BACHELOR' S THESIS / BIOMEDICAL ENGINEERING 2019

A comparative study of different nonlinear methods for localising seizure onset zone in epilepsy patients from intracranial electroencephalographic recordings

Cristian Galán Augé

upf. Universitat Pompeu Fabra Barcelona

A comparative study of different nonlinear methods for localising seizure onset zone in epilepsy patients from intracranial electroencephalographic recordings.

Cristian Galán Augé

Bachelor Thesis UPF / 2019

Thesis Supervisor

Professor Ralph Gregor Andrzejak

Department: Department of Information and Communication Technologies

Universitat Pompeu Fabra
Barcelona

Acknowledgement

I would like to express my sincere gratitude to my supervisor Ralph G. Andrzejak for his patient and his predisposition for helping me when I needed it. I always found the door of his office opened whenever I had a doubt. Thanks to him I have been able to learn a lot in the field of signal analysis and how to face the problems I found during the development of the thesis. Thanks to the PhD student Cristina González who provided me a very valuable feedback since I began to write the thesis. Without her and her feedback, this thesis would have not be possible. Thanks also to Anaïs Espinoso who also provided me with feedback about my thesis from the very beginning. She always cheered me up with her comments and her positive side of things. I am sure that you will have a brilliant future. Thanks also to Dra. Irene Malvestio for spending a part of her time in explain me about predictability score. Her help has also been key. I would like to thank the PhD students Giulia Ruzzene and Angela Ochugboju for their feedback and attention during the Journal Clubs and Writing Clubs we made in the General Meetings.

Finally, I would like to thank to my friends and colleagues of degree Marc Vargas, Ariadna Serrano, Èlia Lleal and Yu Weng for their company and the mutual support we have been providing to ourselves during the last weeks of the thesis.

Abstract

Most of the epilepsy cases are controlled using medication. Unfortunately, medication sometimes does not provide a sufficient control of seizures. For these cases, epilepsy surgery can be an option which needs the detection of the zone where seizures are generated. This involves measuring the electrical activity with electroencephalography (EEG). Intracranial EEG (iEEG) recordings allows a long-term monitoring and it can provide detailed information about the seizure onset zone (SOZ) in epilepsy patients. However, not all characteristics of the EEG can be captured by visual inspection, it may be time demanding and it has user variability. Thus, some analytic methods are used. In particular, in this study we apply and compare four measures in three different iEEG databases from epilepsy patients. These databases contains EEG recordings in the hemisphere where SOZ was located and in the contralateral hemisphere. Two of these methods test determinism in point processes by evaluating predictability: nonlinear prediction error (E) and rank-based nonlinear predictability score (S). The other two are symbolic methods: they analyse qualitative information rather than quantitative. These symbolic methods measure the occurrences of ordinal patterns, which are based on the ranking of a group of samples which are sorted by their amplitude. We found that the use of the predictability measures offers a better differentiation between focal and nonfocal signals than using the symbolic methods.

Keywords

Electroencephalography; signal analysis; symbolic methods; ordinal patterns; permutation entropy; predictability.

Index

List of figures

List of tables

1. Introduction

1.1 Motivation

1.1.1 Epilepsy and its treatments

Epilepsy is one of the most common diseases of the central nervous system. Globally, it affects more than 50 million people [1]. This disease causes seizures: a transient symptom caused by "an abnormal enhanced synchrony of neurons" [2]. They provoke different manifestations depending on many factors such as the onset location or the propagation of the seizure [2].

One way to classify seizures is depending on their focus. Seizures that begin in one or more areas from the same hemisphere (focal hemisphere) are named focal seizures [3, 4], while seizures where both hemispheres are affected from the beginning are named generalised seizures [4]. We name as seizure onset zone (SOZ) the cortex areas where seizures are initiated. Around 70% of patients become seizure-free using medication [5] while in the remaining 30% medication does not provide a sufficient seizure control to the patients. They suffer from pharmacoresistant epilepsy. Most of these 30% of patients have focal seizures [6, 7, 8] and, in some of them, resective surgery can be a potential therapy when the cortex area that generates the SOZ can be completely resected or disconnected [9]. To do this surgery it is not only necessary to fully determine the SOZ but also to guarantee that functional areas are not going to be damaged or, at least, to be able to predict this functional loss [9]. As part of the pre-surgical planning to localise the SOZ, non-invasive techniques such as imaging and electroencephalography are needed. In this thesis we focus on electroencephalography which is crucial for the determination of the different electrical activity that happens before, during and after the seizure. They are named preictal, ictal and postictal activity, respectively [10, 11]. They all together are named *periictal activity* [10]. Finally, interictal activity are the findings between seizures which are shaped by *interictal epileptiform discharges*.

Sometimes the use of non-invasive scalp electroencephalogram (EEG) is not enough

because recordings are too noisy, they have lower resolution than what is needed to correctly define the SOZ or they contain artefacts because of the distance to the cortex or the number of barriers between the electrode and the brain [9]. Some other times they simply are not useful because the affected area is not reachable with this type of EEG. In these cases, intracranial EEG (iEEG) is a useful acquisition technique. The iEEG is recorded with electrodes placed directly on the cortical surface from a certain area of the brain. The direct contact with the cortical surface makes the obtained signals less noisy and of higher resolution than the ones obtained using scalp EEG [9]. Physicians use different electrodes depending on the lobe or structure they need to record: electrodes can go directly to the brain surface (subdural electrodes) or into inner structures such as the hippocampus or the amygdala which are located in the medial temporal lobe. Then, also depending on the area that physicians want to cover, they can use different types of electrodes such as laminar electrodes, which are set in different and distant cortical layers. If they want to precisely cover a small area there are electrodes composed by micro-contacts rather than only macro-contacts [12].

1.1.2 Ordinal patterns as a quantitative technique in EEG

While some features of an EEG from an epilepsy patient can be directly discerned by an expert physician, there are others that need to be quantified. For this reason, analysis of signals with linear and non-linear techniques are performed with the aim of complementing the physicians' diagnosis. In this thesis we focus on ordinal patterns. As it is explained in *Schindler et al*, "ordinal patterns are derived from the rank order of short sequences of consecutive EEG values" [13]. The analysis of ordinal patterns in a signal is what is named a *symbolic method* because it analyses qualitative information whose symbols are the ordinal patterns[14]. Concretely, ordinal patterns are defined as a set of samples from the signal sorted according to their amplitudes. The position of the samples in this sorted list corresponds to the rank of this sample with regard to its amplitude. The temporal sequence of these ranks is referred to as an ordinal pattern [15]. As epileptic seizures are characterised by a non-random process [2], a high degree of stereotype is present not only during seizures but also in

seizure-free focal signals [16]. Thus, some ordinal patterns are generated recurrently during the EEG while others never appear [13]. This last group are the *forbidden ordinal patterns* described by *Amigó et al* [15].

Different authors analysed iEEG recordings using forbidden ordinal pattern algorithms in rats [17] and also in human periictal iEEG [13]. In the second article, authors showed the evolution of the number of forbidden patterns during a periictal recording [13]. Concretely, they found that the number of forbidden pattern increases during the first third of the seizure and then it decreases beyond seizure termination [13]. EEG signals can last hours, and one pattern can appear once by chance. Thus, they cannot be longer considered a forbidden ordinal pattern. For this reason, we calculate the permutation entropy (H): a complexity measure based on the relative frequencies of each ordinal pattern [18, 17, 19, 20].

1.2 Objectives

In this thesis we are going to apply the two symbolic methods above explained and two nonlinear predictability measures: *nonlinear prediction error* (E) and *rank-based nonlinear predictability score* (S). These last two measures are quite used in focus localisation studies with good results [21, 22, 23, 24, 25]. The main objective of the thesis is to see which of them works better in terms of localisation of the SOZ. As a first part of the study, we evaluate and compare the robustness against noise of the number of forbidden patterns, H, E and S in a deterministic dynamic as it is the Lorenz system (described in Section 2). In the second part, we determine if the information obtained with the number of forbidden patterns and permutation entropy gives more hints of a possible focal signal rather than only using E or S and, by consequence, if it helps in the SOZ delimitation. To do so, we evaluate three different iEEG databases with the four above-mentioned methods. With this part of the study we analyse different recordings and we are going to look at values of number of forbidden patterns and permutation entropy on segments of signals rather than the temporal evolution from a whole signal as it is done in *Schindler et al* [13].

2. Materials and Methods

2.1 Materials

2.1.1 Pre-processing of signals

We used three different iEEG datasets. All databases contain recordings from patients with focal epilepsy. Signals were recorded in the focal and nonfocal hemispheres during seizure-free periods. We describe below the exact content of each database and how we processed their information.

1. **Bern-Barcelona 2012 database:** This database is composed by 7500 seizure-free pairs of signals of 20 s. 3750 pairs of signals were measured from the focal hemisphere and the remaining 3750 were measured from the nonfocal hemisphere [24]. Each signal was originally obtained with a sampling frequency $f_s = 512$ Hz, with a length of 10240 sampling points. We discarded the second signal of each pair and then we applied to the remaining signals an eighth-order Butterworth filter with a cut-off frequency at 40 Hz. Next, we down-sampled the signals by a factor four, obtaining signals of 2560 sampling points at 128 Hz [24], that means temporal resolution is 7.8 ms.

