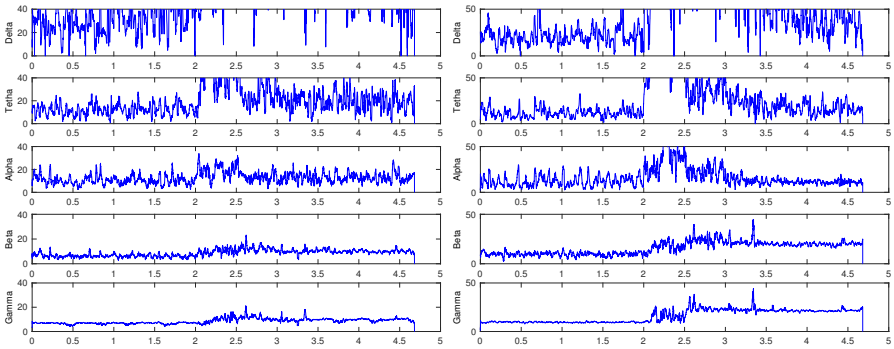
(a) Channel 1: Fp1-F7

(b) Channel 22: FT10-T8



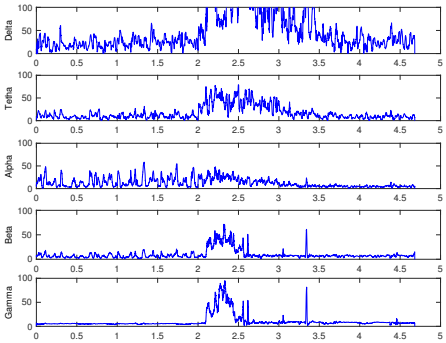
(c) Channel 5: FP1-F3

(d) Channel 15: T8-P8

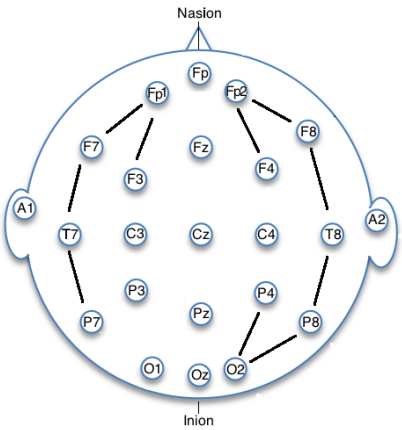


(e) Channel 16: T8-P8

(f) Channel 21: FT10-T8

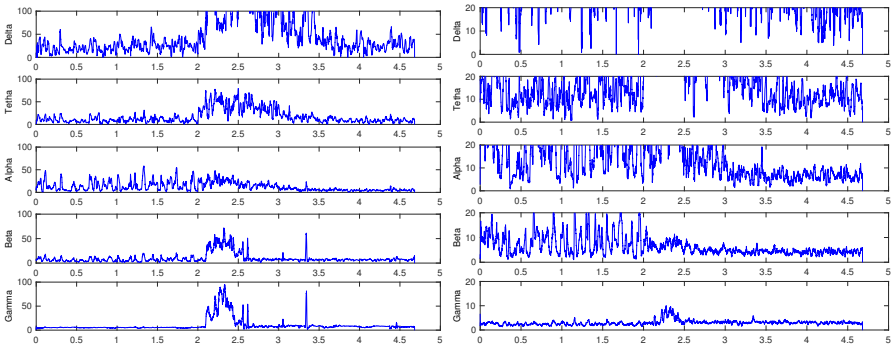


(g) Channel 13: FP2-F8



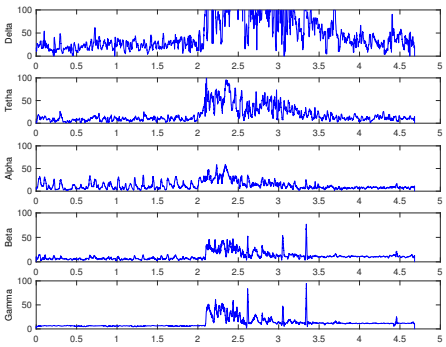
(h) Alpha spread

FIGURE 3.5: Alpha Bands Spread: EEG electrodes array, suggests alpha activity spread across both hemispheres of the cerebral cortex, predominating the right side. Amplitude (y-axis) in mV and time (x-axis) in min.

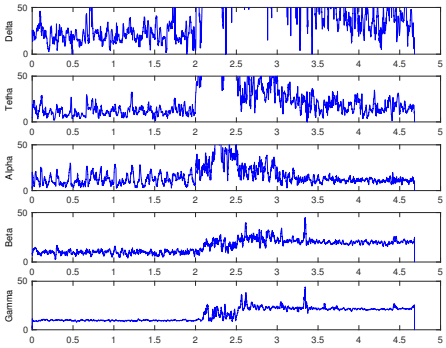


(a) Channel 13: FP2-F8

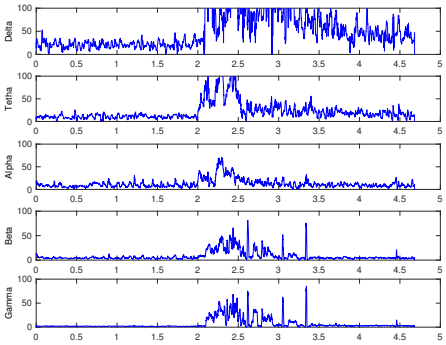
(b) Channel 17:FZ-CZ



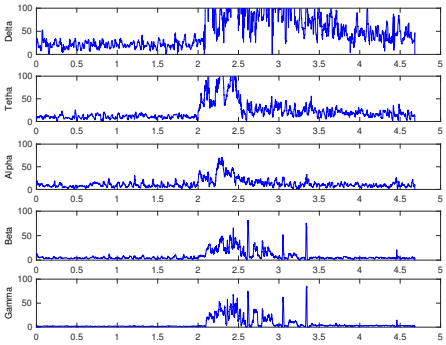
(c) Channel 14: F8-T8



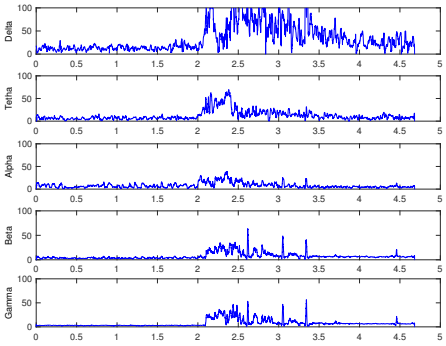
(d) Channel 21: FT9-FT10



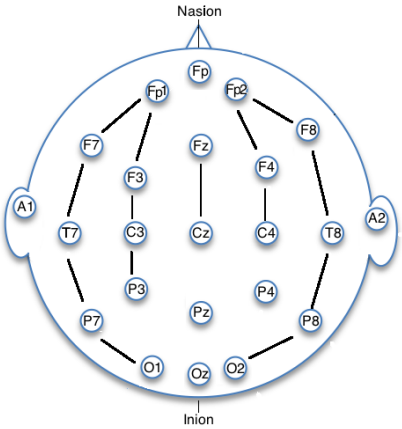
(e) Channel 23: T8-P8



(f) Channel 15: T8-P8

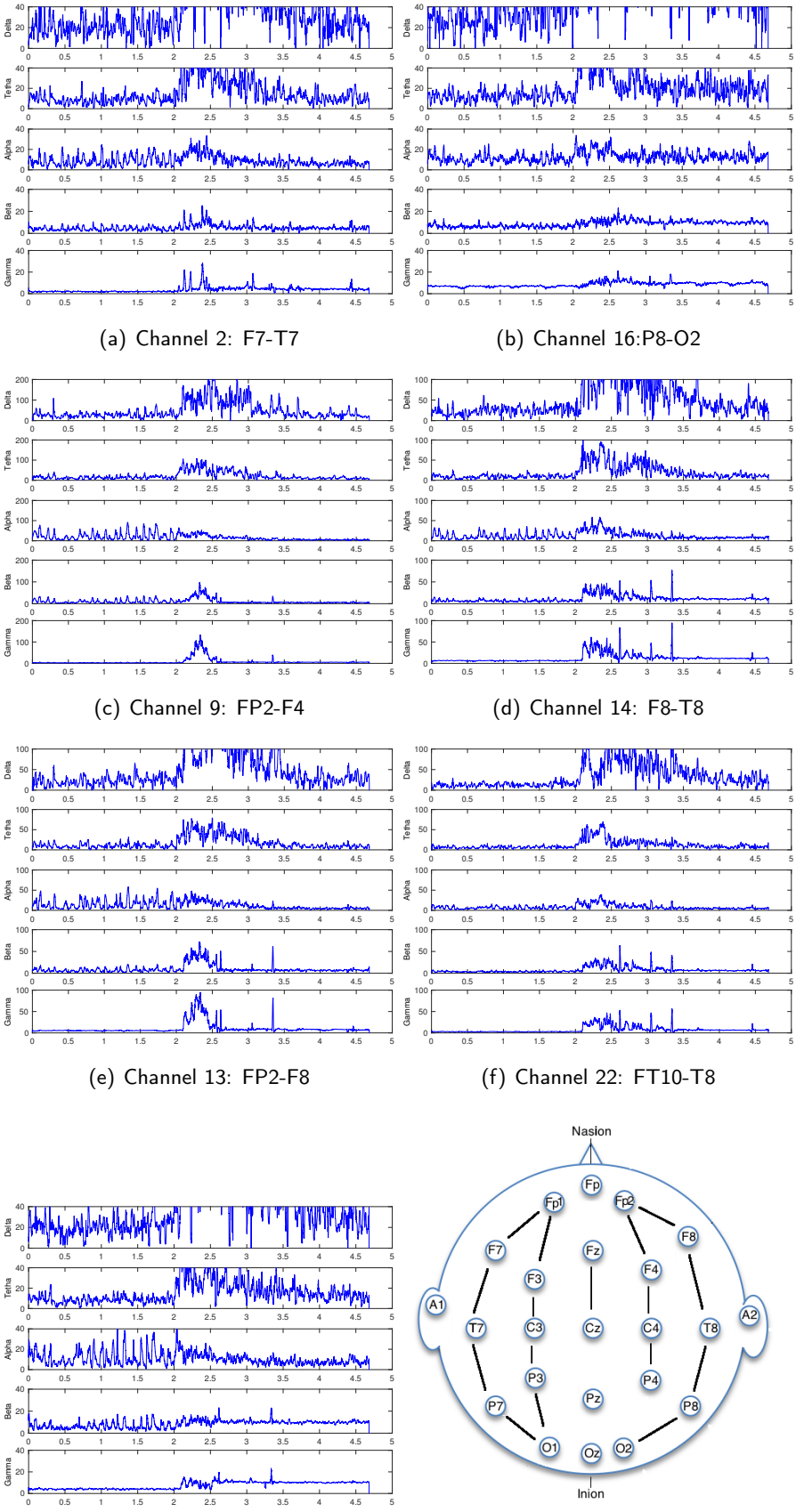


(g) Channel 22: FT10-T8



(h) Beta spread

FIGURE 3.6: Beta Band Spread: EEG electrodes array, suggests beta activity spread across both hemispheres of the cerebral cortex. Amplitude (y-axis) in mV and time (x-axis) in min.



(a) Channel 2: F7-T7

(b) Channel 16:P8-O2

(c) Channel 9: FP2-F4

(d) Channel 14: F8-T8

(e) Channel 13: FP2-F8

(f) Channel 22: FT10-T8

(g) Channel 11: C4-P4

(h) Gamma spread

FIGURE 3.7: Gamma Band Spread: EEG electrodes array, suggests gamma activity spread across both hemispheres of the cerebral cortex. Amplitude (y-axis) in mV and time (x-axis) in min.

The observation model of EEG signals $\boldsymbol{\Phi}$ is given by the equation:

$$\boldsymbol{\Phi} = \boldsymbol{K}\ \boldsymbol{J} + \boldsymbol{\eta} \tag{3.1}$$

where $\boldsymbol{J}$ is a matrix representing the sources, $\boldsymbol{K}$ is the so-called lead field or gain matrix, and $\boldsymbol{\eta}$ is additive noise. Estimating $\boldsymbol{J}$ requires solving the inverse problem (3.6). It has been shown that this can be done by minimizing the following regularized criteria:

$$F_{\boldsymbol{\lambda}}(\boldsymbol{J}) = \|\boldsymbol{K}\boldsymbol{J} - \boldsymbol{\phi}\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{J}\|_p \tag{3.2}$$

where $\boldsymbol{\lambda}$ is the regularization parameter ℓ_p-norm in the interval $1 \leq \ell_p \leq 2$ and $1 \leq p \leq 2$, see (A.8).

sLORETA (see Section A.6.5 from appendix A) uses the Tikhonov regularization ($p = 2$) [149] and solves the resulting cost function

$$min\left[\|\boldsymbol{\phi} - \boldsymbol{K}\boldsymbol{J}\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{J}\|_2^2\right] \tag{3.3}$$

A possible solution to this inverse problem can be expressed as

$$\boldsymbol{j}_i = \boldsymbol{K}_i^T[\boldsymbol{K}_i\boldsymbol{K}_i^T + \boldsymbol{\lambda}_i\boldsymbol{I}]^{-1}\boldsymbol{\phi} = \boldsymbol{R}_i\boldsymbol{J} \tag{3.4}$$

where $\boldsymbol{j}_i$ indicates the possible source candidate at voxel i and $\boldsymbol{R}_i$ is the resolution matrix given by

$$\boldsymbol{R}_i = \boldsymbol{K}_i^T[\boldsymbol{K}_i\boldsymbol{K}_i^T + \boldsymbol{\lambda}_i\boldsymbol{I}]^{-1}. \tag{3.5}$$

The observation model of EEG signals $\boldsymbol{\Phi}$ is given by the equation:

$$\boldsymbol{\Phi} = \boldsymbol{K}\ \boldsymbol{J} + \boldsymbol{\eta} \tag{3.6}$$

where $\boldsymbol{J}$ is a matrix representing the sources, $\boldsymbol{K}$ is the so-called lead field or gain matrix, and $\boldsymbol{\eta}$ is additive noise. Estimating $\boldsymbol{J}$ requires solving the inverse problem (3.6). It has been shown that this can be done by minimizing the following regularized criteria:

$$F_{\boldsymbol{\lambda}}(\boldsymbol{J}) = \|\boldsymbol{K}\boldsymbol{J} - \boldsymbol{\phi}\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{J}\|_p \tag{3.7}$$

where $\boldsymbol{\lambda}$ is the regularization parameter ℓ_p-norm in the interval $1 \leq \ell_p \leq 2$ and $1 \leq p \leq 2$, see (A.8). sLORETA (see Section A.6.5 from appendix A) uses the Tikhonov regularization ($p = 2$) [149] and solves the resulting cost function

$$min\left[\|\boldsymbol{\phi} - \boldsymbol{K}\boldsymbol{J}\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{J}\|_2^2\right] \tag{3.8}$$

A possible solution to this inverse problem can be expressed as

$$\boldsymbol{j}_i = \boldsymbol{K}_i^T[\boldsymbol{K}_i\boldsymbol{K}_i^T + \boldsymbol{\lambda}_i\boldsymbol{I}]^{-1}\boldsymbol{\phi} = \boldsymbol{R}_i\boldsymbol{J} \tag{3.9}$$

where $\boldsymbol{j}_i$ indicates the possible source candidate at voxel i and $\boldsymbol{R}_i$ is the resolution matrix given by

$$\boldsymbol{R}_i = \boldsymbol{K}_i^T [\boldsymbol{K}_i \boldsymbol{K}_i^T + \boldsymbol{\lambda}_i \boldsymbol{I}]^{-1}. \tag{3.10}$$

See section A.6.5 from appendix A for more details about this algorithm. For illustration Figure 3.8 shows the EEG and MRI image for one subject using sLORETA. This information allows the reconstruction of a model of the brain and the determination of the location where the seizure originated, as well as its temporal spread throughout the brain. This information is compared with the spread sequence estimated by our algorithm from EEG data.

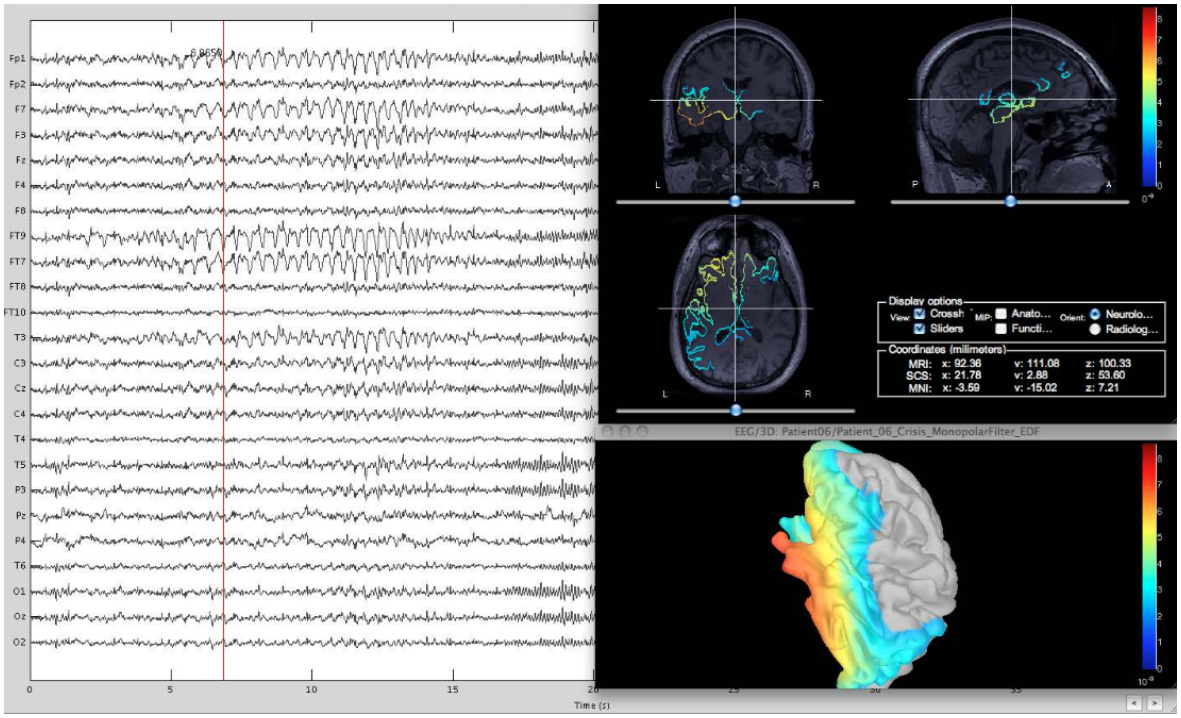


FIGURE 3.8: For illustration, the figure shows the EEG of a patient (left) with his corresponding MRI (right top). A reconstruction of the brain is shown with the corresponding seizure spread (right bottom). On the right of each image, there is a color map which indicates the intensity of each activation source.

3.6 Conclusions

The proposed algorithm is useful for onset detection and temporal spread estimation. Performing the analysis at the level of the brain activity rhythm bands can improve the identification of the area of the brain affected, as seen in the interesting correlation with Table 1.1. This preliminary study shows a plausible path for seizure onset detection and its spread in both focal and generalized seizures in epileptic EEG signals. Therefore, a strong quantitative validation with control of false positive rates in future work is necessary, in order to implement the algorithm in automatic real processing systems.

Chapter 4
Model-based classification for seizure onset detection

4.1 Introduction

EEG signal classifiers play a particularly important role in EEG signal processing. Classification is based on features extracted from single channels, multiple channels or a combination of these. In this chapter, we study four different classifiers: linear discriminant, multivariate Bayesian, ensemble bagging and logistic regression into the broad framework of machine learning. They are trained off-line for each brain rhythm. Each classifier runs in real time to detect seizure onsets from EEG recordings as follows:

Classifier	Wavelet Decomposition	Features	Epochs from the database
Linear discriminant	1D	σ, τ, ν	39
Bayesian	2D	σ, τ	36
Ensemble bagging	1D	σ, τ, ϵ	105
Logistic Regression	1D	λ, ς	36

TABLE 4.1: Classification methods developed independently for each brain rhythm. The scale (σ) and shape (τ) correspond to the generalized Gaussian distribution parameters, ν is the variance, ϵ is the entropy from the generalized Gaussian, λ are the largest Lyapunov exponents from the Analysis of Independent Components (ICA), and ς are the λ-scaling between their $\pm$ standard deviation.

Note that, in this chapter, we use different epochs from the database because all studies correspond to different stages of the research, see Table 4.1. Linear discriminant and multivariate Bayesian classifiers are based on multilevel 1D and 2D wavelet decomposition respectively. They both use the parameters of the generalized Gaussian distribution of the wavelet coefficients, independently for each brain rhythm. The ensemble bagging classifier uses the entropy of the generalized Gaussian distribution associated with the multilevel 1D wavelet coefficients. While the logistic regression classifier is applied to Lyapunov exponents and their scaling given by its $\pm$ standard deviation, from Independent Component Analysis (ICA) of multilevel 1D wavelet decomposition, independently for each brain rhythm. The four classifiers are tested and compared using a challenging pediatric dataset containing both epileptic events and normal brain function signals,

see Section 2.2. The high performance in terms of classification sensitivity and specificity permits to discriminate between *seizure* and *non-seizure* in EEG signals.

In this chapter, we are going to concentrate on the linear discriminant classifier because this method has three main strengths: it has low computational cost making it suitable for real-time implementation in EEG devices; it performs detection separately for each brain rhythm, following the current medical practices; and it can be trained using reasonably small datasets, which is key in clinical problems where there is limited annotated data available. This is in sharp contrast with modern approaches based on machine learning techniques, which achieve very high sensitivity and specificity but require large training sets with expert annotations that may not be available.

4.2 Seizure onset detection (SOD) classifiers

Epileptic seizure detection methods based on EEG signals stem from the observation that EEG signal descriptors allow discriminating normal from abnormal brain activity. This practice originated half a century ago with works by Viglione et al. [41], Liss et al. [42], Ktonas et al. [43] and Gotman et al. [44]; and continued with Iasemidis et al. [45, 46] mainly in the medical literature and by using analog EEG devices, see Section 1.6 for more details. Later, the adoption of digital signal processing in EEG systems stimulated the development of pattern recognition methods to detect and analyze abnormal brain activity automatically. The main practical advantage of EEG technology is its economic accessibility. This has significantly contributed to the wide adoption of EEG in developing countries, whereas other more advanced modalities, such as magnetoencephalography (MEG), are expensive and have not been widely adopted.

There is currently a wide range of EEG signal processing methods to detect brain seizures accurately. Most methods use classification techniques from the supervised machine learning literature, such as support vector machines [150, 151] and discriminant analysis [12], and differ mainly in terms of their feature extraction methods and the features classification approaches. Many methods use time-frequency descriptors, either explicitly (e.g., short-term Fourier or wavelet representations) [37, 152, 153, 154, 151, 155], empirical mode decomposition [156, 157, 158], implicitly by learning neural networks [14, 159] or by using component analysis or common spatial patterns (see for example [13, 160, 161]). Some also use statistical descriptors such as signal entropy [162, 163, 164, 154, 151, 165] or fractal dimension [166, 167].

The main approaches from the state of the art are summarised in Table 4.2, together with their detection performance on a test dataset. Observe that most modern methods perform remarkably well and achieve true positive rates (TPR) or sensitivities of the order of $95\% - 99\%$, and true negative rates or specificities of the order of $85\% - 95\%$, depending on the specific method and dataset considered. This good performance is achieved by using advanced signal processing techniques that are generally very computationally intensive. As a result, state-of-the-art detection methods cannot be incorporated into EEG devices to perform detection in real time. For example, the method [13] uses common spatial patterns that require estimating covariance matrices and performing singular value decompositions at each detection step. This limitation is motivating the development of detection methods that use cloud computing technology to perform detection on a

high-performance computing server that is accessed remotely (see for example [161]). This strategy is potentially very interesting in some settings, but it would be difficult to implement in developing countries where many hospitals still have limited Internet access and poor IT infrastructure.

Another limitation of state-of-the-art methods is that they pull information from all spectral bands to improve detection performance [13]. While beneficial in terms of classification accuracy, this can be problematic in many clinical applications where the current practice is to detect seizures independently in each physiological spectral band or *brain rhythm* (these bands are specified in Section 2.4). Finally, state-of-the-art methods also rely increasingly on large training datasets, which is a drawback in clinical applications where there is limited annotated data available. Also, many existing methods use feature-based classification techniques, with a significant number of features in order to handle the inherent variability of such features.

This chapter seeks to address these limitations of the existing methods by developing an automatic EEG detection technique that has a low computational cost, that performs detection independently in each brain rhythm following current clinical practice, and that can be trained with reasonably small datasets, with a detection performance that is similar to that of state-of-the-art algorithms. In contrast to existing methods, the proposed method adopts a model-based classification approach. Model-based classification has been used in various applications [168, 169, 170]. The idea is to capture the statistical properties of the signal using the parameters of a probabilistic model, see Section 2.5.1. This approach is interesting compared to feature-based classification, especially when features are numerous or exhibit large variability. It can be viewed as an interesting dimensionality reduction technique facing the curse of dimensionality and leading to low computational cost classification, see Section 2.7. Despite its interest, this approach has not been widely investigated in EEG signal processing. Precisely, the linear classification method (explained later) is driven by a parametric statistical model that captures the statistical properties of the signals and their evolution in time, with the model parameters acting as classification features, see Section 2.8.2. This approach is an interesting alternative to the non-parametric features (e.g., signal power spectrum, variance, entropy, etc.) commonly used in the literature because the parametric structure of the model acts as a dimensionality reduction mechanism that regularizes the classification problem and consequently improves the stability and robustness of the classification while reducing significantly computational cost. Despite its advantages, to the best of our knowledge, this promising approach has not been investigated for EEG signal classification. Note however that statistical approaches have been successfully applied to other challenging EEG processing problems (see for example [171, 172]).

Classification method	Features	Test data	Performance	Ref.
Learning vector quantization	Signal entropy from wavelet coefficients	400 epochs from 5 normal subjects and 5 epileptic patients	TPR:98%	[154]
Support Vector Machine	Matching pursuit algorithm	133 EEG from Rigshospitalet University Hospital database (Copenhagen, Denmark)	TPR:78%, TNR:84%	[150]
Support Vector Machine	Spectral and entropy analysis	3 datasets from EEG University Hospital Bonn database	TPR:90%	[151]
Fuzzy classification	Amplitude, frequency and entropy descriptors	56 iEEG from 20 patients from University of Freiburg database	TPR:95.8%, TNR:74%	[173]
Hidden Markov Model	Segmentation of topographic maps of time varying spectral	10 EEG patients from EPILEPSIAE [174]	TPR:94.59%, TNR:92.22%	[155]
Support Vector Machine	Third-order tensor discriminant analysis: spectral, spatial, and temporal domains	36 EEG patients from **Children's Hospital Boston database**	TPR:98%, TNR:94%	[12]
K-means clustering	Spatiotemporal analysis as morphological filter	10 EEG patients from University of Florida Hospital database	TPR:87.4%	[175]
Support Vector Machine	Fractional linear prediction	100 single channel EEG segments from The Bern-Barcelona EEG database	TPR:96%, TNR:95%	[176]
Least Squares Support Vector Machine	Phase space representation	100 segments from the EEG University Hospital Bonn	TPR:100%, TNR:96%	[177]
Support Vector Machine	Empirical mode decomposition	51 EEG segments from 17 patients from University of Freiburg (Germany)	TPR:98.6%, TNR:88.6%	[178]
1-Nearest Neighbor	1D-local binary patterns from bank of Gabor filters	100 ECoG segments from University Hospital Bonn database	TPR:98.33%	[179]
Support Vector Machine	Common spatial Pattern (is a method that uses a linear transform to project multichannel EEG data into a low-dimensional spatial-subspace projection)	36 EEG patients from Children's Hospital Boston database	TPR:100%	[13]
Relevance Vector Machine	Multifractal	21 EEG patients from the Epilepsy Center of the University Hospital of Freiburg	TPR:92.94%, TNR:97.47%	[180]
Regression neural network	Statistical descriptors of dual-tree complex wavelet transform coefficients	100 segments from University of Bonn database and 21 patients from Sir Ganga Ram Hospital (New Delhi)	TPR:92%, TNR:98%	[159]
K-Nearest Neighbor, linear discriminant analysis, naive Bayesian, logistic regression and Support Vector Machine	Time, frequency, time-frequency and nonlinear features	100 segments from University Hospital Bonn database	TPR:99.25%	[181]
1-Nearest Neighbor	Regularization, learning rate and momentum from a convolutional neural network	5 patients from the EEG University Hospital Bonn	TPR:95%, TNR:88.67%	[160]
Random Forest, C4.5, Functional tree, Bayesian-network, Naive-Bayes and K-nearest neighbours	Mean of joint instantaneous amplitude, Mean and variance of monotonic absolute change from empirical wavelet transform	36 EEG patients from **Children's Hospital Boston database**	TPR:97.91%, TNR:99.57%	[182]
Support Vector Machine	Pyramid of difference of Gaussian filtered signals and local binary patterns	100 segments from the EEG University Hospital Bonn	TPR:100%, TNR:100%	[183]
Support Vector Machine	Random subspace ensemble method and Infinite Independent Component Analysis	208 ECoG from University of Pennsylvania and the Mayo Clinic	TPR:98%, TNR:96%	[161]
Least-Square Support Vector Machine	Time-frequency representation based on the improved eigenvalue decomposition of Hankel matrix and Hilbert transform	100 segments from the EEG University Hospital Bonn	TPR:100%, TNR:100%	[184]

TABLE 4.2: State-of-the-art methods to perform seizure detection automatically in EEG signals, summarized in terms of the classification techniques and features used and their reported performance on a test dataset. The performance metrics are the True Positives Rate or Sensitivity (TPR), the True Negative Rate or Specificity (TNR).

4.3 Cross-validation and leave-one-out cross-validation

In many biomedical applications the supply of data for training and testing can be limited, so if we want to build good models, we must use as much of the available data as possible for training. However, if the validation set is small, it will give a relatively noisy estimate of predictive performance. One solution to this dilemma is to use *cross-validation* [185], which is a model validation technique for evaluating how the results of a statistical analysis algorithm can be generalized to an independent data set. Applications of this method in epilepsy date back to the 70s [73] with template matching, see [3] for more details.

Cross-validation is a model validation technique to evaluate how the results of a statistical analysis algorithm can be generalized to an independent data set. This is done by partitioning a dataset and using a subset to train the algorithm and the remaining data for testing. Each round of cross-validation involves randomly partitioning the original dataset into a training set and a testing set. The training set is then used to train a supervised learning algorithm and the testing set is used to evaluate its performance. This process is repeated several times and the average cross-validation error is used as a performance indicator.

In *leave-one-out cross-validation* technique, the data partitions use the k-fold approach where k is equal to the total number of observations in the data [186, 187]. Leave-one-out cross-validation technique has numerous applications in science, engineering and EEG signal processing [188, 189, 190, 191, 192].

4.4 Methodology

The general methodology used in this chapter is presented in the following subsections.

4.4.1 Problem statement

Let $\boldsymbol{X} \in \mathbb{R}^{M\times N}$ denote a time-discretized EEG signal recorded by an array composed of M channels over a period of T seconds, and using a sampling period of T/N seconds. Each row of $\boldsymbol{X}$ is associated with one channel of the array and contains all the sampling points corresponding to the EEG signal recorded by that channel, whereas each column is associated with a sampling point and contains the vector signal acquired by the full array at that time instant. Moreover, to analyse the different frequency components of $\boldsymbol{\theta}^{(i)}$, we denote by $\boldsymbol{\theta}_{\delta}^{(i)}$, $\boldsymbol{\theta}_{\theta}^{(i)}$, $\boldsymbol{\theta}_{\alpha}^{(i)}$, $\boldsymbol{\theta}_{\beta}^{(i)}$, and $\boldsymbol{\theta}_{\gamma}^{(i)}$ the spectral components related to the δ (0-4 Hz), θ (4-8 Hz), α (8-16 Hz), β (16-32 Hz), and γ (32-64 Hz) frequency bands. As mentioned previously, each of these bands is related to different neurological functions and is therefore associated with specific neurological disorders, see Section 1.3 for more details.

This thesis considers the problem of detecting epileptic seizure activity in EEG signals in real-time and identifying the frequency bands where the seizure occurs. Formally, for any time instant $n \in \{1, N\}$, define band-specific binary labels $\omega_{\delta}(n)$, $\omega_{\theta}(n)$, $\omega_{\alpha}(n)$, $\omega_{\beta}(n)$, and $\omega_{\gamma}(n)$ that take value 1 to indicate the presence of an epileptic seizure at their spectral band, and 0

to indicate normal activity. Given some expert annotated training data $\{\boldsymbol{X}_0^{(i)}\}_{i=1}^{I_0}$ and $\{\boldsymbol{X}_1^{(i)}\}_{i=1}^{I_1}$ corresponding to short EEG recordings of healthy and epileptic seizure activity, we consider the supervised classification problem of estimating the values of $\omega_\delta(n)$, $\omega_\theta(n)$, $\omega_\alpha(n)$, $\omega_\beta(n)$, and $\omega_\gamma(n)$ in real-time as $\boldsymbol{X}$ is acquired by the EEG array. This is motivated by our interest in clinical applications where this information is required in real-time, we focus therefore on classifiers that have low computational complexity.

4.4.2 Proposed method

The proposed method has a pipeline structure composed of the following three steps: a filter bank that separates $\boldsymbol{X}$ into its $\boldsymbol{X}_\delta$, $\boldsymbol{X}_\theta$, $\boldsymbol{X}_\alpha$, $\boldsymbol{X}_\beta$, and $\boldsymbol{X}_\gamma$ spectral components, followed by a statistical dimensionality reduction step that maps these components into a low-dimensional representation where pathological brain activity is easily detected, and finally a classification step based on a thresholding approach.

4.4.3 Spectral decomposition by wavelet filter bank

We use a Dauchebies (Db4) wavelet filter bank to separate $\boldsymbol{X}$ into the five spectral components $\boldsymbol{X}_\delta$, $\boldsymbol{X}_\theta$, $\boldsymbol{X}_\alpha$, $\boldsymbol{X}_\beta$, and $\boldsymbol{X}_\gamma$. Because our data is acquired at a 256 Hz sampling rate, in our experiments we use a wavelet filter through tree-based topology, with six scales. The upper five scales match with the spectral bands of interest (the remaining scale related to the 64-128 Hz band has very poor signal-to-noise ratio and is discarded, see Section 2.4.3). The output of this stage are 5 sets of wavelet coefficients $\boldsymbol{\theta}_\delta$, $\boldsymbol{\theta}_\theta$, $\boldsymbol{\theta}_\alpha$, $\boldsymbol{\theta}_\beta$, $\boldsymbol{\theta}_\gamma$ (please note that this approach can be straightforwardly generalized to higher sampling rates by using or discarding any additional bands), see Section 2.4 for more details.

4.4.4 Statistical model of the spectral components

Designing a classifier to detect pathological brain activity directly from the EEG signals (or their wavelet representation) is very challenging due to the high-dimensionality of the data, and because it would require a large training set and a complex classification methodology. To detect abnormal brain activity with limited annotated training data, particularly in the context of classifiers with low computational complexity suitable for real-time implementations, it is necessary to map the EEG data to a meaningful compact representation that highlights the information able to discriminate normal and abnormal activities. A successful representation should also provide the low-dimensional structure and favorable regularity properties that enable a simple classification scheme, such as threshold-based methods.

Here we construct this representation by using a parametric statistical model to summarize the empirical distribution of the wavelet coefficients associated with each spectral band. Precisely, a sliding window approach was adopted and fit a parametric statistical model to the wavelet coefficients associated with the last 2 seconds of $\boldsymbol{X}$. Because the signals considered in the experiments are acquired with $M = 23$ channel array, the 2-second window corresponds respectively to the

coefficients of $\boldsymbol{C}_{\delta}^{(i)}$, $\boldsymbol{C}_{\theta}^{(i)}$, $\boldsymbol{C}_{\alpha}^{(i)}$, $\boldsymbol{C}_{\beta}^{(i)}$, and $\boldsymbol{C}_{\gamma}^{(i)}$. We model each set of wavelet coefficients with zero-mean generalized Gaussian distribution (GGD) with density given by

$$f(x;\sigma,\tau) = \frac{\tau}{2\sigma\Gamma(\tau^{-1})} \exp\left(-\left|\frac{x}{\sigma}\right|^{\tau}\right) \tag{4.1}$$

where $\sigma \in \mathbb{R}^{+}$ is a scale parameter, $\tau \in \mathbb{R}^{+}$ is a shape parameter that controls the density tail, and $\Gamma(\cdot)$ is the Gamma function. We estimate the values of σ and τ for each spectral band by maximum likelihood estimation, see eq. (2.25), which we solve straightforwardly by using a Newton-Raphson algorithm, see Sections 2.8 and 2.8.2 for more details. We obtain the parameter-vector $\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}$

$$\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}} = \left[\sigma_j^{(i)}, \tau_j^{(i)}\right]^T = \underset{[\sigma,\tau]^T}{\operatorname{argmax}}\, f_{GGD}(\boldsymbol{C}_j^{(i)};\sigma,\tau) = \left[\boldsymbol{\theta}_{\gamma}^{\boldsymbol{C}_j^{(i)}}, \boldsymbol{\theta}_{\beta}^{\boldsymbol{C}_j^{(i)}}, \boldsymbol{\theta}_{\alpha}^{\boldsymbol{C}_j^{(i)}}, \boldsymbol{\theta}_{\theta}^{\boldsymbol{C}_j^{(i)}}, \boldsymbol{\theta}_{\delta}^{T}\right]^T. \tag{4.2}$$

4.5 Linear discriminant classifier

4.5.1 Classifier parameters

Linear classification parameters for onset detection are summarized in Figure 4.1.

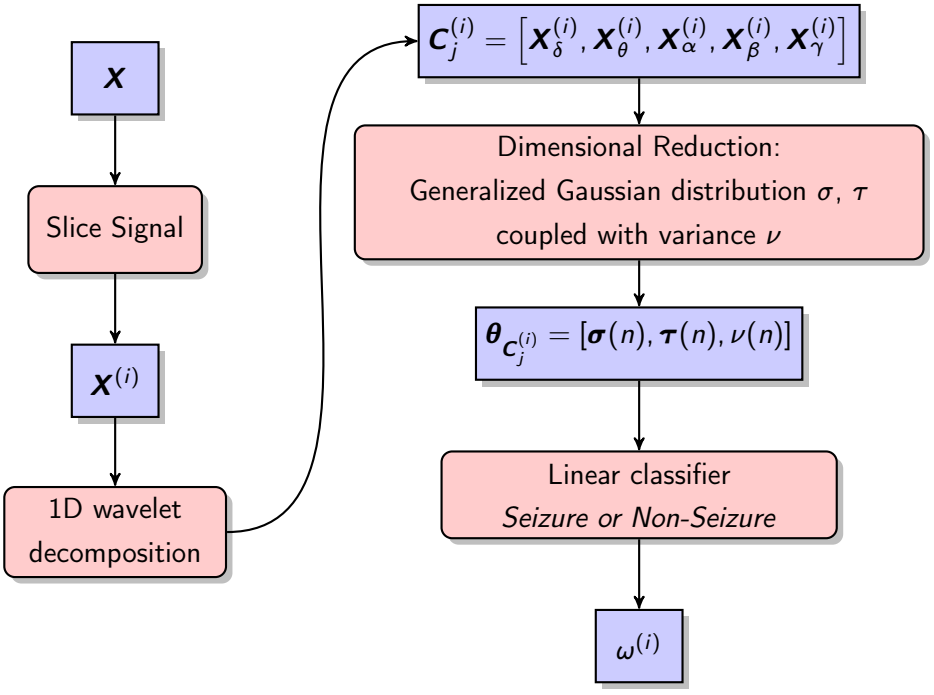


FIGURE 4.1: Algorithm used in linear classifier.

4.5.2 SOD by linear discriminant analysis classification

Consider the classification into two possible classes: ω_s for *seizure* and ω_{ns} for *non-seizure*. For a feature vector $\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}$ belonging either to the class ω_s or to the class ω_{ns}, we assume that $\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}$ has a normal distribution with mean value $\boldsymbol{\mu}_s$ (or $\boldsymbol{\mu}_{ns}$) and covariance matrix $\boldsymbol{\Sigma}_s = \boldsymbol{\Sigma}_{ns}$, then

$$p\left(\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}\middle|\omega_s\right) = \frac{1}{\sqrt{(2\pi)^k\,|\boldsymbol{\Sigma}_s|}}\exp\left[-\frac{1}{2}\left(\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}-\boldsymbol{\mu}_s\right)^T\boldsymbol{\Sigma}_s^{-1}\left(\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}-\boldsymbol{\mu}_s\right)\right] \tag{4.3}$$

$$p\left(\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}\middle|\omega_{ns}\right) = \frac{1}{\sqrt{(2\pi)^k\,|\boldsymbol{\Sigma}_{ns}|}}\exp\left[-\frac{1}{2}\left(\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}-\boldsymbol{\mu}_{ns}\right)^T\boldsymbol{\Sigma}_{ns}^{-1}\left(\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}-\boldsymbol{\mu}_{ns}\right)\right] \tag{4.4}$$

where k is the dimension of the vector $\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}$ and $p(\cdot)$ is the density conditioned by an event.

For linear discriminant analysis, we estimate the mean ($\boldsymbol{\mu}_s$ or $\boldsymbol{\mu}_{ns}$) and the covariance ($\boldsymbol{\Sigma}_s$ or $\boldsymbol{\Sigma}_{ns}$) of each class from observations.

