Application of informational measures combined with kernel density estimator to describe epileptic seizures via Scalp EEG

PhD Candidate: Antonio Squicciarini

9 June 2023

Email: a.squicciarini@alumnos.upm.es
University: Universidad Politécnica de Madrid - UPM ETSII
Department: Departamento de Matemática Aplicada a la Ingeniería Industrial DMAII
Research Group: Teoría de Aproximación Constructiva y Aplicaciones GI TACA
IMEIO program - Ingeniería Matemática, Estadística e Investigación Operativa

Teams that contributed to the work

Prof. Elio Valero Toranzo[a], Prof. Alejandro Zarzo Altarejos[b], Prof. Carlos E. González Guillén[b]

a)Departamento de Matemática Aplicada, Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, España, (elio.vtoranzo@urjc.es)

b)GI-TACA, Departamento de Matemática Aplicada a la Ingeniería Industrial, Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, España, (alejandro.zarzo@upm.es)

Outline

Outline

Seizure problem and Scalp EEG

Seizure Definition

An epileptic seizure is defined as a transient occurrence of signs and/or symptoms caused by abnormal excessive or synchronous neuronal activity in the brain [5]

- Convulsions
- Loss of consciousness
- Muscle stiffness or rigidity

Scalp EEG (Electroencephalography)

Record of electrical brain activity using electrodes placed on the scalp. It is a **non-invasive** technique. It is a multichannel signal.

Seizure problem and Scalp EEG

Seizure Definition

An epileptic seizure is defined as a transient occurrence of signs and/or symptoms caused by abnormal excessive or synchronous neuronal activity in the brain [5]

- Convulsions
- Loss of consciousness
- Muscle stiffness or rigidity

Scalp EEG (Electroencephalography)

Record of electrical brain activity using electrodes placed on the scalp. It is a **non-invasive** technique. It is a multichannel signal.

Seizure problem and Scalp EEG

Seizure Definition

An epileptic seizure is defined as a transient occurrence of signs and/or symptoms caused by abnormal excessive or synchronous neuronal activity in the brain [5]

- Convulsions
- Loss of consciousness
- Muscle stiffness or rigidity

Approximately 50 million people worldwide are affected by epilepsy

Scalp EEG (Electroencephalography)

Record of electrical brain activity using electrodes placed on the scalp. It is a **non-invasive** technique. It is a multichannel signal.

Seizure problem and Scalp EEG

Seizure Definition

An epileptic seizure is defined as a transient occurrence of signs and/or symptoms caused by abnormal excessive or synchronous neuronal activity in the brain [5]

- Convulsions
- Loss of consciousness
- Muscle stiffness or rigidity

Approximately 50 million people worldwide are affected by epilepsy

Scalp EEG (Electroencephalography)

Record of electrical brain activity using electrodes placed on the scalp. It is a **non-invasive** technique. It is a multichannel signal.

The scope of our work

State of Art gap

- Manual detection of electrographic seizures is a time-consuming process that requires experts for data interpretation [3]
- Machine-learning algorithms lack explainability [8, 9]

Main scope of the study

Investigate whether combining differential information measures (IMs) can detect and describe distinct trends during ictal phases in order to:

- Distinguish the region of attack from the region of non-attack
- Identifying the onset of the epileptic seizure
- Compare performances of the differents IMs

The scope of our work

State of Art gap

- Manual detection of electrographic seizures is a time-consuming process that requires experts for data interpretation [3]
- Machine-learning algorithms lack explainability [8, 9]

Main scope of the study

Investigate whether combining differential information measures (IMs) can detect and describe distinct trends during ictal phases in order to:

- Distinguish the region of attack from the region of non-attack
- Identifying the onset of the epileptic seizure
- Compare performances of the differents IMs

Methodology

Methodology Flowchart

1. Transform signal in a sequence of ordered probability density functions (PDFs)
   1. Splitting the signal into an ordered sequence of windows
   2. Inference succession of PDFs
2. Compute a set of information measures (IMs)
3. Find the most informative channel according to each IM
4. Analysis results with IM time plots and Information Planes