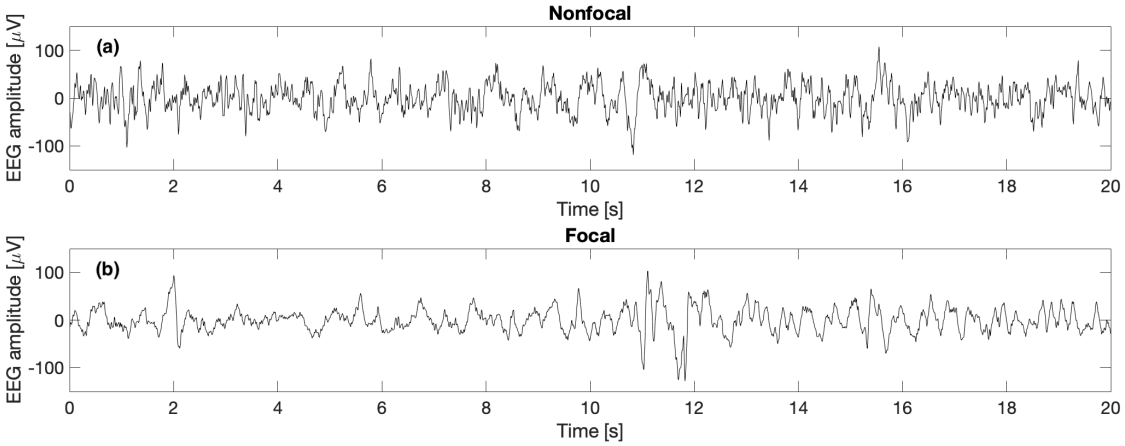

Figure 1: ***Example of Bern-Barcelona signals.*** *Panel (a) shows a nonfocal signal and in (b) a focal signal.*

2. **Bonn 2001 database**: Bonn database contains five sets of signals, each of them contains 100 single-channel EEG with a duration of 23.6 s. The sets A

and B correspond to surface EEG from healthy volunteers with open and closed eyes, respectively. Sets C, D and E correspond to focal, nonfocal and seizure iEEG signals from five epilepsy patients, respectively [21]. Signals were pre-processed with a Butterworth filter of eighth-order in 40 Hz. The recordings were obtained originally at $f_s = 173.61$ Hz and no down-sampling was carried out [21].

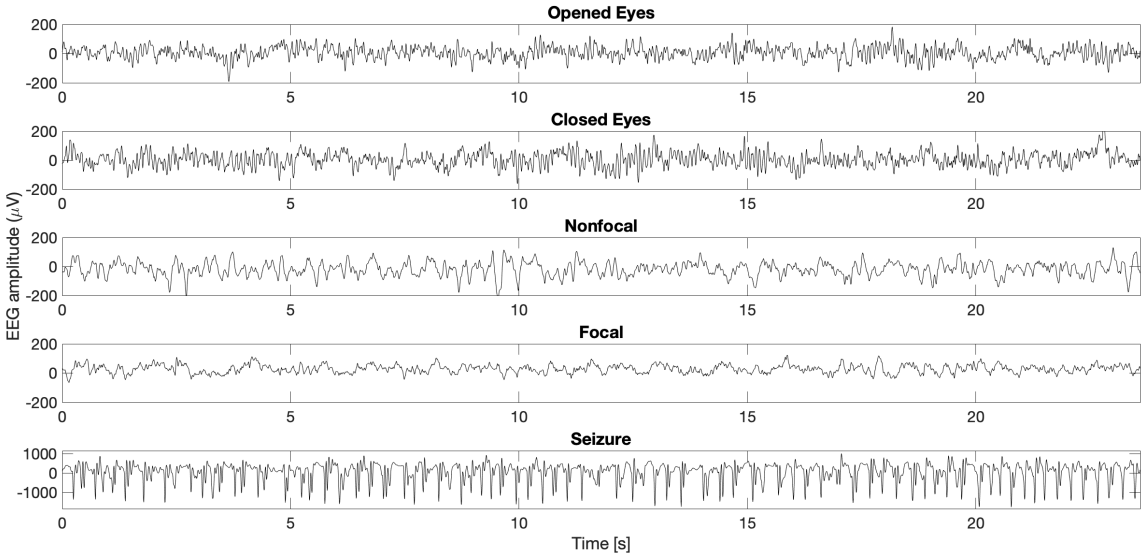

Figure 2: ***Example of Bonn 2001 signals.*** *From top panel to the bottom panel are represented the signals from sets A to E, respectively.*

3. **Bonn 2019 database:** These iEEG signals were provided by Cristina González Martínez, PhD student at the Nonlinear Sime Series Analysis group. This database contains recordings obtained with *hybrid-depth electrodes*. These are electrodes composed by eight macro-contacts and eight micro-wires at their tip. Each of them records a signal. Macro-contacts recordings were re-referenced by using a *bipolar montage*. The bipolar montage gets a "sharper spatial localisation and a reference-free representation of phenomena under observation" [26]. We subtract the activity from each contact with its adjacent contact. We end up with seven channels from macro-contact recordings. Micro-contacts recordings were averaged and we subtracted this average from each wire recording value. Signals were originally obtained at $f_s = 2048$ Hz and they had a time length of 32 s. We down-sampled them by a factor eight, what makes $f_s = 256$

Hz (time resolution is 3.9 ms). Signals were also filtered from 0.5 Hz to 40 Hz and we only analysed the first 16 s.

Recordings come from three different patients. We will name them patient 1, 2 and 3. The patient 1 has the focus in the left hemisphere and the patients 2 and 3 have the focus in the right hemisphere.

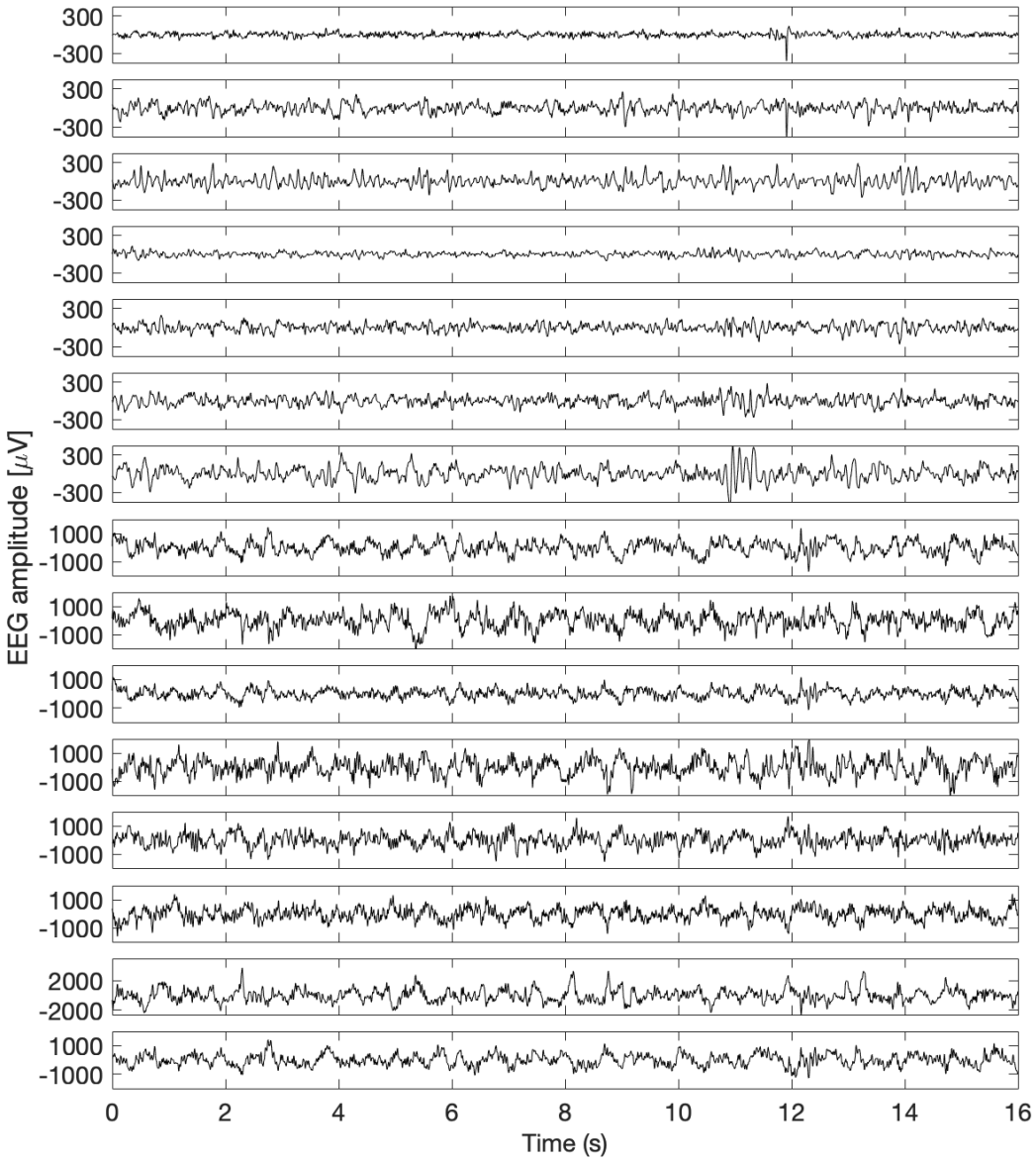

Figure 3: ***Example of re-referenced Bonn 2019 signals belonging to patient 1 in the focal hemisphere (left).*** *The first seven panels correspond to the macro-contacts recordings and last eight to the micro-wire recordings. In the figure both are already re-referenced.*

2.2 Nonlinear time series analysis measures

2.2.1 Forbidden ordinal patterns (F)

In deterministic dynamics we can predict the future state with the knowledge of the present state. Moreover, in deterministic dynamics from a total ensemble of ordinal patterns some of them appear recurrently but others never do it. The ordinal patterns that not appear are named forbidden patterns. On the other hand, the evolution of stochastic dynamics is not constrained by their present state and some random process underlies them making, in comparison to a deterministic signal, more ordinal patterns appear.

The applied algorithm evaluates the number of forbidden ordinal patterns (F) in a signal of length N [18, 15, 13] . The original algorithm takes d subsequent sampling points (which is equivalent to say that we are taking sampling points with a time delay $\tau = 1$ in respect with the previous sampling point), creating a window of length d. These d sampling points are sorted according to their amplitudes in an ascending order and ranked by their temporal sequence obtaining the ordinal patterns of length d. Then, the window is shifted along the signal until its end. Note that each ordinal pattern of length d is one of the $d!$ possible permutations (see Figure 4).