The linear discriminant for these classification problem is given by

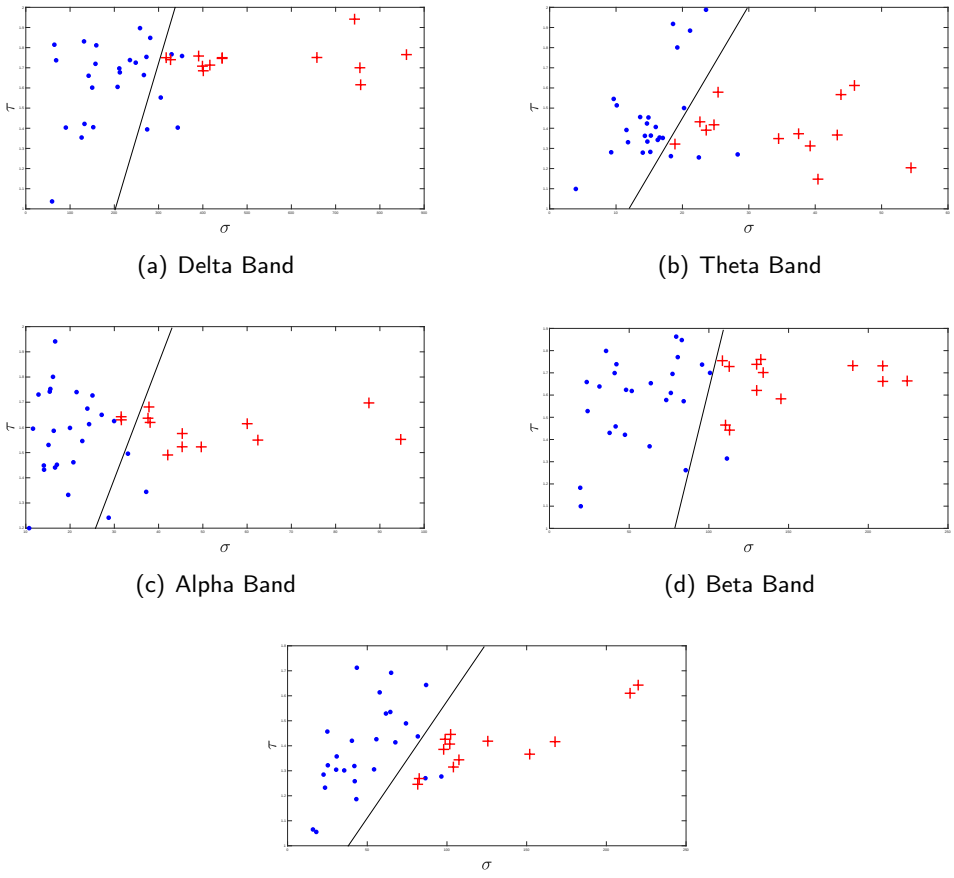
$$\begin{aligned}\ln\frac{p\left(\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}\middle|\omega_s\right)}{p\left(\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}\middle|\omega_{ns}\right)} &= (\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}-\boldsymbol{\mu}_s)^T\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}-\boldsymbol{\mu}_s)+\ln|\boldsymbol{\Sigma}_s| \\ &\quad-(\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}-\boldsymbol{\mu}_{ns})^T\boldsymbol{\Sigma}_{ns}^{-1}(\boldsymbol{\theta}_{\boldsymbol{C}_j^{(i)}}-\boldsymbol{\mu}_{ns})-\ln|\boldsymbol{\Sigma}_{ns}|.\end{aligned} \tag{4.5}$$

Following the methodology used in Section 4.4, the proposed seizure detection pipeline is a classifier that labels the statistical parameters associated with each spectral band as *seizure* or *non-seizure*. Precisely, five independent two-parameter classifiers are used in parallel to classify the pairs $[\sigma_\delta(n),\tau_\delta(n)]$, $[\sigma_\theta(n),\tau_\theta(n)]$, $[\sigma_\alpha(n),\tau_\alpha(n)]$, $[\sigma_\beta(n),\tau_\beta(n)]$, and $[\sigma_\gamma(n),\tau_\gamma(n)]$ generated by the statistical dimensionality reduction step. This allows to simultaneously identify seizure activity and the spectral bands where it occurs. For simplicity, a linear classifier derived from a linear discriminant analysis is used. Just, a supervised approach was adopted where each classifier is band-specific and has been trained by performing a linear discriminant analysis on expert annotated data. The discriminant analysis perform on an augmented vector $[\sigma,\tau,\nu]\in\mathbb{R}^3$, where $\nu = \sigma^2\Gamma(3/\tau)/\Gamma(1/\tau)$ is the variance parameter. Including ν in the discriminant analysis embeds (σ,τ) in a non-linear manifold into $\mathbb{R}^3$ where a better linear classification is possible (note that ν available for free as a by-product of the Newton-Raphson method that estimates σ and τ, hence this augmentation does not introduce any additional computational cost). The resulting linear classifiers are specified by three parameters (a,b,c) defining a plane that splits $\mathbb{R}^3$ in two regions related to *seizure* and *non-seizure* events, and which essentially operate as a three-dimensional threshold for the triplets σ,τ,ν. Lastly, similarly to the choice of the statistical model, it is possible to consider more advanced classifications schemes. However, such classifiers also involve more parameters and hence are more prone to over-fitting and more computationally expensive.

4.5.3 Experimental results

For the experiments, we used data from the Children's Hospital Boston database, see section 2.2. From this database, we used 13 seizure signals or *epochs* selected by an experienced neurologist, see Table 2.1. These correspond to 13 seizure events from 9 different subjects and are between 1 and 5 minutes long (the other data exhibited strong artifacts related to muscle activity and were discarded as a consequence). The resulting dataset consisted therefore of 39 signal segments related to 13 *seizures* and 26 *non-seizure* signals, and of variable length in the range of 1 to 5 minutes.

To illustrate the capacity of the statistical parameters σ and τ to discriminate *seizure* events and *non-seizure* signals, Figure 4.2 shows scatter plots for each spectral band constructed using the signals in the database and the expert annotations (*non-seizure* signals are represented using blue circles and *seizure* signals using red crosses). Observe that this representation provides a very good linear discrimination of the *seizure* and *non-seizure* groups. In particular, one notices that the scale parameter σ is particularly useful for discrimination, see also discrimination tests in subsection 2.8.2.



(a) Delta Band (b) Theta Band

(c) Alpha Band (d) Beta Band

(e) Gamma Band

FIGURE 4.2: Scatter plots for the statistical parameters σ and τ for *seizure* signals (red crosses) and *non-seizure* signals (blue circles) for each spectral band, showing the good discrimination properties of the proposed representation.

Moreover, to assess the performance of the proposed methodology, we adopted a supervised testing approach by using the 39 signal segments described above to train and test the method. Because the dataset is relatively small we used an exhaustive cross-validation technique based on a leave-one-out approach. Precisely, at each iteration of the cross-validation process, we trained the 5 classifiers (each defined by 3 parameters) with data from 13 *seizure* signals and 26 *non-seizure* signals, and then assessed classification performance on the remaining 3 signals (these are 1 *seizure* and 2 *non-seizure* signals). In each iteration of the cross-validation process, the classification performance was assessed by splitting the test signals into sequences of 2 seconds and classifying each sequence individually; these results were then used to assess classification performance. Precisely, we measure the method's true positive rate (TPR) or *sensitivity*, false positive rate (FPR), true negative rate (TRN) or *specificity*, and overall accuracy (ACC), expressed as the rate of good classification. For each figure of merit, we report the mean value and the standard deviation. These results are reported in Table 4.3 below. Average latency (time delay) is also reported between the annotated seizure onset and the detection by the method in Table 4.4. Classification accuracy and latency were compared with the state-of-the-art methods [12, 13, 14], which also report classification performance and latency for the Children's Hospital Boston database. We emphasize again that these state-of-the-art methods are significantly more computationally expensive than the proposed method. For example, [12] uses a third-order tensor discriminant analysis, [14] a stack of neural networks combined with a logistic classifier, and [13] computes singular value decompositions of covariance matrices at each detection step. Neither of these methods can be implemented in real-time in a standard EEG system as a consequence.

Observe from Table 4.3 that, despite the computational simplicity, the proposed method achieves an excellent sensitivity of the order of $97\% - 99\%$ for all spectral bands. This is close to the state-of-the-art performances of $98\% - 100\%$ reported in [12, 13, 14] for this dataset. Moreover, the specificity of the proposed method is approximately 90%. This is slightly bellowed the 94% specificity of [12] (the works [13, 14] do not report specificity). However, notice that to achieve this higher specificity, the method [12] pulls together all spectral bands, and as a result, it does not discriminate between seizures in different bands. Our method performs classification independently on each band because this is useful in clinical practice, at the expense of a slightly lower specificity.

Furthermore, observe from Table 4.4 that the method proposed in this thesis achieves an average latency of approximately 4 seconds for all spectral bands, outperforming the state-of-the-art methods [12, 14] and close to the fastest available method [13]. Ii is important to emphasize at this point that all the latency values reported in the literature measure the delay of the detection algorithm offline, without taking into account any overhead related to the methods' computing times. Therefore, the fact that different methods achieve similar latency does not indicate that they have similar computational complexity.

Finally, note that computing times are not reported for these experiments for two reasons. First, because these proof-of-concept tests were conducted in MATLAB, and processing each 2-second EEG signal window required less than 50 milliseconds. Second, because we do not have access to the implementations of [12], [13] and [14], and therefore the comparisons would not be fair. However, as explained previously, these methods clearly have a significantly higher computational complexity because of the sophisticated mathematical operations involved (e.g., third-order tensor discriminant analysis, singular value decompositions of covariance matrices, stacked neural networks, etc.). A real-time implementation of the proposed method is currently under development.

Metric	Delta Band (δ)	Theta Band (Θ)	Alpha Band (α)	Beta Band (β)	Gamma Band (γ)
TPR	0.97 $\pm$0.06	0.99 $\pm$0.01	0.99 $\pm$0.02	0.97 $\pm$0.05	0.99 $\pm$0.01
TNR	0.92 $\pm$0.07	0.79 $\pm$0.23	0.91 $\pm$0.08	0.90 $\pm$0.10	0.91 $\pm$0.08
FPR	0.08 $\pm$0.07	0.21 $\pm$0.23	0.09 $\pm$0.08	0.10 $\pm$0.10	0.09 $\pm$0.08
ACC	0.95 $\pm$0.17	0.92 $\pm$0.29	0.96 $\pm$0.11	0.94 $\pm$0.22	0.97 $\pm$0.11

TABLE 4.3: Seizure detection performance by using linear discriminant analysis classification for each brain rhythm and for 39 events (13 *seizure* and 26 *non-seizure*) of the Children's Hospital Boston database, in terms of: TPR = True Positives Rate or Sensitivity; TNR = True Negative Rate or Specificity; FPR = False positive Rate; ACC = Accuracy; and [$\pm$ standard deviation].

Proposed					state-of-the-art		
Delta band (δ)	Theta band (θ)	Alpha band (α)	Beta band (β)	Gamma band (γ)	[12]	[13]	[14]
4.3	3.9	4.1	4.0	4.1	4.5	3.4	7.2

TABLE 4.4: Average latency between seizure onset and detection (in seconds), for the proposed method on each spectral band, and for the state-of-the-art methods. [12, 13, 14].

4.5.4 Pearson's product moment correlation

Pearson's product moment correlation coefficient is a measure of the strength of a linear association between two variables or classes, *non-seizure* and *seizure* in our case. Consider the classification into two possible classes: ω_s for *seizure* and ω_{ns} for *non-seizure* from linear discriminant from (4.5), then the linear association from the Pearson product-moment correlation coefficient r is expressed as

$$r = \frac{\sum(\omega_s - \overline{\omega_s})(\omega_{ns} - \overline{\omega_{ns}})}{\sqrt{\sum(\omega_s - \overline{\omega_s})^2 \sum(\omega_{ns} - \overline{\omega_{ns}})^2}} \tag{4.6}$$

where $\overline{\omega_s}$ and $\overline{\omega_{ns}}$ are the means of each class, and r can take a range of values from +1 to -1. A value of 0 indicates that there is no association between the two variables. A value greater than 0 indicates a positive association; that is, as the value of one variable increases, so does the value of the other variable. A value less than 0 indicates a negative association; that is, as the value of one variable increases, the value of the other variable decreases.

The Pearson's correlation coefficient of the two classes (ω_s for *seizure* and ω_{ns} for *non-seizure*) for each brain rhythm are reported in Table 4.5, where p-value is the significance level of the T-test [134]. This shows that for bands *delta*, *theta*, *alpha* and *beta* the correlation for ω_{ns} is high with r near to one, p-values < 0.001, and a good confidence interval; while for ω_s the bands are not correlated with r near to zero. These results suggest that our model based on generalized Gaussian distribution coupled with the linear discriminant classifier can be defined by scale value between $[-1, 1]$ to discriminate between *seizure* and *non-seizure* events.

Bands	r	p-value	IC95%
Delta ω_{ns}	0.88	< 0.001	0.70 0.95
Delta ω_s	0.39	0.11	-0.01 0.72
Theta ω_{ns}	0.81	< 0.001	0.55 0.92
Theta ω_s	0.51	0.03	0.06 0.79
Alpha ω_{ns}	0.80	< 0.001	0.53 0.92
Alpha ω_s	0.45	0.06	-0.02 0.76
Beta ω_{ns}	0.72	< 0.001	0.38 0.89
Beta ω_s	0.15	0.56	-0.34 0.58
Gamma ω_{ns}	0.58	0.01	0.15 0.82
Gamma ω_s	-0.11	0.66	-0.55 0.38

TABLE 4.5: Pearson's product moment correlation coefficient comparison between ω_s for *seizure* and ω_{ns} for *non-seizure* events, over the proposed epilepsy classification model for each brain rhythm; where $r = 1$ is total positive correlation, $r = 0$ is no correlation, $r = -1$ is total negative correlation and $IC95\%$ is the 95 percent confidence interval. The model presents high correlation for all brain rhythms, except for gamma band in *non-seizure* events. This suggests that Pearson's product moment correlation coefficient can be used to estimate changes between *seizure* and *non-seizure* events.

4.6 Multivariate Bayesian Classifier

Methods to analyze epileptic seizure signals can be classified into univariate or multivariate approaches. Univariate approaches analyze the state of a single brain region, while multivariate approaches analyze. many regions simultaneously as well as their interactions [193]. In this classifier, we extend the approach used in the previous Section 4.5 by using a multilevel 2D wavelet representation coupled with a Bayesian classification scheme [194, 195, 196, 197] to operate with multivariate EEG signals so as to analyze several brain regions simultaneously.

4.6.1 Classifier parameters

Multivariate Bayesian classification for onset detection is summarized in Figure 4.3. It is very similar to linear classifier parameters from subsection 4.5.1. The more important difference is the use of multilevel 2D wavelet decomposition in order to analyze the EEG, see subsection 2.4.2 of Section 2.4. The feature vector associated with each time segment is classified by using a Bayesian classifier as *seizure* or *non-seizure*. Here we use a multivariate Gaussian classifier [101]

which provides a robust second-order approximation with a more general Bayesian classification method [198, 199] which has the important advantage of requiring little training data.

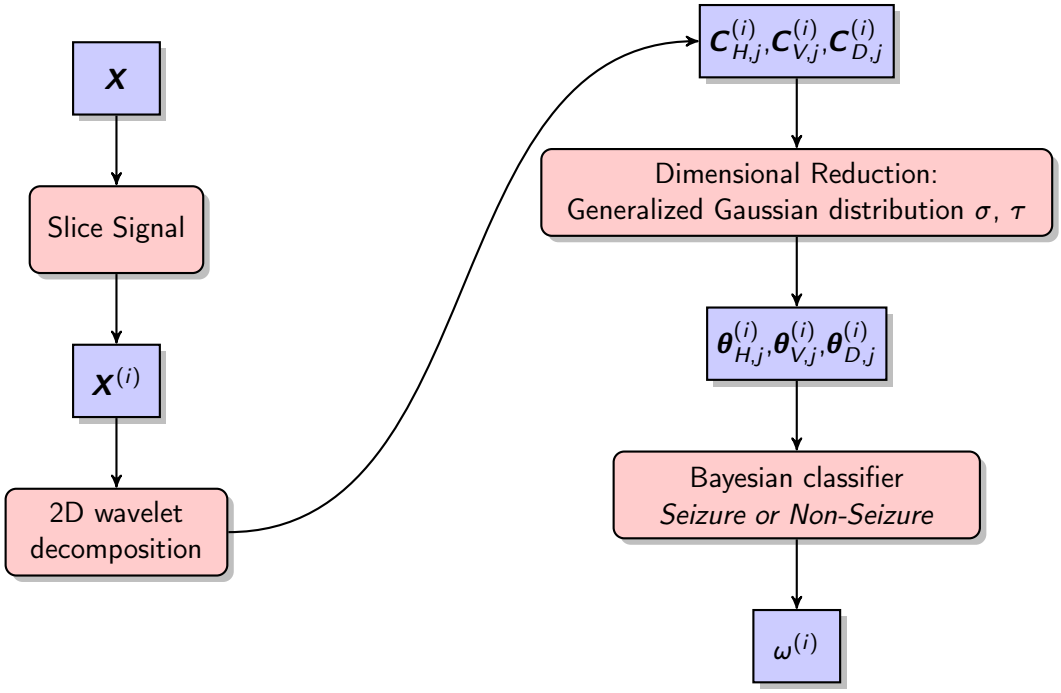


FIGURE 4.3: Algorithm used in Bayesian classifier. The coefficients vector $\boldsymbol{C}_j^{(i)}$ and the vector of parameters estimated $\boldsymbol{\theta}_j^{(i)}$, from different brain rhythms are composed by the details H, V, and D, which refer to horizontal, vertical and diagonal respectively, see Section 2.4.

4.6.2 SOD by Bayesian Classifier

Consider a classification into J possible classes $\omega_1, \ldots, \omega_J$. For a feature vector $\boldsymbol{\theta}^{(i)}$ belonging to the class ω_j, we assume that $\boldsymbol{\theta}^{(i)}$ has a multivariate normal distribution with mean value $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$, so that

$$p\left(\boldsymbol{\theta}^{(i)}\middle|\omega_j\right) = \frac{\exp\left[-\frac{1}{2}(\boldsymbol{\theta}^{(i)} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{\theta}^{(i)} - \boldsymbol{\mu}_j)\right]}{(2\pi)^{K/2}|\boldsymbol{\Sigma}_j|^{1/2}} \tag{4.7}$$

where $p(\cdot)$ is the probability of a particular event, and K is the size of the vector $\boldsymbol{\theta}^{(i)}$.

The Bayes decision rule states that the estimated class $\hat{\omega}^{(i)}$ corresponding to $\boldsymbol{\theta}^{(i)}$ is

$$\hat{\omega}^{(i)} = \arg\max_j p\left(\boldsymbol{\theta}^{(i)}\middle|\omega_j\right) p(\omega_j) \tag{4.8}$$

or equivalently using the logarithmic likelihood we obtain the equivalent rule

$$g_j\left(\boldsymbol{\theta}^{(i)}\right) = \log p\left(\boldsymbol{\theta}^{(i)}\middle|\omega_j\right) + \log p(\omega_j) \tag{4.9}$$

$$\hat{\omega}^{(i)} = \arg\max_j g_j\left(\boldsymbol{\theta}^{(i)}\right) \tag{4.10}$$

where $g_j(\cdot)$ is the so called discriminant function.

From (4.7) and (4.9) the discriminant functions becomes

$$g_j\left(\boldsymbol{\theta}^{(i)}\right) = -\frac{1}{2}\left(\boldsymbol{\theta}^{(i)} - \boldsymbol{\mu}_j\right)^T \boldsymbol{\Sigma}_j^{-1}\left(\boldsymbol{\theta}^{(i)} - \boldsymbol{\mu}_j\right) - \frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}_j| + \log p\left(\omega_j\right). \tag{4.11}$$

4.6.3 Experimental results

The goal is to use EEG data to train off-line and subsequently test the capacity of our classification scheme to identify *seizure* and *non-seizure* signals. Table 4.8 reports the performance of each classification method assessed by using a leave-one-out cross-validation approach to calculate the confusion matrix [185]. These performance matrices are composed of the following measures that characterize the different aspects of the classifiers: the sensitivity or true positives rate (TPR); the false positive rate (FPR); the sensitivity or true negative rate (TNR); and the overall classification accuracy (ACC), calculated as the total number of correct classifications out of 36 events (18 seizure and 18 non-seizure). Notice that the classification results are performed and reported separately for each brain rhythm or frequency band because this information is relevant to neurologists and allows discriminating clinical events of different nature. We observe from tables 4.6 and 4.7 the independent contributions for each 2D wavelet decomposition. The Table 4.6 correspond to the independent contribution between before the seizure and the seizure, where the best sensitivity or true positives rate (TPR) is for Delta band, Theta horizontal band, Alpha vertical band, Alpha diagonal band, all Beta bands and all Gamma bands, while the best sensitivity or true negative rate (TNR) is for Delta band, Beta horizontal band and Gamma horizontal band. The Table 4.7 correspond to the independent contribution between the seizure and after seizure, where the best sensitivity or true positives rate (TPR) is for Alpha vertical band, Alpha diagonal band, Beta horizontal band and Gamma horizontal band, while the best sensitivity or true negative rate (TNR) is for Theta vertical band, Theta diagonal band, all Alpha bands, all Beta bands and all Gamma bands. Table 4.8 reports the average of all 2D wavelet decomposition contributions. We observe that the method detects correctly in terms of overall accuracy (ACC), sensitivity (TPR) and specificity (TNR) for most frequency bands. In addition, the average latency (time delay) is also reported between the annotated seizure onset and the detection by the method in Table 4.9.

Figures 4.4 to 4.8 shows scatter plots of the generalized Gaussian parameters σ and τ for *seizure* events (red stars) and *non-seizure* events (blue squares and black diamonds) observed through the all frequency bands. We observe that the proposed representation, based on a generalized Gaussian model for the wavelet coefficients, leads to a very clear discrimination of *seizure* and *non-seizure*

Metric	δ	θH	θV	θD	αH	αV	αD	βH	βV	βD	γH	γV	γD
TPR	0.94	0.78	0.94	0.56	0.67	0.89	1.00	1.00	1.00	1.00	1.00	0.89	0.83
TNR	0.94	0.72	0.72	0.67	0.78	0.67	0.61	0.94	0.56	0.72	0.94	0.67	0.61
FPR	0.06	0.28	0.28	0.33	0.22	0.33	0.39	0.06	0.44	0.28	0.06	0.33	0.39
ACC	0.94	0.75	0.83	0.61	0.72	0.77	0.80	0.97	0.77	0.86	0.97	0.77	0.72

TABLE 4.6: Seizure detection performance for each independent contribution between before the seizure and the seizure by using multivariate Bayesian classification for each brain rhythm and for 36 events (18 *seizure* and 18 *non-seizure*) of the Children's Hospital Boston database, in terms of: TPR = True Positives Rate or Sensitivity; TNR = True Negative Rate or Specificity; FPR = False positive Rate; and ACC = Accuracy [$\pm$ standard deviation]. δ: Delta band, θH: Theta horizontal band, θV: Theta vertical band, θD: Theta diagonal band, αH: Alpha horizontal band, αV: Alpha vertical band, αD: Alpha diagonal band, βH: Beta horizontal band, βV: Beta vertical band, βD: Beta diagonal band, γH: Gamma horizontal band, γV: Gamma vertical band, γD: Gamma diagonal band. To simplify the visual interpretation we highlight in red the metric that achieves the highest sensitivity, specificity, and overall accuracy for each frequency band

Metric	δ	θH	θV	θD	αH	αV	αD	βH	βV	βD	γH	γV	γD
TPR	0.72	0.72	0.72	0.78	0.78	0.83	0.83	0.89	0.72	0.67	0.89	0.78	0.67
TNR	0.72	0.72	1.00	1.00	0.72	1.00	1.00	0.94	1.00	1.00	1.00	0.94	0.94
FPR	0.28	0.28	0.00	0.00	0.28	0.00	0.00	0.06	0.00	0.00	0.00	0.06	0.06
ACC	0.72	0.72	0.86	0.88	0.75	0.91	0.91	0.91	0.86	0.83	0.94	0.86	0.80

TABLE 4.7: Seizure detection performance for each independent contribution between the seizure and after the seizure by using multivariate Bayesian classification for each brain rhythm and for 36 events (18 *seizure* and 18 *non-seizure*) of the Children's Hospital Boston database, in terms of: TPR = True Positives Rate or Sensitivity; TNR = True Negative Rate or Specificity; FPR = False positive Rate; and ACC = Accuracy [$\pm$ standard deviation]. δ: Delta band, θH: Theta horizontal band, θV: Theta vertical band, θD: Theta diagonal band, αH: Alpha horizontal band, αV: Alpha vertical band, αD: Alpha diagonal band, βH: Beta horizontal band, βV: Beta vertical band, βD: Beta diagonal band, γH: Gamma horizontal band, γV: Gamma vertical band, γD: Gamma diagonal band. To simplify the visual interpretation we highlight in red the metric that achieves the highest sensitivity, specificity, and overall accuracy for each frequency band

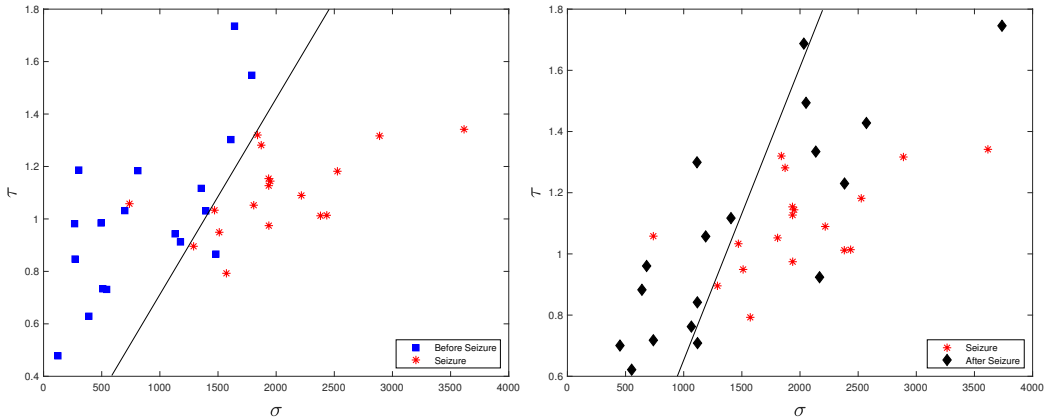
events. In particular, notice that by using this representation it is possible to discriminate events with separating line or hyper-plane, which is essentially what is achieved by using the multivariate Gaussian classifier. Remember that, once the seizure finished, there are several medical pathological factors that cause the signal to take time to stabilize, see Section 2.9.1. This explains why after the seizure, the black diamonds, don't have good discrimination with respect to the seizure, see for example Approximation Delta Band from Figure 4.4 or Horizontal Theta Band from Figure 4.5.

Metric	Delta Band (δ)	Theta Band (Θ)	Alpha Band (α)	Beta Band (β)	Gamma Band (γ)
TPR	0.83 $\pm$0.16	0.75 $\pm$0.12	0.83 $\pm$0.11	0.88 $\pm$0.15	0.84 $\pm$0.11
TNR	0.83 $\pm$0.16	0.81 $\pm$0.15	0.80 $\pm$0.17	0.86 $\pm$0.18	0.85 $\pm$0.17
FPR	0.17 $\pm$0.16	0.19 $\pm$0.15	0.20 $\pm$0.17	0.14 $\pm$0.18	0.15 $\pm$0.17
ACC	0.83 $\pm$0.16	0.77 $\pm$0.10	0.80 $\pm$0.08	0.86 $\pm$0.07	0.83 $\pm$0.10

TABLE 4.8: Seizure detection performance by using multivariate Bayesian classification for each brain rhythm and for 36 events (18 *seizure* and 18 *non-seizure*) of the Children's Hospital Boston database, in terms of: TPR = True Positives Rate or Sensitivity; TNR = True Negative Rate or Specificity; FPR = False positive Rate; and ACC = Accuracy [$\pm$ standard deviation].

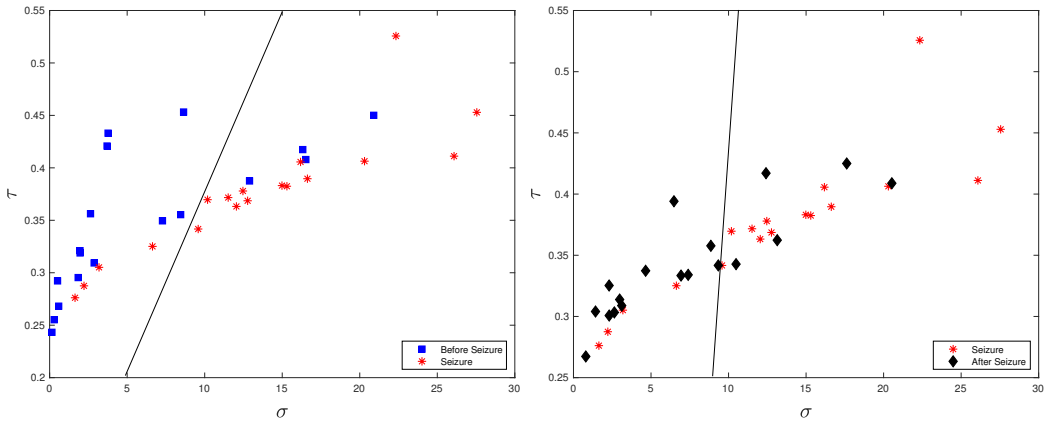
Proposed					state-of-the-art		
Delta band (δ)	Theta band (θ)	Alpha band (α)	Beta band (β)	Gamma band (γ)	[12]	[13]	[14]
4.2	4.1	4.2	4.2	4.2	4.5	3.4	7.2

TABLE 4.9: Average latency between seizure onset and detection (in seconds), for the proposed method on each spectral band, and for the state-of-the-art methods. [12, 13, 14].

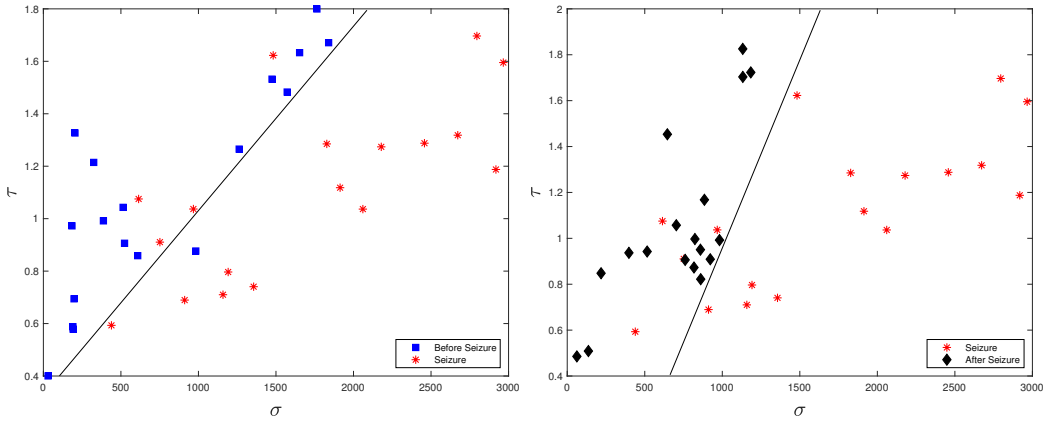


(a) Aproximation Delta band (b) Aproximation Delta band

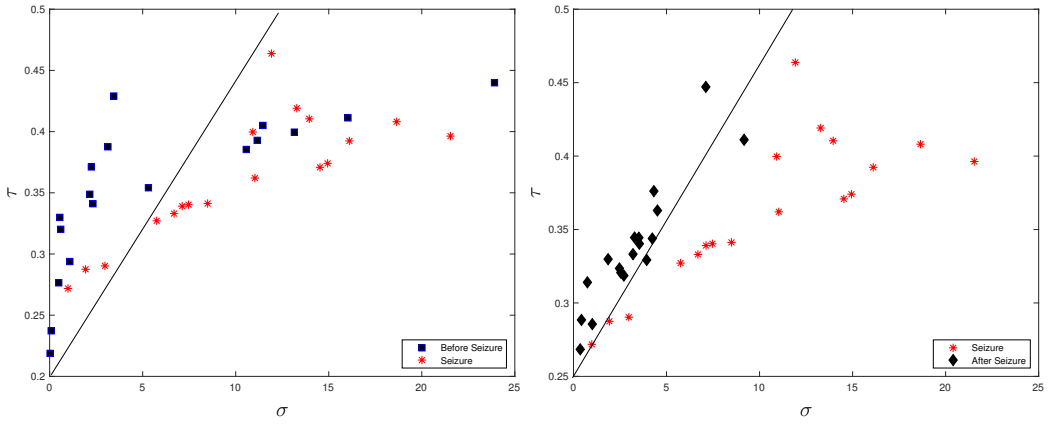
FIGURE 4.4: Scatter plots for the generalized Gaussian parameters σ and τ for seizure events (red stars) and non-seizure events (blue squares for before the seizure and black diamonds for after seizure) observed through the *Delta* frequency band, showing the good linear discrimination power of the proposed approach.



(a) Horizontal Theta band (b) Horizontal Theta band

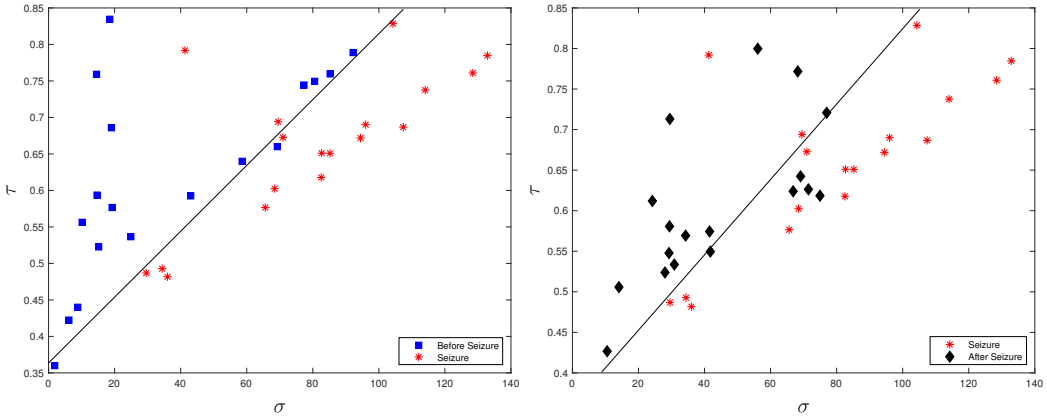


(c) Vertical Theta band (d) Vertical Theta band

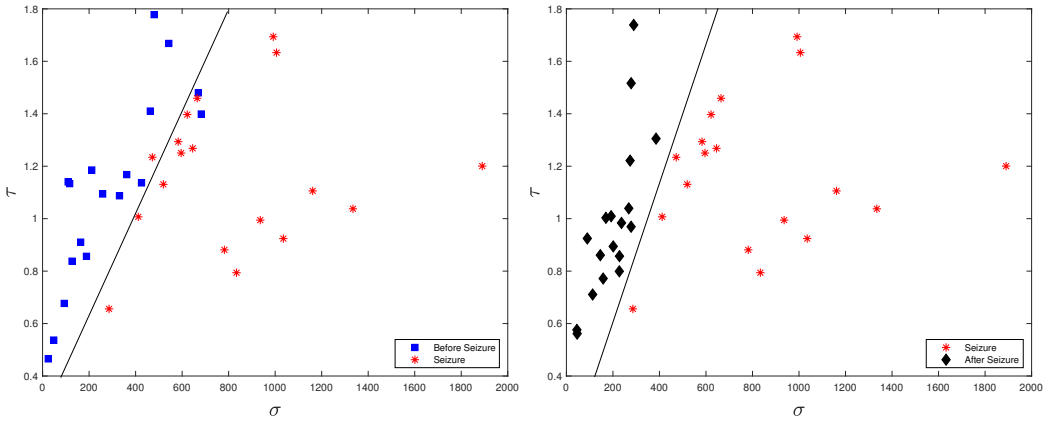


(e) Diagonal Theta band (f) Diagonal Theta band

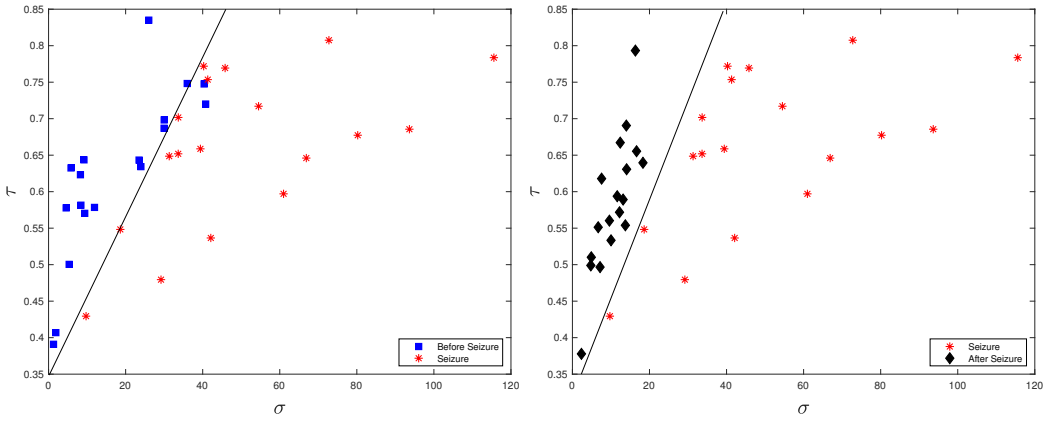
FIGURE 4.5: Scatter plots for the generalized Gaussian parameters σ and τ for seizure events (red stars) and non-seizure events (blue squares for before the seizure and black diamonds for after seizure) observed through the *Theta* frequency band, showing the good linear discrimination power of the proposed approach.



(a) Horizontal Alpha Band (b) Horizontal Alpha Band

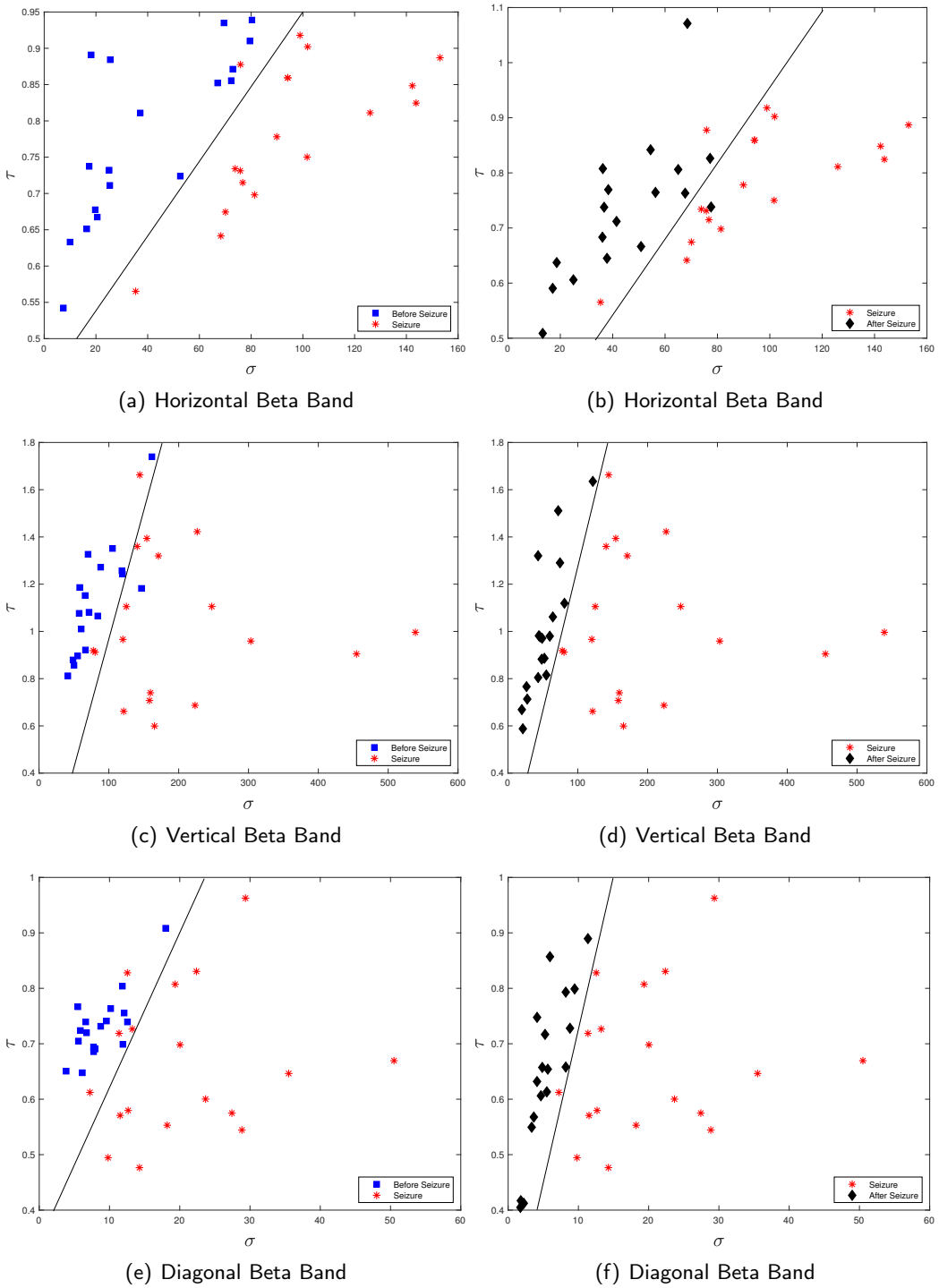


(c) Vertical Alpha Band (d) Vertical Alpha Band



(e) Diagonal Alpha Band (f) Diagonal Alpha Band

FIGURE 4.6: Scatter plots for the generalized Gaussian parameters σ and τ for seizure events (red stars) and non-seizure events (blue squares for before the seizure and black diamonds for after seizure) observed through the *Alpha* frequency band, showing the good linear discrimination power of the proposed approach.

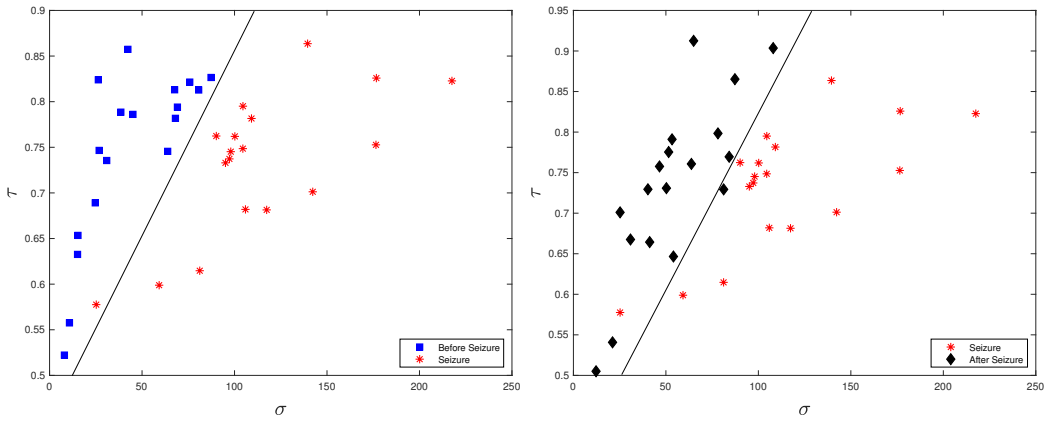


(a) Horizontal Beta Band (b) Horizontal Beta Band

(c) Vertical Beta Band (d) Vertical Beta Band

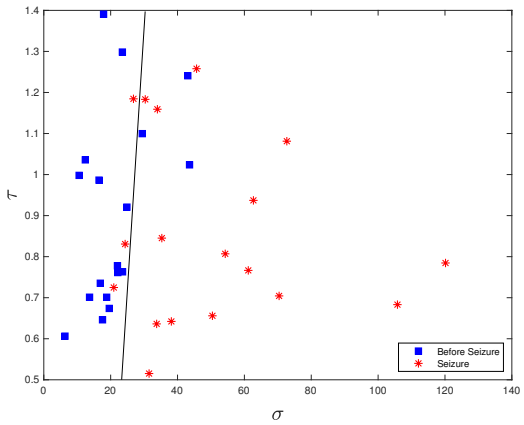
(e) Diagonal Beta Band (f) Diagonal Beta Band

FIGURE 4.7: Scatter plots for the generalized Gaussian parameters σ and τ for seizure events (red stars) and non-seizure events (blue squares for before the seizure and black diamonds for after seizure) observed through the *Beta* frequency band, showing the good linear discrimination power of the proposed approach.

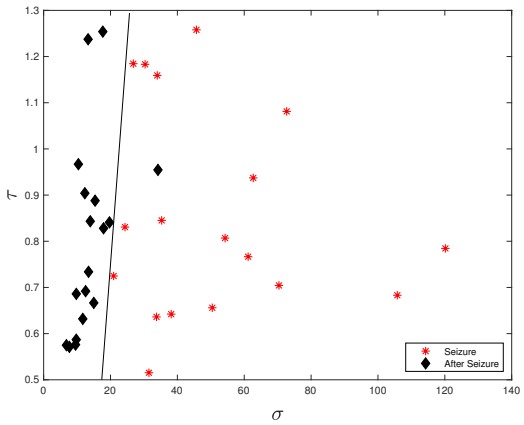


(a) Horizontal Gamma Band

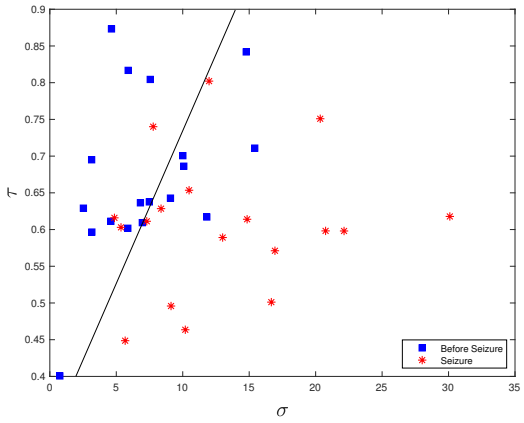
(b) Horizontal Gamma Band



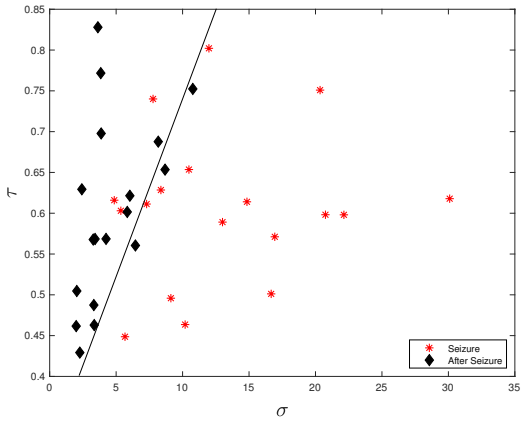
(c) Vertical Gamma Band



(d) Vertical Gamma Band



(e) Diagonal Gamma Band



(f) Diagonal Gamma Band

FIGURE 4.8: Scatter plots for the generalized Gaussian parameters σ and τ for seizure events (red stars) and non-seizure events (blue squares for before the seizure and black diamonds for after seizure) observed through the *Gamma* frequency band, showing the good linear discrimination power of the proposed approach.