Outline

Information Measures general definition

X be a continuous random variable
Probability density function (PDF) $f_X : \Lambda \longrightarrow \mathbb{R}^+ \cup \{0\}$ such that $\int_\Lambda f_X(x)\, dx = 1$

Information Measure (IM) [13]

Functional to quantify the amount of information conveyed by a random variable X

$$IM[f_X(x)] = -\int_\Lambda \phi(f_X(x))\, dx \qquad (1)$$

where ϕ is the so-called entropic functional

Entropy Measures

Shannon Entropy

$$S[f_X] = -\int_\Lambda f_X(x)\log[f_X(x)]\,dx \tag{2}$$

Tsallis Entropy

$$T_q[f_X] = \frac{1}{q-1}\left(1 - \int_\Lambda [f_X(x)]^q\,dx\right), \qquad q \in \mathbb{R} \tag{3}$$

Rényi Entropy

$$R_q[f_X] = \frac{1}{1-q}\log\left(\int_\Lambda [f_X(x)]^q\,dx\right), \qquad q \in \mathbb{R} \tag{4}$$

Entropy Proprieties

- Both Rényi and Tsallis entropies reduce to the Shannon entropy [11, 2] in the limit $q \to 1$:

$$\lim_{q\to 1} R_q[f_X] = S[f_X] \quad \text{and} \quad \lim_{q\to 1} T_q[f_X] = S[f_X],$$

- Shannon and Rényi entropies are **extensive measures** [11]. For iid random variables X and Y, they fulfil the additivity law:

$$S[f_{XY}] = S[f_X] + S[f_Y] \quad \text{and} \quad R_q[f_{XY}] = R_q[f_X] + R_q[f_Y],$$

- Tsallis entropy [2] is non-extensive, and fulfils the pseudo-additivity law:

$$T_q[f_{XY}] = T_q[f_X] + T_q[f_Y] + (1-q)T_q[f_X]T_q[f_Y],$$

- Rényi and Tsallis entropies are strictly decreasing with respect to the order parameter, q, on $(0,\infty)$, i.e.,

$$T_{q_1}[f_X] \geq T_{q_2}[f_X] \quad \text{and} \quad R_{q_1}[f_X] \geq R_{q_2}[f_X], \quad q_1 < q_2 \quad \text{with} \quad q_1, q_2 > 0.$$

Power Entropy

In order to simplify the interpretation of results, it is customary to compute the Shannon, that is a strictly monotonically increasing transformation defined as

Shannon Entropy Power

$$N_S[f_X] = \frac{1}{2\pi e} e^{2S[f_X]} \tag{5}$$

Réyni Entropy Power

$$N_R[f_X] = \frac{1}{2\pi e} e^{2R_q[f_X]}, \tag{6}$$

With $N_S[f] \geq 0$ and $N_R[f_X] \geq 0$, $\forall\, x \in \Lambda$
In case of Gaussian PDF: $N_S[\mathcal{N}(X = x|\mu, \sigma)] = \sigma^2$.

Fisher Information I

The complementary information measure considered is the differential Fisher information, introduced by Fisher in 1925 [4] in the context of statistical estimation:

Parametric Fisher Information [4]

$$F[f_{X,\theta}] = -\int_{\Lambda} \left[\frac{\partial^2}{\partial \theta^2} \log f(x|\theta) \right] f(x|\theta)\, dx = -\mathbb{E}_\theta \left[l''(x|\theta) \right], \quad (7)$$

with $f(x|\theta)$ the PDF associated to a X, θ is an unknown parameter and $l(x|\theta) = \log f(x|\theta)$ the log-likelihood function. Specifically, Fisher information quantifies the information carried by the random variable about the parameter θ.

non-parametric Fisher Information

$$F[f_X] = \int_{\Lambda} \frac{\left(\frac{d}{dx} f_X(x)\right)^2}{f_X(x)} dx = \mathbb{E}\left[\left(\frac{\partial}{\partial x} \log f_X(x)\right)^2\right], \tag{8}$$

When θ is a location parameter [4]

where we assume that $f_X(x)$ is differentiable and both $f_X(x)$ and $\frac{d}{dx} f_X(x)$, are quadratically integrable on $\mathbb{R}$[7].