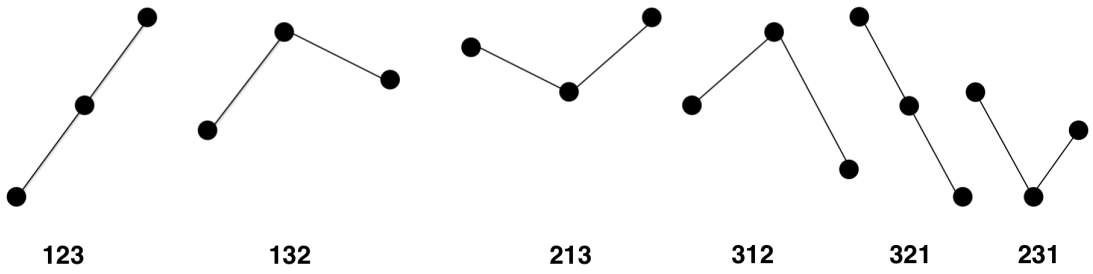

Figure 4: ***Example of the theoretical ordinal patterns when d = 3. We always have*** $d!$ ***theoretical ordinal patterns.***

Afterwards we computed all the ordinal patterns of the signal, we obtained F by observing which permutations do not appear. The more random is a signal, the lower F is expected and vice versa. In our study, we shifted the window with maximal

overlap getting a number of permutations (repeated or not) equal to $N - d - 1$. To guarantee that all the ordinal patterns can theoretically happen, $N - d - 1 > d!$ must be fulfilled [13].
Noise can modify F, according to [18]. For this reason, we evaluated the distribution of all patterns and we computed the permutation entropy described in the next subsection.

In the three databases we use $d = 6$. Scripts and functions for the computation of F were developed in MATLAB 2018b (MathWorks, Natick, MA, U.S.A.) by the author, Cristian Galán Augé.

2.2.2 Permutation Entropy (H)

Permutation entropy is a complexity measure equivalent to the Shannon entropy but applied to the resulting probability distribution of the ordinal patterns. It was defined by Bandt and Pompe [19] as:

$$H(p) = -\sum_{i=1}^{d!}(p_i log(p_i)) \tag{2.1}$$

In our study we name p_i the frequency of each ordinal pattern i. Forbidden patterns gives a binary output: an ordinal pattern that appears once has the same importance than a pattern that appears frequently. Evaluating the frequency of each ordinal pattern we complement the results of F.

If a signal monotonically increases or decreases, the obtained ordinal pattern will always be the same and, in consequence, $H(p) = 0$. The other extreme is random uncorrelated noise or white noise, when $H(p) \approx log(d!)$.

As for F, we use $d = 6$ in the three databases. Scripts and functions for the computation of H were developed in MATLAB 2018b (MathWorks, Natick, MA, U.S.A.) by the author, Cristian Galán Augé.

2.2.3 Nonlinear Prediction Error (E)

The nonlinear Prediction Error (E) tries to predict a future state from a present state and its similarity with the aim of quantifying the committed error. The objective is to distinguish between stochastic and deterministic dynamics. Sometimes we would not have all the information needed to obtain these nearest neighbours because we do not know the underlying dynamics. For this reason, and following the implementation of E described in [22], we need to do a reconstruction of the state space. We compute the *delay vectors*:

$$x_i = (s_i, s_{i-\tau}, ..., s_{i-(m-1)\tau})$$

where s_i are the samples of the original time series with range $i = 1, ..., N$ and x_i takes as indices $i = \eta+1, ..., N$; where N is the length of the signal. The *delay vectors* are computed with an *embedding window* $\eta = (m-1)\tau$ where m is the *embedding dimension* and τ is the time delay. With the reconstruction of the state space done, we calculate the *Euclidean distance* (see Equation 2.2) between states x_i and we take the k *nearest neighbours*, those ones with the smallest distances.

$$D(x_i, x_j) = \sqrt{\sum_{l=1}^{q} (x_{i,l} - x_{j,l})^2} \tag{2.2}$$

In Equation 2.2, $i, j = \eta + 1, ..., N$ and q are the components of the evaluated dynamics. There is the possibility of taking points that are not temporally separated enough from the reference. For this reason, we select a window of samples of length W known as *Theiler correction* that will assure that the compared points are enough separated in temporal terms. Given two points x_{i_A} and x_{i_B}, it must be fulfilled that: $|i_A - i_B| > W$.

Then, we take the average of the future states from the k nearest neighbours at time $i_g + h$, where $g = 1, ...k$, h is the prediction horizon and $i_0 = \eta, ..., N - h$, and we compare this average with the future state from our reference and we obtain the error

(ε) that we are committing:

$$\varepsilon = s_{i_{0,0}+h} - \frac{1}{k}\sum_{g=1}^{k} s_{i_{0,g}+h}, \tag{2.3}$$

where $s_{i_{0,0}}$ is the reference state and $s_{i_{0,g}}$ are the nearest neighbours' state for $g = 1, ..., k$.

The last step is the computation of the root-mean square error from ε.

$$E = \sqrt{\frac{1}{N-h-\eta+1}\sum_{i_0=\eta}^{i_0=N-h} \varepsilon_{i_0}^2} \tag{2.4}$$

When the signal is completely predictable and free of noise, ideally, $E = 0$. When the dynamics are chaotic E tends to higher values than for deterministic dynamics. For white noise time series we obtain a value around $E = 1.1$. We cannot distinguish between correlated noisy or chaotic dynamics by only looking at signal values. To do so, we need to compute surrogates (see Section 2.4).

Due to the different sampling frequencies of each database, we used different parameters for each database (see Section 2.1). In the *Bern-Barcelona 2012* database we used as parameters: $m = 8$, $\tau = 4$, $k = 5$, $h = 4$, $W = 19$ as used in [23]. We took these values as standard and derived the values used for the other databases accordingly. Specifically, we changed the time-dependent parameters (τ, h and W) by doing a ratio of the sampling frequencies of each database. For *Bonn 2001* this ratio was $r = 173.61/128$ and for *Bonn 2019* it was $r = 256/128$. Such as:

$$m = 8$$
$$\tau = r \cdot 4$$
$$k = 5$$
$$h = r \cdot 4$$
$$W = r \cdot 19$$

In case of *Bonn 2001* we rounded the resulting values to the closest integer number. Then, in this database we used $m = 8$, $\tau = 5$, $k = 5$, $h = 5$, $W = 24$. Finally, in

Bonn 2019 we used $m = 8$, $\tau = 8$, $k = 5$, $h = 8$, $W = 38$.

2.2.4 Rank-based nonlinear predictability score (S)

The rank-based nonlinear predictability score S quantifies the degree of predictability as the method E explained above in Section 2.2.3. The main difference between both measures is that S evaluates ranks computed from amplitude differences between the reference and its nearest neighbours rather than amplitude values (see [23] and references therein). We reconstruct the state space as for E. Here, the distances in Equation 2.5 are used for finding the concrete rank.

$$u_{i,j} = |x_i - x_j| \tag{2.5}$$

The rank of each distance $u_{i0,j0}$ are stated as $g_{i0,j0}$. We represent the total number of differences in $\{u_{i0,j}\}$ as $M_{i_0} = N - 2W - 1$ (for details, see [23]), where N is the length of the evaluated signal. For testing the predictability, now we compute the mean rank of the amplitude differences for the k nearest neighbours:

$$R_{i_0} = \frac{1}{k} \sum_{r=1}^{k} g_{i_{0,0}+h, j_{0,r}+h} \tag{2.6}$$

When the signal is completely predictable, we would choose the mean of the k lowest ranks, which is independent of i_0:

$$R^L = \frac{k+1}{2}$$

On the other side, when the signal cannot be predicted, the ranks we get for our k nearest neighbours are taken from an uniform distribution that goes from 1 to the total number of differences M_{i_0}. Thus:

$$R^U_{i_o} = \frac{M_{i_0} + 1}{2}$$

Finally, the rank-based prediction score is stated as:

$$S = \frac{1}{N-\eta-h} \sum_{i_0=1+\eta}^{N-h} \frac{R_{i_0}^U - R_{i_0}}{R_{i_0}^U - R^L} \tag{2.7}$$

When $R_{i_0} = R^L$, $S_{max} = 1$ meaning that the signal is completely predictable and when the signals are completely random the expected value is $S_{min} \approx 0$.

The range of values for the parameters were the same than for E (see Section 2.2.3). The script for the computation of E and S were developed by Daniel Naro [23].

2.3 Lorenz dynamics for testing robustness against noise

One of the objectives is to analyse the robustness against noise of the four measures previously explained. We tested their robustness in a deterministic dynamics such as Lorenz dynamics (see Figure 5) which are determined by the following ordinal differential equations:

$$\begin{aligned} \dot{y_1} &= 10(y_2(t) - y_1(t)) \\ \dot{y_2} &= 39y_1(t) - y_2(t) - y_1(t)y_3(t) \\ \dot{y_3} &= y_1(t)y_2(t) - (8/3)y_3(t) \end{aligned}$$

where y_i are the components of the time series and the coefficients 10, 39 and 8/3 are taken from [22]. In order to integrate the differential equations, we used a Runge-Kutta algorithm of order four with an integration time step of 0.03. We computed 50 signals of 2000 points with a previous pre-allocation of 10000 iterations in order to avoid transient states for the computation of signals at the beginning. Then, we applied increasing levels of observational white Gaussian noise. Concretely, we began adding samples of Gaussian noise with amplitude values reduced at a 0.1% of a normal standard distribution. Noise level was progressively increased in a logarithmic scale. The higher is the noise level, the higher is the impact of the noise in the system.