4.7 Ensemble Bagging Classifier

Ensemble machine learning methods have been developed to enhance the performance of individual classifiers [200]. The underlying principle consists in combining a collection of *weak* classifiers through a suitable manner. The more popular combination schemes are arithmetic or geometric averaging rule, stacking and majority voting rule [201]. Ensemble bagging (stands for **B**ootstrap **Agg**regat**ing**) relies on bootstrap replicates of the training set [202]. The classifier outputs are combined by the plurality vote. This technique allows increasing the size of the training set, decreasing the variance, and increasing the accuracy and narrowly tuning the prediction to expected outcome [200]. Such classifiers can be optimal in terms of stability and predictive accuracy for datasets with imbalanced class distributions, unstable models or for data mining [203, 204, 205]. Ensemble bagging is widely used in bioinformatics, particularly in protein prediction [206, 207] and recently was used in automatic detection of iEEG bad channels [203]. In this section, we study the Shannon entropy of each brain rhythm, based on the probability density function (PDF) of the generalized Gaussian distribution (GGD). Brain rhythms are obtained through wavelet decomposition. An ensemble bagging method is used to classify EEG signals as *seizure* or *non-seizure*. The classification parameters use the entropy and the scale and shape parameters from the GGD. The motivation relates to the fact that averaging measurements can lead to a more stable and reliable estimate, as the influence of random fluctuations in single measurements is reduced. By building an ensemble of slightly different models from the same training data, we might be able to similarly reduce the influence of random fluctuations in single models [208]. The random fluctuations in epilepsy can be modeled according to spontaneous neural or chaotic activity by using the entropy. The idea is to characterize the dynamic EEG signal by determining the sudden changes in the epileptic signals [209, 210]. Therefore, the random fluctuations that are typical of the variation of the uncertainty can be determined when the entropy is used [101]. In this study, we train decision trees having low bias and high variances to discriminate between seizure and non-seizure [185, 208]. To accurately predict responses, we combine these tree by an ensemble technique in order to reduce the variance and maintain the bias interchangeably low.

4.7.1 Classifier Parameters

Ensemble bagging classification for onset detection is summarized in Figure 4.9. It is very similar to linear classifier parameters from subsection 4.5.1, the most important difference is the use of the Shanon entropy ε from generalized Gaussian parameters σ and τ. Rényi entropy [96], for the PDF from eq. (4.1) is defined by

$$\mathcal{J}_R(\zeta) = \frac{1}{1-\zeta} \log \left\{ \int f^{\zeta}(x;\sigma,\tau)dx \right\} \tag{4.12}$$

where $\zeta > 0$ and $\zeta \neq 1$, then solving the integral of eq. (4.12) for the PDF from eq. (4.1) one obtains

$$\begin{aligned}\int_{-\infty}^{\infty} f^{\zeta}(x;\sigma,\tau)dx &= \int_{-\infty}^{\infty} \frac{\tau^{\zeta}}{(2\sigma)^{\zeta}\Gamma^{\zeta}(\tau^{-1})} \exp\left(-\left|\frac{x}{\sigma}\right|^{\tau}\right) dx \\ &= \frac{\tau^{\zeta}}{(2\sigma)^{\zeta}\Gamma^{\zeta}(\tau^{-1})} \frac{2\sigma\zeta^{-(\tau^{-1})}\Gamma(\tau^{-1})}{\tau} \int_{-\infty}^{\infty} \frac{\tau}{2\sigma\zeta^{-(\tau^{-1})}\Gamma(\tau^{-1})} \exp\left(-\left|\frac{x}{\sigma\zeta^{-(\tau^{-1})}}\right|^{\tau}\right) dx \\ &= \frac{\tau^{\zeta}}{(2\sigma)^{\zeta}\Gamma^{\zeta}(\tau^{-1})} \frac{2\sigma\zeta^{-(\tau^{-1})}\Gamma(\tau^{-1})}{\tau} \end{aligned} \tag{4.13}$$

Thus, eq. (4.12) takes the expression

$$\mathcal{J}_R(\zeta) = \frac{\log\zeta}{\tau(1-\zeta)} - \log\left\{\frac{\tau}{2\sigma\Gamma(\tau^{-1})}\right\}. \tag{4.14}$$

Shannon entropy defined by $E[-\log f(\boldsymbol{X})]$ is the particular case of eq. (4.14) for $\zeta \to 1$. Then limiting in (4.14) and using L'Hopital's rule, one obtains the entropy for the generalized Gaussian distribution PDF

$$\varepsilon = E[-\log f(\boldsymbol{X})] = \tau^{-1} - \log\left\{\frac{\tau}{2\sigma\Gamma(\tau^{-1})}\right\}. \tag{4.15}$$

4.7.2 SOD by ensemble of bagged decision trees classification

Let $\mathcal{M}_t : \omega \to \{0,1\}$ be the t_{th} weak binary class for the tree $t = \{1,\cdots,\mathcal{T}\}$, where 0 is for *non-seizure* event and 1 is for *seizure* event; and $\boldsymbol{p} = [\sigma,\tau,\varepsilon] \in \omega$ the parameters to be classified. Then to combine the outputs $\mathcal{M}_1(\boldsymbol{p}),\cdots\mathcal{M}_{\mathcal{T}}(\boldsymbol{p})$ into a single tree-class prediction, a weight linear combination of the outputs of the weak tree-classifiers, can be used through an ensemble prediction function $\mathcal{M} : \omega \to \{0,1\}$ such that

$$\mathcal{M}(\boldsymbol{p}) = \text{sign}\left(\sum_{t=1}^{\mathcal{T}} w_t \mathcal{M}_t(\boldsymbol{p})\right) \tag{4.16}$$

where $w_1,\cdots,w_{\mathcal{T}}$ is a set of weights of the tree t, according to the average vote from all trees in the ensemble.

Consider a dataset $\mathcal{D} = \{d_1, d_2, .., d_N\}$ with $d_i = (\boldsymbol{p}_i, \omega^{(i)})$, where $\omega^{(i)}$ is a class label, 1 for *seizure* and 0 for *non-seizure*. Bagging algorithm returns the ensemble as a set of models by using decision trees according to each region given by the minimum and maximum values from vector $\boldsymbol{p}$. Combining the predictions $\mathcal{T}$ from the different models by average, see algorithm 2. The class predicted is the class that yields the largest weighted average as

$$\arg\max_{\omega} \langle\, \mathcal{M}(\boldsymbol{p})\rangle \tag{4.17}$$

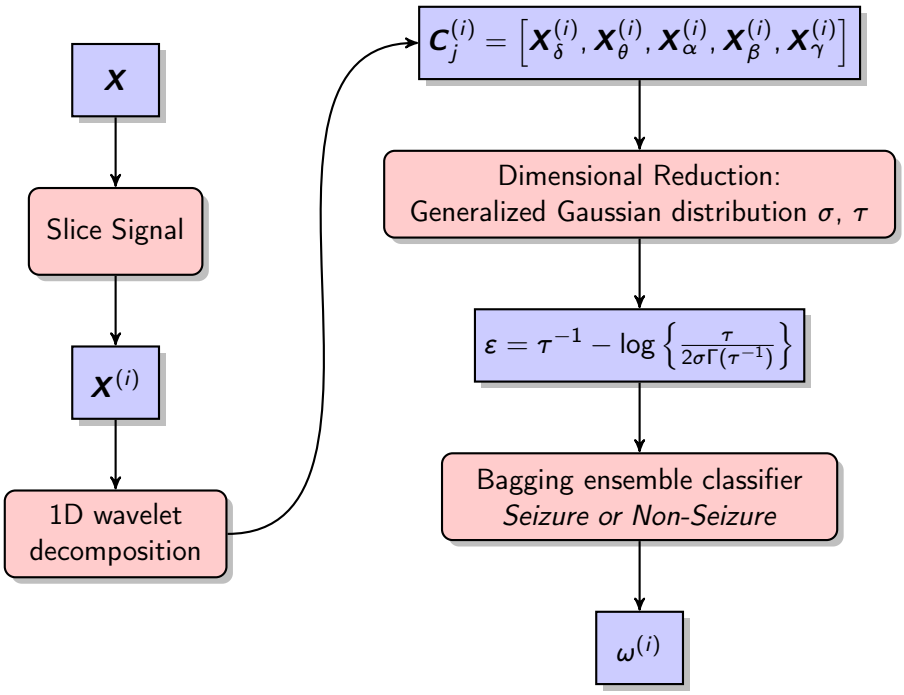


FIGURE 4.9: Algorithm used in bagging ensemble classifier.

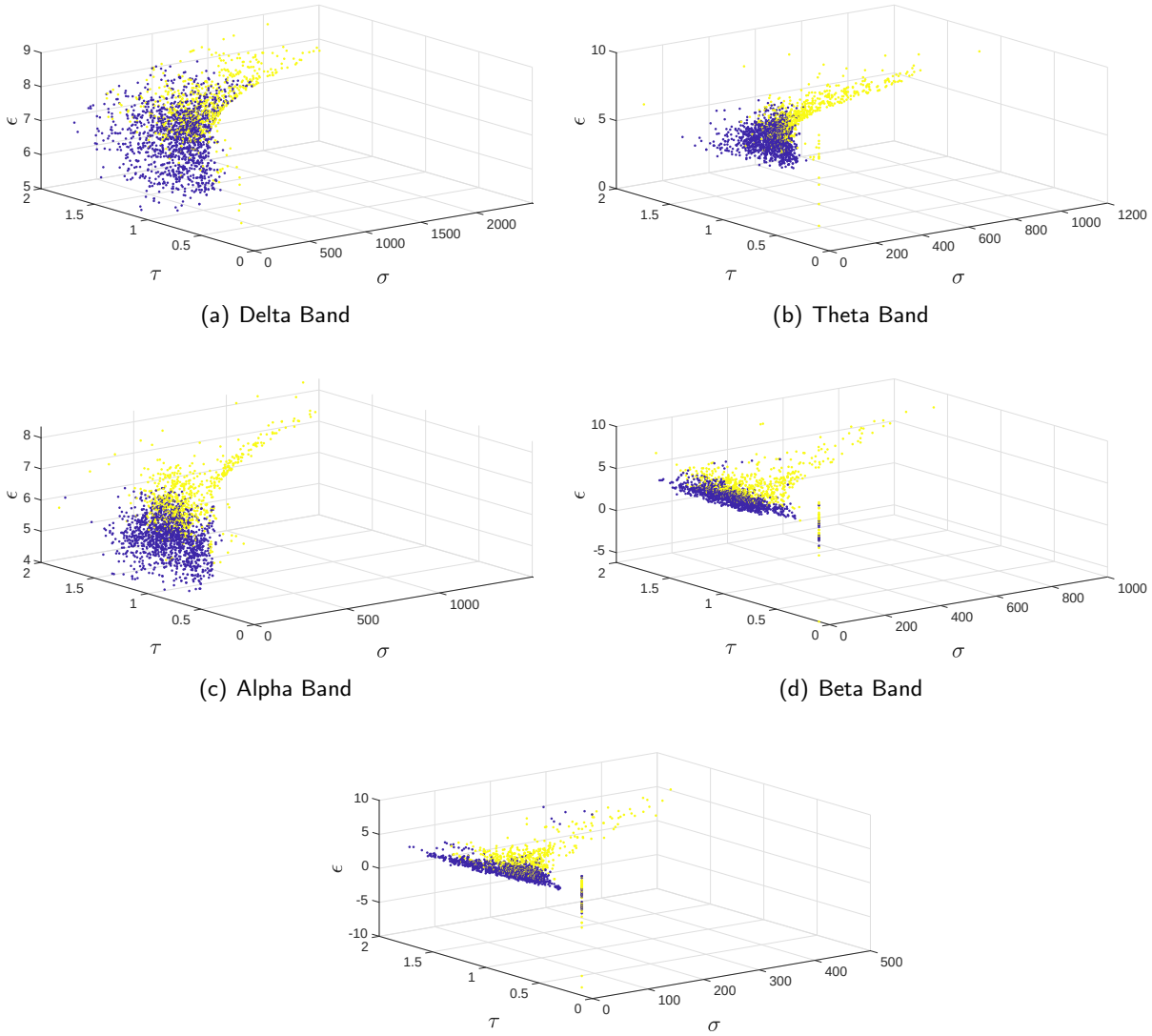
Data: data set $\mathcal{D}$; ensemble size $\mathcal{T}$; learning algorithm $\mathcal{A}$
Result: ensemble of models whose predictions are to be combined by averaging.
for *t=1 to* $\mathcal{T}$ **do**
  build a bootstrap sample $\mathcal{D}t$ from $\mathcal{D}$ by sampling $|\mathcal{D}|$ data points with replacement;
  run $\mathcal{A}$ on $\mathcal{D}t$ to produce a model $\mathcal{M}_t$;
end

Algorithm 2: Bagging($\mathcal{D}$,$\mathcal{T}$,$\mathcal{A}$) – train an ensemble of models from bootstrap samples [208].

4.7.3 Experimental results

Figure 4.10 shows the discrimination properties of the proposed vector representation $\boldsymbol{p} = [\sigma, \tau, \varepsilon] \in \mathbb{R}^3$ from the wavelets coefficients. We can see the direct relation between σ and ε, both increase as they grow in their scale of values for the *seizure* events (yellow circles) with respect to *non-seizure* events (blue circles). For illustration, Figure 4.11 shows the different ranges in the box plots for the entropy, clearly discriminating the two classes, *seizure* or *non-seizure*. For all brain rhythms, except for delta band, the maximum and minimum values for each box together with the quartiles can help to classify based on a thresholding approach.

Table 4.10 reports the mean and standard deviation values for all signals with the proposed vector representation $\boldsymbol{p} = [\sigma, \tau, \varepsilon] \in \mathbb{R}^3$, showing a clear difference between *seizure* events and *non-seizure* events. The 95% confidence interval (IC95%) has different values, which permits to set a threshold for detecting the seizure, with a proper choice, this can help to determine the duration and amplitude between *seizure* events and *non-seizure* events; those are the most important factor affecting the performance of automated detection [211].



(a) Delta Band

(b) Theta Band

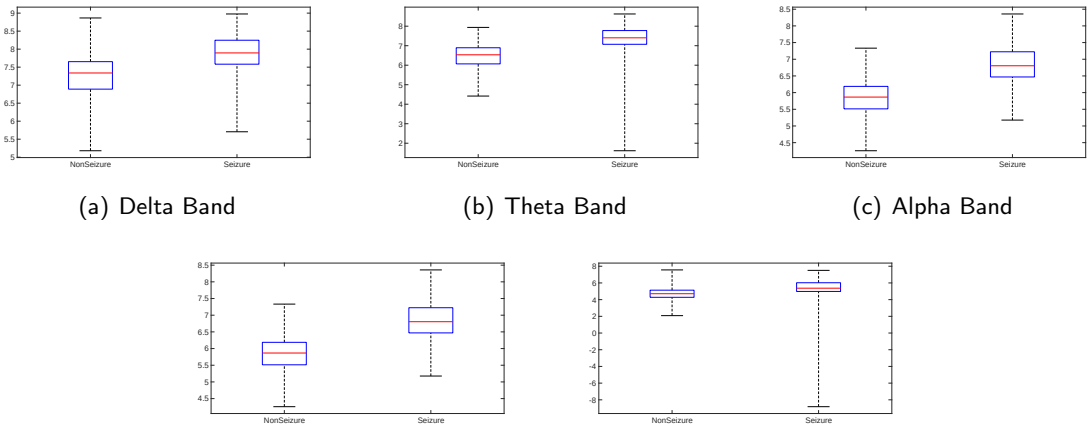
(c) Alpha Band

(d) Beta Band

(e) Gamma Band

FIGURE 4.10: Scatter plots from vector $\boldsymbol{p} = [\sigma, \tau, \varepsilon]$ observed through all brain rhythms using 105 events: 35 *seizures* (yellow dots) and 70 *non-seizures* (blue dots). We can see how the *seizure* event concentrates on high values of σ and ϵ.

To assess the performance of the proposed methodology, a supervised testing approach was adopted and used the 105 events described above to train and test the method with an exhaustive cross-validation technique based on a leave-one-out approach of the vector $\boldsymbol{p} = [\sigma, \tau, \varepsilon] \in \mathbb{R}^3$ with 35 ensemble learnings. Table 4.11 reports the rate of correct classification in terms of: TPR = True Positives Rate or *Sensitivity*; TNR = True Negative Rate or *specificity*; FPR = False Positive Rate; FNR = False Negative Rate; Error Rate; and ACC = Accuracy (ACC). In addition, the average latency (time delay) is also reported between the annotated seizure onset and the detection by the method in Table 4.12.



(a) Delta Band (b) Theta Band (c) Alpha Band

(d) Beta Band (e) Gamma Band

FIGURE 4.11: Box plots of Shannon entropy observed through all brain rhythms using 105 events (35 *seizures* and 70 *non-seizures*). The maximum an minimum values for each box together with the quartiles can help to classify based on a thresholding approach.

	Non-Seizure			Seizure		
Bands	mean	std	IC95%	mean	std	IC95%
Delta	106.23	75.09	[102.28, 110.17]	202.78	122.53	[193.68, 211.89]
Theta	25.84	19.60	[24.81, 26.87]	85.55	67.49	[80.54, 90.56]
Alpha	22.08	14.15	[21.34,22.83]	75.11	67.32	[70.10, 80.11]
Beta	11.96	6.95	[11.59, 12.32]	37.44	44.05	[34.16, 40.71]
Gamma	6.83	6.21	[6.50, 7.15]	35.01	43.57	[31.78, 37.30]

TABLE 4.10: Comparison between *means*, *standard deviations* of the entropy and 95% confidence interval (*IC95%*) of seizure and non-seizure, using 105 events (35 seizures and 70 non-seizures) for each brain rhythm. We can see how one can set a threshold for detecting the seizure

Metric	Delta Band (δ)	Theta Band (Θ)	Alpha Band (α)	Beta Band (β)	Gamma Band (γ)
TPR	0.95 $\pm$0.03	0.97 $\pm$0.03	0.98 $\pm$0.02	0.98 $\pm$0.02	0.99 $\pm$0.01
TNR	0.87 $\pm$0.05	0.94 $\pm$0.05	0.94 $\pm$0.06	0.94 $\pm$0.08	0.95 $\pm$0.07
FPR	0.13 $\pm$0.05	0.06 $\pm$0.05	0.06 $\pm$0.06	0.06 $\pm$0.08	0.05 $\pm$0.07
ACC	0.92 $\pm$0.03	0.96 $\pm$0.02	0.97 $\pm$0.02	0.97 $\pm$0.04	0.97 $\pm$0.04

TABLE 4.11: Seizure detection performance by using ensemble bagged classification for each brain rhythm and for 105 events (35 *seizure* and 70 *non-seizure*) of the Children's Hospital Boston database, in terms of: TPR = True Positives Rate or Sensitivity; TNR = True Negative Rate or Specificity; FPR = False positive Rate; and ACC = Accuracy [$\pm$ standard deviation].

Proposed					state-of-the-art		
Delta band (δ)	Theta band (θ)	Alpha band (α)	Beta band (β)	Gamma band (γ)	[12]	[13]	[14]
4.4	4.2	4.3	4.3	4.2	4.5	3.4	7.2

TABLE 4.12: Average latency between seizure onset and detection (in seconds), for the proposed method on each spectral band, and for the state-of-the-art methods. [12, 13, 14].

4.8 Logistic Regression Classifier

Logistic regression is one of the most common multivariate analysis models used in biomedical applications for analyzing binary outcome data [185, 212]. The choice of the explicative variables that should be included in the logistic regression model is based on prior knowledge of epilepsy and the statistical correlation between the variable and the epileptic event [213, 214], in our case the correlation between the *seizure* and the largest Lyapunov exponent. In recent works, the logistic regression classifier coupled with Cox regression has been used to construct time to first EEG seizure in neonates subjects [215] or to classifier the significant non-antiepileptic drug predictors of psychiatric and behavioral side effects rate [216], or to estimate the average recurrence risk of ictal asystole and its determining factors in people with epilepsy [217].

The EEG signal was decomposed independently for each brain rhythms using ICA to study the epileptic dynamic features of EEG during *seizure* (ictal) and *non-seizure* (interictal) behavior. The difference between typical ictal and interictal feature values enables us to distinguish between the two states, which are identified through the largest Lyapunov exponents (LLE). The results allow us to differentiate the distinctive and appreciable changes during epileptic seizures, discriminating normal from abnormal brain activity.

Independent component analysis (ICA) is a method to find underlying sources (or components) from multivariate or multidimensional statistical data. The main idea of ICA is to find a linear representation of non-Gaussian data in such a way that the components are statistically independent. The advantage of identifying these independent features is that, when used in combination with other methods such as largest Lyapunov exponents (LLE), it makes possible to distinguish between *seizure* and *non-seizure* events in a higher dimensional feature space [32]. ICA has been successfully used by the scientific community and has been applied to numerous signal processing problems in diverse areas such as biomedicine, bioengineering, communications, finance, and remote sensing; and keeps evolving [218]. ICA is widely used in EEG data and its applications are very varied, for example in [219] was demonstrated that ICA can be an efficient approach to separate responses related to epilepsy which are commonly obtained through fMRI studies, or in [220] to select the PROJection onto Independent Components (PROJIC) from EEG data collected during fMRI acquisitions to detect Inter-ictal epileptiform discharges, or by using a new deflation ICA algorithm called penalized semialgebraic unitary deflation (P-SAUD) in order to remove artifacts from interictal epileptic spikes [221].

Largest Lyapunov exponents (LLE) is a time-dependent analysis technique that can be used to infer the properties of a system [222, 223, 224, 225]. In a medical context, they describe the time interval over which the system's evolution diverges, helping to discriminate *seizures* from *non-seizures* in epileptic signals [226, 227]. In recent studies, the Lyapunov coefficients were applied as a filter-noise that can be used as an epilepsy detector [228], as features in order to predict epileptic seizures in synthetic signals [229], coupled with the adaptive Teager energy to seizure detection in long-term signals with a sensitivity of 91% and a specificity of 86% [230], or by using point-process to correlate the heartbeat dynamics with the epileptic signals and SVM classifier with an accuracy of 73.91% [231], for EEG patterns classification based on continuous neural networks

by using a generalization-regularization with an accuracy of 97.2% [232]; as a seizure prediction in intracraneal signals (iEEG) with a sensitivity of 89.8% and a specificity of 96.7% [233], or to detect metabolic encephalopathy by using SVM with a specificity of 100% and a sensitivity of 95.33% [234].

4.8.1 Classifier parameters

The process of logistic regression for onset detection is summarized in Figure 4.12.

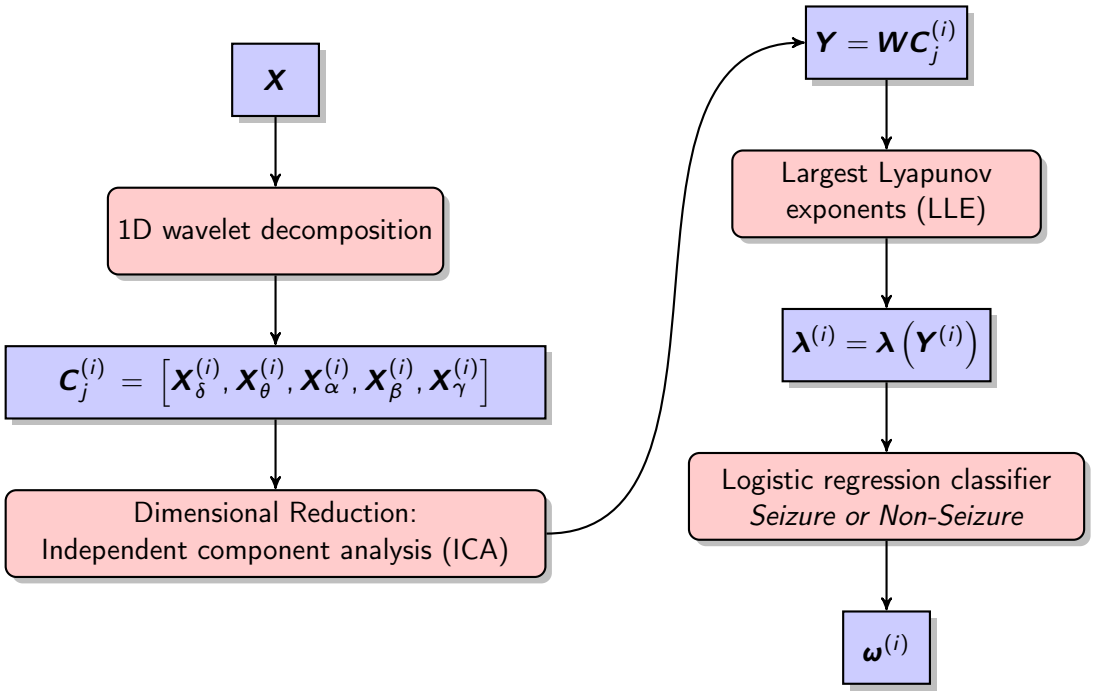


FIGURE 4.12: Algorithm used in logistic regression classifier.

We start from the parameter-vector

$$C_j^{(i)} = \left[X_\delta^{(i)}, X_\theta^{(i)}, X_\alpha^{(i)}, X_\beta^{(i)}, X_\gamma^{(i)}\right] \tag{4.18}$$

The Independent component analysis (ICA) is a representation of a signal (the brain rhythm of eq. (4.18) in this case) through a set of independent constituent components given by the likelihood

$$p\left(C_j^{(i)}|S\right) = \prod_{t=1}^{T} p\left(\chi_t|S_t\right) \tag{4.19}$$

where $p\left(C_j^{(i)}|S\right)$ is the joint probability distribution, $p\left(\chi_t|S_t\right)$ are the marginal distributions, $S \in \mathbb{R}^{T\times N}$ are the unknown sources, χ_t is the observed signal matrix from the wavelet coefficients $C_j^{(i)}$, and T is the number of independent components (see Figure 4.13).

We assume that the source signals arrive at the electrodes at the same time instantaneously, thus the problem of separating sources corresponding to the independent components for each

brain rhythm, of eq. (4.18) is given by

$$\boldsymbol{\chi}_t = \boldsymbol{H}\boldsymbol{s}_t + \boldsymbol{\eta} \tag{4.20}$$

where $\boldsymbol{H}$ is the mixing matrix, $\boldsymbol{s}$ is the source matrix and $\boldsymbol{\eta}$ is the noise.

The separation is performed by means of a matrix $\boldsymbol{W} \in \mathbb{R}^{T\times M}$, the so called unmixing matrix, which uses only the information in $\boldsymbol{\chi}_t$ to reconstruct the original source signals (also known as the independent components) as:

$$\boldsymbol{y}_t = \boldsymbol{W}\boldsymbol{\chi}_t \tag{4.21}$$

where $\boldsymbol{y}_t \in \mathbb{R}^{T\times N}$, $\boldsymbol{W} \in \mathbb{R}^{T\times M}$ and $\boldsymbol{\chi}_t \in \mathbb{R}^{M\times N}$.

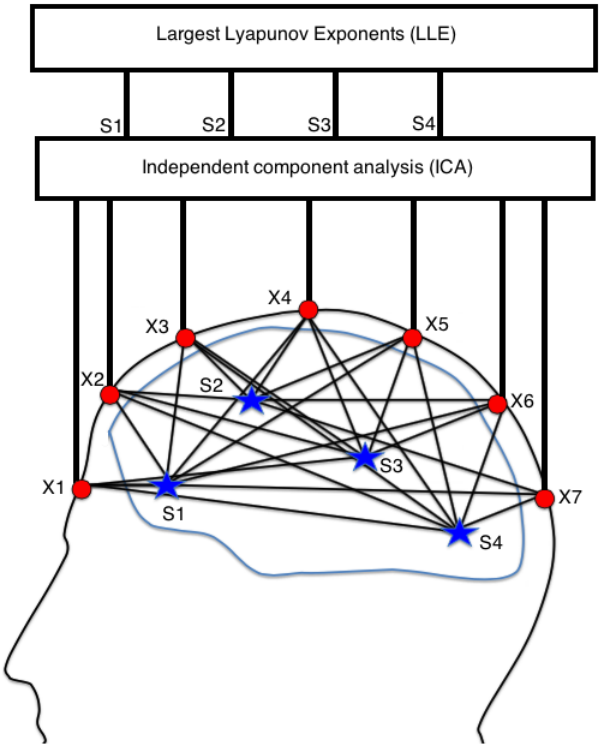


FIGURE 4.13: Example for seven electrodes, namely $\boldsymbol{X}_1, ..., \boldsymbol{X}_7$ and four sources $\boldsymbol{S}_t$ namely $\boldsymbol{s}_1, .., \boldsymbol{s}_4$, representation of assumption that the source signals arrive at the electrodes at the same time instantaneously.

The estimation of the unmixing matrix $\boldsymbol{W}$ in eq. (4.21) is calculated using singular value decomposition (SVD) through the eigenvalue decomposition of the covariance matrix (prewhitening) [235, 236, 237, 238, 239] and the JADE algorithm for real-valued signals [240], to find the best estimation of the independent sources $\boldsymbol{S}$ through

$$\boldsymbol{Y} = \boldsymbol{W}\boldsymbol{C}_j^{(i)} \tag{4.22}$$

The independent sources $\boldsymbol{Y}$ from equation (4.22) for each brain rhythm is split in sets of non-overlapping 2 seconds segments using a rectangular sliding window so that

$$\boldsymbol{Y}^{(i)} = \boldsymbol{\Omega}^{(i)}\boldsymbol{Y} \tag{4.23}$$

See eq. (2.2) for more details about the sliding window.

4.8.2 Largest Lyapunov exponent (LLE)

The nonlinear prediction technique to separate transients from background activity using Lyapunov exponents was first investigated by Leonidas D. Iasemidis and J. Chris Sackellares in [59] where the lowest values of Lyapunov exponents occur during the *seizure*. This gives us an idea of how much the EEG signal background changes when a small perturbation or change occurred during the *seizure* process.

The largest Lyapunov exponent is estimated by means of two-time series, $\boldsymbol{Y}_1^{(i)}$ and $\boldsymbol{Y}_2^{(i)}$ (we would like to remind the reader that $\boldsymbol{Y}^{(i)}$ denotes each segment of the evaluated signal); which originate from the same system and have similar initial conditions [225], defined as a distance vector

$$dist(i) = \left\| \boldsymbol{Y}_1^{(i)} - \boldsymbol{Y}_2^{(i)} \right\| \tag{4.24}$$

and the Lyapunov exponent

$$\boldsymbol{\lambda} = \frac{1}{i} \log \frac{dist(i)}{dist(0)} \tag{4.25}$$

where i is the sample number and $dist(0)$ is the distance between the initial sample points on the two trajectories. A trajectory is a path that the variables trace throughout the phase space. Phase space represents all possible internal states of a system. The divergence value of $\boldsymbol{\lambda}$ magnifies small changes in a trajectory that grow over time [241], this value shows how an increase in distance between trajectories that start from similar conditions become increasingly decorrelated, contrary to convergence. This can be summarized as follows

- If $\boldsymbol{\lambda} > 0$ then the divergence is exponential.
- If $\boldsymbol{\lambda} < 0$ then the convergence is exponential.
- If $\boldsymbol{\lambda} = 0$ then there is no divergence or convergence.

For each segment of eq. (4.23) and each brain rhythm, of eq. (4.18) a largest Lyapunov exponent $\boldsymbol{\lambda}$ is estimated using eq. (4.25) according to the divergence or convergence of the considered value. This allows us to discriminate the divergence or convergence between *seizure* and *non-seizure* in epileptic signals. Two positive Lyapunov exponents were estimated. The presence of a positive exponent is sufficient to detect the seizure [223].

4.8.3 λ-Scaling

Each largest Lyapunov exponent value for each brain rhythm is assigned one scale value between the minimum and the maximum of the standard deviation from LLE, see Table 4.13.

Let $\ell_{sup} = +\lambda_{std}$ and $\ell_{inf} = -\lambda_{std}$, $\lambda_{min} = min(LLE)$ and $\lambda_{max} = max(LLE)$, then the scale value is given by

$$\varsigma = \frac{(\lambda - \lambda_{min})(\ell_{sup} - \ell_{inf})}{\lambda_{max} - \lambda_{min}} + \ell_{inf} \tag{4.26}$$

The proposed seizure detection is a classifier by using logistic regression, that labels each Largest Lyapunov exponents (λ) and their scales (ς) associated with each brain rhythm as seizure or non-seizure. Precisely, five independent two-parameter classifiers are used in parallel to classify the feature vector pairs $\phi_\delta(n) = [\lambda_\delta(n), \varsigma_\delta(n)]$, $\phi_\theta(n) = [\lambda_\theta(n), \varsigma_\theta(n)]$, $\phi_\alpha(n) = [\lambda_\alpha(n), \varsigma_\alpha(n)]$, $\phi_\beta(n) = [\lambda_\beta(n), \varsigma_\beta(n)]$, and $\phi_\gamma(n) = [\lambda_\gamma(n), \varsigma_\gamma(n)]$.

4.8.4 SOD by logistic regression classification

Consider a classification into two possible classes: ω_s for *seizure* and ω_{ns} for *non-seizure*. The posterior probability of class ω_s can be written as

$$p\left(\omega_s|\boldsymbol{Y}^{(i)}\right) = \frac{p(\boldsymbol{Y}^{(i)}|\omega_s)p(\omega_s)}{p(\boldsymbol{Y}^{(i)}|\omega_s)p(\omega_s) + p(\boldsymbol{Y}^{(i)}\omega_{ns})p(\omega_{ns})} \tag{4.27}$$

$$= \frac{1}{1 + \exp(-a)} = \sigma(a) \tag{4.28}$$

$$a = \ln \frac{p(\boldsymbol{Y}^{(i)}|\omega_s)p(\omega_s)}{p(\boldsymbol{Y}^{(i)}i|\omega_{ns})p(\omega_{ns})} \tag{4.29}$$

where $\sigma(.)$ is the logistic sigmoid function, and the class-conditional densities are assumed Gaussian [185]. Then the posterior probability of class ω_s can be written as a logistic sigmoid acting on a linear function of the feature vector ϕ so that

$$p(\omega_s|\phi) = \sigma(\mathrm{w}^T\phi) \tag{4.30}$$

$$p(\omega_{ns}|\phi) = 1 - p(\omega_s|\phi) \tag{4.31}$$

$$\mathrm{w} = \boldsymbol{\Sigma}^{-1}(\mu_1 - \mu_2) \tag{4.32}$$

assuming that all classes share the same covariance matrix $\mathbf{\Sigma}$ and μ are the means of each class. For a data set $\left\{\phi, \omega^{(i)}\right\}$, where $\omega^{(i)} \in \{0,1\}$, 1 for class ω_s and 0 for class ω_{ns}, and $\phi(n) = \phi(\boldsymbol{Y}^{(i)})$ the likelihood can be written

$$p(\boldsymbol{\omega}|w) = \prod_{i=1}^{N} y_i^{\omega^{(i)}} \{1 - y_i\}^{1-\omega^{(i)}} \tag{4.33}$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, .., \omega_N)^T$ and $y_i = p(\omega_s|\phi_b(n))$. It should be noted that the feature vector $\phi_b(n)$ is given by each LLE (λ) and their scales (ς) of each brain rhythm (b). The methodology used can be summarized in four basic steps through next algorithm

Data: Epileptic EEG signals
Result: Seizure and Non-Seizure detection
begin
1. Find all independent brain rhythms using univariate wavelet 1D decomposition;
2. Compute the independent features of each brain rhythms using ICA decomposition;
3. Compute LLE for all independent brain rhythms of 2.;
4. Scale each LLE from step 3. between the minimum and the maximum of the standard deviation;
5. Seizure detection for each pairs [LLE,scale] by using logistic regression;

end

Algorithm 3: Epileptic seizure detection algorithm

4.8.5 Experimental results

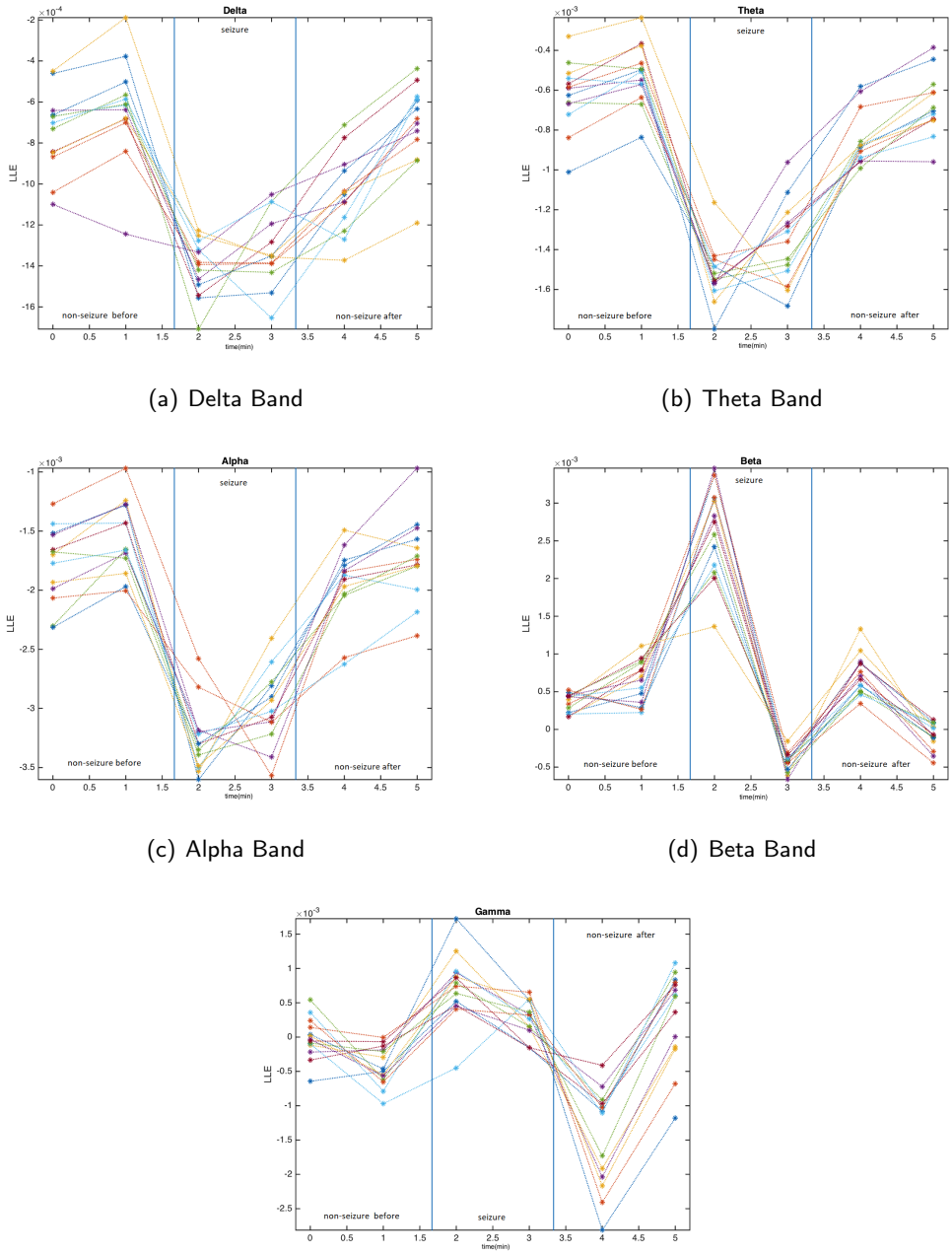
Figures 4.14 show how the EEG signal background changes through the six largest Lyapunov exponents (LLE) from 9 independent components by using ICA. Two LLE before, two LLE during and two after the *seizure* ICA process . In the delta, theta and alpha brain rhythms the largest Lyapunov exponent (LLE) presents the lowest value, while in beta and gamma brain rhythms the opposite happens. This suggests that the algorithm is potentially interesting for epilepsy detection systems because it permits discriminating *seizure* from *non-seizure* in all brain rhythms.

Table 4.13 shows the minimum LLE (λ_{min}), maximum LLE (λ_{max}) and the standard deviation from LLE (λ_{std}) through all the data utilized that permits the use of a threshold approach in order to scale each LLE by using the equation (4.26).

The logistic regression classifier was trained off-line with 20 empirical fold cross-validation. In this experiment, we used two classes: *seizure* and *non-seizure* for each pair $[\lambda_b, \varsigma_b]$ for each brain rhythm (b). Where 264 events correspond to *seizure* and 528 events correspond to *non-seizure*. The performance of the logistic regression classification method through these 792 observations was assessed in terms of overall accuracy classification, and achieves a 100% of sensitivity (True positive rate) and specificity (True false rate) for seizure detection in epilepsy signals with time-delay of 8.9 sec in average for all brain rhythms.

We suggest that these good results in the classification are due to the fact that the LLE coupled with their scaling can discriminate correctly between seizure and non-seizure in all brain rhythms,

as shown in the values of the Table 4.13 and the visual observation of the LLE figures 4.14.



(a) Delta Band

(b) Theta Band

(c) Alpha Band

(d) Beta Band

(e) Gamma Band

FIGURE 4.14: Scatter plot for six largest Lyapunov exponents (LLE) for *seizure* (middle) and *non-seizure* before an after events observed through the different rhythms bands. Lowest valued LLE is in the *seizure* events for delta, theta, and alpha bands, while highest valued LLE are in the *seizure* events for beta and gamma bands. In this example, the start and end of the seizure in this EEG signal were labeled by the neurologist using two lines. The first line divides the EEG signal before the seizure and the second line after the seizure.

Brain rhythm	λ_{min}	λ_{max}	λ_{std}
δ Non-Seizure	-0.00266	0.00028	0.00074
δ Seizure	-0.00328	-0.00003	0.00092
θ Non-Seizure	-0.00249	0.00052	0.00064
θ Seizure	-0.00368	-0.00002	0.00103
α Non-Seizure	-0.00443	0.00117	0.00111
α Seizure	-0.00641	0.00027	0.00147
β Non-Seizure	-0.00311	0.00177	0.00068
β Seizure	-0.00317	0.00346	0.00111
γ Non-Seizure	-0.00281	0.00364	0.00079
γ Seizure	-0.00082	0.00768	0.00149

TABLE 4.13: Minimum, maximum and standard deviation from all LLE for each brain rhythm.