Fisher information is a non-negative functional which quantifies the average of the proportional change of $f_X(x)$ per unit change in x. Hence, Fisher information is able to detect the degree of oscillatory character of a given PDF

For the Gaussian PDF $F[\mathcal{N}(X = x|\mu, \sigma)] = 1/\sigma^2$, showing how this information measure is inversely proportional to the variance of the PDF.

Fisher Shannon Complexity

$$C_{SF}[f_X] = N_S[f_X]F[f_X]. \tag{9}$$

One of its properties is the **so-called** isoperimetric inequality [4, 6] ($C_{SF}[f_X] \geq 1$), **where the** equality stands if and only if the distribution is a Gaussian.

Outline

Signal Structure

Scalp EEG

$$x^{(c)}(t) : [t_0, T] \subset \mathbb{R} \to \mathbb{R} \xrightarrow{\text{discretisation}} \{x^{(c)}\}_{i=1}^{N} = \{x_i^{(c)}(t_i),\ i = 1, \cdots, N,\ x_i^{(c)} \in \mathbb{R}\}.$$

- c channels, each one formed by N equispaced samples
- Time reference is $t_i = t_0 + \frac{i}{f_s}$
  - f_s the sampling frequency
  - t_0 the initial instant of the recording

Sliding Temporal Window

Overlapping Windows definition

Sliding temporal window: $W_j^{(c)}(\delta, \Delta) = \{x_i^{(c)},\ i = 1 + j\delta, \cdots, \Delta + j\delta\}$

- $\Delta \in \mathbb{N}$, $\Delta \leqslant N$ is the window length
- $\delta \in \mathbb{N}$, $\delta \leqslant \Delta$ is the sliding factor
- The subscript, $j \in \left[0, \frac{N-\Delta}{\delta}\right] \cap \mathbb{N}$, refers to the time order of the windows
- Window time reference $t_j = t_0 + \frac{\Delta + j\delta}{f_s}$
- Each window is individually normalised $\tilde{x}_i^{(c)} \in [0, 1] \subset \mathbb{R}$.

Window Labelling

$y_j = 1$ if $t_b < t_j < t_f + \frac{\Delta}{f_s}$

- The timestamps t_b and t_e indicate the beginning and end of the epileptic crisis

Sliding Temporal Window

Overlapping Windows definition

Sliding temporal window: $W_j^{(c)}(\delta, \Delta) = \{x_i^{(c)},\ i = 1 + j\delta, \cdots, \Delta + j\delta\}$

- $\Delta \in \mathbb{N}$, $\Delta \leqslant N$ is the window length
- $\delta \in \mathbb{N}$, $\delta \leqslant \Delta$ is the sliding factor
- The subscript, $j \in \left[0, \frac{N-\Delta}{\delta}\right] \cap \mathbb{N}$, refers to the time order of the windows
- Window time reference $t_j = t_0 + \frac{\Delta + j\delta}{f_s}$
- Each window is individually normalised $\tilde{x}_i^{(c)} \in [0, 1] \subset \mathbb{R}$.