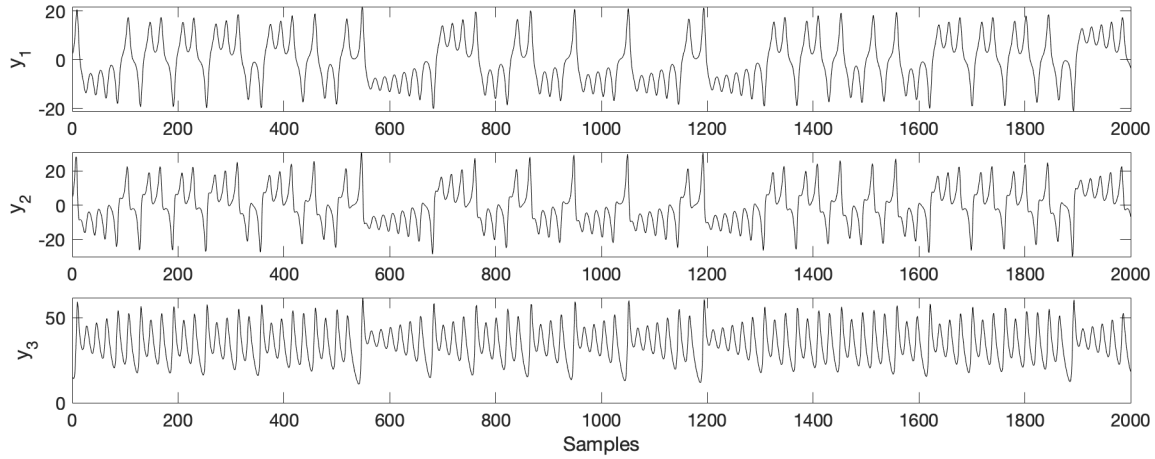

Figure 5: ***Lorenz dynamics.*** *Each panel is a component of the dynamics.*

After the computation of signals, we computed a single phase-randomised surrogate (see next Section 2.4) for each one of the 50 signals. Then, we calculated the four measures for each signal and its surrogate at each noise level. Parameters for E and S were the same used in the Bern-Barcelona 2012 database: $m = 8$, $\tau = 4$, $k = 5$, $h = 4$ and $W = 19$. Afterwards, we took the difference between the measures' values of the original signals and the surrogate for each noise level in order to see at which level of noise the difference is equal to zero.

2.4 Surrogates

We use surrogates when we have doubts about the reliability of the obtained values which can be altered by, for example, measurement noise or by some deterministic structure [22]. We also use them when we want to test some null hypothesis. Surrogates are signals generated from a *controlled randomisation process* where we decide which properties are maintained from the original signal. These properties are selected depending on the null hypothesis we test [22].

We used two types of surrogates. One of them is the phase-randomised surrogates that we compute according to *Theiler et al* [27]. In order to generate these surrogates, we firstly do the Fourier transform of the original time series. Then, we randomise all the phase angle coefficients between 0 and 2π. During this step, we assure symmetry of the periodogram is always maintained, meaning that the frequencies distribution

is the same as the one of the original signal when we compute the inverse Fourier transform of the randomised Fourier coefficients. By maintaining the periodogram, we also maintain the same autocorrelation function of the original signals [22].

The other type of surrogates is the amplitude-adjusted surrogates [27], which introduces an iterative process that allows to adjust the amplitude [28]. In this case, we randomise all the signal features except for autocorrelation and amplitude distribution.

In our case we wanted to determine if the original signals are deterministic or not. In case of the phase-randomised surrogates, the null hypothesis was "that our measured dynamics are a linear, stochastic, stationary and Gaussian amplitude distributed" while for amplitude-adjusted surrogates the null hypothesis is "that our measures comes from a Gaussian linear stochastic and stationary process measured by a static and monotonic, possibly nonlinear, function"[22]. We computed a determined number of surrogates. For each surrogate, we calculated E, S, F and H; and we took the minimum or the maximum values of the measures across all surrogates. We rejected the hypotheses in those cases where the next statements were accomplished:

$$min(E_{surr}) - E_{signal} > 0$$
$$S_{signal} - max(S_{surr}) > 0$$
$$F_{signal} - max(F_{surr}) > 0$$
$$min(H_{surr}) - H_{signal} > 0$$

In the robustness against noise test we created one phase-randomised surrogate for each time series we computed (explained in detail in the following Section 2.3). For each database, we created 19 amplitude-adjusted surrogates with 100 iterations for each signal, obtaining a significance level $\alpha = 0.05$. This means that we expect to reject a 5% of the signals in the case that the null hypothesis was correct. Then, we computed E, S, F and H for each signal and all their surrogates and we tested the null hypothesis. We stored how many signals were rejected in each measure by the null hypothesis.

The corresponding function for the computation of phase-randomised surrogates was

developed by the author, Cristian Galán Augé while the function that computes the amplitude-adjusted surrogates was developed by Prof. Ralph G. Andrzejak [23].

3. Results

3.1 Robustness against noise analysis

We tested the four nonlinear methods (Sections 2.2.1 to 2.2.4) in the Lorenz dynamics. We progressively added observational noise to the dynamics and compared the difference with their surrogates (Figure 7). This noise is white Gaussian noise but its amplitude values are multiplied by a factor (from 0.1% to 103%) that changes the relevance of the noise in the system. The measure is capable to determine a concrete value until a critical level of noise (which depends on the measure). When it overpasses the critical value, the signal changes in a similar way as a sigmoid function (see Figure 6). This critical value changes for each measure and it is lower for the original dynamics than for the surrogate. Surrogate and original dynamics converge to typical values of white noise.

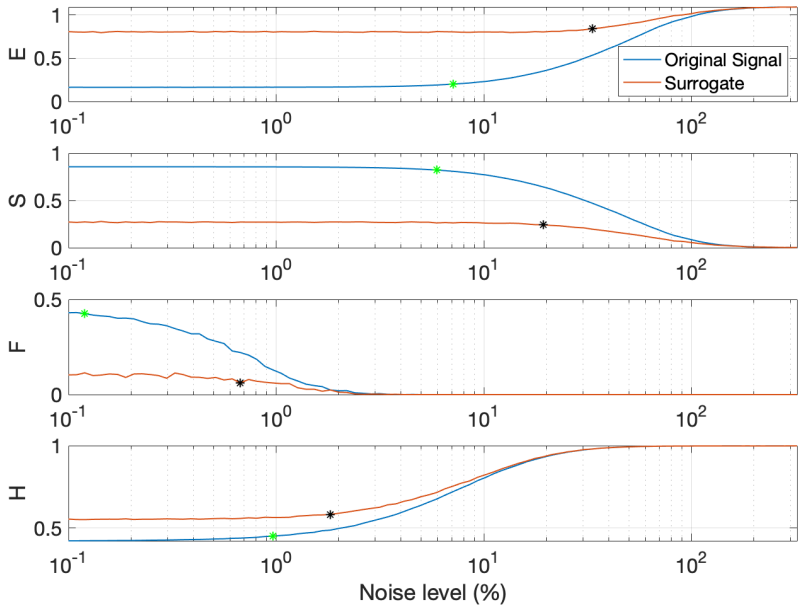

Figure 6: ***Signals and surrogates remain almost constant around a concrete value for each measure until a level of noise when they begin to increase*** *(E* ***and*** *H)* ***or to decrease*** *(S* ***and*** *F).* ***Surrogate and signal converge to the same value.*** *The black star marks the critical value at the surrogate and the green star the critical value at the original signal.*

This behaviour reflects that the Lorenz system is losing their characteristics and the measures are not able to predict the future states (in case of E and S) or they are able to find all the ordinal patterns with the same frequency (in case of F and H).

These are the typical results we would find when evaluating uncorrelated noise.

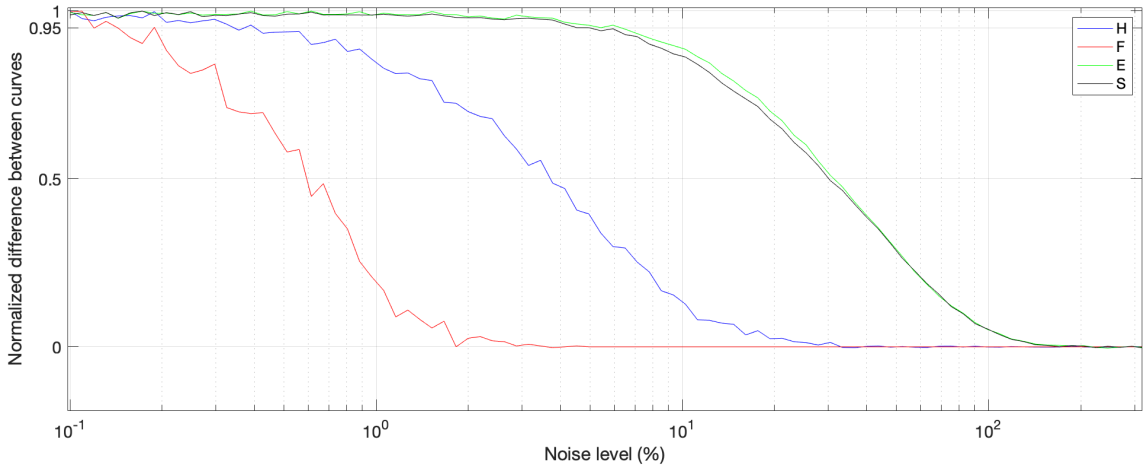

Figure 7: *E and S show the highest robustness against noise while F show the worst. The difference on the noise level is about two orders of magnitude. Differences are normalised between the measure from the original signal and its surrogates for each level of noise in order to be able to compare the four measures in the same scale.*

E and S signal and surrogate values are slightly distinguishable at a noise level of 100%, what means typical values of white Gaussian noise. H has indistinguishable values from the surrogate at a 35% of added noise and F only at a 3%.

3.2 Bern-Barcelona 2012 database

Recall that we computed 19 surrogates for each signal, what sets the significance level to $\alpha = 0.05$. We expect to reject the null hypothesis a 5% by chance. The results presented in Figure 8 show a lower rejection fraction for the methods based on ordinal patterns in comparison with the predictability measures. In the specific case of F, rejections are close to α for focal signals and below α for nonfocal signals. This means that the null hypothesis cannot be rejected at the level of the set of all the signals.