4.9 Conclusions

This chapter presented three new methods to detect epileptic brain activity on-line in EEG signals, with a focus on applications involving real-time constraints and small training datasets. A particularity of the methods is that detection is performed independently for each brain rhythm, following the current medical practices. Detection is achieved by first separating the EEG signals into the five brain rhythms by using a wavelet decomposition, and then using a generalized Gaussian statistical model to map signals onto a low-dimensional representation where classification can be performed efficiently to discriminating between *seizure* and *non-seizure* by linear discriminant analysis, multivariate Bayesian classifier or ensemble bagging classifier by using decision trees (through all brain rhythms) using the entropy from the generalized Gaussian statistical model parameters. The fourth method, using logistic regression of Lyapunov exponents from Independent Component Analysis (ICA) computed independently in each brain rhythms from wavelet decomposition obtained very good results, but computationally, it is very expensive to apply it in real time. All classifiers have similar latency around 4 seconds, except for the Lyapunov exponents around 9 seconds. The four methods are potentially useful for differentiating between *seizure* or *non-seizure* events in epileptic signals and for onset detection, in terms of high sensitivity, specificity, and accuracy. Note that, the model-based classification by using the GGD parameters (scale and shape) permit a correct seizure onset identification in epileptic EEG signals with an acceptable time delay. The best classifier, in general, was the linear discriminant analysis, but the ensemble of bagged decision trees classifier showed the best performance in the specificity. The multivariate Bayesian was the weakest classifier, but it may be possible to improve its performance using a regulation parameter.

Chapter 5
Spike-and-wave epileptiform pattern recognition

5.1 Introduction

This chapter presents three novel methods to detect spike-and-wave discharges (SWD) in EEG signals. The methodology is computationally very efficient, suitable for real-time automation, and can be used to perform the spike-and-wave detection online. The database used was created in Fundación Lucha contra las Enfermedades Neurólogicas Infantiles (FLENI). Method one consist of SWD detection using the generalized Gaussian distribution; method two consists of SWD detection using t-location-scale distribution and method three uses cross-correlation. Methods one and two use k-NN classifiers while method three uses a decision tree classifier.

Spike-and-wave discharge waveform has a regular and symmetric morphology. The information about the morphology and dynamics of EEG signals can be used to accurately identify seizure onset and quantify the severity and dynamical progression of seizure activity. The most relevant EEG features employed to classify epileptogenic abnormality can be categorized in terms of spectral properties, signal morphology and statistical measures [32], see Figure 5.1. In this chapter, we will focus on signal morphology, specifically in spike-and-wave discharge (SWD) pattern in EEG signals.

5.2 Spike-and-Wave discharge (SWD)

A seizure is characterized by the excessive electrical discharges in neurons and such waveforms are known as spikes. Neurologists trained in EEG are able to properly determine an epilepsy diagnosis by analyzing the different types of spikes in the rhythmic activity of the brain. A spike is characterized by short bursts of high amplitude, synchronized and multi-phasic activity, where polarity changes occur several times, which manifest themselves at or around the epileptic focus and stand out from the background EEG.

A spike-and-wave epileptiform is an EEG generalized discharge pattern seen particularly during *absence epilepsy* [242], whose clinical importance lies in cognitive and behavioral disturbances [243]. It is the result of a bilateral synchronous firing of neurons ranging from the neocortex to the thalamus, along with the thalamocortical network [244]. *Absence seizures* are more common

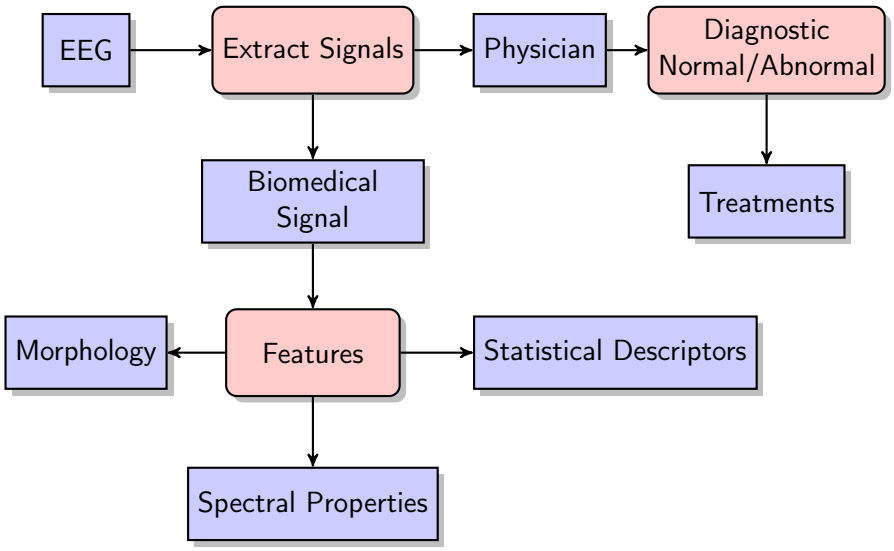


FIGURE 5.1: EEG signals extraction. A physician may determine the patient treatment according to the EEG waveforms; while in a biomedical signal approach the most relevant features employed to classify epileptogenic abnormality can be categorized in terms of spectral properties, signal morphology, and statistical measures.

in children. It causes lapses in awareness, sometimes with staring and it can be so brief they sometimes are mistaken for daydreaming and may not be detected for months. Children between the ages of three and seven exhibit continuous spike-and-wave discharges during slow-sleep. This disorder is found in 0.2%-0.5% of all childhood epilepsy cases. Spike-and-wave activity occupies about 85% of the non-rapid eye movement sleep [245]. This continuous pattern during sleep, like other aspects of the spike-and-wave activity, are not completely understood. However, it is hypothesized that corticothalamic neuronal network involved in oscillating sleep patterns may begin to function as a pathologic discharging source [246].

Spike-and-wave discharge (SWD) is a generalized EEG discharge pattern whose waveform has a regular and symmetric morphology. This morphology can be mathematically described by a Morlet wavelet transform that generates a time-frequency representation of the EEG signal [247, 248, 249, 250]. The *spike* component of a SWD is associated with neuronal firing, whereas the *wave* component is associated with neuronal inhibition or hyperpolarization of neurons [251]. SWD is widely used in mice studies [252, 253, 254, 255], but human testing is limited. Mice have a predisposition for generalized SWD at 7-12 Hz [256] and typically have spontaneous absence-seizure-like-events. The presence of an intact cortex, thalamus and their interconnections is necessary to record them [257, 258].

Existing spike-and-wave discharge detection algorithms can be classified in the following three categories [259]:

1. Algorithms that use the information extracted from changes in the amplitude (magnitude) of the EEG signal when SWD occurs.

2. Detection based on monitoring the energy power in the frequency bands which SWD occupies.

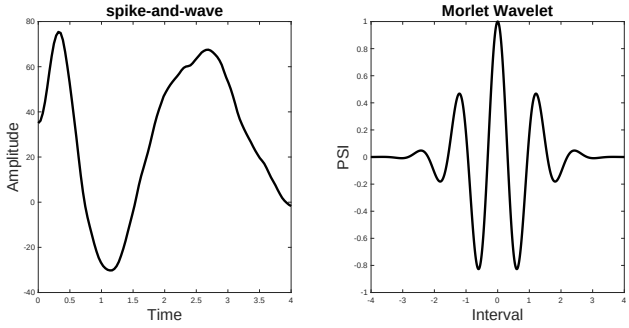


FIGURE 5.2: Spike-and-wave and Morlet wavelet respectively, Note that the scales are different, but for illustration, we can see the symmetric and regular morphology in both signals.

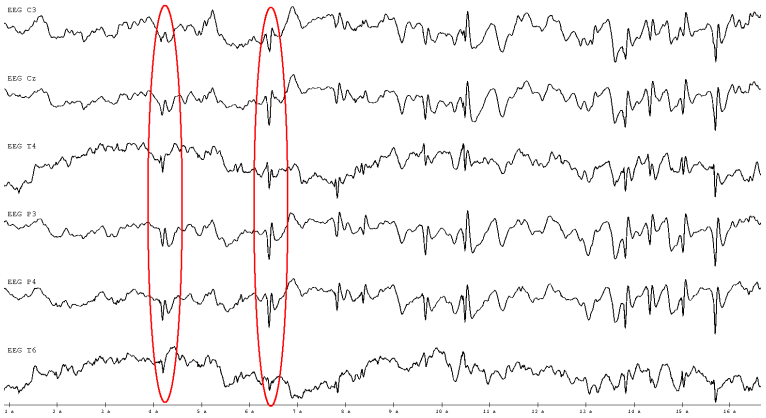


FIGURE 5.3: Example of 6 channels of one monopolar EEG raw, we can see different SWD in all channels.

3. Combination of the first two methods together into labeling the SWD activities in the EEG recordings. The threshold, overlapping window technique and band pass filter are commonly used for enhancing the performance of the detection algorithm.

5.3 Morlet Wavelet

The continuous wavelet transform is given by

$$W_f(t,a,b) = \int_{-\infty}^{\infty} \boldsymbol{X}_t \, \psi^*_{a,b}(\chi) d\chi \tag{5.1}$$

$$\psi^*_{a,b}(\chi) = \frac{1}{\sqrt{a}} \, \psi\left(\frac{\chi - b}{a}\right) \tag{5.2}$$

$$\psi(\chi) = \exp^{\frac{-\chi^2}{2}} \, cos(5\chi) \tag{5.3}$$

where a is the scaling parameter, b is the shifting parameter, $\psi^*_{a,b}(\chi)$ is the mother wavelet function, ψ^* denotes the complex conjugate operation and $\psi(\chi)$ is the analytic expression of the Morlet wavelet. In order to associate the Morlet wavelet as purely periodic signal of frequency F_c,

we use the relationship between scale and frequency

$$F_a = \frac{F_c}{a\Delta} \tag{5.4}$$

where Δ is the sampling period, F_c is the center frequency of Morlet wavelet in Hz and F_a is the pseudo-frequency corresponding to the scale a in Hz. The center frequency-based approximation captures the main wavelet oscillations. Therefore, the center frequency is a convenient and simple characterization of the dominant frequency of the wavelet [260]. Note that the wavelet scale is estimated according to the 1-3 Hz restricted narrow frequency of SWD database, introduced below.

5.4 Database

A database with 780 monopolar 256 Hz signals was created for the off-line training of the classifier: 340 spike-and-wave signals and 440 non-spikes-and-wave signals, measured from six patients from Fundación Lucha contra las Enfermedades Neurológicas Infantiles (FLENI). The spike-and-wave signals have different times and waveforms but their morphology is preserved, while the non-spike-and-wave signals have normal waveforms as Figures 5.2 and 5.3. See section 1.5 for more details.

By analyzing each SWD in the frequency domain, it was observed that they are restricted to a narrow frequency band 1 from 3 Hz. Each EEG was acquired with a 22-channel array using the standard 10/20 system through channels: Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2, Oz, FT10 and FT9. See Figure 3.1 for areas of the brain and Figure 5.4 for electrode positions used.

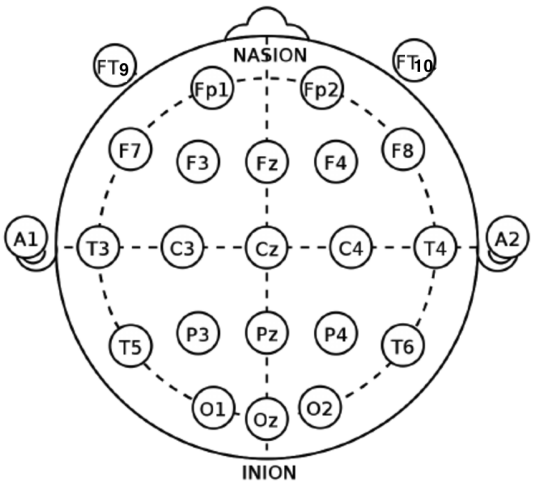


FIGURE 5.4: Electrodes position used with this database.

All new segments to analyze contain different spike-and-waves events. Their onset and duration time has been labeled by an expert neurologist. Here we used the expert annotations to extract a short epoch from each recording such that it is focused on the spike-and-wave in long-time EEG signals (the epochs used have a duration of the order of 1 minute).

It should be noted that, for each new patient to analyze, ten new SWD are selected to be part of the database. This permits to get patient-specific seizure detection.

Table 5.1 summarizes the different methods, features, classifiers, signals for training and testing considered in this chapter. The generalized Gaussian distribution uses the feature vector $[\sigma, \nu, \widetilde{x}]$, where σ is the scale parameter of the generalized Gaussian distribution, ν is the variance parameter from the generalized Gaussian distribution parameters, and $\widetilde{x}$ is the median parameter from the wavelet Morlet coefficients $\boldsymbol{C}^{(i)}$; the t-location-scale distribution uses the feature vector $[\mu, \sigma, \nu]$, which corresponds to the parameters of this distribution, namely location (μ), scale (σ) and shape (ν); and the cross-correlation measure uses the similarity feature vector r. Note that, we use different number of signals from training and test from the database because all studies correspond to different stages of the research.

Method	Features	Classifier	Signals for training	Signals for testing
Generalized Gaussian distribution from $\boldsymbol{C}^{(i)}$	σ, ν, $\widetilde{x}$	k-NN	340 spike-and-waves and 440 non-spike-and-waves	69
t-location-scale distribution from $\boldsymbol{X}^{(i)}$	μ, σ, ν	k-NN	96 spike-and-waves and 96 non-spike-and-waves	46
Cross-correlation from $\boldsymbol{X}^{(i)}$	r	Decision-tress	96 spike-and-waves	46

TABLE 5.1: Methods, features, classifiers, signals for training and testing considered in this chapter. The generalized Gaussian distribution uses the feature vector $[\sigma, \nu, \widetilde{x}]$, where σ is the scale parameter of the generalized Gaussian distribution, ν is the variance parameter from the generalized Gaussian distribution parameters, and $\widetilde{x}$ is the median parameter from the wavelet Morlet coefficients $\boldsymbol{C}^{(i)}$; the t-location-scale distribution uses the feature vector $[\mu, \sigma, \nu]$, which corresponds to the parameters of this distribution, namely location (μ), scale (σ) and shape (ν); and the cross-correlation measure uses the similarity feature vector r.

5.5 Spike-and-wave detection using the generalized Gaussian distribution

5.5.1 Methodology

In this study, we use the same methodology as in Chapter 4, Section 4.4. We start from the parameter-vector $\boldsymbol{\theta}_{\boldsymbol{C}^{(i)}}$ of the wavelet coefficients $\boldsymbol{C}^{(i)}$ such that

$$\boldsymbol{\theta}_{\boldsymbol{C}^{(i)}} = \left[\sigma^{(i)}, \tau^{(i)}\right]^T = \underset{[\sigma,\tau]^T}{\operatorname{argmax}}\, f_{GGD}(\boldsymbol{C}^{(i)}; \sigma, \tau). \tag{5.5}$$

Note that, the eq. (2.25) was estimated by using the discrete wavelet transform (DWT). In this Section, we use the continuous wavelet transform (CWT) by using the Morlet mother wavelet because it can describe mathematically the SWD morphology[247, 248, 249, 250], see the Figure 5.2.

5.5.2 Spike-and-wave detection by k-nearest neighbors classifier

Consider a classification into two possible classes $\omega_{ns} = 0$ and $\omega_s = 1$, then for a feature vector $\boldsymbol{\theta}_{\boldsymbol{C}^{(i)}}$ each class is given by

$$\begin{aligned} p\left(\boldsymbol{\theta}_{\boldsymbol{C}^{(i)}}|\omega_{ns}=0\right) &= \frac{1}{\omega_0}\sum_{\omega\in\omega_{ns}}\mathcal{N}\left(\boldsymbol{\theta}_{\boldsymbol{C}^{(i)}}|\boldsymbol{\theta}^{\omega}_{\boldsymbol{C}^{(i)}},\nu\boldsymbol{I}\right) \\ &= \frac{1}{\omega_0\left(2\pi\nu\right)^{D/2}}\sum_{\omega\in\omega_{ns}}\exp^{-\frac{\left(\theta_{\boldsymbol{C}^{(i)}}-\theta^{n}_{\boldsymbol{C}^{(i)}}\right)^2}{2\nu}} \qquad (5.6) \\ p\left(\boldsymbol{\theta}_{\boldsymbol{C}^{(i)}}|\omega_{s}=1\right) &= \frac{1}{\omega_1}\sum_{\omega\in\omega_{s}}\mathcal{N}\left(\boldsymbol{\theta}_{\boldsymbol{C}^{(i)}}|\boldsymbol{\theta}^{\omega}_{\boldsymbol{C}^{(i)}},\nu\boldsymbol{I}\right) \\ &= \frac{1}{\omega_1\left(2\pi\nu\right)^{D/2}}\sum_{\omega\in\omega_{s}}\exp^{-\frac{\left(\theta_{\boldsymbol{C}^{(i)}}-\theta^{\omega}_{\boldsymbol{C}^{(i)}}\right)^2}{2\nu}} \qquad (5.7) \end{aligned}$$

where D is the dimension of a data-point $\boldsymbol{\theta}_{\boldsymbol{C}^{(i)}}$ and ω_0 or ω_1 are the numbers of training points of class 0 or class 1 respectively, and ν is the variance of $\boldsymbol{\theta}_{\boldsymbol{C}^{(i)}}$.
Using the Bayes rule to classify a new datapoint $\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}$ in class $\omega_{ns} = 0$ the following equation is obtained

$$p\left(\omega_{ws}=0|\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}\right) = \frac{p\left(\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}|\omega_{ns}=0\right)p\left(\omega_{ns}=0\right)}{p\left(\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}|\omega_{ns}=0\right)p\left(\omega_{ns}=0\right)+p\left(\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}|\omega_{s}=1\right)p\left(\omega_{s}=1\right)}. \qquad (5.8)$$

The marginal likelihood $p(\omega_{ns} = 0)$ is $\omega_0/(\omega_0+\omega_1)$, and $p(\omega_s = 1) = \omega_1/(\omega_0+\omega_1)$. An analogous expression to eq. (A.48) can be obtained for $p\left(\omega_s = 1|\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}\right)$. To determine which class is most likely, the ratio between their two expressions is calculated as follows::

$$\frac{p\left(\omega_{ns}=0|\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}\right)}{p\left(\omega_{s}=1|\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}\right)} = \frac{p\left(\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}|\omega_{ns}=0\right)p\left(\omega_{ns}=0\right)}{p\left(\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}|\omega_{s}=1\right)p\left(\omega_{s}=1\right)}. \qquad (5.9)$$

If this ratio is greater than one, $\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}$ is classified as $\omega_{ns} = 0$, otherwise, it is classified as $\omega_s = 1$. It is important to note that in the case where ν is very small in (5.9), then both the numerator as the denominator will be dominated by the term for which the datapoint $\boldsymbol{\theta}^{\omega_0}_{\boldsymbol{C}^{(i)}}$ in class 0 or $\boldsymbol{\theta}^{\omega_1}_{\boldsymbol{C}^{(i)}}$ in class 1 are closest to the point $\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}$, such that

$$\frac{p\left(\omega_{ns}=0|\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}\right)}{p\left(\omega_{s}=1|\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}\right)}=\frac{\exp^{-\frac{\left(\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}-\boldsymbol{\theta}^{\omega_0}_{\boldsymbol{C}^{(i)}}\right)^{2}}{2\nu}}p\left(\omega_{ns}=0\right)/\omega_0}{\exp^{-\frac{\left(\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}-\boldsymbol{\theta}^{\omega_1}_{\boldsymbol{C}^{(i)}}\right)^{2}}{2\nu}}p\left(\omega_{s}=1\right)/\omega_1}=\frac{\exp^{-\frac{\left(\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}-\boldsymbol{\theta}^{\omega_0}_{\boldsymbol{C}^{(i)}}\right)^{2}}{2\nu}}}{\exp^{-\frac{\left(\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}-\boldsymbol{\theta}^{\omega_1}_{\boldsymbol{C}^{(i)}}\right)^{2}}{2\nu}}}. \tag{5.10}$$

On the limit $\nu \to 0$, $\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}$ is classified as class 0 if $\boldsymbol{\theta}^{*}_{\boldsymbol{C}^{(i)}}$ has a point in the class 0 data which is closer than the closest point in the class 1 data. The nearest (single) neighbor method is therefore recovered as the limiting case of a probabilistic generative model [185, 130].

5.5.3 Experimental Results

In the training stage, the annotated database previously exposed in Section 5.4 was utilized. These 780 monopolar signals (340 spike-and-wave signals and 440 non-spikes-and-wave signals), were trained off-line using k-nearest neighbors on a modified vector $[\sigma, \nu, \widetilde{x}] \in \mathbb{R}^3$ collecting the parameters associated with Morlet wavelet coefficients for each 2-second segment, where σ is the scale parameter of the generalized Gaussian distribution, $\nu = \sigma^2\Gamma(3/\tau)/\Gamma(1/\tau)$ is the variance parameter, and $\widetilde{x}$ is the median parameter from the wavelet Morlet coefficients from the feature vector $\boldsymbol{\theta}_{\boldsymbol{C}^{(i)}}$. Remember that scale parameter σ depends on τ and is closely related to the variability of the brain activity, being therefore, a good descriptor for performing seizure detection, see Section 2.8.2 for more details. The variance and the median do not introduce any additional computational cost.

Table 5.2 contains the different bounds for each parameter. Note that both the minimum and the maximum are large for $[\sigma, \nu, \widetilde{x}]$ when SWD and non-SWD signals are compared. This observation suggests that a threshold could be implemented to detect SWD patterns as clear discrimination exists between spike-and-wave and non-spike-and-wave. To illustrate this, Figure 5.5 shows scatter plots for the tree parameters as follows:

1. **Scale parameter (σ) vs variance (ν)**: For class 1 or SWD, we can see the direct relationship between the variance and sigma, both grow proportionally. While for class 0 or non-SWD, both sigma and variance remain in a range of values.

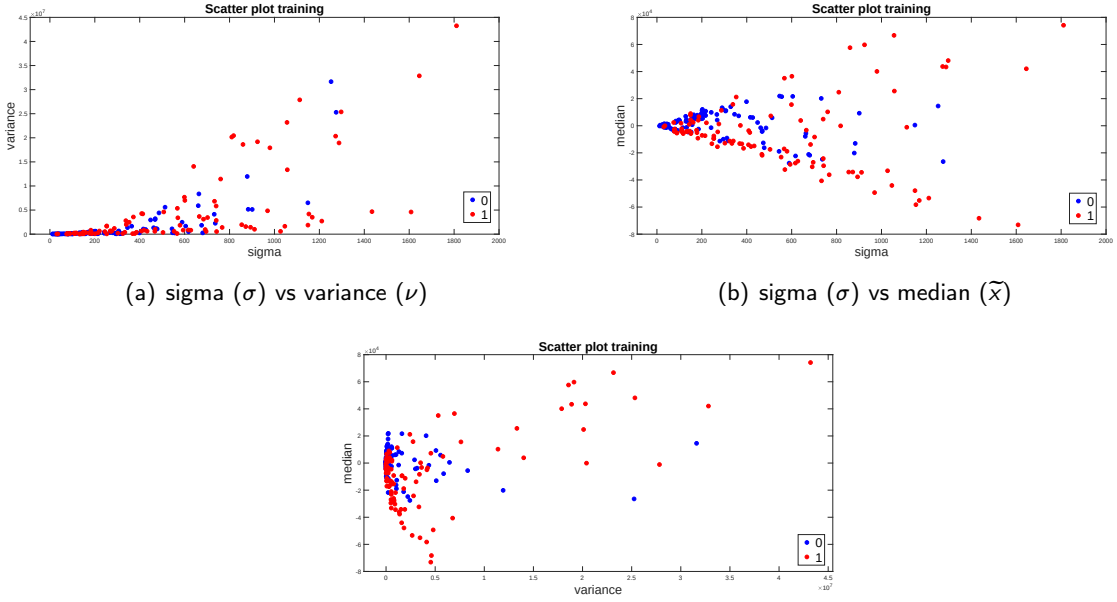
2. **Scale parameter (σ) vs median ($\widetilde{x}$)**: As σ grows, median increases and decreases for both SWD and non-SWD, but is larger for SWD. A cone-shaped pattern can be identified.

3. **Variance (ν) vs median ($\widetilde{x}$)**: As the variance grows, the median increases and decreases for SWD, while for non-SWD, it remains in a small range (cluster).

The performance of the k-nearest neighbors classification method using 10 neighbors with 3 predictors $[\sigma, \nu, \widetilde{x}]$ was evaluated using a dataset consisting of 69 new annotated measurement.

Metric	Sigma (σ)	Variance (ν)	Median ($\widetilde{x}$)
Class 0	[12.19,1275.38]	[946.60,31620526.09]	[-27698.91,21799.08]
Class 1	[31.22,1811.37]	[2715.81,43218940.81]	[-73254.70,74064.46]

TABLE 5.2: Range of values for sigma (σ), variance (ν) and median ($\widetilde{x}$) parameters for class 0 or non-spike-and-wave and for class 1 or spike-and-wave.



(a) sigma (σ) vs variance (ν)

(b) sigma (σ) vs median ($\widetilde{x}$)

(c) variance (ν) vs median ($\widetilde{x}$)

FIGURE 5.5: Scatter plots of the off-line training classification in database signals, for σ, ν and $\widetilde{x}$ parameters for spike-and-waves events (SWD = class 1 = red dots) and non-spike-and-waves events (non-SWD = class 0 = blue dots), showing the data dispersion of the proposed approach. In (a) Scale parameter (σ) vs variance (ν). For class 1 or SWD, we can see the direct relationship between the variance and sigma, both grow proportionally, while for class 0 or non-SWD both sigma and variance remain in a range of values. (b) Scale parameter (σ) vs median ($\widetilde{x}$). As sigma grows, median increases and decreases for both SWD and non-SWD, but is larger for SWD. (c) variance (ν) vs median ($\widetilde{x}$). As variance grows, median increases and decreases for SWD, while for non-SWD, remains in a small range.

The annotated dataset corresponds to 69 segments extracted from six EEG signals of different subjects from Fundación Lucha contra las Enfermedades Neurológicas Infantiles (FLENI). The assessment of the results was performed in terms of the overall accuracy of the classification. The classifier achieved a 100% sensitivity (True Positive Rate) and specificity (True Negative Rate) for SWD detection.

5.6 Spike-and-wave detection using t-location-scale distribution

The t-location-scale distribution or non-standardized Student's t-distribution is a statistical model for univariate and multivariate signals that has three parameters: location, shape and a non-

negative scale. It is useful for modeling data distributions with heavy tails which are more prone to outliers than the normal distribution. The t-location-scale distribution has been applied to different signal processing problems in diverse areas such as radar, watermark, speech and wireless; in medicine and health, it is widely used in genetics and has been recently used in sleep patterns [261]. In Section 5.5, we showed that the generalized Gaussian distribution can be used for SWD pattern recognition. Knowing that the t-location-scale distribution is heavy-tailed and more prone to outliers than the normal distribution, the following question arose: what if this distribution could be used to detect a SWD pattern recognition in epileptic EEG signals?. Question that we will answer in this preliminary study.

The t-location-scale distribution is a statistical model that belongs to location-scale family formed by translation and rescaling of the Student's t-distribution. The probability density function (PDF) of a location-scale distribution, is given by

$$g(x|\mu,\sigma) = \frac{1}{\sigma}\psi\left(\frac{x-\mu}{\sigma}\right). \tag{5.11}$$

The probability density function (PDF) of the Student's t-distribution, is given by

$$\psi(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{\nu+x^2}{\nu}\right]^{-\left(\frac{\nu+1}{2}\right)}. \tag{5.12}$$

Therefore applying (5.12) to (5.11), we have the probability density function (PDF) of the t-location-scale, which is given by

$$f_{\text{tls}}(x|\mu,\sigma,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{\nu+\left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right]^{-\left(\frac{\nu+1}{2}\right)} \tag{5.13}$$

where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter, $\nu > 0$ is the shape parameter, and $\Gamma(.)$ is the Gamma function. The optimization problem of estimating the parameters of this distribution can then be solved using the simplex search method of Lagarias et al. [109].

5.6.1 Methodology

Let $\boldsymbol{X} \in \mathbb{R}^{M\times N}$ denote the matrix gathering M EEG signals $\boldsymbol{x}_m \in \mathbb{R}^{1\times N}$ measured simultaneously on different channels and at N discrete time instants. The proposed methodology is composed of three stages. The first stage splits the original signal $\boldsymbol{X}$ into a set of non-overlapping 1 second segments using a rectangular sliding window $\boldsymbol{X}^{(i)} = \boldsymbol{\Omega}^{(i)}\boldsymbol{X}$. In the second stage the parameters of the t-location-scale distribution for each $\boldsymbol{X}^{(i)}$ are estimated and finally, in the third stage, the feature vector associated with each time segment is classified by using k-nearest neighbors classifier as spike-and-wave/non-spike-and-wave through the feature vector $\boldsymbol{\theta}_{\boldsymbol{C}^{(i)}} = [\mu,\sigma,\nu]^T$ of the t-location-scale parameters: location (μ), scale (σ) and shape (ν), see Section 5.5.2 for more details.

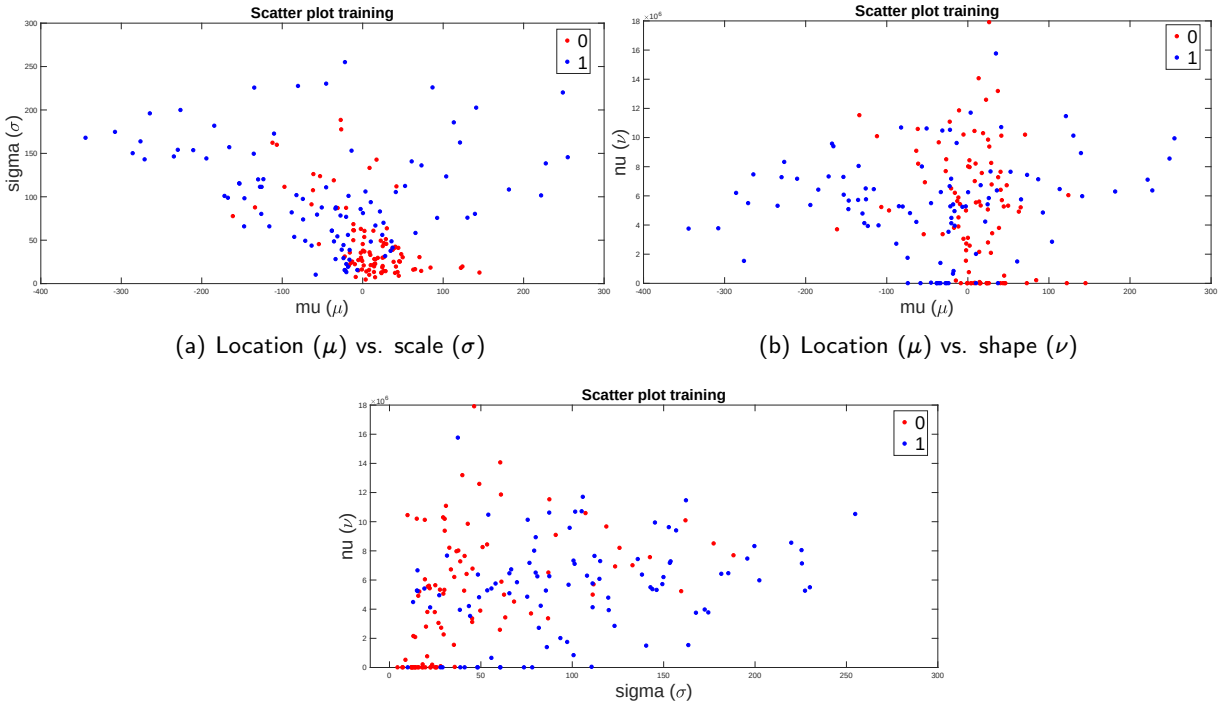
5.6.2 Experimental Results

To asses the performance of the proposed method we used 192 monopolar 256Hz signals for off-line training classifier, 96 spike-and-waves and 96 non-spikes-and-waves from the database described in Section 5.4, and 46 new labeled test signals used for on-line classification. Their onset and duration time have been labeled by an expert neurologist from FLENI. Here we used the expert annotations to extract a short epoch from each recording such that it is focused on the spike-and-wave in long-time signals (the epochs have a duration of the order of 1 minute). Figure 5.6 and Figure 5.8 show scatter plots of the kNN off-line training classifier and on-line classification respectively, using the three t-location-scale distribution parameters: location (μ), scale (σ) and shape (ν).

Data dispersion of spike-and-wave events (label 1: blue dots) and non-spike-and-wave events (label 0: red dots) in Figure 5.6, during the training stage, suggests that in a) and c) spike-and-wave events tend to have a higher scale σ with respect to non-spike-and-wave events, in b) non-spikes-and-wave events tend to have a location μ concentrated between a certain threshold with respect to spike-and-wave events; while the data dispersion in Figure 5.8, during the classification stage, suggests that in a) spike-and-wave events tend towards the center down with respect to non-spike-and-wave events, in b) the trend is not very clear, although there is a great concentration of spike-and-wave events in the center down near zero with respect to non-spike-and-wave events and in c) spike-and-wave events tend to be located towards the right and near zero with respect to non-spike-and-wave events.

From the illustration in Figure 5.7 and Figure 5.9 we can see the different histograms and the perfect group discrimination between spike-and-wave events (label 1) and non-spike-and-wave events (label 0) in off-line training classification for Figure 5.7 and on-line classification Figure 5.9.

The performance of the online k-nearest neighbors classification method using an equal weight distance with the number of neighbors equal to one and the distance metric Euclidean, see Figure 5.8 and Figure 5.9, was assessed in terms of overall accuracy classification, and achieves a 100% of sensitivity (True Positive Rate) and specificity (True Negative Rate) for spike-and-wave detection. These results invite us to study this distribution in deeper (e.g. goodness-of-fit test, parameters estimation, and model-based characterization, similar to Sections 2.6 and 2.8), in order to be used in the future in seizure onset detection and epileptiform patterns recognition in epileptic signals.



(a) Location (μ) vs. scale (σ)

(b) Location (μ) vs. shape (ν)

(c) Scale (σ) vs. shape (ν)

FIGURE 5.6: Scatter plots of the off-line training classification in 192 dataset signals, for the t-location-scale parameters μ, σ and ν for spike-and-waves events (blue dots) and non-spike-and-waves events (red dots), showing the data dispersion of the proposed approach. In a) and c) spike-and-waves tend to have a higher scale σ, in b) non-spikes-and-waves tend to have a location μ concentrated between 0 and 100.

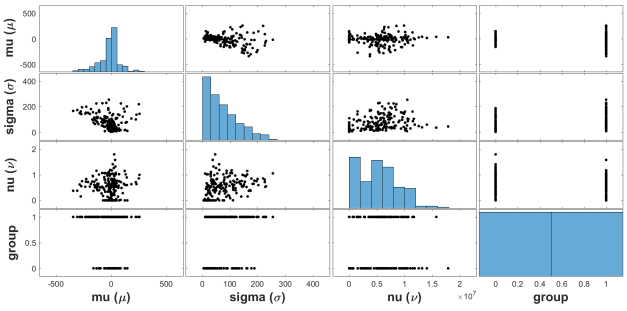
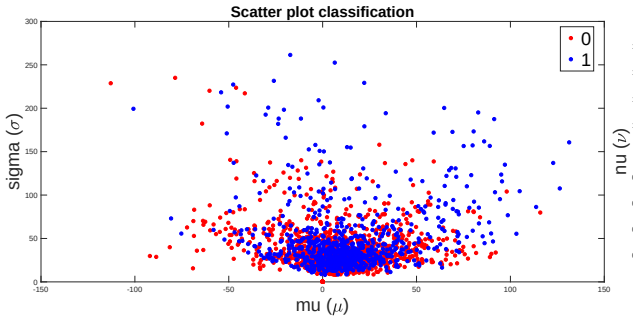
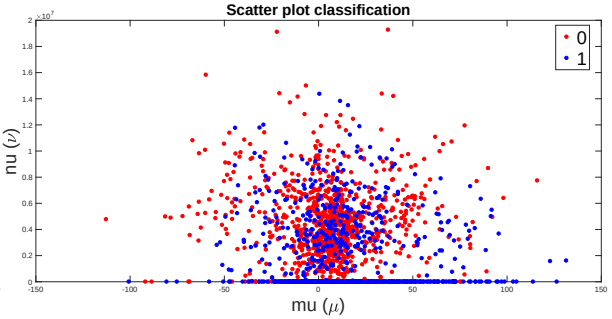


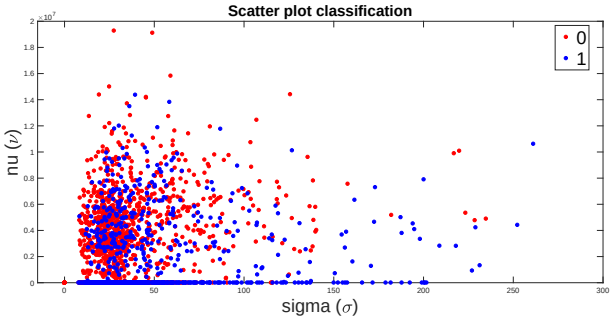
FIGURE 5.7: Scatter plot in off-line training classification in 192 dataset signals for the t-location-scale parameters μ, σ and ν, we can see the correct discrimination between two groups whose size is the same (96 spike-and-waves and 96 non-spikes-and-waves), label 1 for spike-and-wave and label 0 for non-spike-and-wave.



(a) Location (μ) vs. scale (σ)



(b) Location (μ) vs. shape (ν)



(c) Scale (σ) vs. shape (ν)

FIGURE 5.8: Scatter plots in on-line classification in 46 test signals, for the t-location-scale parameters μ, σ and ν for spike-and-waves events (blue dots) and non-spike-and-waves events (red dots), showing the data dispersion of the proposed approach. In a) spike-and-waves tend towards the center down, in b) the trend is not very clear, although there is a great concentration of spike-and-waves in the center down near zero, and in c) spike-and-waves tend to be located towards the right and near zero.

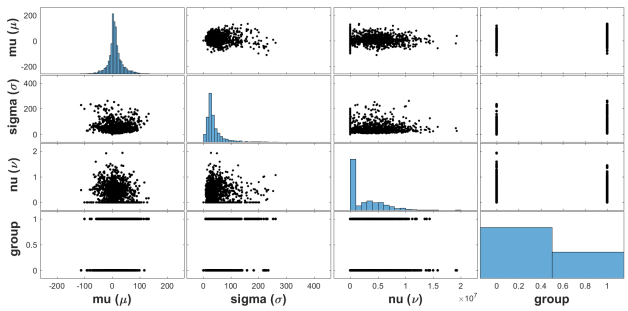


FIGURE 5.9: Scatter plot in on-line classification in 46 test signals for the t-location-scale parameters μ, σ and ν, we can see the correct discrimination between two groups whose size is different (spike-and-waves labeled by an expert neurologist and non-spikes-and-waves), label 1 for spike-and-wave and label 0 for non-spike-and-wave.

5.7 Spike-and-wave detection using cross-correlation and decision-trees

A decision tree is a hierarchical model for supervised learning whereby the local region is identified in a sequence of recursive splits in a smaller number of steps [262, 185, 187]. A decision tree is composed of internal decision nodes and terminal leaves, see Figure 5.10. It is defined in a way that there is a single node, called the root, which has no parents, and all other nodes only have one parent. When a node receives an input a specific test, designed for that particular node, is applied and one of the branches is taken depending on the outcome. This process starts at the root and is repeated recursively until a leaf node is hit, at which point the leaf's value constitutes the output. Each specific test is a simple function which defines a discriminant in the input space dividing it into smaller regions that are further subdivided as we take a path from the root down. In this manner, a complex function is broken into a series of simple decisions by simply writing the tests down as a tree.

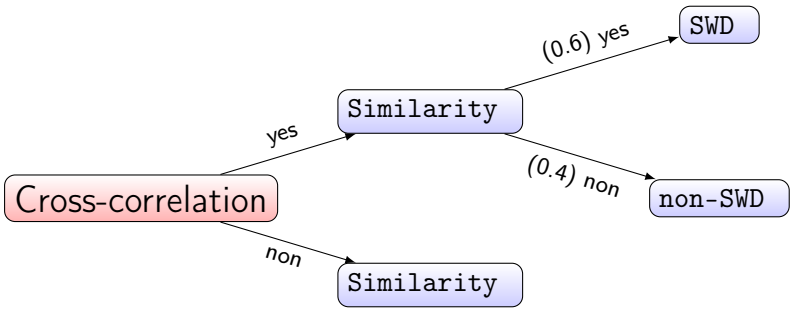


FIGURE 5.10: A decision tree example. Consider the decision problem as to whether or not to go ahead with a cross-correlation similarity. If we go ahead with the similarity and meets the threshold (0.6), then we have a spike-and-wave candidate; on the other hand, if we don't go ahead with the similarity (0.4) then the threshold is not met and therefore we don't have a spike-and-wave candidate. Note that, this tree has only two regions given by the similarity threshold for SWD or non-SWD.

We now introduce the detection trees in general form using the methodology from [185]. The goal is to predict a single target variable t from a D-dimensional vector $r = (r_1, ..., r_D)^T$ of input variables related to the cross-correlation in our study. The training data consists of input vectors $\{r_1, \cdots, r_N\}$ along with the corresponding continuous labels $\{t_1, ..., t_N\}$. If the partitioning of the input space is given, and we minimize the sum-of-squares error function, then the optimal value of the predictive variable within any given region is just given by the average of the values of t_n for those data points that fall in that region, two regions or classes in our case spike-and-wave or non-spike-and-wave, see Figure 5.11. To determine the structure of the decision tree, the first step is start with a single root node, corresponding to the whole input space, and then growing the tree by adding nodes one at a time. At each step there will be some number of candidate regions in input space that can be split, corresponding to the addition of a pair of leaf nodes to the existing tree. For each of these, there is a choice of which of the D input variables to split, as well as the value of the threshold. For a given choice of split variable and threshold, the optimal choice

of predictive variable is given by the local average of the data. This is repeated for all possible choices of the variable to be split, and the one that gives the smallest residual sum-of-squares error is retained. The stopping of the addition of nodes, is related to the number of data points associated with the leaf nodes, to then prune back the resulting tree. The pruning is based on a criterion that balances residual error against a measure of model complexity. For example, if we denote the starting tree for pruning by T_0, then we define $T \subset T_0$ to be a subtree of T_0 if it can be obtained by pruning nodes from T_0. Suppose the leaf nodes are indexed by $\tau = 1, \cdots, |T|$, with leaf node τ representing a region $\mathcal{R}_\tau$ of input space having N_τ datapoints, and $|T|$ denoting the total number of leaf nodes. The optimal prediction for region $\mathcal{R}_\tau$ is then given by

$$y_\tau = \frac{1}{N_\tau} \sum_{r_n \in \mathcal{R}_T} t_n \tag{5.14}$$

and the corresponding contribution to the residual sum-of-squares is given by

$$Q_\tau(T) = \sum_{r_n \in \mathcal{R}_T} \{t_n - y_\tau\}^2. \tag{5.15}$$

The pruning criterion is then given by

$$C(T) = \sum_{\tau=1}^{|T|} Q_\tau(T) + \lambda |T| \tag{5.16}$$

The regularization parameter λ determines the trade-off between the overall residual sum-of-squares error and the complexity of the model as measured by the number $|T|$ of leaf nodes, and its value is chosen by cross-validation. For classification problems, the process of growing and pruning the tree is similar, except that the sum-of-squares error is replaced by a more appropriate measure of performance of the Gini index for a binary classifier, defining $p_{\tau k}$ to be the proportion of data points in region $\mathcal{R}\tau$ assigned to class k, where $k = 1, \cdots, K$, in our case we have two classes spike-and-wave and non-spike-and-wave, see eq. (5.17).

$$Q_\tau(T) = \sum_{k=1}^{K} p_{\tau k}(1 - p_{\tau k}). \tag{5.17}$$

5.7.1 Methodology

Let $\widehat{\boldsymbol{X}} \in \mathbb{R}^{N \times M}$ be an EEG raw signal, measured simultaneously on N different channels with 256 Hz of sample rate and $\widehat{\boldsymbol{SW}} \in \mathbb{R}^{1 \times P}$ a spike-and-wave pattern database gathered from different EEG signals $\widehat{\boldsymbol{X}}$, given by

$$\widehat{\boldsymbol{X}} = [x_1, x_2, ..., x_m, ..., x_N]^T \quad with \quad 1 \leq m \leq N \tag{5.18}$$

$$\widehat{\boldsymbol{SW}} = [sw_1, sw_2, ..., sw_p, ..., sw_P] \quad with \quad 1 \leq p \leq P \tag{5.19}$$

where $N = 23$ channels and $P = 96$ spike-and-waves from database described in 5.4. The proposed methodology is composed of four stages.