Window Labelling

$y_j = 1$ if $t_b < t_j < t_f + \frac{\Delta}{f_s}$

- The timestamps t_b and t_e indicate the beginning and end of the epileptic crisis

Sliding Temporal Window

Overlapping Windows definition

Sliding temporal window: $W_j^{(c)}(\delta, \Delta) = \{x_i^{(c)},\ i = 1 + j\delta, \cdots, \Delta + j\delta\}$

- $\Delta \in \mathbb{N}$, $\Delta \leqslant N$ is the window length
- $\delta \in \mathbb{N}$, $\delta \leqslant \Delta$ is the sliding factor
- The subscript, $j \in \left[0, \dfrac{N-\Delta}{\delta}\right] \cap \mathbb{N}$, refers to the time order of the windows
- Window time reference $t_j = t_0 + \frac{\Delta + j\delta}{f_s}$
- Each window is individually normalised $\tilde{x}_i^{(c)} \in [0, 1] \subset \mathbb{R}$.

Window Labelling

$y_j = 1$ if $t_b < t_j < t_f + \frac{\Delta}{f_s}$

- The timestamps t_b and t_e indicate the beginning and end of the epileptic crisis

Sliding Temporal Window

Overlapping Windows definition

Sliding temporal window: $W_j^{(c)}(\delta, \Delta) = \{x_i^{(c)},\ i = 1 + j\delta, \cdots, \Delta + j\delta\}$

- $\Delta \in \mathbb{N}$, $\Delta \leqslant N$ is the window length
- $\delta \in \mathbb{N}$, $\delta \leqslant \Delta$ is the sliding factor
- The subscript, $j \in \left[0, \dfrac{N - \Delta}{\delta}\right] \cap \mathbb{N}$, refers to the time order of the windows
- Window time reference $t_j = t_0 + \frac{\Delta + j\delta}{f_s}$
- Each window is individually normalised $\tilde{x}_i^{(c)} \in [0, 1] \subset \mathbb{R}$.

Window Labelling

$y_j = 1$ if $t_b < t_j < t_f + \frac{\Delta}{f_s}$

- The timestamps t_b and t_e indicate the beginning and end of the epileptic crisis

KDE

A non-parametric inference solution:

$$\hat{f}_j^{(c)}(x) = \mathcal{K}_h[W_j^{(c)}(\delta, \Delta)] = \frac{1}{h\Delta} \sum_{i=\delta \cdot j}^{\delta \cdot j + \Delta} K\left(\frac{x - x_i}{h}\right), \tag{10}$$

- $K(x)$ is the kernel
- h is the kernel's bandwidth

Kernel density estimation (KDE) Introduction

KDE

A non-parametric inference solution:

$$\hat{f}_j^{(c)}(x) = \mathcal{K}_h[W_j^{(c)}(\delta, \Delta)] = \frac{1}{h\Delta} \sum_{i=\delta \cdot j}^{\delta \cdot j + \Delta} K\left(\frac{x - x_i}{h}\right), \tag{10}$$

- $K(x)$ is the kernel
- h is the kernel's bandwidth

K is an even regular function, with unit variance and zero mean. In our case, we use the Gaussian Kernel.

Bandwidth h optimization

For the optimization, we got inspired by [12, 10, 14].
The goal is to minimize the Mean Integrated Square Error (MISE)

MISE

$$\text{MISE} = \mathbb{E}\left[\int_{-\infty}^{+\infty} \left| f(x) - \hat{f}_j(x) \right|^2 dx\right], \tag{11}$$

The optimal bandwidth $h^* = N^{-1/5} \cdot (J(h))^{-1/5} \cdot (M(K))^{1/5}$ Where:

- $J(h) = \int_{-\infty}^{+\infty} (f''(x))^2 \, dx$
- $M(K) = \int_{-\infty}^{+\infty} K^2(u) \, du$
- f'' the second derivative of f

Bandwidth h optimization

For the optimization, we got inspired by [12, 10, 14].
The goal is to minimize the Mean Integrated Square Error (MISE)

MISE

$$\text{MISE} = \mathbb{E}\left[\int_{-\infty}^{+\infty} \left| f(x) - \hat{f}_j(x) \right|^2 dx\right], \tag{11}$$

The optimal bandwidth $h^* = N^{-1/5} \cdot (J(h))^{-1/5} \cdot (M(K))^{1/5}$ Where:

