Evaluation of EEG Dynamic Connectivity Around Seizure Onset with Principal Component Analysis

Evaluation of EEG dynamic connectivity around seizure onset with principal component analysis

Iris L. Soare and Javier Escudero, *Senior Member, IEEE*

***Abstract*— Seizures represent a brain activity state characterised by extended synchronised firing in multiple regions that prevent normal brain functioning. It is important to develop methods to distinguish between normal and abnormal synchronisation in epilepsy, as well as to localise the networks involved in seizures. To this end, we perform a preliminary investigation in the use of principal components analysis (PCA) to assess the change in dynamic electroencephalogram (EEG) connectivity before and after seizure onset. Source estimation was performed for an openly available EEG dataset from 14 patients with epilepsy. By applying PCA onto the EEG data processed into dynamic connectivity (dFC) matrices, we identified a set of connectivity topologies (eigenconnectivities) that explain high levels of variance in the dynamic connectivity. We compare the dimensionality reduction results obtained on source-level vs. scalp-level connectivity. We identified eigenconnectivities with differences in preictal vs. ictal activity and the brain networks associated with these activations. The work illustrates a data-driven approach for identification of topologies of brain networks that change with seizure onset.**

Clinical relevance **We identified networks that are significantly varying with preictal vs. ictal brain activity, some of which verify preexistent epilepsy markers in a data-driven way.**

I. Introduction

Epilepsy represents a chronic central neural system disorder characterised by a predisposition for recurring abnormal brain activity events – seizures – caused by highly synchronised neural firing [1]. These events can appear or continue in the absence of external stimuli and can be focal or generalised [1]. The interpatient variability in localisation and manifestation of the seizures represents one of the complex challenges associated with studying and treating epilepsy [2]. In contrast, awake healthy brain activity does not display such highly synchronised networks [1]. Nevertheless, it is characterised by highly varying alternating synchronised activity of multiple brain regions [1]. An appropriate level of brain functional connectivity is essential for healthy brain function.

To describe the variability in time and space of functional connectivity, the concept of dynamic functional connectivity (dFC) has been defined [3]. The dFC is a representation and analysis framework to assess how brain regions interact statistically in space and time. The dFC can be abstracted as a continuously transforming set of synchronizing and desynchronizing signals from brain regions whose time and space characteristics depend on the task, mental flexibility, health and other influences [3].

I.L. Soare was supported by an EPSRC PhD studentship.

I.L.S. and J.E. are with the School of Engineering, Institute for Digital Communications; and with the Muir Maxwell Epilepsy Centre, University of Edinburgh, United Kingdom. I.L.Soare@sms.ed.ac.uk

To deal with the fast-changing nature of dFC, we focus on electroencephalogram (EEG) data. This recording has high temporal resolution [4]. Moreover, it is highly relevant to epilepsy, as EEG is used in the standard clinical diagnosis process. Working at the EEG sensor level can be problematic due to problems like volume conduction [4] but source estimation can be done reliably, particularly with appropriate head models [4].

Linear methods such as correlation and coherence have commonly been used to estimate connectivity. Overall, ref. [5] proves that metrics with higher resistance to spatial leakage such as the imaginary part of coherence (iCOH) were more capable to identify true connectivity – than coherence (COH) and amplitude envelope correlation (AEC). The authors of [4] confirmed this result and pointed towards iCOH as being robust to spatial leakage effects while COH is not. While less robust metrics are more consistent across experiments, [4] consider that this consistency is caused by the repeated inclusion of the sources space configuration in the result, due to spatial leakage sensitivity. We conclude that iCOH is a desirable method for our study for its documented resistance to spatial leakage.