The first stage is the filtering of $\widehat{\boldsymbol{X}}$ and $\widehat{\boldsymbol{SW}}$ using two cascade Butterworth IIR filters in $\mathbb{Z}$ domain with an empirical design based on physicians experience, a 2-order lowpass filter with cutoff frequency of 100 Hz and 1-order highpass filter with cutoff frequency of 30 Hz, see eq. 5.20-5.21 respectively

$$\boldsymbol{W_{lp}}(z) = \frac{b}{(1 - az^{-1})^2} \tag{5.20}$$

$$\boldsymbol{W_{hp}}(z) = \frac{b(1 - z^{-1})}{(1 - az^{-1})}. \tag{5.21}$$

Let $\boldsymbol{X}$ and $\boldsymbol{SW}$ be the filtered original signals. Then in the second stage, the filtered signal $\boldsymbol{X}$ is splitted into a set of non-overlapping 1 second segments using a rectangular sliding window so that $\boldsymbol{X}^{(i)} = \boldsymbol{\Omega}^{(i)}\boldsymbol{X}$. In the third stage, a cross-correlation is used to find the best match between the two signals $\boldsymbol{X}^{(i)}$ and $\boldsymbol{SW}_p$. Cross-correlation measures the similarity between $\boldsymbol{SW}_p$ and shifted (lagged) copies of $\boldsymbol{X}^{(i)}$ as a function of the lag. Note that $\boldsymbol{X}^{(i)}$ is an EEG $\mathbb{R}^{N\times M}$ matrix and $\boldsymbol{SW}_p$ is a $\mathbb{R}^{1\times P}$ vector which contains all the spike-and-wave to be analyzed. Assuming that $i = p = n$ then a cross-correlation $r_{\boldsymbol{X},\boldsymbol{SW}}$ for the displacement in time of each EEG channel with respect to each spike-and-wave is given by

$$r_{x,sw}[\tau] = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{X}^{[n-\tau]}\boldsymbol{SW}[n]. \tag{5.22}$$

Then waveforms similarity are classified by the local peaks of the absolute value of $r_{\boldsymbol{X},\boldsymbol{SW}}$. Of which, only the peaks greater than a certain threshold given by eq. (5.23), are considered similar enough. Besides, a minimum distance of 1 second is established between peaks, which means that for this algorithm there could not be more than one SWD per second:

$$\max_{[\boldsymbol{X},\boldsymbol{SW}]^T} \left|r_{\boldsymbol{X},\boldsymbol{SW}[\tau]}\right| - \sigma\left(r_{\boldsymbol{X},\boldsymbol{SW}}[\tau]\right) \tag{5.23}$$

where $|.|$ is the absolute value and σ is the standard deviation. Finally, in the fourth stage, only the SWDs greater than 40% of the threshold of the total of coincidences by using the eq. (5.23), are chosen as spikes-and-waves by each channel. Last two stages are using as a D-dimensional input vector r for decision trees classifier in two regions namely spike-and-wave and non-spike-and-wave, see Section 5.7.

The proposed methodology can be summarized by using the next algorithm:

Data: EEG raw
Result: SWD detection
for *each SWD* **do**
  for *each* $\boldsymbol{X}^{(i)}$ *for each channel* **do**
    1. Cross-correlation estimation between each SWD and $\boldsymbol{X}^{(i)}$, see eq. (5.22);
    2. SWDs candidates selection: based-on the waveform similarity and the distance between peaks of 1 second greater than a threshold given by eq. (5.23);
    3. SWDs: Only the SWDs greater than 40% of the total of coincidences are chosen, see eq. (5.23) and Figure 5.11;
    4. Steps 2. and 3. are using as a D-dimensional input vector r for decision trees classifier in two regions namely spike-and-wave and non-spike-and-wave, see Figure 5.10;
  end
end

Algorithm 4: SWD detection by using cross-correlation and decision trees.

5.7.2 Experimental results

We evaluate the performance of the proposed seizure detector in with 46 segments between 40 and 60 seconds from 23 channels, see Figure 5.11, which correspond to sleep long-term epileptic signals recordings of one patient from Fundación contra las Enfermedades Neurológicas Infantiles (FLENI).

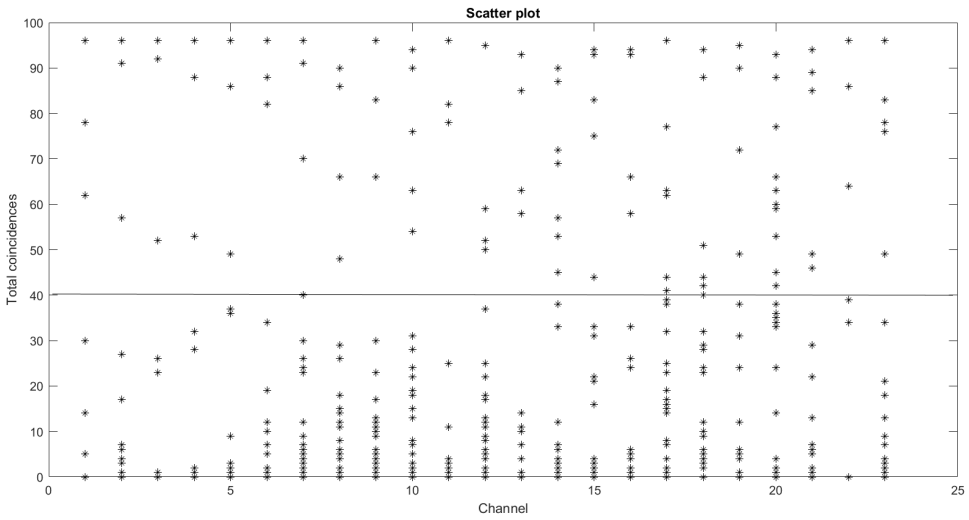


FIGURE 5.11: Scatter plot example between all 23 channels (x axis) and the total coincidences (y axis) into spike-and-wave signals from database, the line in 40 is the threshold used.

We compare the medical annotated data with our cross-correlation classifier and using 10 and 20 empirical K-fold cross-validation through the decision tree to evaluate how our results can be generalized to an independent data set. We found an Area Under Curve (AUC) of 97% in 874 predictors corresponding to all database candidates from two EEG epochs with 23 channels, with 86% sensitivity and 98% specificity for spike-and-waves detection in long-term epileptic signals, see classifier performance in ROC Figure 5.12.

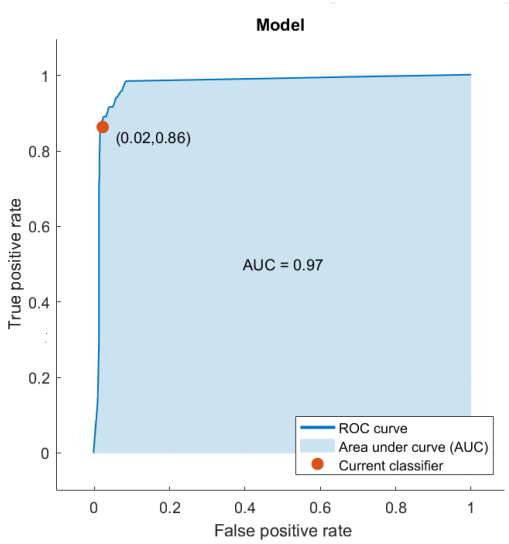


FIGURE 5.12: Receiver operating characteristic curve (ROC) in 874 predictors

5.8 Conclusions

In this chapter, we studied three different methods to detect spike-and-wave discharges in EEG signals using a database created in Fundación Lucha contra las Enfermedades Neurológicas Infantiles (FLENI). For each new patient, ten new SWD patterns were selected to be part of the database before training. Once the entire new database is trained, the prediction transforms into a patient-specific seizure detection.

The main method used the generalized Gaussian distribution (GGD) coupled with the k-NN classifier. In the second method, we used the t-location-scale distribution with a similar methodology as with GGD. Both methods obtained an accuracy of 100%. Finally, in method three, we used a cross-correlation coupled with decision trees classifier getting 98% sensitivity (True positive rate) and 86% specificity (True negative rate) for spike-and-wave detection.

This research experience suggests that the proposed methods are potentially useful for spike-and-wave detection in EEG long-term signals in epilepsy.

Chapter 6
Conclusions, contributions, advantages, limitations and research perspectives

6.1 Conclusions and contributions

This thesis originated with the Dynamic Brain SticAmSud project (2012-2014): "Dynamic image reconstruction and segmentation for brain tissue characterization". The interdisciplinary research was conducted by the following partners: Hadj Batatia and Jean Yves Tourneret from IRIT-ENSEEIHT Laboratory of the University of Toulouse, France; José Bermudez and Marcio Costa from LPDS/EEL Laboratory of the Universidade Federal de Santa Catarina, Brazil; Marcelo Pereyra from School of Mathematics of the University of Bristol, UK; Carlos D'Giano from FLENI, Argentina; and Marcelo Risk and Antonio Quintero-Rincón from ITBA, Argentina. All institutions with funding source by Centre national de la recherche scientifique (CNRS) from France, Coordenação de aperfeiçoamento de pessoal de nivel superior (CAPES) from Brazil and Ministerio de Ciencia, Tecnología e Innovación Productiva (MinCyT) from Argentina.

In this thesis, the main methodological and theoretical aspects of EEG data processing have been covered, from an accurate solution of detection, quantification, and characterization of epilepsy seizures to effective approaches to correct classification between seizures and non-seizures.

Our work with real data led us to the analysis and comparison of state-of-the-art with similar results, domain to which we contributed by introducing a statistical model that acts as a strong dimension reduction mechanism yielding a significantly lower computational complexity. This makes feasible a fast online implementation of an onset detection algorithm using a linear or Bayesian classifier.

This work was motivated by the real need to detect epileptic seizures in clinical practices for two reasons: manually marking the pattern is time consuming and the visual detection may be difficult and error-prone. This topic was investigated from the exploration of the data to the construction of principled methodology that allowed us to obtain promising results. It was possible to achieve two methods: a new onset detection method in epileptic signals with a low computational complexity that can be trained using a reasonably small dataset, but remains an open issue; and the other one, a spike-and-wave epileptiform pattern recognition method.

Our interest in the research area motivated us to address some hard and still open questions in the field. Going beyond detection, we proposed a temporal spread estimation algorithm working

on the *scale* parameter of the generalized Gaussian distribution that offers interesting perspectives for the investigation of the variability of the brain activity. We applied this method to epileptic signals processing which demonstrated that such an approach could provide a good descriptor for performing seizure onset detection.

Another topic addressed during this thesis relates to the analytical development of Kullback-Leibler divergence (KLD) using the generalized Gaussian distribution parameters (scale and shape), to distinguish between seizure and non-seizure in epileptic signals.

Pattern recognition was addressed with three novel methods to detect spike-and-wave discharges (SWD) in long-term EEG signals. Recently, this work was selected for a national innovation competition in Argentina named "INNOVAR 2018".

The contributions are basically methodological and applied. Throughout this thesis, we tried to make the right mathematical choices to model the problems of interest. We believe this enabled us to propose appropriate and efficient algorithms so that we could finally tackle challenging epilepsy problems.

To summarize:

- We contributed by providing to the EEG community an excellent seizure onset detector in epileptic signals.

- We presented mathematical details of the statistical model based on the generalized Gaussian distribution that acts as a strong dimension reduction mechanism yielding a significantly lower computational complexity. This makes feasible an online implementation of an onset detection algorithm.

- We developed a Kullback-Leibler-divergence-based methodology to distinguish between seizure and non-seizure in epileptic signals.

- We contributed to set up a full experimental study from protocol design and data exploration to the construction of a data analysis pipeline that offers promising results for the study of EEG signals.

- We proposed a novel algorithm to address the hard problem of temporal spread estimation. We believe that this contribution can be a valuable tool to investigate inter-trial variabilities, which is of major interest in epilepsy studies.

- We proposed three novel methods to detect spike-and-wave discharges (SWD) following the methodology used in this thesis. SWD is an EEG generalized discharge pattern seen particularly during absence epilepsy, whose clinical importance lies in cognitive and behavioral disturbances.

- We generated challenging research activity for bioengineering students at ITBA. This allowed them to be trained in research and generating different projects on the exciting world of EEG signal processing.

- We generated collaborations between different international universities from France, U.K., Brazil, Germany and Tunisia.

These contributions are reflected in the following publications:

International Journal Papers

1. Antonio Quintero-Rincón, Marcelo Pereyra, Carlos D'Giano, Marcelo Risk and Hadj Batatia. *Fast statistical model-based classification of epileptic EEG signals*. Biocybernetics and Biomedical Engineering, Vol. 38, No. 4, pages 877-889, 2018 [4].

2. Antonio Quintero-Rincón, Carlos D'Giano and Marcelo Risk. *Epileptic seizure prediction using Pearson's product-moment correlation coefficient of a linear classifier from generalized Gaussian modeling*. Neurología Argentina, Vol. 10, Issue. 4, pages 201-217, 2018 [5].

3. Antonio Quintero-Rincón, Marcelo Risk, Carlos D'Giano, Valeria Muro, Jorge Prendes, Hadj Batatia. *A novel spike-and-wave automatic detection in EEG signals*. International Journal of Signal and Imaging Systems Engineering. Vol. x, No. x, pages X, 2019 (In press) [10].

4. Antonio Quintero-Rincón, Catalina Carenzo, Joaquin Ems, Lourdes Hirschson, Valeria Muro, Carlos D'Giano. *Spike-and-wave epileptiform discharge pattern detection based on Kendall's Tau-b coefficient*. Applied Medical Informatics. Vol. x, No. x, pages X, 2019 (In press) [263].

5. Antonio Quintero-Rincón, Carlos D'Giano and Hadj Batatia. *Curve fitting based on two-point central difference to detect epileptic EEG seizures.* Journal of Biomedical Research. Vol. x, No. x, pages X, 2019 (In press) [264].

International Proceedings Papers

1. Antonio Quintero-Rincón, Carlos D'Giano, Hadj Batatia. *Seizure onset detection in EEG signals based on entropy from generalized Gaussian PDF modeling and ensemble bagging classifier*. European Association for Predictive, Preventive and Personalised Medicine (EPMA), Springer book series, Vol. x, No. x, pages X, 2019 (In press), DOI: 10.1007/978-3-030-11800-6, https://www.springer.com/gp/book/9783030117993#aboutBook [7].

2. Antonio Quintero-Rincón, Marcelo Pereyra, Carlos D'Giano, Hadj Batatia and Marcelo Risk. *A visual EEG epilepsy detection method based on a wavelet statistical representation and the Kullback-Leibler divergence*. International Federation for Medical and Biological Engineering (IFMBE) Proceedings Springer book series, Vol. 60, pages 13-16, 2017. https://doi.org/10.1007/978-981-10-4086-3_4 [2]

3. Antonio Quintero-Rincón, Marcelo Pereyra, Carlos D'Giano, Hadj Batatia and Marcelo Risk. *A New algorithm for Seizure Onset Detection and Spread in Epilepsy Signals*. Journal of Physics: Conference Series, Vol. 705. No 1, page 012032, 2016. DOI:10.1088/1742-6596/705/1/012032 [3]

National Journal Papers

1. Antonio Quintero-Rincón, Máximo Flugelman, Jorge Prendes and Carlos D'Giano. *Study on epileptic seizure detection in EEG signals using largest Lyapunov exponents and logistic regression*. Revista Argentina de Bioingeniería, Bioengineering Argentinian Society. Vol. x, No. x, pages X, 2019. (In press) [8].

2. Ivanna Zorgno, Maria Cecilia Blanc, Simon Oxenford, Francisco Gil Garbagnoli, Carlos D'Giano and Antonio Quintero-Rincón. *Epilepsy seizure onset detection applying 1-NN classifier based on statistical parameters*. Argentina Biennial Congress ARGENCON 2018, San Miguel de Tucumán, 6-8 June 2018. DOI: 10.1109/ARGENCON.2018.8646234 [265].

3. Antonio Quintero-Rincón, Manuela Alanis, Valeria Muro and Carlos D'Giano. *Spike-and-Wave detection in epileptic signals using cross-correlation and decision trees*. Revista Argentina de Bioingeniería, Bioengineering Argentinian Society, 22(4):3-6, 2018 [11].

International Conference Papers

1. Bassem Bouaziz, Lotfi Chaari, Hadj Batatia, Antonio Quintero-Rincón. *Epileptic seizure detection using a Convolutional Neural Network.* International conference on digital health technologies (ICDHT). October 15-16, 2018 - Sfax, Tunisia, [266].

2. Antonio Quintero-Rincón, Jorge Prendes, Valeria Muro and Carlos D'Giano. *Study on Spike-and-wave detection in epileptic signals using t-location-scale distribution and the k-nearest neighbors classifier*. IEEE URUCON 2017 Congress on Electronics, Electrical Engineering and Computing. Montevideo, Uruguay, 23-25 October 2017. DOI: 10.1109/URUCON.2017.8171869 [9].

3. Antonio Quintero-Rincón, Jorge Prendes, Marcelo Pereyra, Hadj Batatia and Marcelo Risk. *Multivariate Bayesian Classification of Epilepsy EEG Signals*. The 2016 IEEE Image Video and Multidimensional Signal Processing (IVMSP) workshop (IVMSP). Bordeaux, France. 11-12 July 2016:1-5. DOI: 10.1109/IVMSPW.2016.7528180 [6].

4. Antonio Quintero-Rincón, Hadj Batatia, Marcelo Pereyra and Marcelo Risk. *Detection of Onset in Epilepsy Signals using Generalized Gaussian Distribution*. Fifth International Conference on Advances in New Technologies, Interactive Interfaces and Communicability. Huerta Grande, Córdoba, Argentina, 10-12 November 2014. (*Special Mention*) ISBN: 978.88.96.471.37.1, DOI: 10.978.8896471/371 Blue Herons Editions [1].

International Technical Reports

1. Antonio Quintero-Rincón and Marcelo Risk. *Estimation and Regularization of Inverse Problem in EEG*. STIC-Amsud Technical Report -12STIC-03 DynBrain, EEG dynamic image reconstruction and segmentation for brain tissue characterization, Toulouse, France, 21 January 2014.

2. Antonio Quintero-Rincón and Marcelo Risk. *Head Models: Review of the State of the Art*. STIC-Amsud Technical Report -12STIC-03 DynBrain, EEG dynamic image reconstruction and segmentation for brain tissue characterization, Toulouse, France, October 1st 2012.

National Congress

1. Antonio Quintero-Rincón. *Detección de Crisis en Señales Epilépticas usando la Distribución Gaussiana Generalizada*. Terceras Jornadas Interdisciplinarias de Análisis Avanzado de Imágenes y Señales (JIAAIS), Universidad Tecnológica Nacional de Buenos Aires, May 11-12, 2017.

2. Antonio Quintero-Rincón, Alberto Tablón, Marcelo Pereyra and Marcelo Risk. *Spatial Regularization for Head Models using EEG and MRI*. XIX Argentinean Bioengineering Society Congress, SABI 2013 (XIX Congreso Argentino de Bioingeniería y VIII Jornadas de Ingeniería Clínica), Tucumán, Argentina, september 4-6, 2013 [267].

3. Antonio Quintero-Rincón, Sergio Liberczuk and Marcelo Risk. *EEG preprocessing with Hampel filters*. Biennial Congress of IEEE Argentina, ARGENCON 2012, Córdoba Argentina, June 13-15, No 89, Vol. 2012 [18].

4. Sergio Liberczuk, Antonio Quintero-Rincón and Marcelo Risk. *Evaluación de un mapa Auto-Organizado aplicado a una Interfaz Cerebro Computadora*. Biennial Congress of IEEE Argentina, ARGENCON 2012, Córdoba Argentina, June 13-15, No 126, Vol. 2012 [268].

Finally, we hope that this thesis elucidates some aspects of EEG data processing in order to improve the understanding but also the use of new methodological tools in the community. Consequently, we hope that such a better understanding will improve the quality of results obtained with EEG signal processing in order for these brain studies to have a higher impact on both basic neuroscience and clinical studies.

6.2 Advantages and limitations

Through the use of a statistical model-based classification technique, the proposed method has three main advantages. First, it requires only estimating and classifying two scalar parameters for seizure onset detection or a few scalar parameters for spike-and-wave detection, allowing it to be implemented in dedicated real-time hardware. Second, they can be trained using a reasonably small dataset due precisely to the fact that it used only a few classification parameters. This contrasts with methods using a number of features that would require large training datasets. Third, it allows seizure detection simultaneously in the different brain rhythms, complying with current medical practices.

Nevertheless, the proposed methods have three main limitations. First, due to the very high dynamics of epileptic signals, defining the sliding time-window and the overlap of epochs is difficult. Second, it needs defining regularization parameters for the training stage in order to take into consideration random peaks, noise and artifacts that might lead to false positives. Third, seizures

have variable and dynamic offsets corresponding to the complex nature of different epilepsy types. As an example, when brain waves slow down, change from *seizure* to *non-seizure* is difficult to track and can generate classification errors.

Lyapunov exponents (LLE) are not instant for seizure detection because their coefficients are quite computationally time-consuming. It is important to conduct further studies in order to quantify the variation of LLE coefficients in the different brain states throughout all brain rhythms. This new information may have the potential to correlate with different characteristics of the seizure event and eventually provide new insights to evaluate epileptic treatments.

6.3 Research perspectives

In this thesis, we approached the challenging problems of seizure onset detection and spike-and-wave epileptiform pattern recognition in patients suffering from epilepsy. These topics have a major interest especially in real time monitoring on EEG long-term signals where the inter-trial variability can provide valuable information such as feasibility, anticipation, time, delay and source location of epilepsy seizure. In the near future, we plan to apply our existing tools for source location estimation. The idea is to use source location information to characterize the spatio-temporal patterns and its connectivity of epileptic activity in intracranial and extracranial records.

The next methodological step would be to extend our approach to different kinds of inter-trial variabilities. This work is currently starting in collaboration with Centro Integral de Epilepsia y Telemetría from Fundación contra las Enfermedades Neurológicas Infantiles (FLENI), IRIT-INP-ENSEEIHT from University of Toulouse, School of Mathematical and Computer Sciences (Heriot-Watt University) and Max-Planck Institute for Empirical Aesthetics from Frankfurt (Germany).

Also, we will focus on other epileptic waveforms patterns and implementing a medical-friendly interface with automatic epileptiforms count with an amplitude cerebral area map, robust non-parametric statistical methods application; convolutional neural networks (CNN) approaches and an increase of the database of spike-and-waves in on-line EEG long-term signals detection. It is important to improve the latency time of seizure detection and conduct a detailed study about the reliability prediction as the seizure develops across time. Other interesting research questions to explore in more depths, are related with the analysis of the frequency bands detected before the onset seizure, and which channels are active despite not participating directly in the detection of the seizure onset.

We would also be interested in the marked point process (MPP) [269, 270] using the formalism of generating probability functions (GPF) describing the space-time organization for large space and time windows in EEG signals by taking account of the preexisting heterogeneity of spontaneous seizures in epilepsy signals. Because EEG data assimilation is routinely employed as the optimal way to combine noisy observations with prior model information for obtaining better estimates of a state and, thus, better forecasts that can be achieved by ignoring data uncertainties.

Appendix A
Estimation and regularization of source localization using EEG

A.1 Dipole assumption

A typical model used in neuroscience and appropriated for a single pattern interpretation from magnetic field or potential over the electrodes $\boldsymbol{B}$, is the dipole current. For a dipole at $\boldsymbol{K}$ location, the magnetic field observed at electrode i in the position $\boldsymbol{R}(i)$, is given by

$$\boldsymbol{B}(i) = \frac{\mu}{4\pi} \frac{\boldsymbol{Q} \times (\boldsymbol{R}(i) - \boldsymbol{K})}{|\boldsymbol{R}(i) - \boldsymbol{K}|}, \text{ for } i = 1, ..., N_E, \tag{A.1}$$

For m-dipoles, the magnetic field in the location j can be obtained by

$$\boldsymbol{B}(i) = \frac{\mu}{4\pi} \sum_{i}^{m} \frac{\boldsymbol{Q}_j \times (\boldsymbol{R}(i) - \boldsymbol{K}_j)}{|\boldsymbol{R}(i) - \boldsymbol{K}_j|}, \text{ for } i = 1, ..., N_E \tag{A.2}$$

where $\times$ represents the vector product, $\boldsymbol{Q}$ is the dipole moment, N_E is the number of electrodes, $\boldsymbol{K}_j$ is the j-th dipole location and the magnetic field $\boldsymbol{B}$ can be considered as $\boldsymbol{B} = [B(1), B(2), ..., B(N_E)]$, see Figure (A.1).

Factoring the dipole moments and normalizing with respect to $\frac{\mu}{4\pi}$, into the product of their unit orientation moments and strengths, the magnetic field can be expressed as $\boldsymbol{B} = GQ$, where $\boldsymbol{Q} = [Q_1, Q_2, ...Q_m]^T$ and $\boldsymbol{G} = [g(1), g(2), ..., G(m)]$ is the propagating medium matriz. Therefore B can be expressed as $\boldsymbol{B = GMJ}$, where $\boldsymbol{GM}$ is a function of location and orientation $H(L, M)$, resulting $\boldsymbol{B = H(K, M)J}$.

The initial solution to this problem used a least-square approach (LS) [271], that minimizes the difference between the estimate and the measurement data, which is given by:

$$\boldsymbol{J}_{LS} = \|\phi - \boldsymbol{H(K, M)J}\|_F^2 \tag{A.3}$$

where ϕ is the magnetic field or the potential over the electrodes. The parameters to be estimated are location, dipole orientation and magnitude for each dipole. This is subject to knowing the number of sources or dipoles.

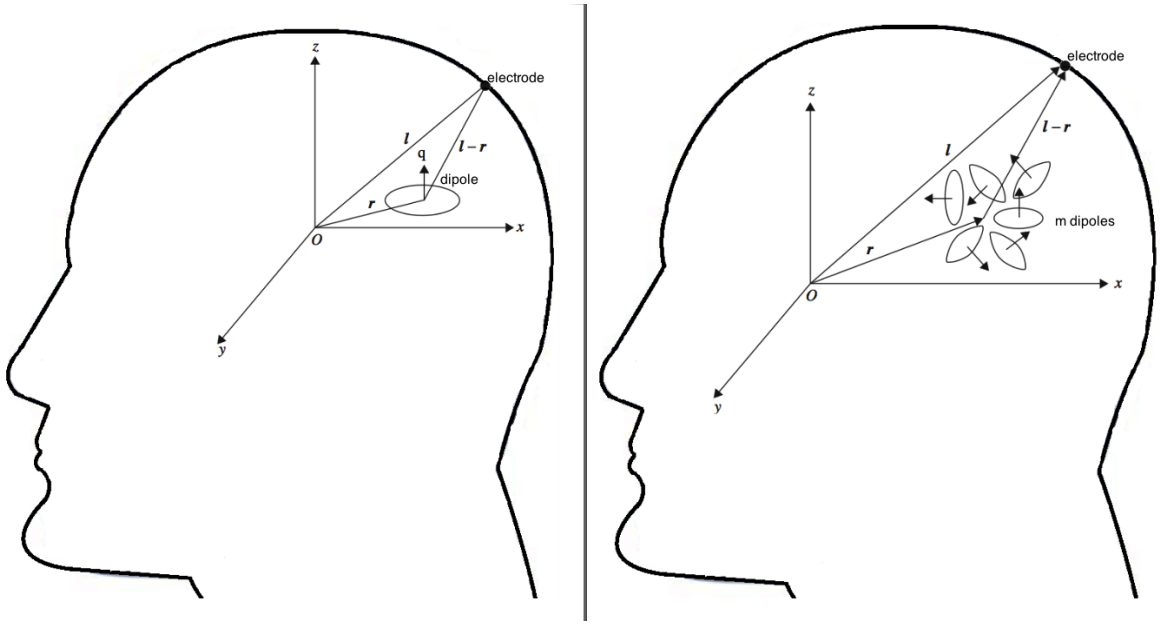


FIGURE A.1: The magnetic field $\boldsymbol{B}$ at each electrode is computed with respect to the dipole(s) moment(s) and the distance between the center of the dipole volume and the electrode.

The application of the dipole moment at the brain activity location, assumes implicitly that the current source is located in a small area or in several places separated in several dipoles models. These require a priori knowledge about the number of sources, which are usually not known. Misleading results can be obtained if these assumptions are not valid. If too few dipoles are selected, the resulting parameters are influenced by the missing dipoles. If too many dipoles, the accuracy will be reduced, because some of them are not valid brain sources. In addition, the computational cost is high because the parameters optimization are made simultaneously [272, 273]. One way to solve this is through the projection of the minimization problem:

$$\boldsymbol{J}_{LS} = \|\phi - \boldsymbol{H}(\boldsymbol{K}, \boldsymbol{M})\boldsymbol{J}\|_F^2 = \|\boldsymbol{P}_H^{\perp}\phi\|_F^2 \tag{A.4}$$

The $P_H^{\perp}$ matrix projects the data on to the orthogonal complement of the column space of $H(K, M)$, F is the Frobenius norm.

A.2 Lead-Field

The distribution of an electromagnetic field on the head or sources from EEG measurements is described by the linear Poisson

$$\bigtriangledown \cdot (\sigma \bigtriangledown \phi) = \bigtriangledown \cdot \boldsymbol{J}^s, \text{ in } \Omega \tag{A.5}$$

with no-flux Neumann boundary conditions on the scalp:

$$\sigma(\bigtriangledown\phi)\cdot = 0, \text{ on } \Gamma_{\Omega} \tag{A.6}$$

where σ is the electrical conductivity tensor, $\boldsymbol{\phi}$ is the electric potential and $\boldsymbol{J}^s$ are the electric current sources. The mapping from electrical sources in the skull for scalp recordings can be represented by a linear operator $\boldsymbol{K}$ and given a particular configuration of sources $j \in \boldsymbol{J}$, the resultant recordings $\boldsymbol{\phi} \in \Phi$, and the noise in the system ν, represent the inverse problem approach as

$$\phi = \boldsymbol{K}\boldsymbol{J} + \nu. \tag{A.7}$$

where $\boldsymbol{K}$ is the lead-field or kernel of the response of the system, his size is N_E x 3_{N_V} and contains information about the geometry and conductivity from the model. The $\boldsymbol{K}$ matrix represents the direct transmission of the coefficients of each source to source array. The model construction is easy by simple geometries such as spheres or for cases in which there is an analytical solution for the direct problem, but it is difficult for geometries based on real data from patients. $\boldsymbol{J}$ is an unknown data that represents the dipole moment matrix or electrical current sources. The perturbation matrix ν is the Gaussian noise, and the data matrix $\boldsymbol{\phi}$ can be found in the literature as: observation model, likelihood, signal temporal data, recordings or measurements.

The goal in *source image localization* problem is: Given a set of recordings $\boldsymbol{\phi}$ and knowing $\boldsymbol{K}$ (a priori), in necessary to make certain assumptions about ν, in order to determine the set of sources $\boldsymbol{J}$, which generated the recordings. The inverse problem solution basically consists of two steps: building the matrix $\boldsymbol{K}$ using the forward problem, then find the solution to the inverse problem from $\boldsymbol{K}$, which is required to estimate the $\hat{\boldsymbol{J}}$ magnitude dipolar matrix, given the positions of the electrodes and the EEG recordings in the scalp, using the $\boldsymbol{K}$ gain matrix calculated in the forward problem.

A.3 Head models

The accuracy of a realistic head model by using EEG partly depends on head tissues geometry and strongly affects the reliability of the source reconstruction process [274, 275, 276]. In the modeling method, two practical considerations must be taken into account. First, to reduce the sensitivity to noise, both in the measured voltages and the measured geometry, the number of independent measurements at the body surface usually must greatly exceed the number of variables in the source model. The overspecified equations are then solved using least-squares approximation eq. (A.3) and possibly other constraints to achieve greater stability. Second, noise sensitivity increases greatly with a growth in the number of degrees of freedom. For example, although greater brain region information could be obtained with a greater number of multiple dipoles, results could become useless if too large a number were selected [17]. A head model solves the eq. (A.6) using the forward and/or inverse problem, see Figure (A.2).

The *forward problem* is the problem in which the source and the conducting medium are known but the field is unknown and must be determined, see Figure A.2. The forward problem has a unique solution and always is possible to calculate the field with an high accuracy. This is limited only by the accuracy with which the source and the volume can be described. Forward models accuracy for EEG partly depends on head tissues geometry and strongly affects the reliability of the source reconstruction process, but it is not yet clear which brain regions are more sensitive to the choice of different model geometry [274, 275, 277].

The *inverse problem* is the problem in which the field and the conductor are known but the source is unknown. To find the source given the measured field, a unique solution cannot be found based on external measurements alone, see Figure A.2. For example in medical applications, specifically in the bioelectric phenomena, the inverse problem is very important in clinical diagnosis, because it trying to determine the source of the measured bioelectric or biomagnetic signals. Therefore the possible pathology related to the source provides the base of the diagnostic decision. The inverse problem may be solved by modeling the source of the bioelectric or biomagnetic signal and the volume conductor in the following way [17]:

1. A model is constructed from the signal source. The model should have a limited number of independent variables yet still have good correspondence with the physiology and anatomy associated with the actual source distribution.

2. A model is constructed from the volume conductor. The conductor model accuracy must be as good as or better than that of the source model.

3. At least as many independent measurements are made as the model as independent variables.

In practice, four types of head volume conductor models are the most used, but they are not the only: a homogeneous sphere head volume conductor model, a boundary element method (BEM) model, a finite element method (FEM) model and Dipole fitting methods (DFM).

A.3.1 Homogeneous sphere head volume conductor model

The inner skull surface is usually chosen in the single compartment case because the skull conductivity is much lower with respect to the cerebrospinal fluid of the brain. Although the currents outside the skull are much smaller, some studies suggest that additional layers can have an important contribution to the external field [278, 279, 280].

The volume conductor head models have been successfully considered to be a series of concentric spherical regions (e.g. brain, skull, and scalp) with results that correspond reasonably well to measurements [281, 282, 283, 284, 277], but although the head has been modeled with a sphere approximating the local inner curvature of the skull, the assumption of sphericity is poor. For example, when temporal and frontal areas are studied or when measurements cover a large area on the head. Even if a brain-shaped homogeneous conductor is considered, the secondary currents on the outer interfaces give only a negligible contribution to the magnetic field outside the head [278, 285, 286, 274, 276]. For head geometry realistic models, the surface Laplacian is estimated

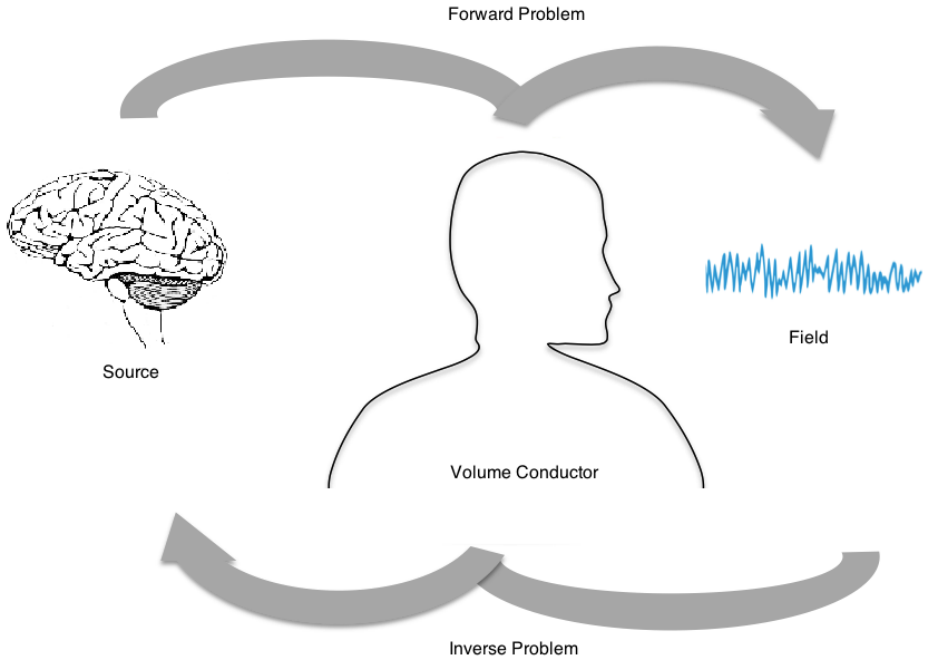


FIGURE A.2: Forward and Inverse Problems

from scalp potential directly into realistic scalp surfaces by using a triangular mesh reconstructed from MRI scans [287].

A.3.2 Boundary element model

The Boundary Element Modeling (BEM) uses the T1-MRI for creates realistically shaped layers of the body tissues. For example, a piece-wise homogeneous conductivity in each layer is assumed to build the subject-specific head volume model. A lot of studies have been shown to be computationally strong, optimum and efficient for many applications of electrophysiological source imaging or to solve the EEG forward or inverse problem [288, 277, 278, 289, 290, 291, 292]. In [293] a realistic BEM head model was constructed to localize sources by introducing the first results of numerical methods for modeling the dynamic structure and evolution of epileptic seizure activity in an intracranial subdural electrode recording.

A.3.3 Finite element model

The Finite Element Method (FEM) divides the head into small elements where the geometry and conductivity can be defined individually. FEM modeling allows handling of conductivity inhomogeneity and also tissue anisotropy. However, is restricted by the complexity and estimation of model construction. Nevertheless, is the best approximation to the real head volume conductor [294, 295, 286, 296]. If the conductivity tensor throughout the head is known, then is possible to obtain an accurate solution using FEM numerical methods [297, 298, 295, 299], without de-

tailed anatomical data for each subject [280]. In [298] the adaptive meshing scheme (wMesh) was introduced. This scheme reflects the electrical properties of the human brain optimally by using the MRI structural information and the fractional anisotropy maps derived from diffusion tensors in the FE-mesh generation process. wMesh produce different forward solutions that are different from conventional regular meshes. These are useful for modeling an individual-specific and high-resolution anisotropic FE head model by incorporating realistic anisotropic conductivity distributions. Allowing a more accurate analysis of a bioelectromagnetic problem.

A.3.4 Dipole fitting model

The Dipole fitting methods (DFM) are used for localizing focal activation with high accurately. DFM has high computational demand and it is necessary to have model assumptions because cannot reconstruct an extended source distribution and cannot localize multiple sources without enough a priori knowledge [300, 277]. Dipole fitting techniques are widely used, but in fMRI and EEG/fMRI studies have been found that spontaneous fluctuations are usually organized as diffused networks [277]. Therefore, one or a few discrete dipoles might not be adequate to represent such large-scale activity. Alternatively, EEG/MEG distributed source imaging serves well for this purpose. A straightforward strategy is to estimate the source distribution instant-by-instant to image brain activity spanning a continuous time period. Such a strategy has been applied to identify large-scale resting-state rhythms, but it is challenged by the low SNR of continuous EEG/MEG signals and the high computational demand. In Epilepsy, the subspace scanning technique for spatiotemporal dipole fitting has been used to reconstruct the ictal activity in short periods. The spatial precision and temporal resolution of which, allowed the identification of a causal relationship between epileptic sources [301]. The comparison between DFM realistic model, BEM model, and the sensor-fitted spherical model, suggest that the realistic geometry can provide a factor of improvement which is particularly important when considering sources placed in the temporal or in the occipital cortex. Template models have been suggested to simplify the analysis pipeline and possibly reduce the computational burden [302, 303, 304]. The performance between the centroid-head models and Thin Plate Spline (TPS)-MNI models (or Montreal Neurological Institute (MNI)-shaped), might be even larger for FEM/DFM models because this type of head modeling is probably much more sensitive to the approximations achieved at finer levels of detail of the image [305].

A.4 Inverse problem

The inverse problem has the imposition that the physiological constraints are based on the information available on the anatomy and physiology of the active tissue [17]. This imposes strong limitations on the number of available solutions because there is no unique solution to the inverse problem. Therefore more than one source configuration will generate fields that are consistent with the measurements. However, it may be possible to select from among these competing solutions one that at the same time meets physiological expectations.

In simplified models, the source and the volume conductor are characterized by only a few degrees of freedom. Because of one only attempts to minimize errors from the last approximation by fitting a sphere. The idea is to go from a uniform conductor to a sphere, and not from approximating the head by a uniform conductor. This is because a fully realistic model would be needed to attempt to reproduce his results with a sphere, which would defeat the purpose of using a spherical model in the first place [280]. The distributed source imaging (DSI), is used for the problems of equivalent dipole modeling. DSI uses the assumption that a source model consisting of a large number of unit dipoles evenly positioned in the brain volume or over the cortical sheet of gray matter (e.g. the cortical current density (CCD) model). Such a distributed source model approximates the biophysical organization and distribution of pyramidal neurons. DSI has the merits of solving a linear inverse problem since the locations of dipoles are fixed [277]. DSI has been developed to obtain an optimal source estimation by adding biophysical and/or physiological constraints to the distributed source imaging inverse problem. For example, the minimum norm estimate (MNE) identifies an optimal solution by using ℓ_2-norm optimization in the sense of most energy efficiency [306], or by applying the weighted-MNE (WMN) method to help to compensate the disfavored deep sources [271], or by using FEM approach [297] or by utilization of low-resolution brain electromagnetic tomography (LORETA) that further consider spatial smoothness of the neural activity [307] and their statistical analysis [308]. However, ℓ_2-norm-based techniques produce blurred images spreading over multiple cortical sulci and gyri, which lack spatial resolution to separate spatially focal sources. Nonlinear techniques based on ℓ_p-norm ($p < 2$) were developed in an attempt to make the distributed source imaging images apply to the distributed focal source, such as the focal underdetermined system solver (FOCUSS) [309], a sparse source imaging based on ℓ-1 norm [310, 311] and ℓ_p-norm iterative sparse [312].

Using MEG source localization is possible investigate the spatiotemporal dynamic estimation in large-scale distributed source spaces with several thousand source locations and hundreds of sensors [313]. The resulting inverse solutions provide substantial performance improvements over static methods by using the dynamic maximum a posteriori expectation-maximization (dMAP-EM) source localization algorithm. This algorithm is useful to estimation of cortical sources and model parameters based on the Kalman filter, the fixed interval smoother, and the EM algorithms. Kalman filter provides a natural framework in order to incorporate dynamic EEG generation models in source localization [314]. The linear spatial filters (e.g. beamformer) explain how the spherical approximation errors can give rise to larger localization differences when all modeling effects are taken into account and with your complex source configurations [280].

A.5 Source localization approaches in EEG inverse problem

The main difficulty of EEG interpretation is the infinity of spatial patterns that result in identical measurements. One possible solution is to select the current distributions, among the infinite available by selecting the more consistent with the a priori information of the problem. This additional information represents some characteristic or restriction of the currents whose cannot be determined directly from the data available. The veracity of this additional information is critical

to the reliability of the solution. Only the a priori knowledge of the source of the data, allows us to make the final decision [22]. EEG measurements along with the solution of the direct problem are possible to locate the brain regions that produced the data, in both space and time. However, due to the physics of the problem, the limited sensors number compared with the possible number of origin locations, and the measurement noise makes ill-posed the inverse problem. The general form of the inverse problem is given by the instantaneous measurements N_E given by the electrodes, and by the amount of N_V voxels in the brain. Typically, the voxels are determined by subdividing uniformly the solution space, which is usually taken as the cortical grey matter volume or surface. At each voxel, there is a point source, which may be a vector with three unknown components (the three dipole moments), or a scalar (e.g. unknown dipole amplitude, known orientation). The estimate of the sources of an EEG electromagnetic field, can be classified into two categories:

A.5.1 Equivalent current dipole approach

The equivalent current dipole approach (ECD) or parametric methods, assume that the EEG signals are generated by a relatively small number of point sources. Typically between 1 and 5 where both, position and the ideal time of each dipole are adjusted by data measuring [315, 316, 317, 318]. The concept of dipole source is a mathematical simplification of the actual distributed current source [319].