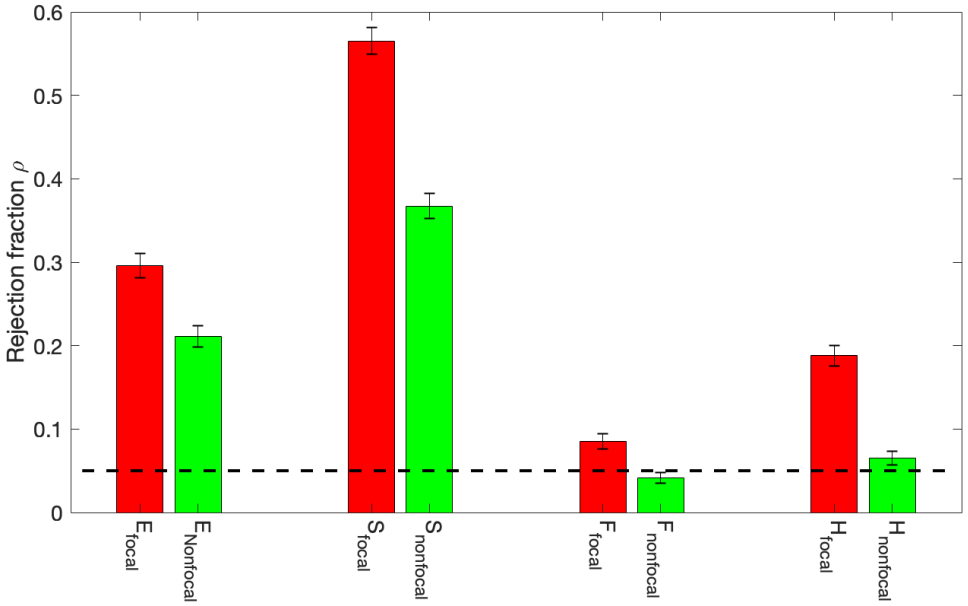

Figure 8: ***Rejection fraction (ρ) of the surrogate null hypothesis obtained by the four measures for focal and nonfocal EEG signals. The rejection fractions of F and H are quite lower than the ones of E and S.*** *Percentages are calculated against the 3750 focal and the 3750 nonfocal signals. The dashed line represents the significance level $\alpha = 0.05$. Error bars show the confidence interval of the 95%. They are calculated by: $[\rho - 1.96\sqrt{\rho(1-\rho)/3750}, \rho + 1.96\sqrt{\rho(1-\rho)/3750}]$ (see [23] and references therein).*

We calculate the distributions for each measure to see if they are capable to distinguish between signals obtained in the focal and the ones obtained in the nonfocal hemisphere (see Figure 9a). While in both predictability measures, E and S, focal and nonfocal distributions are quite distinct by visual inspection, distributions of F are shaped almost equally. The overlap between distributions of H is an intermediate

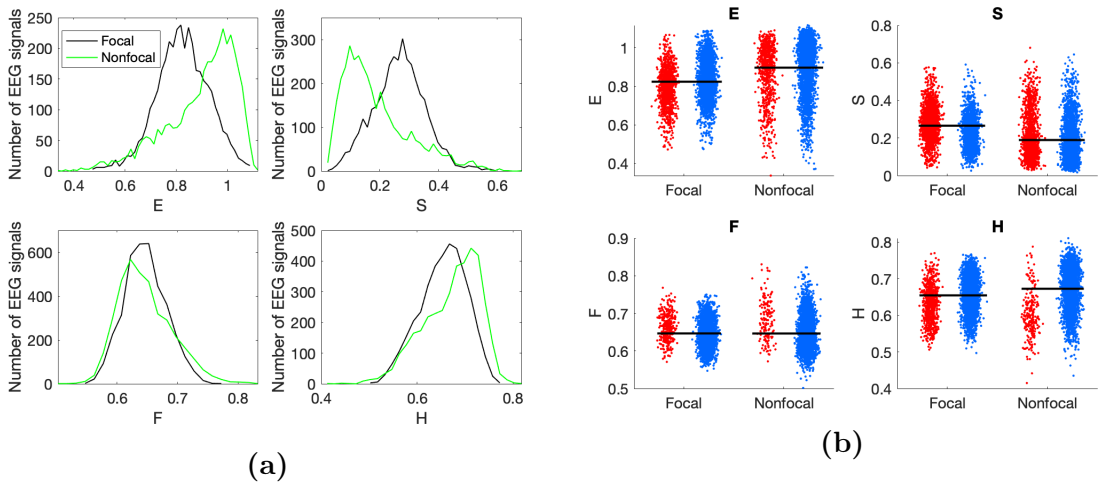

Figure 9: ***Distribution of the measures' values obtained for focal and nonfocal signals and mean values are more distinct for*** E ***and*** S ***rather than for*** H ***and*** E***.*** *In (a), comparison of distributions between focal (black) and nonfocal (green) signals. Bin size is 0.015. In (b), the black line represents the mean value of all signals for each measure; red dots represent the values for the rejected signals and blue dots the values for the non-rejected signals. Dots are spread horizontally to see the accumulation of points in some concrete values.*

case between predictability measures and F.

In Figure 9b, we see the range of values obtained by each rejection in focal and nonfocal signal and also the mean value of all of them. F takes almost equal mean values for focal and nonfocal signals while the rest of measures have different mean values for signals from both hemispheres. Dots do not show huge differences between the rejected and the non-rejected signals. In order to contrast statistically the results for focal and nonfocal signals, we perform a Mann-Whitney test. The null hypothesis for this test is that "data of the measures obtained for focal and nonfocal signals are samples from continuous distributions with equal medians". Outcomes are shown in Table 1.

	E	S	F	H
p-value	$<$0.001	$<$0.001	0.001	$<$ 0.001

Table 1: ***F, H, E and S have statistical difference between the distributions of focal and nonfocal signals.*** *The significance level here is $\alpha = 0.05$, thus we reject the hypothesis when p-value < 0.05.*

With these outcomes we can say that, in all measures, focal and nonfocal distributions

are statistically different.

3.3 Bonn 2001 database

In the Bonn 2001 database we have five different datasets reffered to as A, B, C, D, E. We remember the reader that sets A and B refer to healthy volunteers. Signals from set A are obtained with eyes opened and B with eyes closed. Sets C and D contain recordings from epilepsy patients in the focal and in the nonfocal hemisphere, respectively. Recordings of set E correspond to epileptic signals obtained during seizure.

In Figure 10 we represent the rejection fractions against surrogates. For datasets A, B and C lower rejection fractions against surrogates are obtained than in sets D and E , with the exception of F for the set C. There, C has a higher value than in D. Additionally, in measure H, in datasets A to C we are near or below the significance level α. Notice also that values for set B are bigger than for set C (except for measure

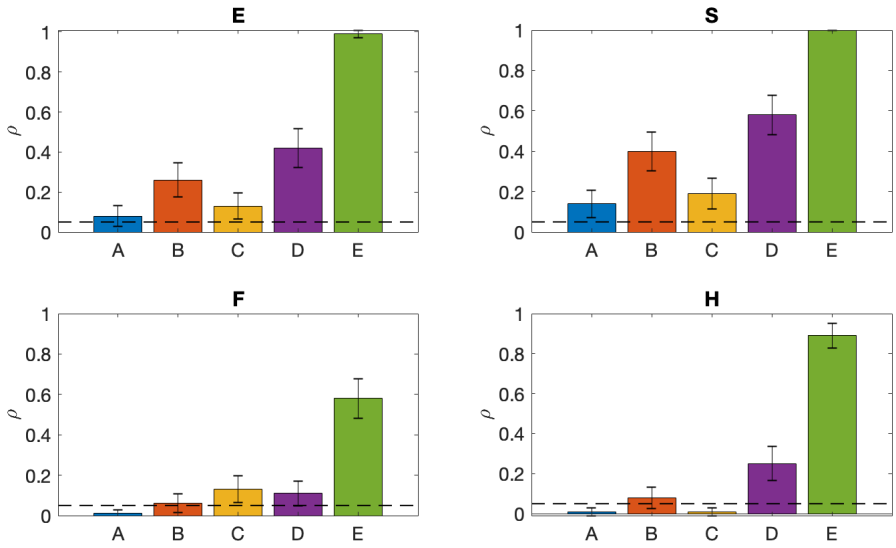

Figure 10: ***Comparison of rejection fraction (ρ) against surrogates in Bonn 2001 database. The rejection fraction of F and H are not significant for the sets A and C.*** *Same as in Figure 8 but percentages are calculated against the 100 focal and the 100 nonfocal signals, respectively. Thus, error bars are calculated by:* $[\rho - 1.96\sqrt{\rho(1-\rho)/100}, \rho + 1.96\sqrt{\rho(1-\rho)/100}]$ *(see [23] and references therein).*

F). Finally in dataset E, that corresponds to seizure signals, we get higher rejection

fractions than for set D in the four measures. In the case of E and S, they are almost 1.

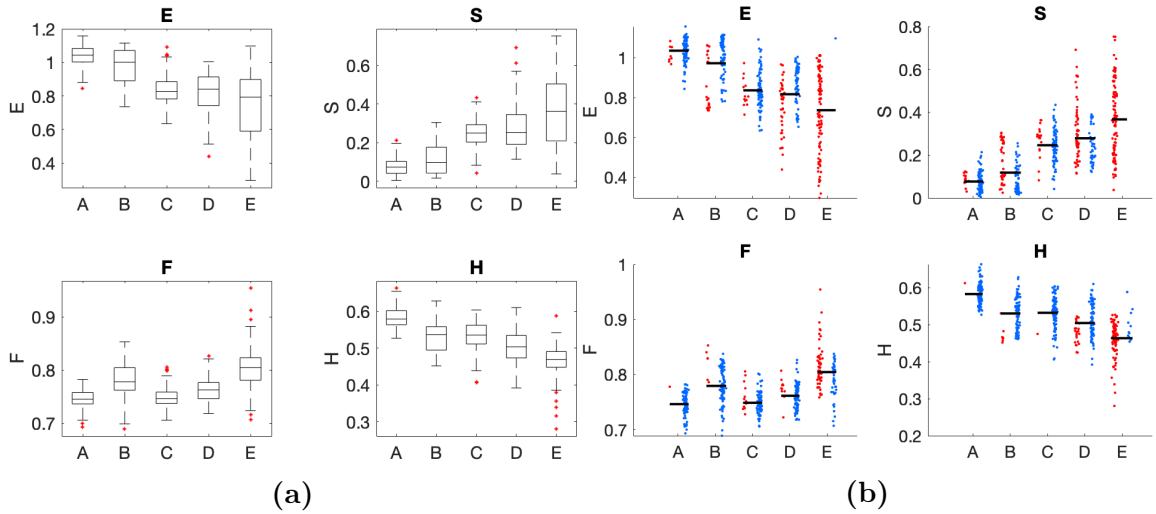

Figure 11: ***In (a), datasets distributions for each measures. In (b), same as Figure 9b but for Bonn 2001.*** *In (a), bottom and top edges of the boxes indicate the first and the third quartile, respectively. Edges inside the boxes represent the median value. The whiskers extend to the most extreme cases that are considered to be part of the series. Finally, outliers are marked with the red crosses.*

As each set of Bonn 2001 is composed by only 100 samples per set and their measure ranges are wide, we do not represent distributions in bins because the information is unreadable. Instead, in Figure 11a we represent the distributions in box plots. The quartiles are slightly different for each dataset: they always intersect partially with another set and we also get some outliers. However, visual inspection is not enough to determine if the distributions are different. Thus, as in Section 3.2, we carry out a Mann-Whitney test between all sets for each measure. In Table 2, we see that the set C is always involved in the rejected sets (but not all evaluations that involve C are rejected).