A characteristic trait of the dFC is its high number of variables: the brain activity changes fast across space and time. Dimensionality reduction algorithms are often used for simplifying and for gaining insight into systems properties [6]. This class of algorithms group data points into a lower dimension representation according to different statistical properties. One such very popular algorithm is principal component analysis (PCA) due to its linearity and cross-field applicability [6]. PCA (in addition to k-means clustering and l1-norm regularized dictionary learning algorithms) has also been used in [7] to obtain patient-wise dFC descriptors from EEG-functional magnetic resonance imaging (fMRI) data. However, we maintain that PCA is optimised for the identification of data variance [6]. In [8] the authors applied PCA on dFC obtained from fMRI data during rest and obtain activity descriptors for subjects with sclerosis vs healthy subjects. Compared to [8], we use a different imaging technique and metric applied to obtain dFC variance descriptors. Most importantly, our analysis compares the change in a small set of high-variance dFC descriptors during two states of brain activity.

In this study, we present a preliminary investigation of a publicly available dataset [9] to explore if we can find preictal vs. ictal activity variability within multiple subjects, as computed on their dynamic connectivities using PCA. We compare the quality of the dimensionality reduction analysis

on EEG data at source-level vs. scalp-level. We aim to understand the differences in connectivity at the two levels using statistical analysis. We obtain brain networks that are significantly varying with seizure onset.

II. Materials and Methods

A. Dataset

We used an openly available EEG dataset containing recordings from 14 patients (8 males, 6 females) with epilepsy [9]. The ages of the patients vary between 20 to 71 years. They experience focal seizures with different localisation. The electrodes, in number of 29 to 32, are arranged as per the 10-20 international system and the signals are sampled at 512 Hz. The dataset contains 47 seizures, annotated by clinicians.

As the openly available EEG dataset does not have MRI scans, we resorted to using MRI templates obtained from subjects of similar age [10]. We imported average 3T MRI templates [10]. We selected templates from both 'Adults' and 'Young adults' categories from the 5-year interval the most similar to each patient's age, while avoiding the ones affected by movement.

B. Preprocessing

For data preprocessing, we used the brain signal analysis software Brainstorm [11] throughout. For both source and sensor-level analysis, we selected the ictal activity segments and 5 minutes of preictal activity preceding each seizure, which we considered sufficient for studying seizure onset.

For the dFC analysis at source level, we marked the fiducial points of the MRI using the automated import option in Brainstorm. The anatomical and functional data of the 14 patients were linked. The raw signals were re-referenced using average referencing then bandpass filtered between 0.5 Hz and 45 Hz. The blinks and saccades were manually marked, identified by their shape on channels 'FP1' and 'FP2', and 'F7' and 'F8' respectively using the visual description in [12]. Some channels showed artifacts throughout and were fully removed. Then independent component analysis (Infomax ICA) was applied thus removing the components responsible for the aforementioned eye artifacts.

We thus developed an unconstrained volume model of the brain that was divided into 133 scouts using the AAL3 parcellation template [13], selected based on the results of [14]. The scouts are a subset of the AAL3 atlas. For the source estimation, we performed a minimum norm (MN) computation using the sLORETA variation for measure. We co-registered the subjects head models, thus reprojecting the sources of all the subjects onto Brainstorm's default cortex from the ICBM152 anatomy. For analysis at scalp-level, we bypassed the volume model computation step.

For both types of analysis, the cleaned EEG data was divided into 2.5 s segments (for time and frequency resolution considerations). Per seizure, this resulted in 120 preictal segments (of which we used 119, excluding the last one that was not consistently selected as 2.5 s) and a varying number of ictal data segments, depending on the seizure length. Of these, a few segments with strong artifacts were removed. A total of 6706 segments resulted, which are used for both source-level and scalp-level dFC analysis. Within each segment, the power spectrum density was estimated using the Welch method in windows of 1 s, with overlap of 75% up to 40 Hz. As the literature suggests that the delta band is one of the most meaningful for seizure analysis [15], only the averaged connectivity matrices between 0.5 Hz and 4 Hz was analysed.