The dipoles locations are found by using the least-squares approximation (LS), see eq. (A.3). LS is a nonlinear optimization method to minimizes the variance of the data with respect to the dipole locations based on the comparison between the maps observed (EEG data) and the theoretical maps generated by the selected dipoles. The method stops when the differences between the two maps are reduced or when they have an acceptable minimum. The high dependence on the initial parameters can cause them to be trapped in the local minima, which do not represent real solutions to the problem. These difficulties are accentuated with an increasing number of sources and therefore the number of dipoles that can reasonably be estimated in practice is less than predicted by the theory. In general, ECD methods have two important limitations. First, the number of dipoles must be specified by the user and second, the optimization algorithm can become trapped in a local minimum, and therefore might not be able to find the optimum dipole location. Indeed, ECD methods are known to be unreliable when used many dipoles [320, 321]. Some methods used to dipoles adjust are brain electrical source analysis (BESA) [316], multiple signal classification (MUSIC) [322] and non-recursive subspace algorithm (FINES) [323].

A.5.2 Linear current distributed approach

The linear current distributed approach, current distributed-source reconstruction (CDR) or non-parametric methods, assume that all possible locations of the sources are simultaneous. A more general model assumes that the EEG measurements are due to a distribution of sources in the brain. As the number of unknown sources is much larger than the number of measurements, additional restrictions are required in order to obtain a unique solution [324, 325, 326, 327, 328, 329].

The source location is equivalent to finding the current amplitudes for all dipoles simultaneously.

This is an ill-posed problem because the number of dipoles is much larger than the number of sensors. However, the use of dipoles fixed location means that the forward problem is linear and the location of the source can be considered as the solution of a linear system with indeterminate equations. Similar to the problems encountered in signal processing and image processing [171]. The optimization of these methods are routed by sung distributed source imaging (DIS) or images methods [277]. The first inverse solution of a distributed model was the minimum-norm least-squares (MNLS) [22], which later developed into weighted minimum-norm (WMN), a solution used to avoid the intrinsic bias toward superficial currents [325, 271].

A.6 Current distributed-source estimation

The inverse problems usually use a solution according to ℓ_p-norm, where the regularization method is to minimize the cost function

$$F_\lambda(J) = \|\boldsymbol{KJ} - \boldsymbol{\phi}\|_2 + \lambda\|\boldsymbol{J}\|_p \tag{A.8}$$

where $\boldsymbol{J}$ is the sources currents vector, $\boldsymbol{K}$ is the lead-field, $\boldsymbol{\phi}$ is the EEG measurements, $\boldsymbol{\lambda}$ is the regularization parameter and $\|.\|$ is the minimum ℓ_p-norm method in the interval $1 \leq \ell_p \leq 2$. The main distributed-source estimates are introduced follow.

A.6.1 Minimum norm estimation

The minimum norm estimation (MNE) is based on finding a unique solution with minimum power. It is necessary to make assumptions about the solution, such assumptions can be formulated as deterministic regularization terms [149]. MNE is used when a minimum a priori information about the J source is available. When no assumptions about current discrete elements, estimates turn out to be current distributions [326], where the dipole activity extends over some areas of the cortical surface (voxel). One approach to minimizes the norm of J under the constraint of the forward problem is

$$min\|\boldsymbol{J}\|_2^2 \text{ subject to } \boldsymbol{\phi} = \boldsymbol{KJ} \tag{A.9}$$

$$\tag{A.10}$$

with a solution as

$$\boldsymbol{J} = \boldsymbol{K}^T(\boldsymbol{K}\boldsymbol{K}^T)^\dagger\boldsymbol{\phi} \tag{A.11}$$

Where † denotes the Moore-Penrose pseudo-inverse. The goal is to find a sparse solution with zero contribution from most of the sources. Therefore by combining the equations (A.9) and (A.11), the cost equation (A.8) becomes

$$F_\lambda(\boldsymbol{J}) = \|\boldsymbol{KJ} - \boldsymbol{\phi}\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{J}\|_2^2 \tag{A.12}$$

$\boldsymbol{J}$ estimation is given by

$$\hat{\boldsymbol{J}}_{\boldsymbol{N_E>N_V}} = (\boldsymbol{K}^T\boldsymbol{K} + \boldsymbol{\lambda} I_{N_V})^{-1}\boldsymbol{K}^T\boldsymbol{\phi} \tag{A.13}$$

$$\hat{\boldsymbol{J}}_{\boldsymbol{N_V>N_E}} = \boldsymbol{K}^T(\boldsymbol{K}\boldsymbol{K}^T + \boldsymbol{\lambda} I_{N_E})^{-1}\boldsymbol{\phi} \tag{A.14}$$

where N_E is the number of electrodes and N_V is the number of dipoles or voxels, T is the inverse operator $\boldsymbol{K}^T(\boldsymbol{K}\boldsymbol{K}^T + \boldsymbol{\lambda} I_{N_E})^{-1}$, so $T\boldsymbol{K}$ is the resolution matrix, ideally the identity matrix I.

The main feature is that MNE penalizes distant sources to the sensors. Therefore the estimation benefits the surface sources. This favors distributions from sources close to the measurement surface, resulting in a poor location capability for deeper sources, also has poor performance source localization in three-dimensional space.

A.6.2 Weighted minimum-norm

The weighted minimum-norm (WMNE) compensates the MNE depth sources. Therefore the method estimates the weak and surface sources. In addition, improves the performance of the three-dimensional location of sources. This is accomplished by introducing spatial weights (e.g. 3x3 weighting voxels/dipoles matrix) which ensure distribution of activity in all brain volume. The norms of the columns of $\boldsymbol{K}$ are normalized, therefore the constraint can be formulated as

$$min\|\boldsymbol{W}\boldsymbol{J}\|_2^2 \text{ subject to } \boldsymbol{\phi} = \boldsymbol{K}\boldsymbol{J} \tag{A.15}$$

$$\tag{A.16}$$

with a solution as

$$\boldsymbol{J} = \boldsymbol{W}^{-1}\boldsymbol{K}^T(\boldsymbol{K}\boldsymbol{W}^{-1}\boldsymbol{K}^T)^{\dagger}\boldsymbol{\phi} \tag{A.17}$$

where $\boldsymbol{W}$ is a diagonal of $3_{N_V} \times 3_{N_V}$ weighting matrix, which compensates for deep sources in the following way:

$$\boldsymbol{W} = \text{diag}\left[\frac{1}{\|\boldsymbol{K}_1\|_2}, \frac{1}{\|\boldsymbol{K}_2\|_2}, ..., \frac{1}{\|K_{3_{N_V}}\|_2}\right] \tag{A.18}$$

where $\|\boldsymbol{K}_i\|_2$ represents the Euclidean norm of the ith column of $\boldsymbol{K}$, $\boldsymbol{W}$ corresponds to the inverse of the distances between the sources and electrodes. Therefore the cost equation (A.8) becomes

$$F_{\boldsymbol{\lambda}}(\boldsymbol{J}) = \|\boldsymbol{K}\boldsymbol{J} - \boldsymbol{\phi}\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{W}\boldsymbol{J}\|_2^2 \tag{A.19}$$

$\boldsymbol{J}$ estimation is given by

$$\|\boldsymbol{J}\| = (\boldsymbol{K}^T\boldsymbol{K} + \boldsymbol{\lambda}\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{K}^T\boldsymbol{\phi} \tag{A.20}$$

$$\tag{A.21}$$

or

$$\|\hat{\boldsymbol{J}}\| = (\boldsymbol{K}\boldsymbol{K}^T)^{-1}\boldsymbol{K}^T(\boldsymbol{K}(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{K}^T + \boldsymbol{\lambda}I_{N_E})^{-1}\phi \tag{A.22}$$

There are many weighting strategies for the WMN solution, the low-resolution electromagnetic tomography (LORETA) algorithm, is a well-known solution for a weight matrix choice capable to the depth compensation with a smooth solution, hence the name low resolution.

A.6.3 LORETA

The low-resolution electromagnetic tomography (LORETA) [307], combines the lead-field matrix normalization $\boldsymbol{K}$ with the spatial Laplacian operator ($\mathcal{L}$) with the aim of reconstructing the surfaces and deeps sources. This operator produces a spatially smooth solution, given an assumption with respect to the neurophysiological a priori. The function of interest is

$$min\|\mathcal{L}\boldsymbol{W}\boldsymbol{J}\|_2^2 \text{ subject to } \boldsymbol{\phi} = \boldsymbol{K}\boldsymbol{J} \tag{A.23}$$

This minimum norm approach produces a smooth topography in which the peaks representing the source locations are accurately located. Therefore the cost equation (A.8) becomes

$$F_{\boldsymbol{\lambda}}(\boldsymbol{J}) = \|\boldsymbol{K}\boldsymbol{J} - \boldsymbol{\phi}\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{\Delta}\boldsymbol{B}\boldsymbol{J}\|_2^2 \tag{A.24}$$

where $\boldsymbol{B} = \hat{\boldsymbol{\Omega}} \otimes I_3$, $\otimes$ denotes the Kronecker product, I_3 is the 3x3 identity matrix and a $\boldsymbol{\Omega}$ is a diagonal matrix for the column normalization of $\boldsymbol{K}$.

$\boldsymbol{J}$ estimation is given by

$$\|\boldsymbol{J}\| = (\boldsymbol{K}^T\boldsymbol{K} + \boldsymbol{\lambda}\boldsymbol{B}\boldsymbol{\Delta}^T\boldsymbol{\Delta}\boldsymbol{B})^{-1}\boldsymbol{K}^T\boldsymbol{\phi} \tag{A.25}$$

or

$$\|\hat{\boldsymbol{J}}\| = (\boldsymbol{B}\boldsymbol{\Delta}^T\boldsymbol{\Delta}\boldsymbol{B})^{-1}\boldsymbol{K}^T(\boldsymbol{K}(\boldsymbol{B}\boldsymbol{\Delta}^T\boldsymbol{\Delta}\boldsymbol{B})^{-1}\boldsymbol{K}^T + \boldsymbol{\lambda}I_{N_E})^{-1}\boldsymbol{\phi} \tag{A.26}$$

LORETA is better than MNE because the sources are distributed in smaller quantities throughout the interior volume of the head. This process generates source distributions with low spatial resolution. Whereas that the deeper sources by using MNE cannot be recovered because the dipoles are placed on the surface of the source space. The depth compensated of the inverse solution is given according to the restriction of smoothly distributed sources based on maximum smoothness of the solution. The LORETA overall average localization error is smaller than one grid unit [308]. LORETA detect relatively strong activations in the thalamus, but a focal source reconstructed by LORETA appears to be a cloud of active sources with the maxima hopefully located at the true source location [312]. However, in many cases a higher spatial resolution is more desirable (e.g. when extracting spatial features for spatiotemporal pattern recognition) [319].

A.6.4 FOCUSS

Focal underdetermined system solver (FOCUSS), it is an iterative process of locating energy to make the solution sparse and localized, a linear function $\boldsymbol{J} = \boldsymbol{W_q}$ is applied to the function of interest as follows

$$min\|\boldsymbol{q}\| \text{ subject to } \boldsymbol{\phi} = \boldsymbol{KW_q} \tag{A.27}$$

FOCUSS is a high-resolution process for underdetermined systems [330, 309], where the $\boldsymbol{k}$-th iteration of the transform $\boldsymbol{W}_k$ is a diagonal matrix constructed by the prior iteration solution $\boldsymbol{J}_{k-1}$, denoted by $\boldsymbol{W}_k = diag(\boldsymbol{J}_{k-1})$. The final solution depends largely on the initial source distribution $\boldsymbol{J}_0$, usually provided by LORETA. In addition, is sensitive to noises and source configurations [331]. During each iteration, a matrix inverse is needed and such an inverse calculation greatly determines the stability and validation of FOCUSS. Current efforts in improvement of FOCUSS are mainly made to improve the calculation of the matrix inverse and various techniques such as singular value decomposition (SVD) truncation and regularization technique are adopted [312].

The basic form of the FOCUSS algorithm is

1. $\boldsymbol{W}_k = \text{diag}(\boldsymbol{J}_{k-1})$
2. $\boldsymbol{q}_k = (\boldsymbol{KW}_k)^{\dagger}\boldsymbol{\phi}$
3. $\boldsymbol{J}_k = \boldsymbol{W}_k\boldsymbol{q}_k$

When the iteration number is larger than the predefined maximum iteration number or when the difference between neighboring iterations is less than the termination tolerance error, then the iteration will be terminated and a sparse and energy localized solution will be achieved. FOCUSS is appropriate for recovering a few focal sources but relies on a robust initialization, additionally will converge to a localized solution with zero on most elements. The FOCUSS result is highly dependent on the initialization of the algorithm.

WMNE-FOCUSS

FOCUSS repeats the procedure of the WMN method (see the previous section A.6.2), recursively adjusting the weighting matrix until most elements of the solution become nearly zero, thus achieving a localized solution. However, the final solution depends, to some degree, on the assumed initial current distribution [319]. The Weighted Minimum Norm (WMNE) compensates for the lower gains of deeper sources by using lead-field normalization [332]. The information from the previous iteration is given by

$$min\|\boldsymbol{CJ}\|_2^2 \text{ subject to } \boldsymbol{\phi} = \boldsymbol{KJ} \tag{A.28}$$

where $\boldsymbol{C} = (\boldsymbol{W}^{-1})^T\boldsymbol{W}^{-1}$ and $\boldsymbol{W}_i = \boldsymbol{W}_{i-1}[\text{diag}(\boldsymbol{J}_{i-1}(1), ..., \boldsymbol{J}_{i-1}(3N_V))]$, the solution of iteration i is given by

$$\hat{\boldsymbol{J}}_i = \boldsymbol{W}_i\boldsymbol{W}_i^T\boldsymbol{K}^T(\boldsymbol{K}\boldsymbol{W}_i\boldsymbol{W}_i^T\boldsymbol{K}^T)^{\dagger}\boldsymbol{\phi} \tag{A.29}$$

The iterations stop when not exists significant change in the estimation. Therefore the cost equation (A.8) becomes

$$F_{\boldsymbol{\lambda}}(\boldsymbol{J}) = \|\boldsymbol{K}\boldsymbol{J} - \boldsymbol{\phi}\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{C}\boldsymbol{J}\|_2^2 \tag{A.30}$$

$\boldsymbol{J}$ estimation is given by

$$\|\hat{\boldsymbol{J}}\| = \boldsymbol{W}_i\boldsymbol{W}_i^T\boldsymbol{K}^T(\boldsymbol{K}\boldsymbol{W}_i\boldsymbol{W}_i^T\boldsymbol{K}^T + \boldsymbol{\lambda}I_{N_E})^{-1}\boldsymbol{\phi} \tag{A.31}$$

where i is the iteration index and $\boldsymbol{W}_i$ is the diagonal matrix computed using

$$\|\boldsymbol{W}_i\| = \boldsymbol{w}_i\boldsymbol{W}_{i-1}\text{diag}(\boldsymbol{J}_{i-1}) \tag{A.32}$$

The diagonal matrix for deeper source compensation is defined as

$$\|\boldsymbol{W}_i\| = \text{diag}(\frac{1}{|\boldsymbol{K}(:,j)|}), j \in [1, 2, ..., N_V], \boldsymbol{K}(:,j)\ \boldsymbol{J}\text{th is the column of } \boldsymbol{K} \tag{A.33}$$

The algorithm is initialized with the $||\hat{J}||$ MNE solution given by

$$\|\boldsymbol{W}_0\| = \text{diag}\|\hat{\boldsymbol{J}}\| = \text{diag}(\hat{\boldsymbol{J}}_{\boldsymbol{0}}(1), \hat{\boldsymbol{J}}_{\boldsymbol{0}}(2), ..., \hat{\boldsymbol{J}}_{\boldsymbol{0}}(3N_V)) \tag{A.34}$$

where $\hat{\boldsymbol{J}}_{\boldsymbol{0}}(n)$ represents the n-th element of the vector $\hat{\boldsymbol{J}}_{\boldsymbol{0}}$. If continued long enough, FOCUSS converges to a set of concentrated solutions equal to the number of electrodes. The localization accuracy is improved impressively in comparison to MNE. However, localization of deeper sources cannot be properly estimated [332]. In practice, the algorithm converges close to the initialization point and may easily become stuck in some local minimum [331].

LORETA-FOCUSS

Is similar to WMNE-FOCUSS, but using LORETA, so both can be combined together according to the following steps:

1. Compute the current density using LORETA, get the smooth solution $\hat{\boldsymbol{J}}$.
2. Construct the $\boldsymbol{W}$ matrix according to (A.32), the initial value of $\boldsymbol{W}$ is given by $\|\hat{\boldsymbol{J}}\|$ of LORETA eq. (A.34).
3. Compute the current density using eq. (A.31), that involving the compute of FOCUSS using WMNE.
4. Repeat steps (2) and (3) until the solution $\hat{J}$ no longer changes, i.e. until convergence.

FOCUSS is not able to accurately reconstruct the time series of active sources. Normally make a solution increasingly sparse during the iteration. Therefore is better to remove the nodes that do not have focal activities or to recover the active nodes that can be discarded by mistake [333].

A.6.5 sLORETA

Standardized low-resolution brain electromagnetic tomography (sLORETA), is different from LORETA because it does not use the Laplacian operator, but is similar to Dale and Sereno Method [334]. This method is based in the inverse MNE solution, which assumes the noise value N and the dipole intensity $\boldsymbol{J}$ are distributed with media no zero and the $\boldsymbol{R}$ and $\boldsymbol{C}$ covariance matrix are proportional to $\boldsymbol{I}$ matrix, is given by:

$$\hat{\boldsymbol{J}} = \boldsymbol{R}\boldsymbol{K}^T(\boldsymbol{K}\boldsymbol{R}\boldsymbol{K}^T + \boldsymbol{C})^{-1}\boldsymbol{\phi} \tag{A.35}$$

In other words, Dale and Sereno [334] proposed a method in which the localization inference is based on a standardization of the current density approach. In particular, the current density estimation is employed given by the minimum norm solution, and standardized by using the expected standard deviation, which is hypothesized to be originated exclusively by measurement noise. This method produces systematic non-zero localization errors, even in the presence of negligible noise. Precisely sLORETA location [307], is based on images of standardized current density approach. This method employs the current density estimation given by the minimum norm solution $||\boldsymbol{J}||$, where instantaneous extracranial measurements satisfy the expression $N_v \gg N_E$ (number of voxels in the brain $\gg$ number of electrodes), see equation (A.13). The localization inference is based on standardized values of the current density estimation, which is defined so the variance of the actual source be $\boldsymbol{S}_D = \boldsymbol{I}_{3dipoles} = \boldsymbol{I}_{3_{N_V}}$ and the noisy variations measures are $\boldsymbol{S}_{\phi}^{Noise} = \boldsymbol{\lambda}\boldsymbol{I}_{N_E}$.

The electrical potential variance is given by

$$\boldsymbol{S}_{\boldsymbol{\phi}} = \boldsymbol{K}\boldsymbol{S}_{\boldsymbol{\phi}}\boldsymbol{K}^T + \boldsymbol{S}_{\phi}^{Noise} \tag{A.36}$$

The variance of the estimated current density is given by

$$\boldsymbol{S}_{\hat{\boldsymbol{J}}} = \boldsymbol{T}\boldsymbol{S}_{\boldsymbol{\phi}}\boldsymbol{T}^T = \boldsymbol{K}^T[\boldsymbol{K}\boldsymbol{K}^T + \boldsymbol{\lambda}I_{N_E}]^{-1}\boldsymbol{K} \tag{A.37}$$

Eq. (A.37) is similar to the $\boldsymbol{T}\boldsymbol{K}$ matrix resolution (see section MNE (A.6.1)). For the EEG, for an unknown current density vector, the standardized current density power estimation is given by

$$\hat{\boldsymbol{J}}_{MNE,I}^T\{[\boldsymbol{S}_{\hat{\boldsymbol{J}}}]_{II}\}^{-1}\hat{\boldsymbol{J}}_{MNE,I} \tag{A.38}$$

where $\hat{\boldsymbol{J}}_{MNE,I} \in \mathcal{R}^{3x1}$ is the current density estimation at the voxel Ith given by MNE and $[\boldsymbol{S}_{\hat{\boldsymbol{J}}}]_{II} \in \mathcal{R}^{3x1}$ is the Ith diagonal block of the resolution matrix $\boldsymbol{S}_{\hat{\boldsymbol{J}}}$. sLORETA returns a unique solution to the inverse problem. Therefore the cost equation (A.8) becomes

$$min\|\boldsymbol{\phi} - \boldsymbol{K}\boldsymbol{J}\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{J}\|_2^2 \tag{A.39}$$

Using the Tikhonov-Phillips regularization, find a possible solution to the inverse problem of the form

$$\boldsymbol{j}_i = \boldsymbol{K}_i^T[\boldsymbol{K}_i\boldsymbol{K}_i^T + \boldsymbol{\lambda}_i I]^{-1}\boldsymbol{\phi} = \boldsymbol{R}_i J \tag{A.40}$$

where $\boldsymbol{j}_i$ indicates the candidate sources, $\boldsymbol{J}$ are the actual sources, $\boldsymbol{R}$ is the resolution matrix given by

$$\boldsymbol{R}_i = \boldsymbol{K}_i^T[\boldsymbol{K}_i\boldsymbol{K}_i^T + \boldsymbol{\lambda}_i I]^{-1} \tag{A.41}$$

The reconstruction of multiple sources performed by the final iteration of sLORETA is used as an initialization for the combined adaptive standardized LORETA-FOCUSS (ALF) (see section A.7.3) and weighted minimum norm (WMN or FOCUSS) algorithms [273]. The number of sources is reduced each time and the equation (A.11) is modified as follow

$$\boldsymbol{j}_i = \boldsymbol{W}_i\boldsymbol{W}_i^T\boldsymbol{K}_f^T[\boldsymbol{K}_f\boldsymbol{W}_i\boldsymbol{W}_i^T\boldsymbol{K}_f^T + \boldsymbol{\lambda} I]^{-1}\phi \tag{A.42}$$

where $\boldsymbol{K}_f$ indicates the final $N_E \times N_V$ *lead-field* matrix, returned by sLORETA. $\boldsymbol{W}_i$ is a diagonal $(3N_V)_f \times (3N_V)_f$ matrix, which is recursively improved based on the current density estimated by the previous step:

$$\boldsymbol{W}_i = diag[\boldsymbol{j}_{i-1}(1), \boldsymbol{j}_{i-1}(2), ..., \boldsymbol{j}_{i-1}((3_{N_E})_f)] \tag{A.43}$$

The resolution matrix given by (A.41) after each iteration changes to

$$\boldsymbol{R}_i = \boldsymbol{W}_i\boldsymbol{W}_i^T\boldsymbol{K}_f^T[\boldsymbol{K}_f\boldsymbol{W}_i\boldsymbol{W}_i^T\boldsymbol{K}_f^T + \boldsymbol{\lambda} I]^{-1}\boldsymbol{K}_f \tag{A.44}$$

The iterations are continued until the solution does not change significantly. sLORETA permits an accurate location without errors when the single sources are reconstructed. The maximum power current density estimation matches with the exact dipole location. While in Dale and Sereno method, the current density estimation is based only on the noise measurement. Also, sLORETA takes into account the variance of the actual source [308].

A.6.6 LPISS

The ℓ_p-norm iterative sparse solution (LPISS) find a sparse solution using ℓ_p-norm iterative with $p \leq 1$, hence the name [312]. LPISS is different to FOCUSS because the sparse solution of an intermediate auxiliary variable $\boldsymbol{q}$, is estimated by using ℓ_p-norm constrained optimization procedure, instead of the matrix inverse estimation. When the algorithm converges, a sparse solution of source $\boldsymbol{J}$ is readily derived from the obtained sparsely by using the next algorithm:

1. Initialization. Set $k = 1$, iteration termination error ϵ and the maximum iteration number $\boldsymbol{T}_{max}$, initialize source distribution $\boldsymbol{J}_{k-1}$ with LORETA solution.

2. Update the diagonal weight matrix: $\boldsymbol{W}_k = \text{diag}(\boldsymbol{J}_{k-1})$.

3. Using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization method [335], which is an iterative method to solve unconstrained nonlinear optimization, to estimate the sparse $\boldsymbol{q}_k$ in:

$$\arg\min \|\phi - \boldsymbol{K}\boldsymbol{W}_k\boldsymbol{q}_k\|_2 + \boldsymbol{\lambda}\|\boldsymbol{q}_k\|_p \tag{A.45}$$

4. Update source distribution: $\boldsymbol{J}_k = \boldsymbol{W}_k\boldsymbol{q}_k$.

5. Judge termination condition. Comparing the difference between the prior and the last source distribution:
   If $\|\boldsymbol{J}_k - \boldsymbol{J}_{k-1}\| \leq \epsilon$ or $k \geq \boldsymbol{T}_{max}$ terminate the iteration and $\boldsymbol{J}_k$ is the final source distribution.
   else
   $k = k + 1$ and jump to step 2, FOCUSS Algorithm (A.6.4) and go on.

A.6.7 Bayesian approach

This method relates the probability functions with both current and data in order to select a highly probable current distribution in a statistical sense. Depending on the complexity of the distribution assumed, may be linear or nonlinear algorithms, which allows incorporations of a priori information more elaborate. The Classic solution to solve the inverse problem, is given by eq. (A.46) previously introduced in (A.7). The idea is to try to estimate $\boldsymbol{J}$, knowing that $\boldsymbol{\phi}$ is the known variable, which can be measured and observed, ν is the perturbation and $\boldsymbol{K}$ is the lead-field that represents the solution to the direct problem. In this approach is essential to take into account any information about the a priori probabilities. In EEG applications the inverse problem equation is given by

$$\boldsymbol{\phi}(n) = \boldsymbol{K}\boldsymbol{J}(n) + \nu(n) \tag{A.46}$$

where $\boldsymbol{\phi}(n)$ is a $N_E \times 1$ vector (where N_E is the number of electrodes) containing the sample values of the EEG in time n, $\boldsymbol{K}$ is a $N_E \times m$ matrix, represents the head model, $\boldsymbol{J}(n)$ are the $m \times 1$ vectors, containing the sample values of the sources at the time n and $\nu(n)$ is a noise sample vector $N_E \times 1$ at instant n. The a priori information about the sources imposes some restrictions on their locations and their temporal properties because there is an infinite amount of $\boldsymbol{J}(n)$. Usually, the goal is to find the $\hat{\boldsymbol{J}}$ estimation that maximize some kind of criteria. The estimation may be performed using a maximum a posteriori (MAP) criterion, in which the estimator tries to find $\boldsymbol{J}(n)$ that maximizes the probability distribution of $\boldsymbol{J}(n)$ given the measurements $\boldsymbol{\phi}(n)$. The estimator is denoted as:

$$\hat{\boldsymbol{J}} = \arg\max_{\boldsymbol{J}}[\boldsymbol{\rho}(J(n)|\boldsymbol{\phi}(n))] \tag{A.47}$$

where $\rho(\boldsymbol{J}|\phi)$ is the refers to the conditional probability density of $\boldsymbol{J}$ given $\boldsymbol{\phi}$. This estimator is the more probably with respect to the measurements and the a priori considerations [332].

The starting point of all Bayesian methods is based on the Bayes theorem. Known that $\rho(\boldsymbol{J}|\boldsymbol{\phi})$ is the a posteriori density of $\boldsymbol{J}$, therefore

$$\rho(\boldsymbol{J}(n)|\boldsymbol{\phi}(n)) = \frac{\rho(\boldsymbol{\phi}(n)|\boldsymbol{J}(n))\boldsymbol{\rho}(\boldsymbol{J}(n))}{\rho(\boldsymbol{\phi}(n))} \tag{A.48}$$

where $\rho(\boldsymbol{\phi}(n)|\boldsymbol{J}(n))$ is the *likelihood*, $\rho(\boldsymbol{\phi}(n))$ is the marginal distribution of the measurements or evidence and $\rho(\boldsymbol{J}(n)$ is the prior probability.

The posterior can be written in terms of energy functions as

$$\rho(\boldsymbol{J}(n)|\boldsymbol{z}(n)) = \frac{1}{\boldsymbol{z}(n)} \exp[-\boldsymbol{U}(\boldsymbol{J}(n))] \tag{A.49}$$

where $\boldsymbol{U}(J(n))$ can be expressed as

$$\boldsymbol{U}(J(n)) = (1-\lambda)\boldsymbol{U}_1(J(n)) + \lambda U_2(J(n)) \tag{A.50}$$

where $\boldsymbol{U}_1$ is the *likelihood*, $\boldsymbol{U}_2$ is the a priori and $0 \leq \boldsymbol{\lambda} \leq 1$. The a priori may be separated into two functions, spatial priors $\boldsymbol{U}_s$ and temporal priors $\boldsymbol{U}_t$. Both reflect a balance between the data fidelity and the function spatiotemporal smoothness $\boldsymbol{\lambda}$.

A.6.8 Spatial estimation

The spatial estimation is a modification of quadratic estimation, where the $\boldsymbol{U}_2(J)$ parameter estimation in the eq. (A.50), is calculated by using the intensity gradient of the dipole [336], which leads to smooth variations in the solution by means of the estimator

$$\hat{\boldsymbol{J}} = (\boldsymbol{K}^T\boldsymbol{K} + \boldsymbol{\lambda} \bigtriangledown^T \bigtriangledown)^{-1}\boldsymbol{K}^T\boldsymbol{\phi} \tag{A.51}$$

or

$$\hat{\boldsymbol{J}} = (\bigtriangledown^T\bigtriangledown)^{-1}\boldsymbol{K}^T(\boldsymbol{K}(\bigtriangledown^T\bigtriangledown)^{-1}\boldsymbol{K}^T + \boldsymbol{\lambda} I_{N_E})^{-1}\boldsymbol{\phi} \tag{A.52}$$

The spatial estimation is an inversion procedure based on a non-quadratic choice for $\boldsymbol{U}_2(J)$ in the eq. (A.50), which makes the estimator becomes nonlinear and more suitable for detecting the intensity of energy [336], this is given by

$$\boldsymbol{U}_s(\boldsymbol{J}) = \sum_{n=1}^{N} \boldsymbol{\Phi}(\bigtriangledown \boldsymbol{J}) \tag{A.53}$$

where $N = N_V \times N$, is the number of dipoles (e.g. voxels) by the number of neighbors of each source $\boldsymbol{J}$; $\bigtriangledown\boldsymbol{J}$ is th-vector element $\bigtriangledown$. The spatial prior function can take into account the smoothness of the spatial variation of the sources. A cost function that determines the spatial

smoothness is

$$\boldsymbol{\Phi}(u) = \frac{u^2}{1 + (\frac{u}{\boldsymbol{K}})^2} = \boldsymbol{\alpha} \times \boldsymbol{\beta} \tag{A.54}$$

where $\boldsymbol{K}$ is the scaling factor that determines the required smoothness, $\boldsymbol{\alpha}$ depends on the distance between a source and his current neighbor and $\boldsymbol{\beta}$ of the discrepancy depends on the orientations of the two sources considered. Thus the function a priori for space constraints can be written as:

$$\boldsymbol{U}_s(\boldsymbol{J}(n)) = \sum_{n=1}^{N} \left[\Phi_k^x(\bigtriangledown_x \boldsymbol{J}(n)|k) + \boldsymbol{\Phi}_k^y(\bigtriangledown_y \boldsymbol{J}(n)|k)\right] \tag{A.55}$$

where the indices x and y correspond to horizontal and vertical gradients respectively. For small $\bigtriangledown$, the local cost is quadratic, producing areas with spatial changes of smooth intensity. Whereas for big $\bigtriangledown$, the local cost is infinite, $\Phi(u) \approx \boldsymbol{K}^2$, thus enabling preservation of discontinuities estimator is given by

$$\hat{\boldsymbol{J}} = \theta(\boldsymbol{K}, \hat{\boldsymbol{J}}_{i-1})\boldsymbol{\phi} \tag{A.56}$$

where $\boldsymbol{\theta}$ is a $N_V \times N_E$ matrix (N_V is the number of dipoles and N_E is the number of electrodes), which depends on $\boldsymbol{K}$ and with priors calculated from the previous estimated source $\hat{\boldsymbol{J}}_{i-1}$. The spatial resolution depends on the signal-noise ratio of the scalp. One approach to achieving a higher resolution is to use ℓ_1-norm instead of ℓ_2-norm. ℓ_1-norm methods can generate more focal solutions and have a more robust behavior with respect to outliers in the measurement data. However, ℓ_1-norm methods require much more computational effort in comparison with ℓ_2-norm methods [282].

A.6.9 Spatiotemporal estimation

The component of the temporal magnitudes are assumed to evolve slowly respect to the sampling frequency, therefore the time restrictions imposed by assuming that the projection of $\boldsymbol{J}(n)$ on the space perpendicular to $\boldsymbol{J}(n-1)$ is small or close.

The temporal prior function can be expressed as

$$\boldsymbol{U}_t(\boldsymbol{J}(n)) = \|P_{n-1}^{\perp} \boldsymbol{J}(n)\|^2 \tag{A.57}$$

where $P_{n-1}^{\perp}$ matrix is the projection onto space perpendicular to $\boldsymbol{J}(n-1)$. Therefore the overall minimization criterion for estimation of $\boldsymbol{J}(n)$ will be

$$\boldsymbol{U}_t(\boldsymbol{J}(n)) = \underset{\boldsymbol{J}}{arg\,min}\left[\|\phi(n) - \boldsymbol{K}\boldsymbol{J}(n)\|^2 + \boldsymbol{\alpha}\sum_{k=1}^{N_E}\left[\Phi_k^x(\bigtriangledown_x \boldsymbol{J}(n)|k) + \Phi_k^y(\bigtriangledown_y \boldsymbol{J}(n)|k)\right] + \boldsymbol{\beta}\|P_{n-1}^{\perp}\boldsymbol{J}(n)\|^2\right] \tag{A.58}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the penalty terms or regularization parameters.

A.7 Shrinking methods

A.7.1 Shrinking LORETA-FOCUSS

This method combines LORETA and FOCUSS by iterative adjustments in the space solution in order to reduce calculation time and increase the resolution of the source. The method begins with a soft LORETA solution that improves the intensity of some relevant dipoles in the solution while accentuates the intensity of those who are not.

The algorithm consists of the following steps:

1. Estimate the current density using LORETA to obtain $\hat{\boldsymbol{J}}$.
2. Construct the weighting matrix $\boldsymbol{W}$ using the solution estimated from the previous step, i.e., update the weighting matrix (A.32), Then, compute the current density according to (A.29), where the initial value is given by the equation (A.34) with $\hat{\boldsymbol{J}}$ of LORETA.
3. Current density $\hat{\boldsymbol{J}}_i$ estimation, which is calculated using the equation that involves the estimation of FOCUSS using WMNE (A.31).
4. In the readjusted smoothing process, the prominent nodes are preserved, i.e. the nodes greater than 1% of maximum value with its neighbors. The values of current density at these nodes, readjust by smoothing, the new values are given by

$$\frac{1}{N_l}\left[\hat{\boldsymbol{J}}(l) + \sum_u \hat{\boldsymbol{J}}(u)\right] \ \forall u \text{ under the constraint } \|r_l - r_u\| \leq d \tag{A.59}$$

   where r_l is the position vector of lth node, N_l is the number of neighbor nodes around the lth node with a distance equal to the minimum distance d between the nodes.
5. Shrinking process: the elements corresponding to $\hat{\boldsymbol{J}}$ and $\boldsymbol{K}$ are retained and the $\boldsymbol{K} = \boldsymbol{J}\hat{\boldsymbol{J}}$ matrix is calculated.
6. Repeat steps 2. to 5, until convergence, can be defined as a threshold or when there is no significant change in the weights of the additional iterations.
7. Let the solution of the last iteration before smoothing be the final solution.

Steps 4. and 5. are stopped if the new solution space has fewer nodes than the number of electrodes or the solution of the current iteration is less sparse than that estimated by the previous iteration.

Once steps 4. and 5. are stopped, the algorithm becomes a FOCUSS process.

The method in simulated data without noise, is able to reconstruct a three-dimensional source distribution with a smaller error of localization and energy, regarding WMNE, ℓ_1-norm and LORETA with FOCUSS is 10 times faster than LORETA-FOCUSS and several thousand to ℓ_1-norm. The method is not able to accurately reconstruct the temporal series of active sources [319, 337].

A.7.2 Standardized shrinking LORETA-FOCUSS (SSLOFO)

The Standardized Shrinking LORETA-FOCUSS (SSLOFO) method combines the characteristics of high resolution given by FOCUSS with the low resolution given by WMNE and sLORETA. Therefore SSLOFO can extract regions of dominant activity, as well as locate multiple sources within the region of interest. For the algorithm is not trapped in a local minimum is performed smoothing process. The algorithm is similar to LORETA-FOCUSS:

1. Current density is calculated using sLORETA to obtain $\hat{\boldsymbol{J}}$.
2. The weighting matrix $\boldsymbol{W}$ is build using the eq. (A.34) with sLORETA $\hat{\boldsymbol{J}}$.
3. Current density $\hat{\boldsymbol{J}}_i$ is calculated using eq. (A.31) of FOCUSS. Estimate the source power is normalized as:

$$\hat{\boldsymbol{J}}_i^T(l)\left\{[\boldsymbol{R}_i]_{ll}\right\}^{-1}\boldsymbol{J}_i(l) \tag{A.60}$$

   where $[\boldsymbol{R}_i]_{ll}$ is the lth diagonal block of matrix $\boldsymbol{R}_i = \boldsymbol{W}_i\boldsymbol{W}_i^T\boldsymbol{K}^T(\boldsymbol{K}\boldsymbol{W}_i\boldsymbol{W}_i^T + \boldsymbol{\lambda}I_{N_E})\boldsymbol{K}$
4. Retain the prominent nodes and their neighboring nodes. Adjust the values on these nodes through smoothing.
5. Redefine the solution space to contain only the retained nodes, i.e. only the corresponding elements in $\boldsymbol{J}$ and the corresponding column in $\boldsymbol{K}$.
6. Update the weighting matrix.
7. Repeat steps 3 to 6 until a stopping condition is satisfied. The stopping condition may be when a threshold is defined, or when there is no negligible change in the weights in further iterations.
8. The final solution is the result of the last step before smoothing.

In simulated data the reconstruction of the temporal waveforms of both source, individual as multiple, were correct. Thus allowing the direct estimation of the dynamics of cortical neuronal sources. The algorithm achieves an excellent localization ability on noise-free data. It is capable of recovering complex source configurations with arbitrary shapes and can produce high-quality images of extended source distributions [337]. SSLOFO shrinks the source space after each iteration of FOCUSS, reducing the computational request. The sources reconstructions obtained are better than WMNE with FOCUSS and sLORETA, even outperforms FOCUSS with many extended sources.

A.7.3 Adaptive standardized LORETA-FOCUSS (ALF)

While the above methods need a complete calculation of the matrix $\boldsymbol{K}$ of SSLOFO, Adaptive Standardized LORETA-FOCUSS (ALF), only requires between 6% and 11% of the of the full-resolution

of the lead-field matrix with a localization accuracy that was not significantly different from an exhaustive search through a fully-sampled source space. ALF It minimizes forward computations through an adaptive procedure that increases source resolution as the spatial extent is reduced [331].

The algorithm consists of the following steps:

1. A set of successive decimation is defined on the set of possible sources, these ratios determine successively higher resolutions, the first ratio is selected so as to produce a selective number of sources selected by the user and the last one relationship produces the full resolution model.

2. Beginning with the first decimation ratio, are retained only the corresponding dipole locations and columns in $\boldsymbol{K}$ of SSLOFO.

3. sLORETA in eq. (A.38) is used to achieve a smooth solution. The source with maximum normalized power is selected as the center point for spatial refinement in the next iteration, in which the next decimation ratio is applied. Successive iterations include sources within a spherical region at successively higher resolutions.

4. Steps 2 and 3 are repeated until the last decimation ratio is reached. The solution produced by the final iteration of sLORETA is used as initialization of the FOCUSS algorithm, where the normalization process, see eq. (A.60), is incorporated.

5. Iterations are continued until there is no change in the solution.

The location accuracy achieved with ALF is not significantly different than that obtained when an exhaustive search is performed for fully sampled source space is made [331]. A multiresolution framework approach can be implemented too [338]. At each iteration of the algorithm, the source space on the cortical surface was scanned at a higher spatial resolution such that at every resolution but the highest, the number of source candidates was kept constant [332].

A.8 Regularization

The regularization techniques or penalty, are approximations of an ill-posed problem in a family neighborhood of well-posed problem. They are used to prevent a large number of degrees of freedom in the source space from being used to over-fit to added noise [337].
The regularization is expressed in terms of the λ operator which is given by

$$F_{\boldsymbol{\lambda}} = \|\boldsymbol{K}\boldsymbol{J} - \boldsymbol{\phi}\|_2 + \boldsymbol{\lambda}\|J\|_p \tag{A.61}$$

This cost equation has two properties

1. Among all possible solutions, only select the best fits that satisfy the set constraints such as math (minimum norm estimates), anatomical, physiological or functional a priori information.

This is based on the understanding of both neurophysiological and biophysics analysis. The goal is to find the best approximation to the solution.

2. The solution is stable.

Thus, a challenge is to develop efficient optimization strategies that can solve the EEG inverse problem with such priors in a short time. The optimum λ value, can be determined using the spectral truncation [149] or the truncated SVD (TSVD) criteria [330], where in general, a high noise power corresponds a greater λ [339, 312, 340, 341]. Different criteria to find λ can be found in the literature, for example by using a visually or empirically approaches [342], using the L-Curve method [343, 344], setting, scaling or adjusting a SNR value [345, 346, 311, 347], by taking the additive noise as a parameter known [348], or by selecting a percent of the Lead-Field matrix, such as 99% of the total power [346] or 10% of maximum singular value [349]. Another way to estimate λ is by using the equation $\lambda = \sigma\sqrt{2logN}$, where σ is the standard deviation of noise and N is the size of the solution space [350].

A.8.1 ℓ_1 and ℓ_2 Regularization

In the literature, we can be found two types of approaches, the least-squares approximation, and the logistic regression. The idea in both is try to find a $\boldsymbol{J}$ matrix, which will minimize a loss function $\|\boldsymbol{KJ} - \boldsymbol{\phi}\|^2$ and the $\boldsymbol{\lambda}$ operator. The linear model using, see section (A.2), is given by

$$\boldsymbol{\phi} = \boldsymbol{KJ} + \nu \tag{A.62}$$

where $\boldsymbol{\phi}$ is the observation matrix such that $\boldsymbol{\phi} \in \mathcal{R}^m$, $\boldsymbol{J}$ is the unknown matrix such that $\boldsymbol{J} \in \mathcal{R}^m$, ν is the noise matrix such that $\nu \in \mathcal{R}^m$ and $\boldsymbol{K}$ is the lead-field matrix such that $\boldsymbol{K} \in \boldsymbol{R}^{m\times n}$. Over-fitting may occur when the number of observations m is not large enough compared with the number of feature variables ν, which tends to occur when large weights are found in $\boldsymbol{J}$. Therefore ℓ_2 *regularization* is used, also it is easy to calculate and does not add too much complexity to existing problems [351]. In this case the regularization term in (A.61), is given by $\lambda\|\boldsymbol{J}\|_2^2$.

The least square problems seek to minimize the equation

$$\|KJ - \boldsymbol{\phi}\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{J}\|_2^2 \tag{A.63}$$

where the regularization term restricts large value components, this is a special case of Tikhonov regularization. This can be computed directly ($\boldsymbol{O}(n^3)$) or by using iterative methods, such as the conjugate gradients method.

The logistic regression problem seek to minimize the equation

$$\boldsymbol{I}_{avg}(\boldsymbol{v}, \boldsymbol{J}) + \boldsymbol{\lambda}\|\boldsymbol{J}\|_2^2 \tag{A.64}$$

where the smooth and convex problem can be solved by using the gradient descent, steepest descent, Newton, quasi-Newton, truncated Newton or conjugate gradients methods.