- $J(h) = \int_{-\infty}^{+\infty} (f''(x))^2 \, dx$
- $M(K) = \int_{-\infty}^{+\infty} K^2(u) \, du$
- f'' the second derivative of f

Bandwidth h optimization: Plug-in method

Plug-in Algorithm

1. Compute $M(K)$ based on the kernel selected.
2. Arbitrary initialization of $J^{(0)}(f)$.
3. With the first value of J compute $h^{*(0)}$ and then $f^{(0)}$.
4. Verify if the stop criteria $\left|\dfrac{h^{*(n)} - h^{*(n-1)}}{h^{*(n-1)}}\right| < \epsilon$ is **fulfilled**.
5. Compute $J^{(n)}(f)$, $h^{*(n)}$ and $f^{(n)}$ until the stop criteria of point 4. is **fulfilled**.

Our adaptation to our case

1. A single optimal h to each channel to increase distinguishability.
2. h optimal computed averaging the results over n_h equally spaced non-ictal windows.

Outline

Children's Hospital Boston (CHB-MIT) Scalp EEG database

CHB-MIT Dataset composition

- 24 cases, 22 pediatric subjects with intractable seizures (5 males, ages 3–22; and 17 females, ages 1.5–19)
- $f_z = 256[H_z]$ with a 16-bit resolution signal
- EEG electrode positions follow the international 10-20 system [1]
- 664 records of one or 4 hours, of which 129 are affected by one or more seizures

The non-seizure signal used in the analysis consists of a duration of 3 minutes both before and after the epileptic crisis.

Main Paramerers

Δ	$32[s] = 2E-13\ points$
δ	$2[s] = 2E-9\ points$
q	0.7, 1.5, 2, 3, and 4
n_h	3

Table: Values of the main parameters associated with the sliding temporal window technique (Δ, δ) and the generalized information measures/products (q)

IMs Summary

$$IM_j^{(c)} = IM[\hat{f}_j^{(c)}(x)] \quad (12)$$

Name	Notation	Formula	Parameter	Character	Range
Shannon entropy	$S[f_X]$	(2)	No	Global	$\mathbb{R}$
Shannon entropy power	$N_S[f_X]$	(5)	No	Global	$\mathbb{R}_+$
Rényi entropy	$R_q[f_X]$	(4)	q	Global	$\mathbb{R}$
Rényi entropy power	$N_R[f_X]$	(6)	q	Global	$\mathbb{R}_+$
Tsallis entropy	$T_q[f_X]$	(3)	q	Global	$\mathbb{R}$
Fisher information	$F[f_X]$	(8)	No	Local	$\mathbb{R}_+$
Fisher-Shannon complexity	$C_{SF}[f_X]$	(8)	No	Local/Global	$[1, +\infty) \subset \mathbb{R}_+$

Table: Summary of the information measures

Channel Selection

To identify the most significant channel that would accelerate the analysis of results

Maximum IM Channel

$$S_{IM}^{(c)} = \frac{max(\{|IM_j^{(c)}|\}_{y_j=1})}{max(\{|IM_j^{(c)}|\}_{y_j=0})}. \tag{13}$$

The best channel is selected for each information measure based on the correspondent maximum score $c_{IM}^{*} = argmax(S_{IM}^{(c)})_{c=1}^{C}$, **being** C the total number of channels.