C. Dynamic connectivity

We define a network or graph as $G = (V, E)$ where V is the set of nodes and E is the set of edges defined on $V \times V$. To characterise its change in time, we define the dynamic connectivity. This represents a set of consecutive descriptors that encompass the change of the graph topology. We describe each graph by its adjacency matrix $\mathbf{A}$, where each entry is the weight signifying the connection strength of a link between the nodes of the graph. We define one adjacency matrix for many consecutive time intervals.

For the purpose of estimating the dynamic connectivity on our graph, we are using the imaginary part of coherence (iCOH) [5], which we found to be robust to spectral leakage. Coherence (COH) is defined as

$$COH_{ij}(f) = \frac{G_{ij}(f)}{\sqrt{G_{ii}(f)G_{jj}(f)}}.$$

The term $G_{ij}(f)$ represents the cross-spectrum density, or the Fourier transform of the correlation between two signals. Finally, we extract the imaginary part of coherence (iCOH), a measure of phase synchronisation:

$$iCOH_{ij}(f) = \Im(COH_{ij}(f)).$$

D. Principal component analysis

We applied principal components analysis (PCA) on a patient-by-patient basis as per [8] to meet two purposes: 1. to reduce the data dimensionality; 2. to summarise the data using a small set of components and then explore them for patterns showing difference in their relationship with preictal vs. ictal activity. PCA represents a nonparametric algorithm that identifies the orthogonal directions of variance of the data (eigenvectors) and the proportion of contribution of each of these directions of variance (eigenvalues).

For this analysis, we extracted the connectivity matrices across the delta band. We had $S = 47$ sets of paired preictal followed by ictal connectivity matrices (one per seizure), that belong to $P = 14$ patients. From each dFC matrix the lower diagonal was removed as iCOH is a symmetric connectivity measure that results in undirected graphs. Next, the remaining data from each matrix was assembled into a vector and the vectors were concatenated together into S structures of the shape $\mathbf{C}_s = [\mathbf{C}_{s,preictal}, \mathbf{C}_{s,ictal}]$, where $s = \{1, 2, .., S = 47\}$. Each $\mathbf{C}_s$ structure contains a variable number of connectivity matrices reassembled into vectors, depending on the seizure length. Next we group the data per patient by concatenating the $\mathbf{C}_s$ formations into $\mathbf{X}_P = [\mathbf{C}_s, \mathbf{C}_{s+1}, ...]$. Therefore, we have a number of time segments M_P for each of the P patients, depending

on the number of seizures and seizure length. The data dimensions go from $N \times N \times M_P$ to $(N(N-1)/2) \times M_P$. The number of scalp or volume probes N varies with the analysis method. Thus, $N = 133$ for the source-level analysis where each probe is one MRI scout, while for the scalp-level analysis $N = 29$, $N = 31$ or $N = 32$ where each probe is an EEG channel. PCA requires normalisation so we subtracted the row-wise mean across each $\mathbf{X}_P$ structure, thus obtaining $\mathbf{X}'_P = [\mathbf{C}_s - \overline{\mathbf{C}}_s, \mathbf{C}_{s+1} - \overline{\mathbf{C}}_{s+1}, ...]$. PCA finds the eigenvectors and eigenvalues of the data covariance matrix: $\mathbf{X}'^{\mathrm{T}}_P \mathbf{X}'_P = \mathbf{U}_P \mathbf{\Lambda}_P \mathbf{U}^{\mathrm{T}}_P$. The data is fed as a number of observations (connectivity values) times variables (time segments). The resulting highest eigenvalues indicate the directions of largest variance. We expand the obtained eigenvectors back into M_P matrices which we call eigenconnectivities [8].

We proceed to estimate to what extent our extracted dFC descriptors show similarity to the most prevalent eigenconnectivities. For each patient, we get a set of weights $\mathbf{W}_P$, as per $\mathbf{X}'_P = \mathbf{U}_{P,1:L} \mathbf{W}_P$, where L is the number of eigenconnectivities we will keep for analysis. These time courses show how our connectivity matrices evolve as points in the PCA-extracted latent space.