Additionally, in Figure 11b, we look at the range of values where we reject our signals. Despite they present some overlap with the values for the non-rejected signals, they have values that show more predictability (lower values for E and H and higher values for S and F). Furthermore, the mean value of the measure decreases from A to E for the measures E and H while it increases for S and F. Notice that the exception is the set B for the measure F. This set has a higher mean value than set C which

	$p-value$			
Test	E	S	F	H
A vs B	0.002	0.008	<0.001	<0.001
A vs C	0.830	<0.001	0.830	<0.001
A vs D	<0.001	<0.001	<0.001	<0.001
A vs E	<0.001	<0.001	<0.001	<0.001
B vs C	<0.001	<0.001	<0.001	0.765
B vs D	<0.001	<0.001	<0.001	<0.001
B vs E	<0.001	<0.001	<0.001	<0.001
C vs D	0.825	0.291	<0.001	<0.001
C vs E	0.003	<0.001	<0.001	<0.001
D vs E	0.010	<0.001	<0.001	<0.001

Table 2: ***Mann-Whitney test for all measures. Null hypothesis is not rejected when distributions of sets A and C; B and C; C and D; and, C and D are compared.***

agrees with what we see in Figure 10.

3.4 Bonn 2019 database

As in the two previous sections, we are looking for signals that can be shaped by determinism. To do so, we determine the percentage of signals for which the surrogate test led to a rejection of the null hypothesis. In this database we see different behaviours for each patient. In patient 1 (Figure 12) we get for all measures more rejections for the recordings from the micro-contacts than for the macro-contacts. We see that E and S are capable to reject more signals than symbolic methods. Concretely for H, we get a number of rejections near or even below $\alpha = 0.05$. The rejection fractions is higher for focal than for nonfocal signals for all measures.

In contrast, in patient 2 (Figure 13) we get more rejections in the macro-contacts. Here E and S are also higher and for F and H we get higher rejection fractions in the micro-contacts than by using E. Notice also that in the macro-contacts, E and S are rejecting more null-hypothesis for nonfocal than focal signals.

Finally, in patient 3 (Figure 14) we get more rejections in the macro-contacts using E and S but by using F and H we get more rejections in the micro-contacts. However, the values of H for focal and nonfocal signals recorded by the macro-contacts are near α.

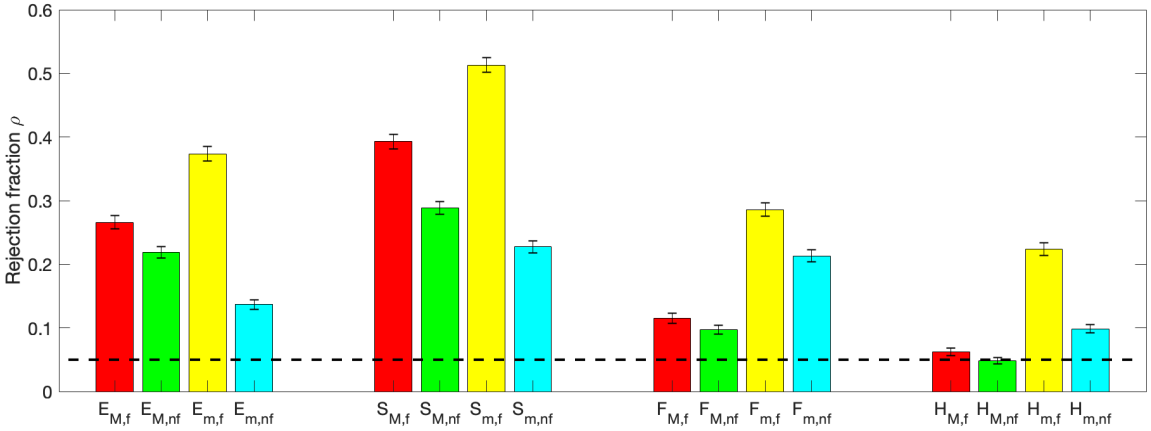

Figure 12: ***Patient 1. We reject more times the null hypothesis for the micro-contact recordings than for the macro-contact recordings.*** *H* ***has a number of rejections near to*** α ***by the macro-contact signals.*** *The index M indicates macro-contact, m micro-contact, f focal and nf nonfocal. In macro-recordings we have* $N_M = 6720$ *signals and in micro-recordings we have* $N_m = 7680$ *signals. Thus, errorbars are computed by using:* $[\rho - 1.96\sqrt{\rho(1-\rho)/N_{M,m}}, \rho + 1.96\sqrt{\rho(1-\rho)/N_{M,m}}]$ *(see [23]).*

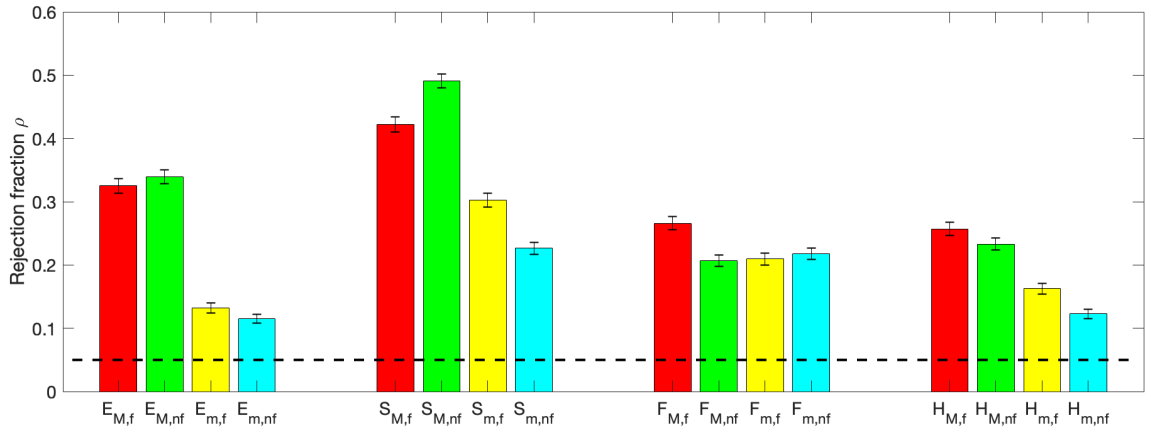

Figure 13: ***Patient 2. We reject more macro-contact recordings than micro-contact recordings.*** *Same as Figure 12 but for patient 2.*

Observing the distributions displayed in Figure 15, we do not see any general pattern of differences in focal and nonfocal signals. In patient 1 we obtain differences in the ordinal measures for the micro-recordings where focal signals have a wider range than nonfocal signals. In micro-recordings of the patient 2, for E and S, we get more separated distributions than in the rest of the results. In patient 3 we do not see, by visual inspection, huge differences on the distributions in any case.

In Figure 16 for all patients we get similar mean values in the four measures for the macro-recordings and similar mean values for the micro-recordings. Rejected signals

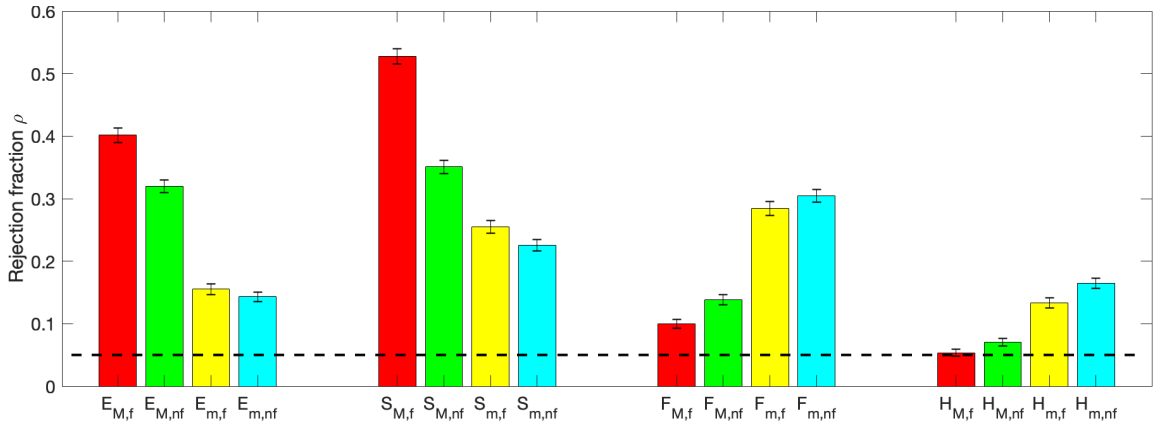

Figure 14: ***Patient 3. For predictability measures*** E ***and*** S ***we reject more macro-contact recordings than micro-contact recordings. Using*** F ***and*** H ***we get the opposite behaviour.*** *Same as Figure 12 but for patient 3.*

seem to have higher values in E and F and lower values in S and H.