PDF Seizure evolution

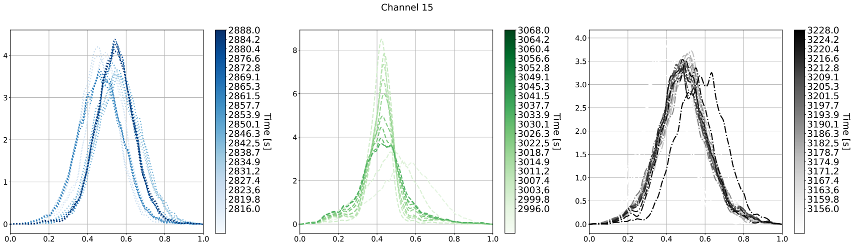
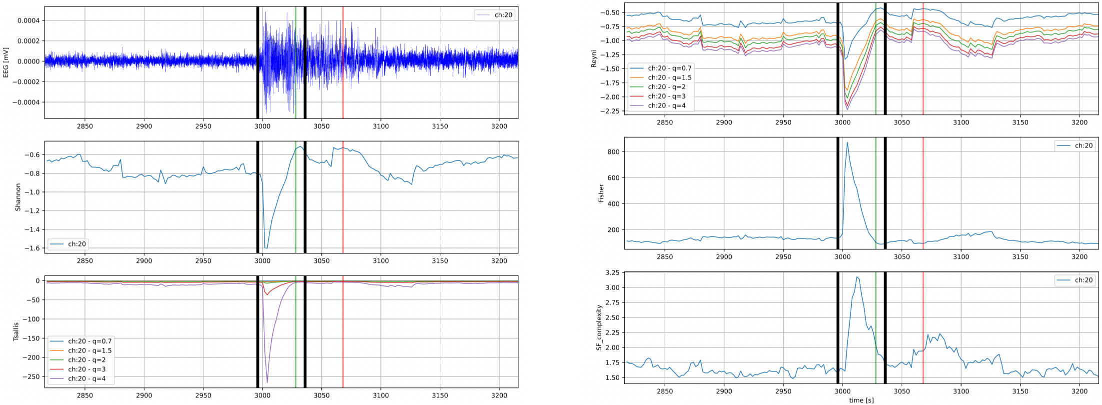


Figure: The images show PDF time evolution generated by three sliding windows located in different parts of the EEG record. A) pre-seizure, B) seizure, C) post-seizure. The time shift is kept constant and the time reference is indicated on the lateral bar. The colour gradient indicates the time sequence of the probability distribution

IM time-dependent plots: chb01 - record 4



The black lines limit the seizure period. The green one indicates when the window is completely inside the epileptic range, and the red one is when it is completely outside.

Information plane: chb01 - record 4

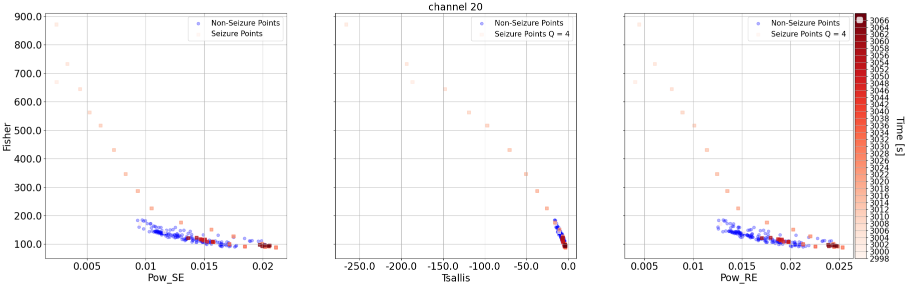
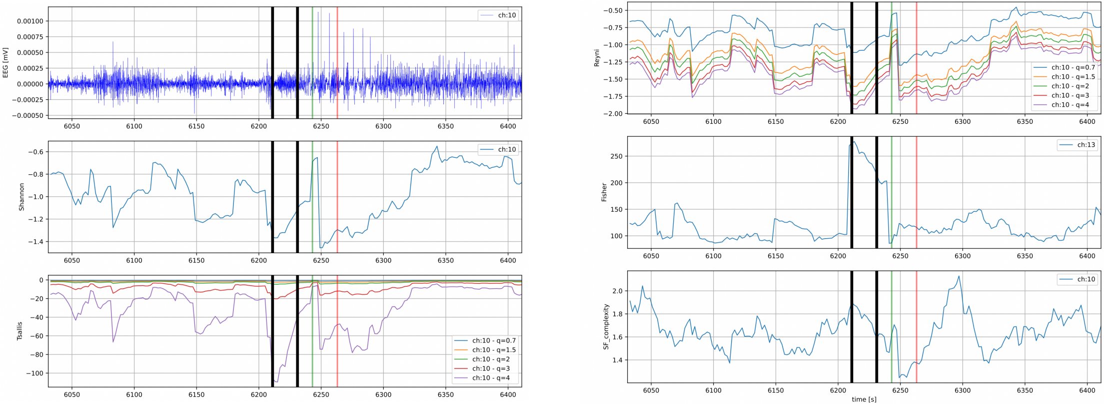