E. Statistical analysis

To identify meaningful eigenconnectivities as descriptors for the variation of the signal before and during seizure, we apply statistical analysis onto the time-varying weights $\mathbf{W}_P$. We apply the following steps twice, for the connectivity descriptors at source-level and scalp-level.

For each patient, we divide the weights associated with the reduced set of eigenconnectivities $E_{P,1:L}$ into two groups: the preictal $W_{P,1:L,preictal}$ and ictal $W_{P,1:L,ictal}$ segments weights. We then apply an effect size test on the two aforementioned categories for each patient for each set of weights:

$$ES_{P,L} = \left| \frac{m_{P,L,preictal} - m_{P,L,ictal}}{\sqrt{s^2_{P,L,preictal} + s^2_{P,L,ictal}}} - 0.5 \right|.$$

We note with m the means of each compared group and with s^2 their variances. If any of the compared weights distributions are significantly different, then the output of that effect size test diverges from 0. The purpose of our experiment is to get any indication of variation between the preictal and the ictal segments weights, to understand along which parameters the variability of our data is the highest.

At the next step, the top three components showing the highest variation at the previous statistical test are concatenated in $W_{P,L_{max},preictal}$ and $W_{P,L_{max},ictal}$. Another effect size test is applied between the two sets of weights. The purpose of this analysis is to understand how well the variation between preictal and ictal connectivity descriptors is identified by the principal components.

III. Results and Discussion

To decide how many descriptors to analyse in more detail, we performed a cumulative sum of the system eigenvalues for each patient. We found that the first eigenvalue explains around 10% in data variability for each patient. The first ten eigenvalues explain between 25% and up to 45% of the variability in data. We continue our analysis using the first $L = 10$ eigenconnectivities only.

To test for meaningful variation in the dFC descriptors at seizure onset, we look for high mean and variance between the preictal and the ictal data, as per the effect size test. The hypothesis is considered as valid for a patient if the resulting distributions do not intersect the origin.

Fig. 1 compares the quantified change in connectivity at source-level (blue) vs. scalp-level (red). We observe that the effect sizes vary with subject and analysis method. Overall, the results at source level indicate lower mean difference between preictal and ictal data, suggesting smaller change in the connectivity with seizure onset as opposed to the scalp-level analysis. At the same time, the clustering on the scalp data appears more sparse.

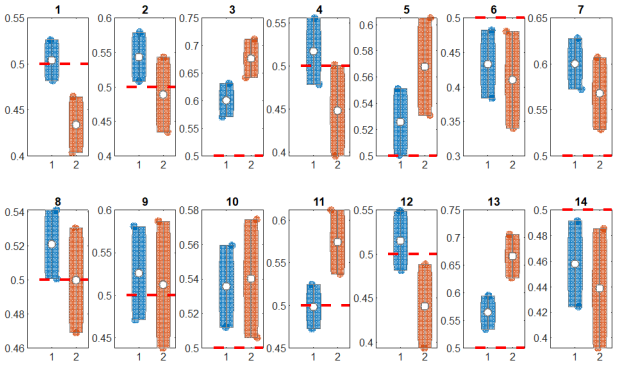

Fig. 1: The variation of preictal vs. ictal weights for the eigenconnectivities with the highest effect size for each patient. In blue, results on source data; in red, results on scalp data. A result is considered meaningful if the confidence interval of effect-size result does not go through the origin.

Some patients show good results for both methods; some for one of either methods. One possibility to be considered is whether seizure can be better localised at source- or scalp-level on a patient-by-patient basis depending on the affected dFC network [17]. For some patients the results are not meaningful altogether which warrants for further analysis, for example by applying other dimensionality reduction techniques.

Further, we look at the eigenconnectivities with highest seizure onset variation for Patient 3, chosen based on the effect size test results as per Fig. 1. We compare the source vs. scalp-level eigenconnectivities in terms of synchronised brain regions. The structures $\mathbf{X}'_P$ were normalised before the principal components were extracted. As such, dFC networks with high brain areas synchronisation levels are identified for both the top 50% positive values and the lowest 50% negative values. This is possible because PCA has a sign ambiguity.