The ℓ_1 *regularization* term in eq. (A.61) is given by $\boldsymbol{\lambda}\|J\|_1$, this creates sparse answers and better approximations in relevant cases [351].

The least square problems seek to minimize the equation

$$\|\boldsymbol{K}\boldsymbol{J} - \boldsymbol{\phi}\|^2 + \|\boldsymbol{J}\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{J}\|_1 \tag{A.65}$$

The logistic regression problems seek to minimize the equation

$$\boldsymbol{\ell}_{avg}(\boldsymbol{v}, \boldsymbol{J}) + \boldsymbol{\lambda}\|\boldsymbol{J}\|_1 \tag{A.66}$$

The regularization term in eq. (A.65) and eq. (A.66) penalize all factors equally which make $\boldsymbol{J}$ sparse. This means that the complexity can be reduced, thus can be viewed as a selection of relevant characteristics and/or importance. In addition, it is a non-differentiable, therefore the problem is NP-complex, which can be transformed into a convex quadratic problem, where it seeks to minimize

$$\|\boldsymbol{K}\boldsymbol{J} - \boldsymbol{\phi}\|^2 + \|\boldsymbol{J}\|_2^2 + \boldsymbol{\lambda}\sum_{i=1}^{n} u_i \text{ subject to: } -u_i \leq \boldsymbol{J}_i \leq u_i \text{ for } i = 1, ..., n \tag{A.67}$$

The solution is by using the standard convex optimization methods that usually cannot handle large practical problems [351]. The regularization assumptions made in terms of deterministic regularization or Bayesian interpretation as the minimum norm estimate (MNE) and minimum current estimate (MCE) [326, 329] correspond to ℓ_1-norm and ℓ_2-norm. In this methods a large $\boldsymbol{\lambda}$ corresponds to a large penalty in the current sources and a small λ emphasizes the reliability of the data. This means that ℓ_1-norm promotes sparse solutions, which that fact is a strong hypothesis. This implies that the solution must have only a small number of coefficients different from zero. While in some cases the minimum ℓ_2-norm may tend to force the solution of the dipoles near the sensors, rather than the true source because the magnetic field follows the inverse square of the distance between the sensor and the source [352]. In other words, this process usually leads to an overestimation of the extension of the focal areas of activation.

ℓ_1-priors are used to promote a priori solution spatially sparse and soft, while ℓ_2-priors are used either for guidance [353] or for time and guidance [311], which leads to an optimization problem convex. The prior most used in the EEG community are based on ℓ_2-norm [271, 334, 307] and the Gaussian distribution as a likelihood measurements from the current sources [329, 354, 355, 356, 357]. In ℓ_2-norm, MNE is post-processed to obtain an interpretable image activation patterns spatiotemporally [358]. MNE locates the source configuration with minimal energy. The regularization parameter or Ridge regularization is estimated by cross-validation producing a linear solution obtained by simple matrix multiplication [149]. This process makes estimation extremely fast. In [339] for example using MEG, the λ parameter was calculated by using the Markov random field (MRF) coupled with the mean field, in order to evaluate the accuracy of the wrong location and the difference between points in E/MEG.

The MNE solution is usually unstable with respect to modeling errors or the location and noise

measurement [271]. Noise measurement can be projected onto the cortical sources, therefore setting the λ parameter is intended to limit this effect [340]. Also, MNE often too diffuse and tend to estimate sources extending over a considerable part of the brain, which is not always physiologically significant, see section (A.6.1). To address these limitations have been proposed many alternatives, for example, in [310] was proposed the regularization of amplitudes from estimated sources with ℓ_1 priors by using the optimization procedure based on the simplex method [359]. This approach was later modified slightly [329], which was called MCE penalized ℓ_1 solutions, the solutions obtained by MCE are often too sparse and tend to scatter around the true sources [360]. LORETA [307] uses a regularization term based on a spatial Laplacian to enforce the smoothing solution [311]. The iteratively reweighed least-squares (IRLS) approach was proposed by [330] to find the source image location. FOCUSS algorithm approximates the solution with ℓ_0 a priori, this involves an iterative weights estimation of MNE solutions with updated weights after each iteration [361]. Some applications using FOCUSS are to find the lead-field in magnetic field tomography (MFT) [362],or to improve the calculation of the gradient vector flow (GVF) in MRI data [363]. Other approaches are by combining algorithms such as shrinking from LORETA and FOCUSS, in order to adjust the weight matrix and the spatial solution, obtaining a low error rate in the energy and focal localization sparse [333], later improved in shape recursive [319] and using with sLORETA in SSLOFO approach [337], Further improvements of the method arose with adaptive improvement, using only between 6% and 11% of the resolution of lead-field, creating sparse signals to locate the source [331], see section (A.7).

Some alternatives using ℓ_1-norm can be found in [353], which proposes the focal vector field reconstruction method (FVR) with a λ fixed, this is calculated by the division between the standard values of the electrodes and the array containing all the raw data, in simulated data the sources were recovered reliably. [364] proposed the sparse source imaging method (SSI), which reconstructs the estimated sources and the active and inactive cortical currents, λ was set to a large enough value so that the probability satisfies $\|\boldsymbol{\phi} - \boldsymbol{KJ}\|_2 \geq \boldsymbol{\lambda}$; [341] proposed the common spatial patterns method (CPS-ℓ_1) with a λ calculated by using the Tikhonov method [149] in order to optimize the outliers to classify the motor imagery in BCI spatial filters. Other methods are prelocation sources of multivariate (MAP), which restricts the solution space to rebuild the focal activity in cortical areas [344] or the mixed integer linear programming (MILP), where $\boldsymbol{\lambda}$ is estimated based on the uncertainty potential measurement. This method obtains a location of the minimum number of streams able to reconstruct the evoked potentials recorded on the scalp.

The Bayesian approach [336] was extended by introducing a spatial and temporal prior information [334], where the regularization parameter $\boldsymbol{\lambda}$ was calculated empirically through the maximum a-posteriori probability (MAP). MAP finds the coefficients that are used to establish a balance between a probability term (the quantification of fidelity to the raw data) and a term a priori on the estimator behavior. As for the marginalization of the regularization parameter, concerning the dependence of the current measurements, in [365] was proposed that instead of calculating a single best solution according to some criterion, is better generate before use, a large number of possible solutions using MCMC, both for the data as for the priors information. λ-hiperapriori can be marginalized so that, is possible to create a connection point between ℓ_2-norm, MCE, FO-

CUSS, sLORETA and minimum beamformer algorithm, to locate sources of E/MEG [366] or also by applying the potential functions resolution in empirical form. Gradients are applied to optimize the anatomical information by using a method based on the multiresolution approach to identify real sparse focal patterns from current density [338].

These approaches are originated from the fact that it is possible that the configurations of realistic sources, have only a limited number of active regions, see section (A.5), for example, when a few brain regions are activated significantly from a particular cognitive task. The source configuration is said to be spatially sparse. This assumption has proven to be relevant for clinical applications and also justifies the location approximation from dipole fitting which is currently the most widely used method in the clinical setting. [358].

However, the above approaches suffer from significant limitations. Solutions to sparsity-inducing priors, are slow when applied to the analysis of real data sets, also algorithms proposed so far, are complex and difficult to implement [358]. These can be calculated in a few hundred milliseconds, but sparse inverse solutions can take a long time to converge when the actual dimensions are used [353, 311]. More however under certain conditions, it has been shown that the sparse can allow perfect resolution of ill-posed problems [367]. On the other hand, when a sparse solution is originated independently at each time instant, is not possible to recover the time trajectories of cortical sources, therefore is important to consider the temporal dynamics from data [366, 368, 311], where hyperparameters can be calculated by using the restricted maximum likelihood (REML) approach [369, 370] or focal vector field reconstruction [353].

In sparse Bayesian learning methods [369, 370, 366], the problem can be reduced to the maximization of a non-convex cost function called *model evidence*. For example, in [353, 311], the problem was approached by using sparsity-Inducing prior, a mixture of ℓ_1-nom and ℓ_2-norm. Estimating mixed standards ($M \times N_E$) have the ability to structure a priori in order to incorporate some additional assumptions on the sources. Is possible to promote spatially focal sources with soft time estimation with a level-2 ℓ_1/ℓ_2, while a level-3 can be used to promote spatially non-overlapping sources between different experimental conditions. Some $M \times N_E$ can be obtained efficiently, for example, by using coordinate descent (based on the proximal estimation) or gradient operators methods. Both based on the current understanding of mathematics and convergence properties of these solutions.

LARS-LASSO algorithm [371, 372], which is a variant of the homotopy method [373, 374], is an extremely powerful method for solving the ℓ_1 problem [375]. Simple coordinate descent [376] or blockwise coordinate descent methods, also called block coordinate relaxation (BCR) [377], are possible strategies. Other methods have been proposed based on the gradient projection and proximity operators [358], sparsely connected sources analysis (SCSA) [357] or adaptive recursive least-squares (RLS) group lasso for real-time [374], in this last, the time sequence is generated of the optimal coefficients from the sparse prediction vectors, with a fixed $0.1 < \boldsymbol{\lambda} < 0.9$ value, typical standard RLS value for homotopy method with a steady state 0.05 error. Another interesting approach is through the structured sparse regularization, by using brain electrical sources (BES) matrix directly in the spatiotemporal source space, without having to rely on selecting a good basis for sparse decomposition techniques [378].

A.8.2 $1 < p < 2$ regularization

Usually, the main neural electrical activities are localized to sparse way, therefore, a reasonable solution must not only explain scalp records that they are localized sparse way [312]. The approach to achieve a sparse inverse problem solution for EEG can assume sparse few sources by using a non-linear optimization method [274] or estimating directly ℓ_p-norm with predefined values, such as $p = 1$ or $p = 2$. ℓ_2-norm produces extensive and extremely smooth estimates, while ℓ_1-norm can be estimated focally. These approaches used the classical regularization methods such as Tikhonov-Arsenin [149] or/and TSVD [361], both combined with the estimation methods such as LORETA [307], sLORETA [308], Bayesian approach [326, 336, 329] or by using ℓ_1-norm and ℓ_2-norm where a large λ corresponds to a large penalty in the current sources and a small λ emphasizes the reliability of the data.

For p values between $1 < p < 2$, which are subject to uncertainty, p must be treated as an unknown variable. One option is by using Bayesian inference, which is a popular method for finding a solution of the electromagnetic inverse problem [365, 354, 336], another option is through the method of Markov Chain Monte Carlo (MCMC) [365, 379], or methods based on the solution of shrinking spaces, shrinking LORETA-FOCUSS [319], standardized shrinking LORETA-FOCUSS (SSLOFO) [337] and adaptive standardized LORETA-FOCUSS (ALF) [331]. These methods using a fuzzy initial solution of the distributed source as MNS, by iterative reduction of the solution space, the solution converges to a relatively sparse, such as self-coherence enhancement algorithm (SCEA) [380] and the focal solution of indeterminate systems using FOCUSS [309], WMNE with FOCUSS [319] or by using methods that promote spatially sparse solutions, taking into account the temporal dynamics of the data [366, 368, 311]. Sparse component analysis (SCA) is a method that permits that allows the decomposition of sparse signals, which takes usually ℓ_p-norm with $p \leq 1$ as the restriction of signals decomposition, criteria used in LPISS Method [381, 382, 312].

Bibliography

[1] A. Quintero-Rincón, H. Batatia, M. Pereyra, and M. Risk, "Detecting onset in epilepsy signals using generalized gaussian distribution," *Blue Herons Editions: Fifth International Conference on Advances in New Technologies, Interactive Interfaces and Communicability*, vol. 2014, 2014.

[2] A. Quintero-Rincón, M. Pereyra, C. D'Giano, H. Batatia, and M. Risk, "A visual eeg epilepsy detection method based on a wavelet statistical representation and the kullback-leibler divergence," *IFMBE proceedings, Springer Singapore*, vol. 60, pp. 13–16, 2017.

[3] A. Quintero-Rincón, M. Pereyra, C. D'Giano, H. Batatia, and M. Risk, "A new algorithm for epilepsy seizure onset detection and spread estimation from eeg signals," *Journal of Physics: Conference Series*, vol. 705, no. 1, p. 012032, 2016.

[4] A. Quintero-Rincón, M. Pereyra, C. D'Giano, M. Risk, and H. Batatia, "Fast statistical model-based classification of epileptic eeg signals," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 4, pp. 877–889, 2018.

[5] A. Quintero-Rincón, C. D'Giano, and M. Risk, "Epileptic seizure prediction using pearson's product-moment correlation coefficient of a linear classifier from generalized gaussian modeling," *Neurología Argentina*, vol. 10, no. 4, pp. 201–217, 2018.

[6] A. Quintero-Rincón, J. Prendes, M. Pereyra, H. Batatia, and M. Risk, "Multivariate bayesian classification of epilepsy eeg signals," *2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, pp. 1–5, 2016.

[7] A. Quintero-Rincón, C. D'Giano, and H. Batatia, "Seizure onset detection in eeg signals based on entropy from generalized gaussian pdf modeling and ensemble bagging classifier (in press)," *Advances in Predictive, Preventive and Personalised Medicine (PPPM), DOI 10.1007/978-3-030-11800-6*, vol. 10, no. X, p. X, 2019.

[8] A. Quintero-Rincón, M. Flugelman, J. Prendes, and C. D'Giano, "Study on epileptic seizure detection in eeg signals using largest lyapunov exponents and logistic regression (in press)," *Revista Argentina de Bioingeniería, Bioengineering Argentinian Society*, 2018.

[9] A. Quintero-Rincón, J. Prendes, V. Muro, and C. D'Giano, "Study on spike-and-wave detection in epileptic signals using t-location-scale distribution and the k-nearest neighbors

classifier.," *IEEE URUCON Congress on Electronics, Electrical Engineering and Computing*, vol. 2017, pp. 1–4, 2017.

[10] A. Quintero-Rincón, M. Risk, C. D'Giano, V. Muro, J. Prendes, and H. Batatia, "A novel spike-and-wave automatic detection in eeg signals (in press)," *International Journal of Signal and Imaging Systems Engineering*, 2018.

[11] A. Quintero-Rincón, M. Alanis, V. Muro, and C. D'Giano, "Spike-and-wave detection in epileptic signals using cross-correlation and decision trees," *Revista Argentina de Bioingeniería, Bioengineering Argentinian Society*, vol. 21, no. 2, pp. 1–4, 2018.

[12] S. Nasehi and H. Pourghassem, "A novel fast epileptic seizure onset detection algorithm using general tensor discriminant analysis," *Journal of Clinical Neurophysiology*, vol. 30, no. 4, pp. 362–370, 2013.

[13] M. Qaraqe, M. Ismail, and E. Serpedin, "Band-sensitive seizure onset detection via csp-enhanced eeg features," *Epilepsy and Behavior*, vol. 50, pp. 77–87, 2015.

[14] A. Supratak, L. Li, and Y. Guo, "Feature extraction with stacked autoencoders for epileptic seizure detection," *Engineering in Medicine and Biology Society (EMBC)*, vol. 2014, no. 4, pp. 4184–4187, 2014.

[15] H. Berger, "Uber das elektroenkephalogram des menschen," *Archives of Psychiatry and Clinical Neuroscience*, vol. 94, no. 16, pp. 16–60, 1931.

[16] G. Buzsaki, *Rhythms of the Brain*. Oxford University Press USA, 2011.

[17] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, 1st ed., 1995.

[18] A. Quintero-Rincón, S. Liberczuk, and M. Risk, "Eeg preprocessing with hampel filters," *IEEE ARGENCON Conferences*, vol. 89, 2012.

[19] C. Sinclair, M. Gasper, and A. Blum, "Basic electronics in clinical neurophysiology," *The Clinical Neurophysiology Primer. Humana Press*, pp. 3–18, 2007.

[20] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*. John Wiley and Sons Ltd, 2008.

[21] A. B. Usakli, "Improvement of eeg signal acquisition: An electrical aspect for state of the art of front end," *Computational Intelligence and Neuroscience*, vol. 2010, no. 630649, 2010.

[22] M. Hamalainen, R. Hari, R. Limoniemi, J. Knuutila, and O. Lounasmaa, "Magnetoencephalography - theory, instrumentation and applications to noninvasive studies on the working human brain," *Reviews of Modern Physics*, vol. 65, no. 2, pp. 413–497, 1993.

[23] S. O'Regan, S. Faul, and W. Marnane, "Automatic detection of eeg artefacts arising from head movements," *Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010.

[24] D. Jaeger and R. Jung, *Encyclopedia of Computational Neuroscience*. Springer-Verlag New York Inc, 2015.

[25] M. F. Bear, B. Connors, and M. Paradiso, *Neuroscience Exploring the Brain*. Lippincott Williams and Wilkins, 2015.

[26] J. J. Newson and T. C. Thiagarajan, "Eeg frequency bands in psychiatric disorders: A review of resting state studies," *Frontiers in Human Neuroscience*, vol. 12, p. 521, 2019.

[27] W. H. Smithson and M. C. Walker, *ABC of Epilepsy*. BMJ Books, 2012.

[28] WHO, "World health organization (who). mental health action plan 2013-2020," 2018. Accessed: 2018-02-02.

[29] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. E. Jr., L. Forsgren, J. A. French, M. Glynn, D. C. Hesdorffer, B. Lee, G. W. Mathern, S. L. Moshe, E. Perucca, I. E. Scheffer, T. Tomson, M. Watanabe, and S. Wiebe, "Ilae official report: A practical clinical definition of epilepsy," *Epilepsy*, vol. 55, no. 4, pp. 475–482, 2014.

[30] H. O. Luders, *Textbook of Epilepsy Surgery*. CRC Press, 2008.

[31] J. Engel-Jr., *Epilepsy Board Quick Review Selected Tables and Figures from Seizures and Epilepsy*. Oxford University Press, 2014.

[32] I. Osorio, H. P. Zaveri, M. G. Frei, and S. Arthurs, *EPILEPSY, The Intersection of Neurosciences, Biology, Mathematics, Engineering, and Physics*. CRC Press, 2011.

[33] S. B. Wilson and R. Emerson, "Spike detection: A review and comparison of algorithms," *Journal of Clinical Neurophysiology*, vol. 113, no. 12, pp. 1873–1881, 2002.

[34] S. Sanei, *Adaptive Processing of Brain*. Wiley, 2013.

[35] EFA, "Epilepsy foundation of america." url:https://www.epilepsy.com, 2018. Accessed: 2018-02-02.

[36] V. E. Johnson, "Improving early seizure detection," *Epilepsy and Behavior*, vol. 22, no. 1, pp. S44–S48, 2011.

[37] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S. T. Treves, and J. Guttagf, "Patient-specific seizure onset detection," *Epilepsy and Behavior*, vol. 5, pp. 483–498, 2004.

[38] F. Rosenow and H. Luders, "Presurgical evaluation of epilepsy," *Brain*, vol. 124, pp. 1683–1700, 2001.

[39] I. Dolezalová, M. Brázdil, M. H. M, I. Horáková, I. Rektor, and R. Kuba, "Intracranial eeg seizure onset patterns in unilateral temporal lobe epilepsy and their relationship to other variables," *Clinical Neurophysiologys*, vol. 124, no. 6, pp. 1079–1088, 2013.

[40] P. Perucca, F. Dubeau, and J. Gotman, "Intracranial electroencephalographic seizure-onset patterns: effect of underlying pathology," *Brain*, vol. 137, no. 1, pp. 183–196, 2014.

[41] S. Viglione, V. Ordon, and F. Risch, "A methodology for detecting ongoing changes in the eeg prior to clinical seizures," *21st Western Institute on Epilepsy*, vol. WD1399, no. A, 1970.

[42] S. Liss, "Apparatus for monitoring and counteracting excess brain electrical energy to prevent epileptic seizures and the like," *US patent No. 3850161*, 1973.

[43] P. Ktonas and J. Smith, "Quantification of abnormal eeg spike characteristics," *Computers in Biology and Medicine*, vol. 4, no. 2, pp. 157–163, 1974.

[44] J. Gotman, J. Ives, and P. Gloor, "Automatic recognition of inter-ictal epileptic activity in prolonged eeg recordings," *Electroencephalography and Clinical Neurophysiology*, vol. 46, no. 5, pp. 510–520, 1979.

[45] L. D. Iasemidis, H. P. Zaveri, J. C. Sackellares, and W. J. Williams, "Linear and nonlinear modeling of ecog in temporal lobe epilepsy," *25th Annual Rocky Mountain Bioengineering Symposium*, 1988.

[46] L. D. Iasemidis, H. P. Zaveri, J. C. Sackellares, and W. J. Williams, "Nonlinear dynamics of electrocorticographic data," *Journal of Clinical Neurophysiology*, vol. 5, no. 339, 1988.

[47] W. P. Birkemeier, A. B. Fontaine, G. G. Celesia, and K. M. Ma, "Pattern recognition techniques for the detection of epileptic transients in eeg," *IEEE Transactions on Biomedical Engineering*, vol. 25, no. 3, pp. 213–217, 1978.

[48] R. Fisher, W. Webber, R. Lesser, S. Arroyo, and S. Uematsu, "High-frequency eeg activity at the start of seizures," *Journal of Clinical Neurophysiology*, vol. 9, no. 3, pp. 441–448, 1992.

[49] N. Wiener, *Cybernetics or control and communication in the animal and the machine*. The MIT Press, 1965.

[50] J. Maier, G. Dagnelie, H. Spekreijse, and B. van Dijk, "Principal components analysis for source localization of veps in man," *Vision Research*, vol. 27, no. 2, pp. 165–177, 1987.

[51] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1004–1034, 1995.

[52] H. Morris and H. Luders, "Long-term monitoring and computer analysis of eeg and epilepsy," *Elsevier Biomedical Press*, pp. 3–23, 1985.

[53] A. Murro, D. King, J. Smith, B. Gallagher, H. Flanigin, and K. Meador, "Computerized seizure detection of complex partial seizures," *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 4, pp. 330–333, 1991.

[54] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "Phase synchronization of chaotic oscillators," *Physical Review Letters*, vol. 76, pp. 1804–1807, 1996.

[55] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Physical Review Letters*, vol. 50, pp. 346–349, 1983.

[56] D. E. Lerner, "Monitoring changing dynamics with correlation integrals: case study of an epileptic seizure," *Physica D: Nonlinear Phenomena*, vol. 97, no. 4, pp. 563–576, 1996.

[57] M. L. V. Quyen, J. Martinerie, M. Baulac, and F. Varela, "Anticipating epileptic seizures in real time by a nonlinear analysis of similarity between eeg recordings," *NeuroReport*, vol. 10, no. 10, pp. 2149–2155, 1999.

[58] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, "Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble," *Physical Review Letters*, vol. 54, pp. 6709–6724, 1996.

[59] L. D. Iasemidis and J. C. Sackellares, "Chaos theory and epilepsy," *Neuroscientist*, vol. 2, pp. 118–126, 1996.

[60] R. Q. Quiroga, J. Arnhold, and P. Grassberger, "Learning driver-response relationships from synchronization patterns," *Physical Review E*, vol. 61, no. 5, pp. 5142–5148, 2000.

[61] L. Hively, P. Gailey, and V. Protopopescu, "Detecting dynamical change in nonlinear time series," *Physics Letters A*, vol. 258, pp. 103–114, 1999.

[62] N. Thomassona, T. J. Hoeppnerb, C. L. W. Jr., and J. P. Zbiluta, "Recurrence quantification in epileptic eegs," *Physics Letters A*, vol. 279, pp. 94–101, 2001.

[63] K. J. Blinowska and M. Malinowski, "Non-linear and linear forecasting of the eeg time series," *Biological Cybernetics*, vol. 66, no. 2, pp. 159–165, 1991.

[64] A. Varsavsky, I. Mareels, and M. Cook, *Epileptic seizures and the EEG Measurement, Models, Detection and Prediction*. CRC Press, 2011.

[65] J. Frost-Jr., "Automatic recognition and characterization of epileptiform discharges in the human eeg," *Journal of Clinical Neurophysiology*, vol. 2, no. 3, pp. 231–249, 1985.

[66] P. Y. Ktonas, "Automated spike and sharp wave (ssw) detection," *Methods of analysis of brain electrical and magnetic signals. Elsevier Amsterdam*, pp. 211–411, 1987.

[67] A. A. Dingle, R. D. Jones, and G. J. C. W. R. Fright, "A multisatge system to detect epileptiform activity in the eeg," *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 12, pp. 1260–1268, 1993.

[68] K. Jerger, T. Netoff, J. Francis, T. Sauer, L. Pecora, S. Weinstein, and S. Schiff, "Early seizure detection. in: Journal of clinical neurophysiology," *Journal of Clinical Neurophysiology*, vol. 18, no. 3, pp. 259–268, 2001.

[69] P. E. McSharry, L. A. Smith, and L. Tarassenko, "Comparison of predictability of epileptic seizures by a linear and a nonlinear method," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, pp. 628–633, 2003.

[70] K. Lehnertza and B. Litt, "The first international collaborative workshop on seizure prediction summary and data description," *Journal of Clinical Neurophysiology*, vol. 116, pp. 493–505, 2005.

[71] M. El-Gohary, J. McNames, and S. Elsas, "User-guided interictal spike detection," *30th Annual International Conference of the IEEE in Engineering in Medicine and Biology Society*, pp. 821–824, 2008.

[72] A. T. Tzallas, M. G. Tsipouras, D. G. Tsalikakis, E. C. Karvounis, L. Astrakas, S. Konitsiotis, and M. Tzaphlidou, *Automated Epileptic Seizure Detection Methods: A Review Study*. Epilepsy - Histological, Electroencephalographic and Psychological Aspects. Dr. Dejan Stevanovic Editions InTech, 2012.

[73] D. Lloyd, C. Binnie, and B. Batchelor, "Pattern recognition in eeg," *Advances in Behavioral Biology*, vol. 5, pp. 153–166, 1972.

[74] K. Arakawa, D. H. Fender, H. Harashima, H. Miyakawa, and Y. Saitoh, "Separation of a nonstationary component from the eeg by a nonlinear digital filter," *Advances in Behavioral Biology*, vol. 33, no. 7, pp. 724–726, 1986.

[75] J. Gotman and P. Gloor, "Automatic recognition and quantification of interictal epileptic activity in the human scalp eeg," *Electroencephalography and Clinical Neurophysiology*, vol. 41, no. 5, pp. 513–529, 1976.

[76] P. G. de Oliveira, C. Queiroz, and F. L. da Silva, "Spike detection based on a pattern recognition approach using a microcomputer," *Electroencephalography and Clinical Neurophysiology*, vol. 56, no. 1, pp. 97–103, 1983.

[77] P. Y. Ktonas, "Automated analysis of abnormal electroencephalograms," *Critical Reviews in Biomedical Engineering*, vol. 9, no. 1, pp. 39–97, 1983.

[78] J. Gotman, "Automatic seizure detection: improvements and evaluation," *Electroencephalography and Clinical Neurophysiology*, vol. 76, no. 4, pp. 317–324, 1990.

[79] R. Walters, J. C. Principe, and S.-H. Park, "Spike detection using a syntactic pattern recognition approach," *Proceedings of the Annual International Conference of the IEEE Engineering in Engineering in Medicine and Biology Society*, vol. 6, pp. 1810–1811, 1989.

[80] R.C.Eberhart, R. Dobbins, and W. Webber, "Eeg wavefonn analysis using casenet," *Proceedings of the Annual International Conference of the IEEE Engineering in Engineering in Medicine and Biology Society. Images of the Twenty-First Century*, vol. 6, pp. 2043–2047, 1989.

[81] C. C. C. Pang, A. R. Upton, G. Shine, and M. V. Kamath, "A comparison of algorithms for detection of spikes in the electroencephalogram," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 4, pp. 521–525, 2003.

[82] J. R. Glover, N. Raghavan, P. Y. Ktonas, and J. D. F. Jr., "Context-based automated detection of epileptogenic sharp transients in the eeg: Elimination of false positives," *IEEE Transactions on Biomedical Engineering*, vol. 36, no. 5, pp. 519–527, 2003.

[83] T. P. Exarchos, A. T. Tzallas, D. I. Fotiadis, S. Konitsiotis, and S. Giannopoulos, "Eeg transient event detection and classification using association rules," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 3, pp. 451–457, 2006.

[84] N. Moghim and D. W. Corne, "Predicting epileptic seizures in advance," *IPlos One*, vol. 9, no. 6, p. e99334, 2014.

[85] B. Davey, W. Fright, G. Caroll, and R. Jones, "Expert system approach to detection of epileptiform activity in the eeg," *Medical and Biological Engineering and Computing*, vol. 27, no. 4, pp. 365–370, 1989.

[86] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, and H. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[87] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1996.

[88] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 2008.

[89] MATHWORKS, "Continuous and discrete wavelet transforms." url:https://www.mathworks.com/help/wavelet/gs/continuous-and-discrete-wavelet-transforms.html, 2018. Accessed: 2018-02-02.

[90] I. Daubechies, *Ten Lectures On Wavelets Daubechies*. SIAM, 1992.

[91] V. J. Samar, A. Bopardikar, R. Rao, and K. Swartz, "Wavelet analysis of neuroelectric waveforms a conceptual tutorial," *Brain and Language 66*, vol. 66, pp. 7–60, 1999.

[92] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of eeg records in an epileptic patient using wavelet transform," *Journal of Neuroscience Methods*, vol. 123, pp. 69–87, 2003.

[93] M. Zijlmans, G. A. Worrell, M. Dumpelmann, T. Stieglitz, A. Barborica, M. Heers, A. Ikeda, N. Usui, and M. L. V. Quyen, "How to record high-frequency oscillations in epilepsy: A practical guideline," *Epilepsia*, vol. 58, no. 8, pp. 1305–1315, 2017.

[94] A. Davison, *Statistical Models*. Cambridge University Press, 2008.

[95] B. S. Everitt and A. Skrondal, *The Cambridge Dictionary of Statistics*. Cambridge University Press, fourth edition ed., 2010.

[96] S. Nadarajah, "A generalized normal distribution," *Journal of Applied Statistics*, vol. 32, no. 7, pp. 685–694, 2005.

[97] M. Novey, T. Adali, and A. Roy, "Complex generalized gaussian distribution characterization, generation, and estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1427–1433, 2010.

[98] M. Bicego, D. Gonzalez-Jimenez, E. Grosso, and J. A. Castro, "Generalized gaussian distributions for sequential data classification," *IEEE International Conference Proceedings on Pattern Recognition (ICPR08)*, vol. 58, no. 3, pp. 1–4, 2008.

[99] Y. Bazi, L. Bruzzone, and F. Melgani, "Image thresholding based on the em algorithm and the generalized gaussian distribution," *Pattern Recognition*, vol. 40, no. 2, pp. 619–634, 2007.

[100] C. Hesse, D. Holtackers, and T. Heskes, "On the use of mixtures of gaussians and mixtures of generalized exponentials for modelling and classification of biomedical signals," *IEEE Benelux EMBS Symposium*, 2006.

[101] E. Niedermeyer and F. L. da Silva, *Electroencephalography Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams and Wilkins, 2010.

[102] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance," *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 146–158, 2002.

[103] F. Pascal, L. Bombrun, J.-Y. Tourneret, and Y. Berthoumieu, "Parameter estimation for multivariate generalized gaussian distributions," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5960–5971, 2013.

[104] B. Zuo and H. Zhendong, "The route optimization in logistic distribution based on improved ant colony algorithm," *3rd International Conference on Advanced Computer Theory and Engineering*, vol. 6, pp. 433–436, 2010.

[105] Z. fang Li and H. Bai, "Multi-ant colony optimization algorithm for the route optimization of logistic distribution," *Second International Conference on Computational Intelligence and Natural Computing Proceedings IEEE*, vol. 1, pp. 141–144, 2010.

[106] C. Piech, "Logistic regression." url:https://web.stanford.edu/class/archive/cs/cs109/cs109.1176/lectures/23-LogisticRegression.pdf, 2015. Accessed: 2018-02-02.

[107] K. El-Darymli, C. Moloney, E. Gill, P. Mcguire, and D. Power, "Nonlinearity and the effect of detection on single-channel synthetic aperture radar imagery," *Oceans IEEE*, pp. 1–7, 2014.

[108] B. Kasiri and V. B. Vakily, "A t-location-scale-based mmse estimator under a combination of stochastic and deterministic speech model," *6th International Conference on Information, Communications and Signal Processing IEEE*, pp. 1–5, 2007.

[109] J. Lagarias, J. A. Reeds, M. Wright, and P. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions," *SIAM Journal of Optimization*, vol. 1, no. 9, pp. 112–147, 1998.

[110] M. Veillette, "Alpha-stable distributions in matlab." url-http://math.bu.edu/people/mveillet/html/alphastablepub.html, 2008. Accessed: 2016-02-02.

[111] D. Salas-Gonzalez, J. Górriz, J. Ramírez, I. Illán, P. Padilla, F. Martínez-Murcia, and E. Lang, "Building a fp-cit spect brain template using a posterization approach," *Neuroinformatics*, vol. 13, no. 4, pp. 391–402, 2015.

[112] H. He, J. Lu, J. Chen, X. Qiu, and J. Benesty, "Robust blind identification of room acoustic channels in symmetric alpha-stable distributed noise environments," *The Journal of the Acoustical Society of America*, vol. 136, no. 2, pp. 693–704, 2014.

[113] M. Pereyra and H. Batatia, "Modeling ultrasound echoes in skin tissues using symmetric alpha-stable processes," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 59, no. 1, pp. 60–72, 2012.

[114] M. Bergamasco and L. Piroddi, "Active noise control of impulsive noise using online estimation of an alpha-stable model," *49th IEEE Conference on Decision and Control*, pp. 36–41, 2010.

[115] H. Bhaskar, L. Mihaylova, and A. Achim, "Automatic object detection based on adaptive background subtraction using symmetric alpha stable distribution," *IET Seminar on Target Tracking and Data Fusion: Algorithms and Applications*, p. 195, 2008.

[116] S. Samorodnitsky and M. Taqqu, *Stable non-Gaussian random processes: stochastic models with infinite variance*. Chapman and Hall, 1994.

[117] J. P. Nolan, "Numerical calculation of stable densities and distribution functions," *Communications in Statistics. Stochastic Models*, vol. 13, no. 4, pp. 759–774, 1997.

[118] M. J. Lombardi and G. Calzolari, "Indirect estimation of α-stable distributions and processes," *The Econometrics Journal*, vol. 11, no. 1, pp. 193–208, 2008.

[119] D. Bloch, "A note on the estimation of the location parameters of the cauchy distribution," *Journal of the American Statistical Association*, vol. 61, no. 316, pp. 852–855, 1966.

[120] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, "Parameter estimation for the cauchy distribution," *19th International Conference on Systems, Signals and Image Processing*, pp. 350–353, 2012.

[121] Y. Wang, Y. Qi, J. Zhu, J. Zhang, Y. Wang, G. Pan, X. Zheng, and Z. Wu, "A cauchy-based state-space model for seizure detection in an eeg monitoring system," *IEEE Intelligent Systems*, vol. 30, no. 1, pp. 6–12, 2015.

[122] V. Nath and D. Hazarika, "Comparison of generalized gaussian and cauchy distributions in modeling of dyadic rearranged 2d dct coefficients," *3rd National Conference on Emerging Trends and Applications in Computer Science*, pp. 89–92, 2012.

[123] I. Dinov, J. Boscardin, M. Mega, E. Sowell, and A. Toga, "A wavelet-based statistical analysis of fmri data: I. motivation and data distribution modeling," *Neuroinformatics*, vol. 3, no. 4, pp. 319–342, 2005.

[124] R. E. Carrillo, T. C. Aysal, and K. E. Barner, "A generalized cauchy distribution framework for problems requiring robust behavior," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, p. 312989, 2010.

[125] R. C. Laha, *Handbook of methods of Applied Statistics Volume I: Techniques of computation, descriptive methods, and statistical inference*. John Wiley and Sons, 1967.

[126] NIST-SEMATECH, "e-handbook of statistical methods." url:http://www.itl.nist.gov/div898/handbook/, 2012. Accessed: 2016-02-02.

[127] T. Anderson, "On the distribution of the two-sample cramer-von mises criterion," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1148–1159, 1962.

[128] J.-J. Ren, "Generalized cramer-von mises test of goodness of fit for doubly censored data," *Annals of the Institute of Statistical Mathematics*, vol. 47, no. 3, pp. 525–549, 1995.

[129] J. Besag, "Statistical analysis of non-lattice data," *The Statistician*, vol. 24, no. 3, pp. 179–195, 1975.

[130] D. Barber, *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

[131] H. D. Nguyen, G. J. McLachlan, P. Orban, P. Bellec, and A. L. Janke, "Maximum pseudolikelihood estimation for model-based clustering of time series data," *arXiv:1602.08787v2 [stat.CO]*, 12016.

[132] L. Fahrmeir and H. Kaufmann, "Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models," *The Annals of Statistics*, vol. 13, no. 1, pp. 342–368, 1985.

[133] Y. Zhiliang, "Asymptotic properties of a maximum likelihood estimator with data from a gaussian process," *Journal of Multivariate Analysis*, vol. 36, pp. 280–296, 1991.

[134] N. S. Software, "T-tests." url:https://www.ncss.com/software/ncss/comparing-means-in-ncss/, 2018. Accessed: 2018-02-02.

[135] V. E. Johnson, "Uniformly most powerful bayesian test," *Annals of statistics*, vol. 41, pp. 1716–1741, 2013.

[136] S. Kullback and R. Leibler, "On information and sufficency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[137] S. Kullback, *Information theory and statistics*. Dover Publications, 1978.

[138] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, 2006.

[139] R. M. Gray, *Entropy and Information Theory*. Springer, 2011.

[140] C. Nafornita, Y. Berthoumieu, I. Nafornita, and A. Isar, "Kullback-leibler distance between complex generalized gaussian distributions," *IEEE 20th European Signal Processing Conference*, pp. 1850–1854, 2012.

[141] R. Q. Quiroga, J. Arnhold, and P. Grassberger, "Kulback-leibler and renormalized entropies: Applications to electroencephalograms of epilepsy patients," *Physical Review E*, vol. 62, no. 6, pp. 8380–8386, 2000.

[142] S. Sabesan, L. B. Good, K. S. Tsakalis, A. Spanias, D. M. Treiman, and L. D. Iasemidis, "Information flow in coupled nonlinear systems: Application to the epileptic human brain," *Data mining in biomedicine, Springer optimization and its applications series, P. Pardalos, V. Boginski, and A. Vazacopoulos, eds., New York: Springer*, pp. 483–504, 2007.

[143] S. Sabesan, L. B. Good, K. S. Tsakalis, A. Spanias, D. M. Treiman, and L. D. Iasemidis, "Information flow and application to epileptogenic focus localization from intracranial eeg," *IEEE Transactions on Neural Systems and Rehabili- tation Engineering*, vol. 17, no. 3, pp. 244–253, 2009.

[144] M. Phothisonothai and M. Nakagawa, "Eeg signal classification method based on fractal features and neural network," *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3880–3883, 2008.

[145] A. Gupta, S. Parameswaran, and C.-H. Lee, "Classification of electroencephalography (eeg) signals for different mental activities using kullback leibler (kl) divergence," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1697–1700, 2009.

[146] H. Wang, "Harmonic mean of kullback-leibler divergences for optimizing multi-class eeg spatio-temporal filters," *Neural Process Letters*, vol. 36, pp. 161–171, 2012.

[147] W. K. Pratt, *Introduction to Digital Image Processing*. CRC Press, 2013.

[148] F. Tadel, S. Baillet, J. Mosher, D. Pantazis, and R. Leahy, "Brainstorm: A user-friendly application for meg/eeg analysis," *Computational Intelligence and Neuroscience*, vol. ID 879716, 2011.

[149] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill Posed Problems*. V.H. Winston and Sons, 1st ed., 1977.

[150] T. L. Sorensen, U. L. Olsen, I. Conradsen, J. Henriksen, T. W. Kjaer, C. E. Thomsen, and H. B. D. Sorensen, "Automatic epileptic seizure onset detection using matching pursuit: a case study," *32nd Annual International Conference of the IEEE EMB*, vol. 2010, pp. 3277–3280, 2010.

[151] S.-F. Liang, H.-C. Wang, and W.-L. Chang, "Combination of eeg complexity and spectral analysis for epilepsy diagnosis and seizure detection," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, p. 853434, 2010.

[152] L. Meng, M. G. Frei, I. Osorio, G. Strang, and T. Q. Nguyen, "Gaussian mixture models of ecog signal features for improved detection of epileptic seizures," *Medical Engineering and Physics*, vol. 26, no. 5, pp. 379–393, 2004.

[153] A. Chan, F. Sun, E. Boto, and B. Wingeier, "Automated seizure onset detection for accurate onset time determination in intracranial eeg," *Clinical Neurophysiology*, vol. 119, no. 12, pp. 2687–2696, 2008.

[154] H. Ocak, "Automatic detection of epileptic seizures in eeg using discrete wavelet transform and approximate entropy," *Expert Systems with Applications*, vol. 36, pp. 2027–2036, 2009.

[155] B. Direito, C. Teixeira, B. Ribeiro, M. Castelo-Branco, F. Sales, and A. Dourado, "Modeling epileptic brain states using eeg spectral analysis and topographicmapping," *Journal of Neuroscience Methods*, vol. 210, no. 2, pp. 220–229, 2012.

[156] R. B. Pachori, "Discrimination between ictal and seizure-free eeg signals using empirical mode decomposition," *Research Letters in Signal Processing*, vol. 293056, 2008.

[157] R. BilasPachori and V. Bajaj, "Analysis of normal and epileptic seizure eeg signals using empirical mode decomposition," *Computer Methods and Programs in Biomedicine*, vol. 104, no. 3, pp. 373–381, 2011.

[158] R. B. Pachori and S. Patidar, "Epileptic seizure classification in eeg signals using second-order difference plot of intrinsic mode functions," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 494–502, 2014.

[159] P. Swami, T. K. Gandhi, B. K. Panigrahi, M. Tripathi, and S. Anand, "A novel robust diagnostic model to detect seizures in electroencephalography," *Expert Systems With Applications*, vol. 56, pp. 116–130, 2016.

[160] U. Acharya, S. L. Oh, Y. Hagiwara, J. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals," *Computers in Biology and Medicine*, vol. 17, pp. 1–9, 2017.

[161] M. Hosseini, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, "Random ensemble learning for eeg classification," *Artificial Intelligence in Medicine*, vol. 17, no. S0933-3657, pp. 30201–30204, 2018.

[162] N. Paivinen, S. Lammi, A. Pitkanen, J. Nissinen, M. Penttonen, and T. Gronfors, "Epileptic seizure detection: a nonlinear viewpoint," *Computer Methods and Programs in Biomedicine*, vol. 79, pp. 151–159, 2005.

[163] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain*, vol. 130, pp. 314–333, 2007.

[164] A. Zandi, G. Dumont, M. Javidan, and R. Tafreshi, "An entropy-based approach to predict seizures in temporal lobe epilepsy using scalp eeg," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 228–231, 2009.

[165] L. Logesparan, E. Rodríguez-Villegas, and A. Casson, "The impact of signal normalization on seizure detection using line length features," *Medical & Biological Engineering and Computing*, vol. 53, no. 10, pp. 929–942, 2015.

[166] R. R. Sharma and R. B. Pachori, "A novel approach to detect epileptic seizures using a combination of tunable-q wavelet transform and fractal dimension," *Journal of Mechanics in Medicine and Biology*, vol. 17, no. 17, p. 2017, 2017.

[167] M. Sharma, R. B. Pachori, and U. R. Acharya, "A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension," *Pattern Recognition Letters*, vol. 94, pp. 172–179, 2017.

[168] C. Fraley and A. E. Raftery., "Model-based methods of classification: using the mclust software in chemometrics," *J. Statistical Planning and Inference*, vol. 18, no. 6, pp. 1–13, 2007.

[169] P. McNicholas, "Model-based classification using latent gaussian mixture models," *Journal of Statistical Planning and Inference*, vol. 140, no. 5, pp. 1175–1181, 2010.