Figure: Information planes associated with IMs. The red-colour gradient indicates the time evolution of the probability distribution inside the ictal phase. The record is chb01 - record 4 - channel 15

IM time-dependent plots: chb 06 - record 4 - seizure 2



In this example, only Fisher information and Tsallis entropy successfully detect the seizure

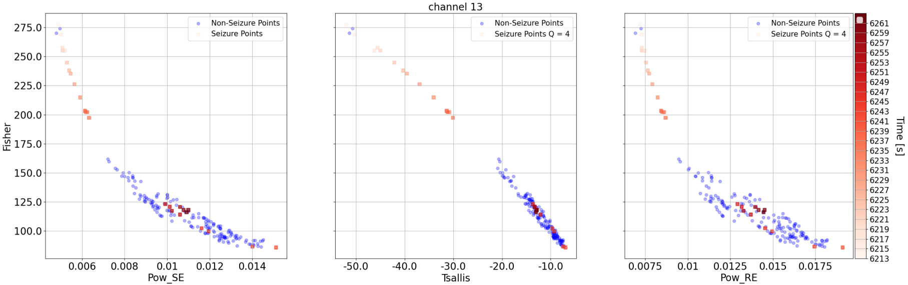
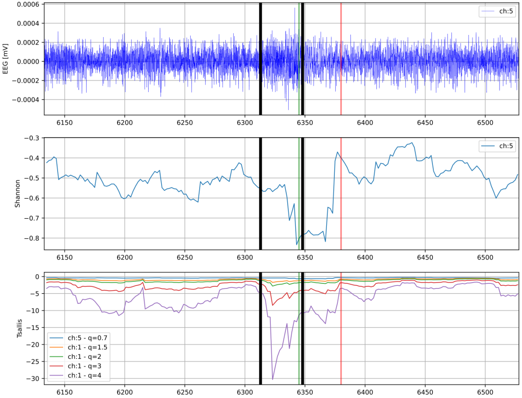
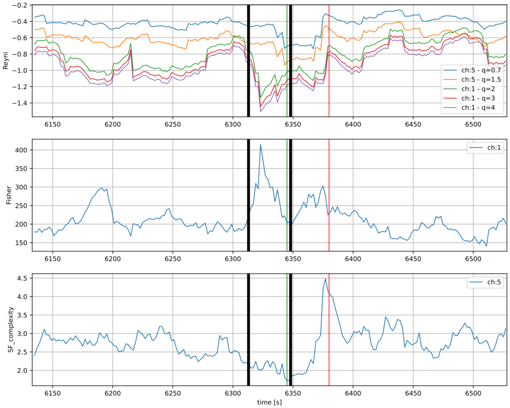


Figure: Information planes associated with IMs. The red-colour gradient indicates the time evolution of the probability distribution inside the ictal phase. The record is chb06 - record 4 - seizure 2