P1	p-value	
	Macro	**Micro**
E	<0.0001	0.0080
S	0.0002	0.6463
F	<0.0001	<0.0001
H	<0.0001	<0.0001

P2	p-value	
	Macro	**Micro**
E	<0.0001	<0.0001
S	<0.0001	<0.0001
F	<0.0001	<0.0001
H	<0.0001	<0.0001

P3	p-value	
	Macro	**Micro**
E	0.2688	<0.0001
S	0.0077	0.0032
F	<0.0001	<0.0001
H	<0.0001	<0.0001

Table 3: ***Values obtained by*** S ***in the micro-contacts for the recordings of both hemispheres of patient 1 are not statistically different. In patient 3, values obtained by*** E ***at the macro-contacts are not statistically different.*** *In the first table we show p-values for patient 1, in the second we show them for patient 2 and in the third we show them for patient 3. In the first column we compare focal macro-recordings with nonfocal macro-recordings. In the second column we compare focal micro-recordings with nonfocal micro-recordings.*

We do a statistical test in order to compare distributions of focal and nonfocal signals in macro-recordings and separately the same for micro-recordings. Hence, we compute again a Mann-Whitney test (see Section 3.2). As we do multiple testing, we apply a Bonferroni correction to $\alpha = 0.05$. We carry out 6 independents tests (3 patients and 2 types of contacts). Thus, $\alpha = 0.05/6 = 0.0083$. For patient 1, we obtain that the distributions that are not significantly different for focal and nonfocal signals are the ones obtained by the measure S in the micro-contacts' recordings. For patient 2 we reject for all measures in macro and micro-contacts. For patient 3, we do not reject the null hypothesis for the distributions obtained by E at the macro-contact signals.

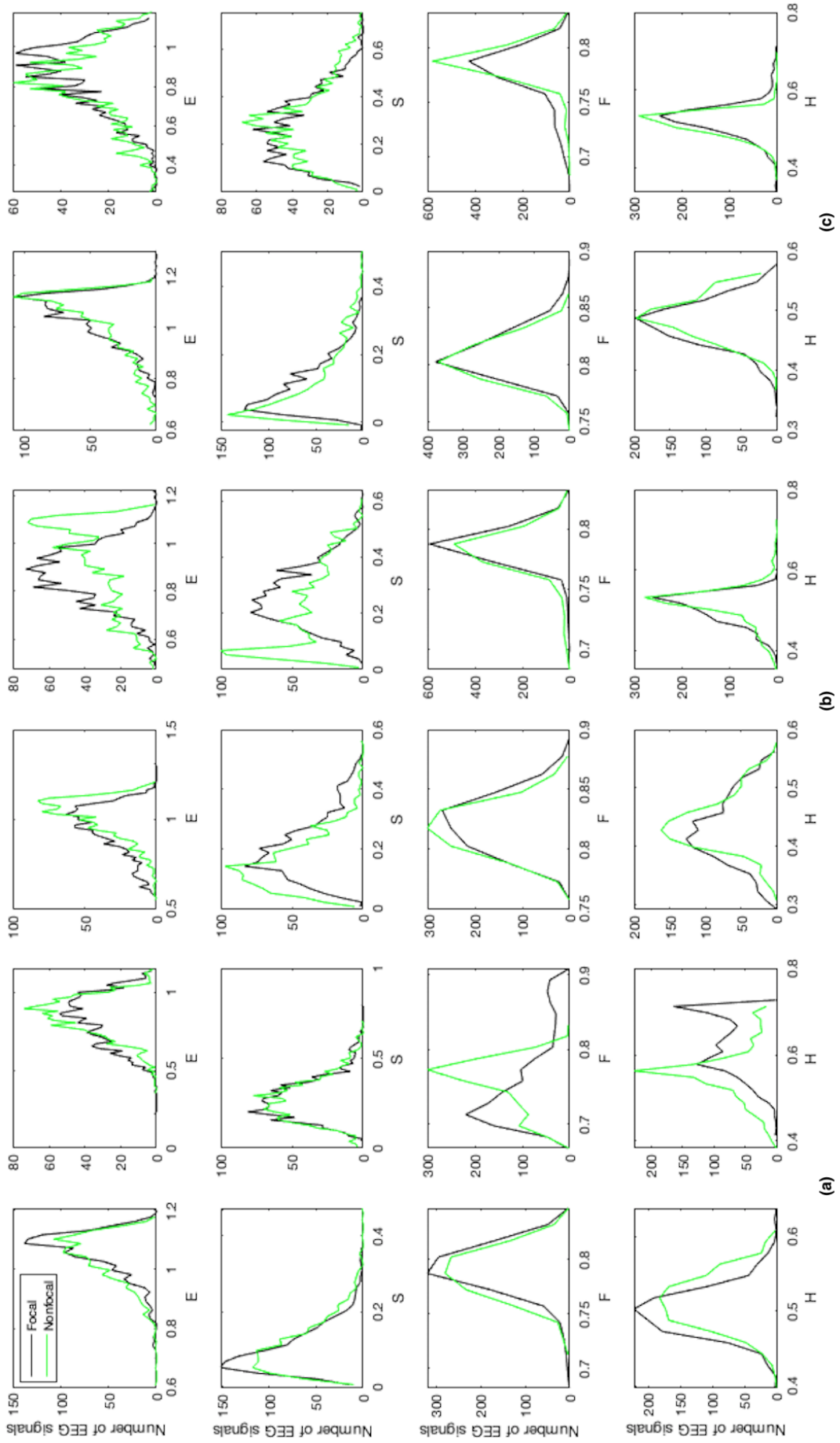

Figure 15: ***Distributions show very similar values for focal and nonfocal signals for all measures and patients. The exception are in patient 1 and patient 2. In patient 1, micro-recordings present different distributions for F and H. In patient 2 the same but for F and H.*** *Distributions are shown for the three patients by the focal (black line) and the nonfocal (green line) hemispheres for the macro-recordings in the odd columns and the micro-recordings in the even columns. In (a), the results for patient 1, in (b) for the patient 2 and in (c) for the patient 3. Bin size is always 0.015.*

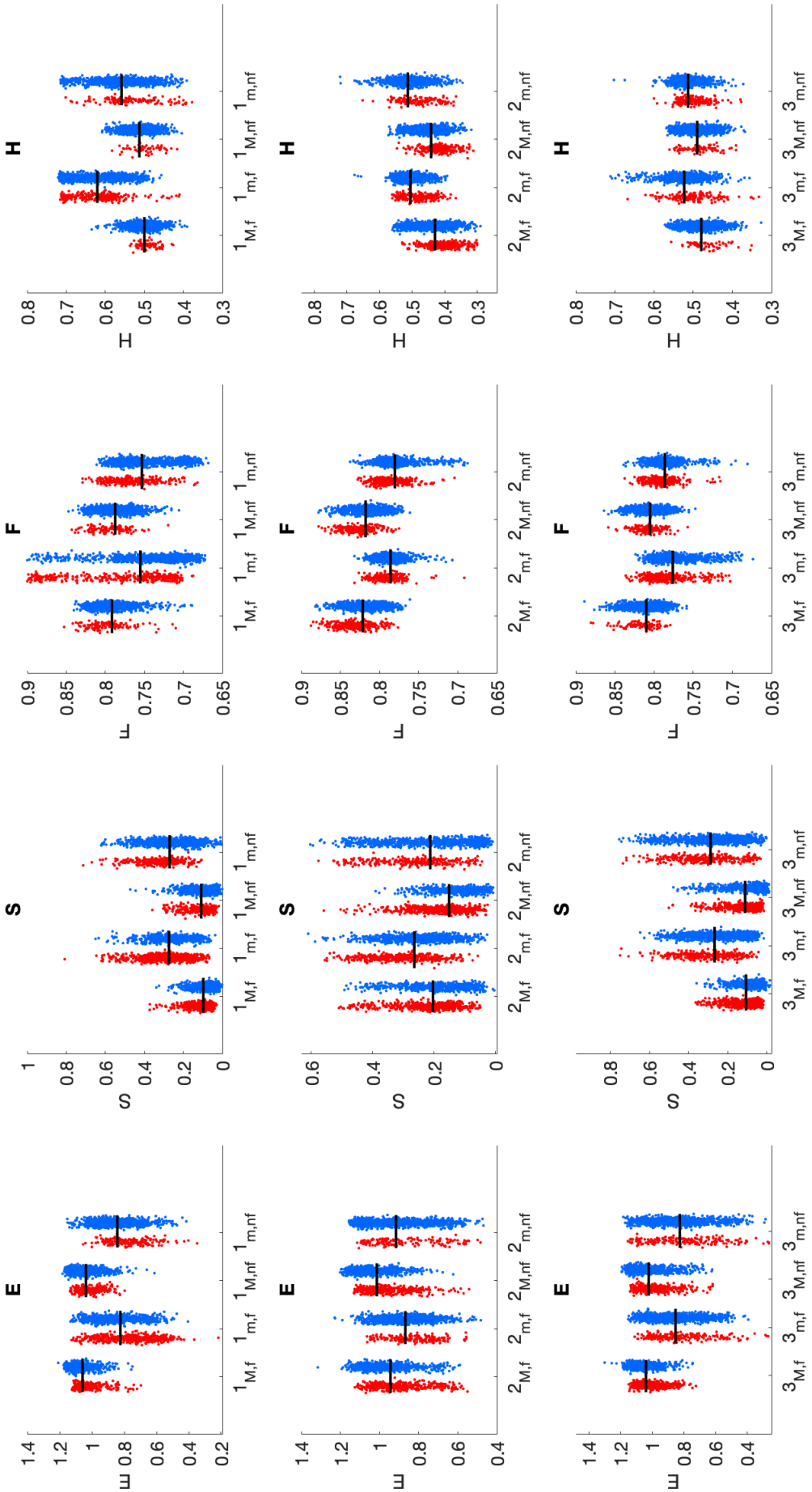

Figure 16: ***Micro-recordings values show more predictability than in the macro-recordings for the three patients.*** *At first row patient 1, at second row patient 2 and in third row patient 3. Colours and features same than Figure 11. Indices of the x-axis as Figure 12*

4. Discussion

The motivation of this thesis began with *Schindler et al* study [13] where they analysed the temporal evolution of the number of forbidden patterns (F) in periictal recordings. They observed that F increases in the first third of the seizure and then it decreases until seizure termination. We wanted to look at the values of F and H for the whole signal and test them against surrogates in order to see if these symbolic methods were able to obtain similar information as the predictability measures E and S, basing our study on the references [21, 22, 23, 24, 25]. We used amplitude-surrogates for testing the null hypothesis that says "our measures come from a Gaussian linear stochastic and stationary process measured by a static and monotonic, possibly nonlinear, function"[22]. If we reject the null hypothesis, it means that our signals could be deterministic caused by a non-random process which happens in epilepsy. But being deterministic is not the only alternative hypothesis we would obtain. It could also be possible that our signals did not accomplish any other condition (or conditions) stated in the null hypothesis.