[170] A. Saez, C. Serrano, and B. Acha, "Model-based classification methods of global patterns in dermoscopic images," *IEEE Trans. Med. Imaging*, vol. 33, pp. 1137–1147, May 2014.

[171] M. Luessi, S. D. Babacan, R. Molina, J. R. Booth, and A. K. Katsaggelos, "Bayesian symmetrical eeg fmri fusion with spatially adaptive priors," *Neuroimage*, vol. 55, no. 1, pp. 113–132, 2011.

[172] J. Cortés, A. López, R. Molina, and A. Katsaggelos, "Variational bayesian localization of eeg sources with generalized gaussian priors," *The European Physical Journal - Plus (EPJP)*, vol. 127:140, pp. 12140–12149, 2012.

[173] A. F. Rabbi and R. Fazel-Rezai, "A fuzzy logic system for seizure onset detection in intracranial eeg," *Computational Intelligence and Neuroscience*, vol. 2012, p. 705140, 2012.

[174] EPILEPSIAE, "Epilepsiae epilepsy project." url:http://www.epilepsiae.eu/, 2018. Accessed: 2018-02-02.

[175] B. krishnan, I. Vlachos, A. Faith, S. Mullane, K. Williams, A. Alexopoulos, and L. Iasemidis, "A novel spatiotemporal analysis of peri-ictal spiking to probe the relation of spikes and seizures in epilepsy," *Annals of Biomedical Engineering*, vol. 42, no. 8, pp. 1606–1617, 2014.

[176] V. Joshi, R. B. Pachori, and A. Vijesh, "Classification of ictal and seizure-free eeg signals using fractional linear prediction," *Biomedical Signal Processing and Control*, vol. 9, pp. 1–5, 2014.

[177] R. R. Sharma and R. B. Pachori, "Classification of epileptic seizures in eeg signals based on phase space representation of intrinsic mode functions," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1106–1117, 2015.

[178] Y. Zheng, J. Zhu, Y. Qi, X. Zheng, and J. Zhang, "An automatic patient-specific seizure onset detection method using intracranial electroencephalography," *Neuromodulation: Technology at the Neural Interface*, vol. 18, no. 2, pp. 79–84, 2015.

[179] T. S. Kumar, V. Kanhanga, and R. B. Pachori, "Classification of seizure and seizure-free eeg signals using local binary patterns," *Biomedical Signal Processing and Control*, vol. 15, pp. 33–40, 2015.

[180] Y. Zhang, W. Zhou, and S. Yuan, "Multifractal analysis and relevance vector machine-based automatic seizure detection in intracranial eeg," *International Journal of Neural Systems*, vol. 25, no. 6, p. 1550020, 2015.

[181] L. Wang, W. Xue, Y. Li, M. Luo, J. Huang, W. Cui, and C. Huang, "Automatic epileptic seizure detection in eeg signals using multi-domain feature extraction and nonlinear analysis," *Entropy*, vol. 19, no. 222, 2017.

[182] A. Bhattacharyya and R. B. Pachori, "A multivariate approach for patient specific eeg seizure detection using empirical wavelet transform," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2003–2015, 2017.

[183] A. K. Tiwari, R. B. Pachori, V. Kanhangad, and B. K. Panigrahi, "Automated diagnosis of epilepsy using key-points based local binary pattern of eeg signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 4, pp. 888–896, 2017.

[184] R. R. Sharma and R. B. Pachori, "Time-frequency representation using ievdhm-ht with application to classification of epileptic eeg signals," *IET Science, Measurement and Technology*, vol. 12, no. 01, pp. 72–82, 2018.

[185] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006.

[186] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2011.

[187] E. Alpaydin, *Introduction to Machine Learning*. The MIT Press, 2014.

[188] E. Combrisson and K. Jerbi, "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy," *The Journal of Neuroscience Methods*, vol. 250, pp. 126–136, 2015.

[189] S. Sargolzaei, M. Cabrerizo, M. Goryawala, A. Eddin, and M. Adjouadi, "Scalp eeg brain functional connectivity networks in pediatric epilepsy," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 158–166, 2015.

[190] Z. Zhang and K. Parhi, "Seizure detection using wavelet decomposition of the prediction error signal from a single channel of intra-cranial eeg," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2014, pp. 4443–446, 2014.

[191] N. Stevenson, J. O'Toole, I. Korotchikova, and G. Boylan, "Artefact detection in neonatal eeg," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 926–929, 2014.

[192] S. Liang, Y. Chen, Y. Wang, P. Chen, C. Yang, and H. Chiueh, "A hierarchical approach for online temporal lobe seizure detection in long-term intracranial eeg recordings," *Journal of Neural Engineering*, vol. 10, no. 4, p. 045004, 2013.

[193] B. Schelter, J. Timmer, and A. Schulze-Bonhage, *Seizure Prediction in Epilepsy: From Basic Mechanisms to Clinical Applications*. Wiley-VCH, 2008.

[194] M. Saab and J. Gotman, "A system to detect the onset of epileptic seizures in scalp eeg," *Clinical Neurophysiology*, vol. 116, pp. 427–442, 2005.

[195] P. LeVan, E. Urrestarazu, and J. Gotman, "A system for automatic artifact removal in ictal scalp eeg based on independent component analysis and bayesian classification," *Cinical Neurophysiology*, vol. 117, no. 4, pp. 912–927, 2006.

[196] P. Valenti, E. Cazamajou, M. Scarpettini, A. Aizemberg, W. Silva, and S. Kochen, "Automatic detection of interictal spikes using data mining models," *Journal of Neuroscience Methods*, vol. 150, no. 1, pp. 105–110, 2006.

[197] J. Halford, R. Schalkoff, J. Zhou, S. Benbadis, W. Tatum, R. Turner, S. Sinha, N. Fountain, A. Arain, P. Pritchard, E. Kutluay, G. Martz, J. Edwards, C. Waters, and B. Dean,

"Standardized database development for eeg epileptiform transient detection: Eegnet scoring system and machine learning analysis," *The Journal of the Acoustical Society of America*, vol. 212, no. 2, pp. 308–316, 2014.

[198] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Journal of Clinical Neurophysiology*, vol. 29, pp. 131–163, 1997.

[199] C. Solomon and T. Breckon, *Fundamentals of Digital Image Processing*. Wiley Blackwell, 2011.

[200] Z.-H. Zhou, *Ensemble Methods Foundations and Algorithms*. Chapman and Hall/CRC, 2012.

[201] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*. Academic Press, 2015.

[202] L. Breiman, "Bagging predictors," *Technical Report No. 421, Department of Statistics, University of California*, vol. 24, no. 2, pp. 123–140, 1996.

[203] V. Tuyisenge, L. Trebaul, M. Bhattacharjee, B. Chanteloup-Foret, C. Saubat-Guigui, I. Mîndruta, S. Rheims, L. Maillard, P. Kahane, D. Taussig, and O. David, "Automatic bad channel detection in intracranial electroencephalographic recordings using ensemble machine learning," *Clinical Neurophysiology*, vol. 29, no. 3, pp. 548–554, 2018.

[204] C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning and Data Mining*. Springer-Verlag New York Inc., 2017.

[205] G. Seni and J. Elder, *Ensemble Methods in Data Mining Improving Accuracy Through Combining Predictions*. Morgan and Claypool Publishers, 2010.

[206] Z. Yu, Z. Deng, H. Wong, and L. Tan, "Identifying protein-kinase-specific phosphorylation sites based on the bagging-adaboost ensemble approach," *IEEE Transactions on NanoBioscience*, vol. 9, no. 2, pp. 132–143, 2010.

[207] H. Ashtawy and N. Mahapatra, "Bgn-score and bsn-score: bagging and boosting based ensemble neural networks scoring functions for accurate binding affinity prediction of protein-ligand complexes," *BMC Bioinformatics*, vol. 16, no. 4:S8, 2015.

[208] P. Flach, *Machine Learning*. Cambridge University Press, 2012.

[209] P. E. Rapp, I. D. Zimmerman, A. M. Albano, G. C. deGuzman, N. N. Greenbaun, and T. R. Bashore, *Experimental Studies of Chaotic Neural Behavior: Cellular Activity and Electroencephalographic Signals*, pp. 175–205. Berlin, Heidelberg: Springer Berlin Heidelberg, 1986.

[210] B. J. West, *Fractal physiology and chaos in medicine*. World Scientific Publishing Company, 2013.

[211] M. L. V. Quyen and A. Bragin, "Analysis of dynamic brain oscillations methodological advances," *TRENDS in Neurosciences*, vol. 30, no. 7, pp. 365–373, 2007.

[212] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. Wiley, 2013.

[213] A. Alkan, E. Koklukaya, and A. Subasi, "Automatic seizure detection in eeg using logistic regression and articial neural network," *Journal of Neuroscience Methods*, vol. 148, pp. 167–176, 2005.

[214] A. Subasi and E. Ercelebi, "Classication of eeg signals using neural network and logistic regression," *Computer Methods and Programs in Biomedicine*, vol. 78, pp. 87–99, 2005.

[215] A. Sansevere, K. Kapur, J. Peters, I. Fernández, T. Loddenkemper, and J. Soul, "Seizure prediction models in the neonatal intensive care unit," *Journal of Clinical Neurophysiology*, 2019.

[216] B. Chen, H. Choi, L. Hirsch, A. Katz, A. Legge, R. Buchsbaum, and K. Detyniecki, "Psychiatric and behavioral side effects of antiepileptic drugs in adults with epilepsy," *Epilepsy and Behavior*, vol. 76, pp. 24–31, 2017.

[217] K. Hampel, R. Thijs, C. Elger, and R. Surges, "Recurrence risk of ictal asystole in epilepsy," *Neurology*, vol. 89, no. 8, pp. 785–791, 2017.

[218] T. Adali, M. Anderson, and G. S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 18–33, 2014.

[219] D. Maziero, M. Sturzbecher, T. Velasco, C. Rondinoni, A. Castellanos, D. Carmichael, and C. Salmon, "A comparison of independent component analysis (ica) of fmri and electrical source imaging (esi) in focal epilepsy reveals misclassification using a classifier," *Brain Topography*, vol. 28, no. 6, pp. 813–831, 2015.

[220] R. Abreu, M. Leite, A. Leal, and P. Figueiredo, "Objective selection of epilepsy-related independent components from eeg data," *Journal of Neuroscience Methods*, vol. 258, no. 67-78, pp. 1–8, 2016.

[221] H. Becker, L. Albera, P. Comon, A. Kachenoura, and I. Merlet, "A penalized semialgebraic deflation ica algorithm for the efficient extraction of interictal epileptic signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 94–104, 2017.

[222] 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.

[223] M. T. Rosenstein, J. J. Collins, and C. D. Luca, "A practical method for calculating largest lyapunov exponents from small data set," *Physica D: Nonlinear Phenomena*, vol. 65, no. 1-2, pp. 117–134, 1993.

[224] C. Diks, *Nonlinear Time Series Analysis, Methods And Applications*. World Scientific Publishing Co., 1999.

[225] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge University Press, 2004.

[226] O. A. Rosso and M. L. Mairal, "Characterization of time dynamical evolution of electroencephalographic epileptic records," *Physica A*, vol. 312, pp. 469–504, 2002.

[227] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A wavelet-chaos methodology for analysis of eegs and eeg subbands to detect seizure and epilepsy," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 2, pp. 205–211, 2007.

[228] T. Khoa, N. Huong, and W. Toi, "Detecting epileptic seizure from scalp eeg using lyapunov spectrum," *Computational and Mathematical Methods Medicine*, vol. 2021, no. 847686, 2012.

[229] F. Shayegh, S. Sadri, R. Amirfattahi, and K. Ansari-Asl, "A model-based method for computation of correlation dimension, lyapunov exponents and synchronization from depth-eeg signals," *Computer Methods and Programs in Biomedicine*, vol. 113, pp. 323–337, 2014.

[230] V. Venkataraman, I. Vlachos, A. Faith, B. Krishnan, K. Tsakalis, D. Treiman, and L. Iasemidis, "Brain dynamics based automated epileptic seizure detection," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 946–949, 2014.

[231] G. Valenza, A. Romigi, L. Citi, F. Placidi, F. Izzi, M. Albanese, E. Scilingo, M. Marciani, A. Duggento, M. Guerrisi, N. Toschi, and R. Barbieri, "Predicting seizures in untreated temporal lobe epilepsy using point-process nonlinear models of heartbeat dynamics," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 985–988, 2016.

[232] M. Alfaro-Ponce, A. Arguelles, and I. Chairez, "Pattern recognition for electroencephalographic signals based on continuous neural networks," *Neural Networks*, vol. 79, pp. 88–96, 2016.

[233] A. Aarabi and B. He, "Seizure prediction in patients with focal hippocampal epilepsy," *Clinical Neurophysiology*, vol. 128, no. 7, pp. 1299–1307, 2017.

[234] J. Jacob, A. Cherian, K. Gopakumar, T. Iype, D. Yohannan, and K. Divya, "Can chaotic analysis of electroencephalogram aid the diagnosis of encephalopathy?," *Neurology Research International*, no. 8192820, pp. 1–8, 2018.

[235] J. Karhunen, E. Oja, and A. Hyvarinen, *Independent Component Analysis*. John Wiley and Sons Press, 2001.

[236] J. V. Stone, *Independent Component Analysis: A Tutorial Introduction*. The MIT Press, 2004.

[237] P. Comon and C. Jutten, *Handbook of Blind Source Separation - Independent Component Analysis and Applications*. Academic Press, 2010.

[238] E. Moreau and T. Adali, *Blind Identification and Separation of Complex-valued Signals*. Wiley, 2013.

[239] G. Strang, *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 2016.

[240] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEE Proceedings F - Radar and Signal Processing*, vol. 6, no. 140, pp. 362–370, 1993.

[241] J. L. Semmlow and B. Griffel, *Biosignal and Medical Image Processing.* CRC Press, 1st ed., 2014.

[242] W. Blume and J. Lemieux, "Morphology of spikes in spike-and-wave complexes," *Electroencephalography and Clinical Neurophysiology*, vol. 69, no. 6, pp. 508–515, 1988.

[243] R. Appleton, A. Nicolson, D. Smith, D. Chadwick, and J. MacKenzie, *Atlas of epilepsy*. CRC Press, 2007.

[244] O. C. Snead-III, "Basic mechanisms of generalized absence seizures," *Annals of Neurology*, vol. 37, no. 2, pp. 146–157, 1995.

[245] P. Veggiotti, M. P. F. Teutonico, D. Brazzo, U. Balottin, and C.A.Tassinari, "Therapy of encephalopathy with status epilepticus during sleep (eses/csws syndrome): An update," *Epileptic Disordes*, vol. 1, no. 14, pp. 1–11, 2012.

[246] T. Loddenkemper, I. Sanchez-Fernandez, and J. M. Peters, "Continuous spike and waves during sleep and electrical status epilepticus in sleep," *Journal of Clinical Neurophysioly*, vol. 28, no. 2, pp. 154–2164, 2011.

[247] A. Subasi, A. Alkana, E. Koklukayab, and M. K. Kiymik, "Analysis of epileptic seizure. detection methods based on parameter estimation, power spectrum density and morlet wavelet transform," *Neural Networks*, vol. 18, pp. 985–997, 2005.

[248] P. Xanthopoulos, C. C. Liu, J. Zhang, E. R. Miller, S. P. Nair, B. M. Uthman, K. Kelly, and P. M. Pardalos, "A robust spike and wave algorithm for detecting seizures in a genetic absence seizure model," *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2184–2187, 2009.

[249] E. Sitnikova, A. E. Hramov, A. A. Koronovsky, and G. van Luijtelaar, "Sleep spindles and spike-wave discharges in eeg: Their generic features, similarities and distinctions disclosed with fourier transform and continuous wavelet analysis," *Journal of Neuroscience Methods*, vol. 180, pp. 304–316, 2009.

[250] C. Richard, A. Tanenbaum, B. Audit, A. Arneodo, A. Khalil, and W. Frankel, "Swdreader a wavelet-based algorithm using spectral phase to characterize spike-wave morphological variation in genetic models of absence epilepsy," *Journal of Neuroscience Methods*, vol. 0, pp. 127–140, 2015.

[251] D. A. Pollen, "Intracellular studies of cortical neurons during thalamic induced wave and spike," *Electroencephalography and Clinical Neurophysiology*, vol. 17, pp. 398–404, 1964.

[252] P. V. Hese, J. Martens, L. Waterschoot, P. Boon, and I. Lemahieu, "Automatic detection of spike and wave discharges in the eeg of genetic absence epilepsy rats from strasbourg," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 706–717, 2009.

[253] A. Ovchinnikov, A. Luttjohann, A. Hramov, and G. van Luijtelaar, "An algorithm for real-time detection of spike-wave discharges in rodents," *Journal of Neuroscience Methods*, vol. 94, no. 1, pp. 172–178, 2010.

[254] R. Bergstrom, J. Choi, A. Manduca, H. Shin, G. Worrell, and C. Howe, "Automated identification of multiple seizure-related and interictal epileptiform event types in the eeg of mice," *Scientific Reports*, vol. 3, no. 1483, 2013.

[255] K. Rodgers, F. Dudek, and D. Barth, "2. progressive, seizure-like, spike-wave discharges are common in both injured and uninjured sprague-dawley rats: Implications for the fluid percussion injury model of post-traumatic epilepsy," *Journal of Neuroscience*, vol. 35, no. 24, pp. 9194–9204, 2015.

[256] P. Pearce, D. Friedman, J. Lafrancois, S. Iyengar, A. Fenton, N. Maclusky, and H. Scharfman, "Spike wave discharges in adult sprague dawley rats and their implications for animal models of temporal lobe epilepsy," *Epilepsy and Behavior*, vol. 32, pp. 121–131, 2014.

[257] H. Blumenfeld, "Cellular and network mechanisms of spike-wave seizures," *Epilepsia*, vol. 46, no. 9, pp. 21–33, 2005.

[258] Massimo, "A brief history on the oscillating roles of thalamus and cortex in absence seizures," *Epilepsia*, vol. 53, no. 5, pp. 779–789, 2012.

[259] P. Xanthopoulos, S. Rebennack, C. C. Liu, J. Zhang, G. L. Holmes, B. M. Uthman, and P. M. Pardalos, "A novel wavelet based algorithm for spike and wave detection in absence epilepsy," *2010 IEEE International Conference on BioInformatics and BioEngineering*, pp. 14–19, 2010.

[260] P. Abry, *Ondelettes et turbulence. Multirésolutions, algorithmes de décomposition, invariance d'échelles*. Diderot Editeur, Paris, 1997.

[261] J. C. Ong, D. Hedeker, J. K. Wyatt, and R. Manber, "Examining the variability of sleep patterns during treatment for chronic insomnia: Application of a location-scale mixed model," *Journal of Clinical Sleep Medicine*, vol. 6, no. 12, pp. 797–804, 2016.

[262] J. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 19, no. 1, pp. 81–106, 1986.

[263] A. Quintero-Rincón, C. Carenzo, J. Ems, L. Hirschson, V. Muro, and C. D'Giano, "Spike-and-wave epileptiform discharge pattern detection based on kendall's tau-b coefficient," *Applied Medical Informatics (In press)*, vol. X, no. X, p. X, 2019.

[264] A. Quintero-Rincón, C. D'Giano, and H. Batatia, "Curve fitting based on two-point central difference to detect epileptic eeg seizures curve fitting based on two-point central difference to detect epileptic eeg seizures curve fitting based on two-point central difference to detect epileptic eeg seizures," *Journal of Biomedical Research (In press)*, vol. X, no. X, p. X, 2019.

[265] I. Zorgno, M. C. Blanc, S. Oxenford, F. G. Garbagnoli, C. D'Giano, and A. Quintero-Rincón, "Epilepsy seizure onset detection applying 1-nn classifier based on statistical parameters," *IEEE Argentina Biennial Congress ARGENCON 2018, DOI: 10.1109/ARGENCON.2018.8646234*, vol. 2018, pp. 1–5, 2018.

[266] B. Bouaziz, L. Chaari, H. Batatia, and A. Quintero-Rincón, "Epileptic seizure detection using a convolutional neural network (in press)," *Advances in Predictive, Preventive and Personalised Medicine (PPPM), DOI 10.1007/978-3-030-11800-6*, vol. 10, no. X, p. X, 2019.

[267] A. Quintero-Rincón, A. Tablón, M. Pereyra, and M. Risk, "Spatial regularization for head models using eeg and mri," *SABI 2013 - XIX Argentine Congress of Bioengineering*, vol. 2013, no. 142, 2013.

[268] S. Liberczuk, A. Quintero-Rincón, and M. Risk, "Evaluación de un mapa auto-organizado aplicado a una interfaz cerebro computadora," *IEEE ARGENCON Conferences*, vol. 126, 2012.

[269] M. Jacobsen, *Point process theory and applications Marked point and piecewise deterministic processes*. Birkhauser, 2006.

[270] M. van Lieshout, "A j-function for marked point patterns," *AISM*, vol. 58, no. 2016, pp. 235–259, 2006.

[271] J. Wang, S. Williamson, and L. Kaufman, "Magnetic source images determined by a lead-field analysis: The unique minimum-norm least squares estimation," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 7, pp. 665–675, 1992.

[272] C. H. Muravchik and A. Nehorai, "Eeg/meg error bounds for a static dipole source with a realistic head model," *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 330–333, 2001.

[273] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Wiley-Interscience, 1st ed., 2007.

[274] N. von Ellenrieder, C. H. Muravchik, and A. Nehorai, "Effects of geometric head model perturbations on the eeg forward and inverse problems," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 3, pp. 421–429, 2006.

[275] N. Von-Ellenrieder, C. H. Muravchik, M. Wagner, and A. Nehorai, "Effect of head shape variations among individuals on the eeg/meg forward and inverse problems," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 587–597, 2009.

[276] F. Vatta, F. Meneghini, F. Esposito, S. Mininel, and F. D. Salle, "Realistic and spherical head modeling for eeg forward problem solution: a comparative cortex-based analysis," *Computational Intelligent and Neuroscience*, vol. 2010, no. 972060, 2010.

[277] B. He, L. Yang, C. Wilke, and H. Yuan, "Electrophysiological imaging of brain activity and connectivity challenges and opportunities," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 7, pp. 1918–1931, 2011.

[278] M. Hamalainen and J. Sarvas, "Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data," *IEEE Transactions on Biomedical Engineering*, vol. 36, no. 2, pp. 165–171, 1989.

[279] G. Dassios, S. N. Giapalaki, A. N. Kandili, and F. Kariotou, "The exterior magnetic field for the multilayer ellipsoidal model of the brain," *The Quarterly Journal of Mechanics and Applied Mathematics*, vol. 60, pp. 1–25, 2007.

[280] M. Lalancette, M. Quraan, and D. Cheyne, "Evaluation of multiple-sphere head models for meg source localization," *Physics in Medicine and Biology.*, vol. 56, no. 17, pp. 5621–5635, 2011.

[281] B. N. Cuffin, "A method for localizing eeg," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 1, pp. 68–71, 1995.

[282] M. Fuchs, R. Drenckhahn, H. Wischmann, and M. Wagner, "An improved boundary element method for realistic volume-conductor modeling," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 8, pp. 980–997, 1998.

[283] R. Uitert and C. Johnson, "Can a spherical model substitute for a realistic head model in forward and inverse meg simulations," *Proceedings of the 13th International Conference on Biomagnetism*, pp. 798–800, 2002.

[284] D. Gutierrez, A. Nehorai, and C. H. Muravchik, "Estimating brain conductivities and dipole source signals with eeg arrays," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 12, pp. 2113–2122, 2004.

[285] B. He and T. .Musha, "Effects of cavities on eeg dipole localization and their relations with surface electrode positions," *International Journal of Bio-Medical Computing at Science Direct*, vol. 24, no. 4, pp. 269–282, 1989.

[286] F. Darvas, J. Ermer, J. Mosher, and R. Leahy, "Generic head models for atlas-based eeg-source analysis," *Human Brain Mapping*, vol. 27, no. 2, pp. 129–143, 2006.

[287] S. Deng, W. Winter, S. Thorpe, and R. Srinivasan, "Improved surface laplacian estimates of cortical potential using realistic models of head geometry," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 11, pp. 2979–2985, 2012.

[288] B. He, "Electric dipole tracing in the brain by means of the boundary element method and its accuracy," *IEEE Transactions on Biomedical Engineering*, vol. 34, no. 6, pp. 406–414, 1987.

[289] Z. Akalin-Acar and N. G. Gencer, "An advanced boundary element method (bem) implementation for the forward problem of electromagnetic source imaging," *Physics in Medicine and Biology*, vol. 49, no. 21, pp. 5011–5028, 2004.

[290] J. Kybic, M. Clerc, O. Faugeras, R. Keriven, and T. Papadopoulo, "Generalized head models for meg/eeg: boundary element method beyond nested volumes," *Physics in Medicine and Biology*, vol. 7, no. 51(5), pp. 1333–1346, 2006.

[291] N. Cao, I. Yetik, A. Nehorai, C. Muravchik, and J. Haueisen, "Parametric surface-source modeling and estimation with electroencephalography," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2414–2424, 2006.

[292] L. Peng, M. Ming, J. Yujian, and X.Anhuai, "Comparison of eeg forward solutions in two head models," *International Conference on Multimedia and Signal Processing, CMSP*, vol. 2, pp. 193–197, 2011.

[293] Z. Akalin-Acar, J. Palmer, G. Worrell, and S. Makeig, "Electrocortical source imaging of intracranial eeg data in epilepsy," *Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, EMBC*, pp. 3909–3912, 2011.

[294] N. P. Yan Y and H. RT, "Finite-element model of the human head: Scalp potentials due to dipole sources," *Medical and Biological Engineering and Computing, MBEC*, vol. 29, no. 5, pp. 475–481, 1991.

[295] R. V. Uitert, C. Johnson, and L. Zhukov, "Influence of head tissue conductivity in forward and inverse magnetoencephalographic simulations using realistic head models," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 12, pp. 2129–2137, 2004.

[296] Y. Zhang, L. Ding, W. van Drongelen, K. Hecox, D. M. Frim, and B. He, "A cortical potential imaging study from simultaneous extra and intracranial electrical recordings by means of the finite element method," *IEEE Transactions on Biomedical Engineering*, vol. 31, no. 4, pp. 1513–1524, 2006.

[297] F. Lucka, S. Pursiainen, M. Burger, and C. Wolters, "Hierarchical bayesian inference for the eeg inverse problem using realistic fe head models: Depth localization and source separation for focal primary currents," *Neuroimage*, vol. 61, no. 4, pp. 1364–1382, 2012.

[298] W. Lee and T. Kim, "Methods for high-resolution anisotropic finite element modeling of the human head: automatic mr white matter anisotropy-adaptive mesh generation," *Medical Engineering and Physics*, vol. 34, no. 1, pp. 85–98, 2012.

[299] S. Lew, C. Wolters, T. Dierkes, C. Roer, and R. Macleod, "Accuracy and run-time comparison for different potential approaches and iterative solvers in finite element method based eeg source analysis," *Applied Numerical Mathematics*, vol. 59, no. 8, pp. 1970–1988, 2009.

[300] H. Hallez, B. Vanrumste, P. V. Hese, Y. DAsseler, I. Lemahieu, and R. V. de Walle, "A finite difference method with reciprocity used to incorporate anisotropy in electroencephalogram dipole source localization," *Physics in Medicine and Biology*, vol. 50, no. 16, pp. 3787–806, 2005.

[301] L.Ding, G.A.Worrell, T.D.Lagerlund, and B.He, "Ictal source analysis: localization and imaging of causal interactions in humans," *Neuroimage*, vol. 34, no. 2, pp. 575–586, 2007.

[302] Z. Akalin-Acar and S. Makeig, "Neuroelectromagnetic forward modeling toolbox," *Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, EMBC*, pp. 3991–3994, 2008.

[303] R. Henson, J. Mattout, C. Phillips, and K. Friston, "Selecting forward models for meg source-reconstruction using model-evidence," *Neuroimage*, vol. 15, no. 46(1), pp. 168–176, 2009.

[304] M. Fuchs, J. Kastner, M. Wagner, S. Hawes, and J. Ebersole, "A standardized boundary element method volume conductor model," *Journal of Clinical Neurophysiology*, vol. 113, no. 5, pp. 702–712, 2002.

[305] P. A. Valdes-Hernandez, N. von Ellenrieder, A. Ojeda-Gonzalez, S. Kochen, Y. Aleman-Gomez, C. Muravchik, and P. A. Valdes-Sosa, "Approximate average head models for eeg source imaging," *Journal of Neuroscience Methods*, vol. 185, pp. 125–132, 2009.

[306] M. Hamalainen and R. Ilmoniemi, "Interpreting magnetic fields of the brain: minimum norm estimates," *Medical and Biological Engineering and Computing, MBEC*, vol. 32, no. 1, pp. 35–42, 1994.

[307] R. Pascual-Marqui, D. Lehmann, T. Koenig, K. Kochi, M. Merlo, D. Hell, and M. Koukkou, "Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain," *International Journal of Psychophysiology*, vol. 18, no. 1, pp. 49–65, 1994.

[308] R. Pascual-Marqui, "Standardized low-resolution brain electromagnetic tomography (sloreta): Technical details," *Methods and Findings in Experimental and Clinical Pharmacology*, vol. 24, no. Suppl D:5-12, 2002.

[309] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[310] K. Matsuura and Y. Okabe, "Selective minimum-norm solution of the biomagnetic inverse problem," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 6, pp. 608–615, 1995.

[311] W. Ou, M. Hamalainen, and P. Golland, "A distributed spatio-temporal eeg/meg inverse solver," *Neuroimage*, vol. 1, no. 44, pp. 932–946, 2009.

[312] P. Xu, Y. Tian, H. Chen, and D. Yao, "Lp norm iterative sparse solution for eeg source localization," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 3, pp. 400–409, 2007.

[313] C. Lamus, M. Hamalainen, S. Temereanca, E. Brown, and P. Purdon, "A spatiotemporal dynamic distributed solution to the meg inverse problem," *Neuroimage*, vol. 63, no. 2, pp. 894–909, 2011.

[314] M. Barton, P. R. P.A., S. Kumar, A. Galka, H. Durrant-Whyte, J. Guivant, and T. Ozaki, "Evaluating the performance of kalman-filter-based eeg source localization," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 1, 2009.

[315] M. Scherg and J.S.Ebersole, "Brain source imaging of focal and multifocal epileptiform eeg activity," *Neurophysiologie Clinique*, vol. 24, no. 1, pp. 51–60, 1994.

[316] W. Miltner, C. Braun, R. Johnson, G. Simpson, and D. Ruchkin, "A test of brain electrical source analysis (besa): a simulation study," *Electroencephalography and Clinical Neurophysiology*, vol. 4, pp. 295–310, 1994.

[317] M. Scherg, T. Bast, and P. Berg, "Multiple source analysis of interictal spikes: goals, requirements, and clinical value," *Neurophysiologie Clinique*, vol. 16, no. 3, pp. 214–224, 1999.

[318] C. Aine, M. Huang, J. Stephen, and R. Christner, "Multistart algorithms for meg empirical data analysis reliably characterize locations and time courses of multiple sources," *Neuroimage*, vol. 12, no. 2, pp. 159–172, 2000.

[319] H. Liu, X. Gao, P. Schimpf, F. Yang, and S. Gao, "A recursive algorithm for the three-dimensional imaging of brain electric activity: Shrinking loreta-focuss," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 10, pp. 1794–1802, 2004.

[320] J. Yao and J. Dewald, "Evaluation of different cortical source localization methods using simulated and experimental eeg data," *Neuroimage*, vol. 25, no. 2, pp. 369–382, 2005.

[321] A. Blenkmann, G. Seifera, J. P. Princich, D. Consalvo, S. Kochen, and C. Muravchikc, "Association between equivalent current dipole source localization and focal cortical dysplasia in epilepsy patients," *Epilepsy Research*, vol. 66, no. 98, pp. 223–231, 2012.

[322] J. C. Mosher and R. M. Leahy, "Recursive music: A framework for eeg and meg source localization," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 11, pp. 960–966, 1998.

[323] X. Xu, B. Xu, and B. He, "An alternative subspace approach to eeg dipole source localization," *Physics in Medicine and Biology*, vol. 49, no. 2, pp. 327–343, 2004.

[324] G. Backus and F. Gilbert, "Uniqueness in the inversion of inaccurate gross earth data," *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 266, no. 1173, pp. 123–192, 1970.

[325] J. Sarvas, "Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem," *Physics in Medicine and Biology*, vol. 32, no. 1, pp. 11–22, 1987.

[326] M. Hamalainen and R. Ilmoniemi, "Interpreting magnetic fields of the brain: Minimum norm estimates," *Medical and Biological Engineering and Computing (MBEC)*, vol. 32, no. 1, pp. 35–42, 1994.

[327] R. G. de Peralta Menendez and S. G. Andino, "Backus and gilbert method for vector fields," *Human Brain Mapping*, vol. 7, no. 3, pp. 161–165, 1999.

[328] R. D. Pascual-Marqui, "Review of methods for solving the eeg inverse problem," *Biolectromagnetism*, vol. 1, no. 1, pp. 75–86, 1999.

[329] K. Uutela, M. Hamalainen, and E. Somersalo, "Visualization of magnetoencephalographic data using minimum current estimates," *Neuroimage*, vol. 10, no. 2, pp. 173–180, 1999.

[330] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with focuss: a recursive weighted minimum norm algorithm," *Electroencephalography and Clinical Neurophysiology*, vol. 95, no. 4, pp. 231–251, 1995.

[331] P. H. Schimpf, H. Liu, C. Ramon, and J. Haueisen, "Efficient electromagnetic source imaging with adaptive standardized loreta/focuss," *Clinical Neurophysiology*, vol. 52, no. 5, pp. 901–908, 2005.

[332] R. Grech, T. Cassar, J. Muscat, K. P. Camilleri, S. G. Fabri, M. Zervakis, P. Xanthopoulos, V. Sakkalis, and B. Vanrumste, "Review on solving the inverse problem in eeg source analysis," *NeuroEngineering and Rehabilitation*, vol. 5, no. 25, pp. 1–33, 2008.

[333] H. S. Liu, F. Yang, X. Gao, and S. Gao, "Shrinking loreta-focuss: A recursive approach to estimating high spatial resolution electrical activity in the brain," *Proceedingsof the 1st International IEEE EMBS Conference on Neural Engineering*, pp. 545–548, 2003.

[334] A. M. Dale and M. I. Sereno, "Improved localization of cortical activity by combining eeg and meg with mri cortical surface reconstruction," *Journal of Cognitive Neuroscience*, vol. 5, no. 2, pp. 162–176, 1993.

[335] D.-H. Li and M. Fukushima, "A modified bfgs method and its global convergence in nonconvex minimization," *Journal of Computational and Applied Mathematics*, vol. 1-2, no. 129, pp. 15–35, 2001.

[336] S. Baillet and L. Garnero, "A bayesian approach to introducing anatomo-functional priors in the eeg/meg inverse problem," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 5, pp. 374–385, 1997.

[337] H. Liu, P. H. Schimpf, G. Dong, X. Gao, F. Yang, and S. Gao, "Standardized shrinking loreta-focuss (sslofo): A new algorithm for spatio-temporal eeg source reconstruction," *Transactions on Biomedical Engineering IEEE*, vol. 52, no. 10, pp. 1681–1691, 2005.

[338] L. Gavit, S. Baillet, J.-F. Mangin, J. Pescatore, and L. Garnero, "A multiresolution framework to meg/eeg source imaging," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 10, pp. 1080–1087, 2001.

[339] A. K. Liu, A. M. Dale, and J. W. Belliveau, "Monte carlo simulation studies of eeg and meg localization accuracy," *Human Brain Mapping*, vol. 16, no. 1, pp. 47–62, 2002.

[340] J. Fruitet and M. Clerc, "Reconstruction of cortical sources activities for online classification of electroencephalographic signals," *Annual International Conference of the IEEEEngineering in Medicine and Biology Society*, pp. 6317–6320, 2010.

[341] H. Wang, Q. Tang, and W. Zheng, "L1-norm-based common spatial patterns," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 3, pp. 653–662, 2012.

[342] M. Kowalski and A. Gramfort, "A priori par normes mixtes pour les problémes inverses, application á la localisation de sources en m/eeg," *Traitement du signal et de l'image*, vol. 37, no. 1, pp. 51–76, 2010.

[343] C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: numerical aspects of linear inversion*. siam, Society for Industrial and Applied Mathematics, 1th ed., 1997.

[344] J. Mattout, J. Daunizeau, M. Pélégrini-Issac, L. Garnero, and H. Benali, "Better conditioning the meg/eeg inverse problem: the multivariate source prelocalization approach," *2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pp. 1361–1364, 2004.

[345] T. Waberski, R. Gobbele, G. Herrendorf, B. Steinhoff, R. Kolle, M. Fuchs, W. Paulus, and H. Buchner, "Source reconstruction of mesial-temporal epileptiform activity: comparison of inverse techniques," *Epilepsia*, vol. 41, no. 12, pp. 1574–1583, 2000.

[346] F.-H. Lin, J. W. Belliveau, A. M. Dale, and M. S. Hamalainen, "Distributed current estimates using cortical orientation constraints," *Human Brain Mapping*, vol. 27, no. 1, pp. 1–13, 2006.

[347] M. Stenroos and O. Hauk, "Minimum-norm cortical source estimation in layered head models is robust against 2skull conductivity error," *Neuroimage*, vol. 81, pp. 265–272, 2013.

[348] A. Gramfort and M. Kowalski, "Improving m/eeg source localizationwith an inter-condition sparse prior," *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, vol. 2011, pp. 141–144, 2009.

[349] S. Baillet, J. Mosher, and R. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, vol. 18, no. 6, pp. 14–30, 2001.

[350] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[351] Y. A. Park, "Lecture notes for cse 291," *Top/Computer Sci and Engineering with Saul at UC San Diego (UCSD)*, 2007.

[352] B. Jeffs, R. Leahy, and M. Singh, "An evaluation of methods for neuromagnetic image reconstruction," *IEEE Transactions on Biomedical Engineering*, vol. 34, no. 9, pp. 713–723, 1987.

[353] S. Haufe, V. Nikulin, A. Ziehe, K. Muller, and G. Nolte, "Combining sparsity and rotational invariance in eeg/meg source reconstruction," *Neuroimage*, vol. 42, no. 2, pp. 726–738, 2008.

[354] J. Phillips, R. Leahy, and J. Mosher, "Meg-based imaging of focal neuronal current sources," *IEEE Transactions on Medical Imaging*, vol. 16, no. 3, pp. 338–348, 1997.

[355] C. Phillips, M. D. Rugg, and K. J. Friston, "Systematic regularization of linear inverse solutions of the eeg source localization problem," *Neuroimage*, vol. 17, no. 1, pp. 287–301, 2002.

[356] R. Henson, J. Mattout, K. Singh, G. Barnes, A. Hillebrand, and K. Friston, "Population-level inferences for distributed meg source localization under multiple constraints application to face-evoked fields," *Neuroimage*, vol. 38, no. 3, pp. 422–438, 2007.

[357] S. Haufe, R. Tomioka, G. Nolte, K. Muller, and M. Kawanabe, "Modeling sparse connectivity between underlying brain sources for eeg meg," *Neuroimage*, vol. 57, no. 8, pp. 1954–1963, 2010.

[358] A. Gramfort, M. Kowalski, and M. Hamalainen, "Mixed-norm estimates for the m/eeg inverse problem using accelerated gradient methods," *Physics in Medicine and Biology*, vol. 57, no. 7, pp. 1937–1961, 2012.

[359] S. Finke, R. M. Gulrajani, and J. Gotman, "Conventional and reciprocal approaches to the inverse dipole localization problem of electroencephalography," *Nature Reviews Neuroscience*, vol. 50, no. 6, pp. 657–666, 2003.

[360] S. Haufe, T. Dickhaus, C. Sanelli, B. Blankertz, G. Nolte, and K. Muller, "Large-scale eeg meg source localization with spatial flexibility," *Neuroimage*, vol. 54, no. 2, pp. 851–859, 2011.

[361] Y. Li, "A globally convergent method for lp problems," *Computer Science Technical Reports Cornell University*, 1991.

[362] J. Taylor, A. Ioannides, and H. Muller-Gartner, "Mathematical analysis of lead field expansions," *IEEE Transactions on Medical Imaging*, vol. 18, no. 2, pp. 151–163, 1999.

[363] Q. Wu, L. Wang, Y. Li, S. He, and W. Yan, "A computer simulation system of the electric activity in brains," *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 990–993, 2003.

[364] L. Ding, "A novel sparse source imaging in reconstructing extended cortical current sources," *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4555–4558, 2008.

[365] D. M. Schmidt, J. S. George, and C. Wood, "Bayesian inference applied to the electromagnetic inverse problem," *Human Brain Mapping*, vol. 7, no. 3, pp. 195–212, 1997.

[366] D. Wipf and S. Nagarajan, "A unified bayesian framework for meg/eeg source imaging," *Neuroimage*, vol. 44, no. 3, pp. 947–966, 2009.

[367] E. Candes and T. Tao, "The dantzig selector: Statistical estimation when p is much larger than n," *Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2005.

[368] P. Valdes-Sosa, M. Vega-Hernández, J. Sánchez-Bornot, E. Martínez-Montes, and M. Bobes, "Eeg source imaging with spatio-temporal tomographic nonnegative independent component analysis," *Human Brain Mapping*, vol. 30, no. 6, pp. 1898–1910, 2009.

[369] C. Phillips, J. Mattout, M. D. Rugg, P. Maquet, and K. J. Friston, "An empirical bayesian solution to the source reconstruction problem in eeg," *Neuroimage*, vol. 24, no. 4, pp. 997–1011, 2005.

[370] K. Friston, L. Harrison, J. Daunizeau, S. Kiebel, C. Phillips, N. T. Barreto, R. Henson, G. Flandin, and J. Mattout, "Multiple sparse priors for the m/eeg inverse problem," *Neuroimage*, vol. 39, no. 3, pp. 1104–1120, 2008.

[371] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[372] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[373] M. Osborne, B. Presnell, and B. Turlach, "A new approach to variable selection in least squares problems," *Numerical Analysis (IMA)*, vol. 20, pp. 389–404, 2000.

[374] Y. Chen and A. O. Hero, "Recursive l1-inf group lasso," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 3978–3987, 2012.

[375] M. Schmidt, "Least squares optimization with l1-norm regularization," *CS542B Project Report, UBC Department of Computer Science*, 2005.

[376] J. Friedman, T. Hastie, H. Hofling, and R. Tibshirani, "Pathwise coordinate optimization," *The Annals of Applied Statistics*, vol. 1, no. 2, pp. 302–332, 2007.

[377] S. Sardy, A. G. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparamatric signal denoising," *Computational and Graphical Statistics*, vol. 9, no. 2, pp. 36–379, 2000.

[378] J. Montoya-Martínez, A. Artés-Rodríguez, L. K. Hansen, and M. Pontil, "Structured sparsity regularization approach to the eeg inverse problem," *3rd International Workshop on Cognitive Information Processing (CIP)*, pp. 1–6, 2012.

[379] W. Kincses, C. Braun, S. Kaiser, W. Grodd, H. Ackermann, and K. Mathiak, "Reconstruction of extended cortical sources for eeg and meg based on a monte carlo markov chain estimator," *Human Brain Mapping*, vol. 18, no. 2, pp. 100–110, 2003.

[380] J.Yao and B. He, "A self-coherence enhancement algorithm and its application to enhancing 3d source estimation from eegs," *Annals of Biomedical Engineering*, vol. 29, no. 11, pp. 1019–27, 2001.

[381] P. Bofill, P. B. L. B, and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Neuroimage*, vol. 81, no. 11, 2001.

[382] T. Auranen, A. Nummenmaa, M. Hamalainen, I. Jaaskelainen, J. Lampinen, A. Vehtari, and M. Sams, "Bayesian analysis of the neuromagnetic inverse problem with l(p)-norm priors," *Neuroimage*, vol. 1, no. 26, pp. 870–884, 2005.