IM time-dependent plots: chb10 - record 12 - seizure 1





Information plane: chb10 - record 12 - seizure 1

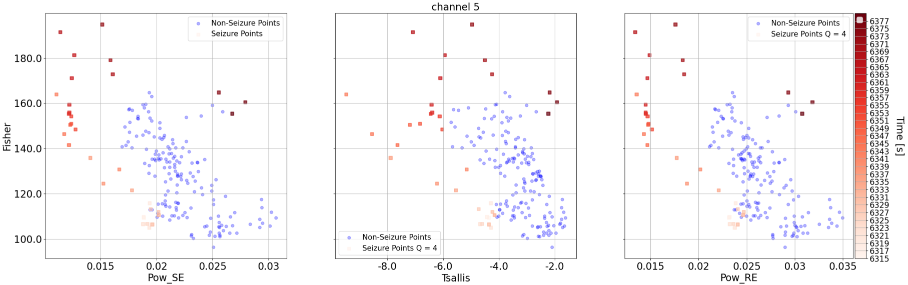


Figure: Information planes associated with IMs. The red-colour gradient indicates the time evolution of the probability distribution inside the ictal phase. The record is ch010 - record 12 - seizure 1

Thank you!

The work was made possible thanks to the Programa Propio de la Universidad Politécnica de Madrid UPM

UNIVERSIDAD POLITÉCNICA DE MADRID

POLITÉCNICA

INDUSTRIALES ETSII | UPM

DMAII

Santander Universidades

Outline

References I

13. Electroencephalography.
http://www.bem.fi/book/13/13.htm#03.

Possible generalization of boltzmann-gibbs statistics.

Poomipat Boonyakitanont, Apiwat Lek-uthai, Krisnachai Chomtho, and Jitkomut Songsiri.
A review of feature extraction and performance evaluation in epileptic seizure detection using EEG.
Biomedical Signal Processing and Control, 57:101702, March 2020.

Thomas M Cover and Joy A Thomas.
ELEMENTS OF INFORMATION THEORY.
page 774, 2006.

References II

Robert S. Fisher, Walter van Emde Boas, Warren Blume, Christian Elger, Pierre Genton, Phillip Lee, and Jerome Engel.
Epileptic seizures and epilepsy: Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE).
Epilepsia, 46(4):470–472, April 2005.

Fabian Guignard, Mohamed Laib, Federico Amato, and Mikhail Kanevski.
Advanced Analysis of Temporal Data Using Fisher-Shannon Information: Theoretical Development and Application in Geosciences.
Frontiers in Earth Science, 8:255, July 2020.

Bercher J.-F.
On escort distributions, q-gaussians and fisher information.
In *AIP Conference Proceedings*, volume 1305, March 2011.

References III

Qiong Li, Jianbo Gao, Qi Huang, Yuan Wu, and Bo Xu.
Distinguishing Epileptiform Discharges From Normal Electroencephalograms Using Scale-Dependent Lyapunov Exponent.
Frontiers in Bioengineering and Biotechnology, 8, 2020.

Khansa Rasheed, Adnan Qayyum, Junaid Qadir, Shobi Sivathamboo, Patrick Kwan, Levin Kuhlmann, Terence O'Brien, and Adeel Razi.
Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review.
IEEE Reviews in Biomedical Engineering, 14:139–155, 2021.

Vikas Chandrakant Raykar and Ramani Duraiswami.
Fast optimal bandwidth selection for kernel density estimation.
pages 524–528, April 2006.

References IV

Alfréd Rényi.
On measures of entropy and information.
In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, volume 4, pages 547–562. University of California Press, 1961.

Molka Troudi, Adel M. Alimi, and Samir Saoudi.
Analytical Plug-In Method for Kernel Density Estimator Applied to Genetic Neutrality Study.
EURASIP Journal on Advances in Signal Processing, 2008(1):739082, December 2008.

References V

Irene Valero-Toranzo, Steeve Zozor, and Jean-Marc Brossier.
Generalization of the de Bruijn Identity to General ϕ -Entropies and ϕ -Fisher Informations.
IEEE Transactions on Information Theory, 64(10):6743–6758, October 2018.

Adriano Z. Zambom and Ronaldo Dias.
A Review of Kernel Density Estimation with Applications to Econometrics.
International Econometric Review, 5(1):20–42, April 2013.