In the first part of the study we evaluated the robustness against noise for all measures using Lorenz dynamics. We increased the level of noise until we got a ratio of noise where the original signal was indistinguishable from the surrogate. Robustness in F is about two orders of noise magnitude below than for E and S. In the case of H, it is almost an order of magnitude below E and S. The two predictability measures show similar robustness. So, E and S are more robust than F and H against noise. The next step was to test the measures in real databases. We discuss each database individually.

We firstly analyse the results obtained in Bern-Barcelona 2012. Here, the rejection fraction against surrogates (ρ) is considerably higher in S and E than in F and H. Moreover, ρ is below the significance level $\alpha = 0.05$ in F for nonfocal signals. These lower percentages of the symbolic methods are in agreement with the study of robustness against noise in Section 3.1. In any case, the rejections in F and H are higher for focal than for nonfocal EEG signals. This is consistent with previous studies [23, 24, 21]. The differences between focal and nonfocal distributions and

their mean value for each measure (see Figure 9) are consistent between measures: we have a higher difference between focal and nonfocal in the values of E and S than in the values of F and H. Distributions of F are almost overlapped and mean values are practically equal. For this reason, we think that F does not provide with a proper differentiation between focal and nonfocal signals. However, the results obtained in the statistical test, where our hypothesis is that "data of the used measures in focal and nonfocal signals are samples from continuous distributions with equal medians" is rejected because they are not completely equal. In any case, predictability measures have a higher difference between distributions and their mean values. This together with the fact that we obtain higher ρ means that, in this database, E and S have a higher sensitivity than F and H.

The database from Bonn 2001 is composed of five subsets of 100 signals [21]. Subsets A and B are signals from healthy patients with closed eyes and opened eyes, respectively. Subsets C and D were obtained during a presurgical planning and they correspond to nonfocal and focal signals, respectively. Subset E is composed of seizure signals also obtained during the presurgical planning [22].

The pathology of epilepsy imposes constraints in the neurological activity included in absence of seizures. The fact of closing the eyes might also impose a constraint in the activity in comparison to set A (see [21] and references therein). Thus, we expected a progressive increase of ρ from subset A to E because the more we advance, the more probability to have signals shaped by a non-random process.

Despite we expected a higher number of rejections in subset C than in subset B, we obtain the other way around. In measure E and S we used different parameters for analysing the signal as compared to the original study [22]. This could be one of the explanations of the increment of the rejections for this subset B or the low increment for subset C. Mean values of E and H progressively decrease while in S and F increase. This means that determinism increases from subset A to subset E. In any case, in this database the sensitivity of E and S is also higher than the sensitivity of F and H given the higher ρ.

We compare the specific values of the signals where the null hypothesis is rejected with the ones where it is not rejected. If we look at the specific values of the signals

with the rejected null hypothesis, E and H are lower than for the non-rejected ones while in S and F signals where we rejected the null hypothesis are higher than in the non-rejected ones. This was expected because epilepsy signals are shaped by a non-random process [2].

Finally, results from the database Bonn 2019 present a higher degree of variability between the three patients. A common behaviour in the three patients is that the lowest values of ρ are found in H. This is in contrast with what we observe in the robustness study and in the Bern-Barcelona 2012 database where F had the worst results. In general, ρ for H and F are lower than for the measures based on predictability. But in micro-recordings from patients 2 and 3 we get higher rates with F and H than with E. So, in these cases symbolic methods have a better performance. However, in the cases of F for these two patients and H for the patient 3, ρ is higher for the nonfocal signals than for the focal signals in contrast to the expected result given previous studies (see [24, 23] and references therein). We see that, overall, E and S have again a higher sensitivity but each patient behaves in a different way and we do not see a clear pattern.
We are not able to distinguish a common pattern in all patients. Moreover, we observe that mean values are almost equal between focal and nonfocal signals for the same type of electrode and values of the signals where we reject the null hypothesis are not very different from the signals where we do not reject it. The differences in rejection arise when we compare the macro and micro-recordings. The higher rates of rejection for the micro-electrodes could mean that they are located in the area where the focus is.

To conclude, symbolic methods present rates below significance level for each database except for micro-recordings of patients 2 and 3 of the last database, where they have a better performance than E. Moreover, methods based on predictability have a higher sensitivity than the symbolic methods in the three analysed databases. The exception is for patient 3 in Bonn 2019 database where we reject more times the null hypothesis by using F than by using E and S. F and H are not able to obtain the same amount of rejections than E and S. Recalling Schindler paper [13], while F and

H provide new information by looking at their temporal evolution, when we look at their value at the whole signal they provide less information than E and S.

As a future work, the use of a predictability measure together with H could work as a way of double-checking if our focal and nonfocal signals are deterministic. The procedure would be based on identifying which of the signals that are rejected in H are also rejected on E or S. In this way, we would have two measures that are not based on the same principle which may obtain similar results. Another approach could be the analysis of temporal segments of the signals with H as a first step for selecting starting points where it seems that determinism appears.

Bibliography

[1] World Health Organization: Epilepsy (2019). URL https://www.who.int/news-room/fact-sheets/detail/epilepsy. Web; visited 27-02-2019.

[2] Fisher, R. S. *et al.* Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ilae) and the international bureau for epilepsy (ibe). *Epilepsia* **46**, 470–472 (2005).

[3] Lehnertz, K. & Dickten, H. Assessing directionality and strength of coupling through symbolic analysis: an application to epilepsy patients. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **373**, 20140094 (2015).

[4] Fisher, R. S. *et al.* Operational classification of seizure types by the international league against epilepsy: Position paper of the ilae commission for classification and terminology. *Epilepsia* **58**, 522–530 (2017).

[5] Kwan, P. & Brodie, M. J. Early identification of refractory epilepsy. *New England Journal of Medicine* **342**, 314–319 (2000).

[6] Keränen, T., Sillanpää, M. & Riekkinen, P. J. Distribution of seizure types in an epileptic population. *Epilepsia* **29**, 1–7 (1988).

[7] Devinsky, O. Patients with refractory seizures. *New England Journal of Medicine* **340**, 1565–1570 (1999).

[8] Cascino, G. D. *Mayo Clinic Proceedings*, vol. 65, chap. Intractable partial epilepsy: evaluation and treatment, 1578–1586 (1990).

[9] Rosenow, F. & Lüders, H. Presurgical evaluation of epilepsy. *Brain* **124**, 1683–1700 (2001).

[10] Mula, M. & Monaco, F. Ictal and peri-ictal psychopathology. *Behavioural neurology* **24**, 21–25 (2011).

[11] Badawy, R., Macdonell, R., Jackson, G. & Berkovic, S. The peri-ictal state: cortical excitability changes within 24 h of a seizure. *Brain* **132**, 1013–1021 (2009).

[12] NYU ECoG: Department of Neurology, NYU School of Medicine, New York. Intracranial Electrodes (2019). URL `https://ecog.med.nyu.edu/intracranial-electrodes/`. Web; visited 1-05-2019.

[13] Schindler, K. *et al.* Forbidden ordinal patterns of periictal intracranial eeg indicate deterministic dynamics in human epileptic seizures. *Epilepsia* **52**, 1771–1780 (2011).

[14] Amigó, J. M., Keller, K. & Unakafova, V. A. Ordinal symbolic analysis and its application to biomedical recordings. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **373**, 20140091 (2015).

[15] Amigó, J., Kocarev, L. & Szczepanski, J. Order patterns and chaos. *Physics Letters A* **355**, 27 – 31 (2006).

[16] Fisher, R. S., Scharfman, H. E. *et al.* How can we identify ictal and interictal abnormal activity? In *Issues in Clinical Epileptology: A View from the Bench*, 3–23 (Springer, 2014).

[17] Ouyang, G., Dang, C., Richards, D. A. & Li, X. Ordinal pattern based similarity analysis for eeg recordings. *Clinical Neurophysiology* **121**, 694–703 (2010).

[18] Amigó, J., Zambrano, S. & Sanjuán, M. Detecting determinism in time series with ordinal patterns: A comparative study. *International Journal of Bifurcation and Chaos* **20** (2009).

[19] Bandt, C. & Pompe, B. Permutation entropy: a natural complexity measure for time series. *Physical Review Letters* **88**, 174102 (2002).

[20] Schindler, K., Gast, H., Goodfellow, M. & Rummel, C. On seeing the trees and the forest: Single-signal and multisignal analysis of periictal intracranial eeg. *Epilepsia* **53**, 1658–1668 (2012).

[21] Andrzejak, R. G. *et al.* Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E* **64**, 061907 (2001).

[22] Osario, I. & Andrzejak, R. G. *Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Engineering and Physics*, chap. 9 Nonlinear Time Series Analysis in a Nutshell, 125 – 138 (CRC Press, 2011).

[23] Naro, D., Rummel, C., Schindler, K. & Andrzejak, R. G. Detecting determinism with improved sensitivity in time series: Rank-based nonlinear predictability score. *Physical Review E* **90**, 032913 (2014).

[24] Andrzejak, R. G., Schindler, K. & Rummel, C. Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients. *Physical Review E* **86**, 046206 (2012).

[25] Mormann, F., Andrzejak, R. G., Elger, C. E. & Lehnertz, K. Seizure prediction: the long and winding road. *Brain* **130**, 314–333 (2006).

[26] Zaveri, H. P., Duckrow, R. B. & Spencer, S. S. On the use of bipolar montages for time-series analysis of intracranial electroencephalograms. *Clinical Neurophysiology* **117**, 2102–2108 (2006).

[27] Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena* **58**, 77–94 (1992).

[28] Schreiber, T. & Schmitz, A. Improved surrogate data for nonlinearity tests. *Physical Review Letters* **77**, 635 (1996).