We compare the eigenconnectivity with the greatest effect size value for both source- and scalp-level analysis, as per Fig. 2A and B, namely $E_{3,2}$. At source-level, $E_{3,2}$ indicates one network formed of the left calcarine (47), left cuneus (49) and the left occipital lobe (53, 55, 57). displayed

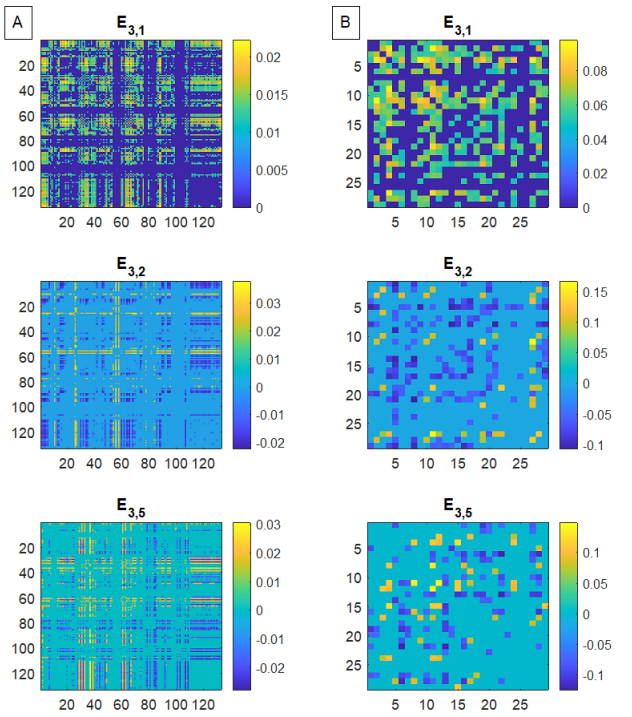

Fig. 2: The top three eigenconnectivities encompassing the preictal-ictal data variability for Patient 3 ($E_{3,1}, E_{3,2}, E_{3,5}$), according to the effect size test. Only the weights within 50% of the maximum and minimum values are shown. A. Source-level eigenconnectivities. The scouts are ordered alphabetically by names, left region followed by right region, and represent a subset of the AAL3 atlas as extracted with our head model from Brainstorm [13], update. B. Scalp-level eigenconnectivities. The 29 channels are ordered as per the original clinical files [9].

with positive values; another one is formed of the anterior cingulate cortex (151-156), sections of the left frontal lobe (9, 11, 21), left orbital gyri (25, 27, 29, 31), left putamen (77), the amygdala (45-46), the hippocampus (41-42). The sensor-level $E_{3,2}$ shows one network of synchronisation between channels 'F3', 'C3', 'F7', 'FC5', 'CP1', 'P4', 'CP6', 'F10', which consists of mostly frontal and central left hemisphere regions of and central-pariental right hemisphere regions. Another network was identified between channels 'FP1', 'P3', 'O1', 'F7', 'T5', 'FC1', 'Fz', Cz', 'Pz', 'F4', 'P4', 'O2', and 'FP2' which shows activity more evenly distributed across the scalp. Such activity could represent high dFC synchronisation.

Overall, the eigenconnectivities identified with the two methods seem different. The sensor-level analysis is invariably less precise due to the lower number of channels, the volume conduction effect, and the distance from deep brain sections. Source-level analysis includes other sources of error, such as potentially inexact volume models resulting from EEG data co-registration with generic MRI templates. For further insight, a statistical analysis of dFC seizure descriptors correspondence should be undertaken across all patients. However, such differences between source and sensor-level analysis are found in the literature [16]. Further work in this direction is required.

The literature on seizure localisation is vast, as epilepsy is characterised by high between-patient variability. Clinically [17] and analysis-wise [18], the frontal lobe, the temporal lobe, the parietal lobe, the occipital lobe, the cingulate gyrus and others are attested as seizure locations. In terms of connectivity patterns, some studies show increased, others decreased, and others show complex changes in connectivity during epileptic seizures [2]. Our results include some of the well-documented epileptic dFC seizures descriptors for one patient, while further analysis could provide deeper insight.

To conclude with, our analysis pipeline for EEG dFC identified networks that are significantly varying with seizure onset, some of which indicate similarity with preexistent epilepsy markers. For our analysis pipeline, PCA performed overall better on scalp-level EEG data. Going forward, we plan to apply other dimensionality reduction techniques on epilepsy dFC EEG data to confirm our results. Moreover, the activity as a set of points evolving through a latent space will be further researched, including the causality between data points in this space.

References

[1] J. Velazquez and R. Wennberg, "Metastability of brain states and the many routes to seizures: Numerous causes, same result," *Recent Res. Devel. Biophys.* Jan. 2004.

[2] P. van Mierlo et al., "Network Perspectives on Epilepsy Using EEG/MEG Source Connectivity," *Front. Neurol.*, vol. 10, p. 721, Jul. 2019.

[3] U. Gschwandtner et al., "Dynamic Functional Connectivity of EEG: From Identifying Fingerprints to Gender Differences to a General Blueprint for the Brain's Functional Organization," *Front. Neurosci.*, vol. 15, 2021.

[4] K. Mahjoory et al., "Consistency of EEG source localization and connectivity estimates," *NeuroImage*, vol. 152, pp. 590–601, 2017.

[5] M. Yu, "Benchmarking metrics for inferring functional connectivity from multi-channel EEG and MEG: A simulation study," *Chaos*, vol. 30, no. 12, p. 123124, Dec. 2020.

[6] J. P. Cunningham and Z. Ghahramani, "Linear Dimensionality Reduction: Survey, Insights, and Generalizations," *J. Mach. Learn. Res.* p. 42, 2015.

[7] R. Abreu et al., "Identification of epileptic brain states by dynamic functional connectivity analysis of simultaneous EEG-fMRI: a dictionary learning approach," *Sci. Rep.*, vol. 9, p. 638, Jan. 2019.

[8] N. Leonardi et al., "Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest," *NeuroImage*, vol. 83, pp. 937–950, Dec. 2013.

[9] P. Detti. Siena Scalp EEG Database (version 1.0.0), PhysioNet, 2020.

[10] P. T. Fillmore et al., "Age-specific MRI brain and head templates for healthy adults from 20 through 89 years of age," *Front. Aging Neurosci.*, vol. 7, 2015.

[11] F. Tadel et al., "Brainstorm: A User-Friendly Application for MEG/EEG Analysis," *Comput. Intell. Neurosci.*, vol. 2011, pp. 1–13, 2011.

[12] O. G. Lins et al., "Ocular artifacts in EEG and event-related potentials I: Scalp topography," *Brain Topogr.*, vol. 6, no. 1, pp. 51–63, Sep. 1993.

[13] E. T. Rolls et al., "Automated anatomical labelling atlas 3," *NeuroImage*, vol. 206, p. 116189, Feb. 2020.

[14] A. Y. Revell et al., "A Framework for Brain Atlases: Lessons from Seizure Dynamics," bioRxiv 2021.06.11.448063, Oct. 2021.

[15] Y. Liang et al., "Altered Functional Connectivity after Epileptic Seizure Revealed by Scalp EEG," *Neural Plast.*, vol. 2020, p. 8851415, Nov. 2020.

[16] M. Lai et al., "A comparison between scalp- and source-reconstructed EEG networks," *Sci. Rep.*, vol. 8, p. 12269, Aug. 2018.

[17] F. A. Chowdhury et al., "Localisation in focal epilepsy: a practical guide," *Pract. Neurol.*, vol. 21, no. 6, pp. 481–491, Dec. 2021.

[18] D. F. Abbott et al., "Editorial: Functional Brain Mapping of Epilepsy Networks: Methods and Applications," *Front. Neurosci.*, vol. 13, 2019.