Epilepsy seizure onset detection applying 1-NN classifier based on statistical parameters

Ivanna Zorgno[‡§1], María Cecilia Blanc[‡§2], Simon Oxenford[‡§3], Francisco Gil Garbagnoli[‡§4], Carlos D'Giano [†5] and Antonio Quintero-Rincón [‡6]

[§]*All authors contributed equally to this work*
[‡]*Department of Bioengineering, Instituto Tecnológico de Buenos Aires (ITBA)*
Avenida Eduardo Madero 399, C1106ACD CABA, Argentina
[1]izorgno@itba.edu.ar
[2]mblanc@itba.edu.ar
[3]soxenford@itba.edu.ar
[4]fgilgarbagnoli@itba.edu.ar
[6]aquinter@itba.edu.ar
[†]*Fundación Lucha contra las Enfermedades Neurológicas Infantiles (FLENI)*
Montañeses 2325, C1428AQK CABA, Argentina
[5]cdigiano@fleni.org.ar

***Abstract*—Epilepsy is a disease caused by an excessive discharge of a group of neurons in the cerebral cortex. Extracting this information using EEG signals is an ongoing challenge in biomedical signal processing. In this paper, a new method is proposed for onset seizure detection in epileptic EEG signals based on parameters from the t-location-scale distribution coupled with the variance and the Pearson correlation coefficient. The 1-nearest neighbor classifier achieved a 91% sensitivity (True positive rate) and 95% specificity (True Negative Rate) with a delay of 4.5 seconds (on average) in the 45 signals analyzed, which suggests that the proposed methodology is potentially useful for seizure onset detection in epileptic EEG signals.**

***Resumen*— La epilepsia es una enfermedad causada por una descarga excesiva de un grupo de neuronas en la corteza cerebral. Obtener esta información a partir de un EEG es un desafío continuo en el procesamiento de señales biomédicas. En este artículo se propone un nuevo método para la detección del comienzo de una crisis epiléptica en señales de EEG basado en la distribución *t-location-scale* junto con la varianza y el coeficiente de correlación de Pearson. La clasificación 1-vecino más cercano utilizado alcanzó una sensitividad (Verdaderos positivos) del 91% y una especificidad (Verdaderos negativos) del 95% con un retraso en promedio de 4.5 segundos en las 45 señales analizadas, lo que sugiere que es una metodología potencialmente útil para la detección del comienzo de una crisis epiléptica en señales de EEG.**

I. Introduction

Neurons generate electrochemical impulses that act on other neurons, glands and muscles to produce human thought, feelings and action. In epilepsy the normal pattern of neuronal activity is disturbed causing strange sensations, emotions and behaviors and can sometimes lead to convulsions, muscle spasms and loss of consciousness [1].

Epilepsy is a chronic brain disorder that results from the hyper-excitability of neurons. It is the tendency to have recurrent, unprovoked seizures [2]. Electroencephalography (EEG) is a non-invasive and widely available biomedical modality that is used to make a diagnosis of the latter.

According to the International League Against Epilepsy (ILAE) [3] an epileptic seizure is a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain. Elements defining an epileptic seizure include: mode of onset and offset, clinical manifestations and abnormal enhanced synchrony [4].

The t-location-scale distribution is a statistical model for univariate and multivariate signals that describe its features through three parameters estimated by maximum likelihood: location (μ), shape (ν) and a non-negative scale (σ). This distribution was recently applied in spike-and-wave pattern recognition in epileptic signals [5]. Based on the results of this study and [9], [11], [12], the following question arose: *what if this distribution could be used to detect a seizure onset in epileptic EEG signals?*

A k-nearest neighbor classifier (kNN) has been successful in a large number of classification problems such as handwritten digits, satellite, image signal processing and biosignal patterns. kNN fits a specific point in the EEG data with the k-nearest neighbor EEG data points in the training set. For 1NN this point depends only of 1 single other EEG point. Therefore the similarity between two EEG points is established based on the fitting of the features extracted from the EEG signals. See [5]–[8] for some recent works in epileptic EEG signals and [10] for a fast approximate automatic algorithm configuration.

In this paper, a new method is proposed based on parameters of the t-location-scale distribution, the variance (σ^2) and the Pearson correlation coefficient (ζ) as features for onset seizure detection in epileptic EEG signals. In addition the 1NN classifier was incorporated based on its classification performance; it is simple, accurate, fast, and has low bias [14].

The remaining paper is organized as follows; Section II describes the methodology, in Section III the methodology is applied to real EEG data and Sections III and IV include the results, discussion, conclusions and future works.

II. METHODOLOGY

Let $\boldsymbol{X} \in \mathbb{R}^{N\times M}$ denote the matrix gathering M EEG signals $\boldsymbol{x}_m \in \mathbb{R}^{N\times 1}$ measured simultaneously on different channels and at N discrete time instants. In this research the EEG signals of nine patients were analyzed, twenty three bipolar channels per patient. The proposed methodology is composed of three stages. To begin with, 5 channels per patient, where the epileptic seizure was most visible, were selected by an expert neurologist. Since a matrix $\boldsymbol{X}$ exists per patient, M=5 and N is determined by each individual patient according to the length of the seizure. Secondly, two rectangular sliding empirical windows $\boldsymbol{\omega}(n)_i = 1$, each with a length of 3 seconds ($n \in [i : i + 3]$) and an overlap of 50% were created and applied to each signal, such that $\boldsymbol{x}_i = \boldsymbol{\omega}_i \boldsymbol{x}_m$ and $\boldsymbol{x}_{(i+1.5)} = \boldsymbol{\omega}_{(i+1.5)} \boldsymbol{x}_m$, with $1 \leq m \leq M$. The resulting time segmentation of the original channel was analyzed utilizing 5 feature parameters: μ, σ, ν from t-location-scale distribution, σ^2 (variance) and ζ (Pearson correlation coefficient). Therefore, there are 5 parameters that correspond to each window applied. The third stage includes the implementation of the 1-nearest-neighbor classifier using the feature predictor vector $\boldsymbol{\theta} = [\mu, \sigma, \nu, \sigma^2, \zeta]$ associated with each window time segment and the response vector for each signal, both composed of seizure (1) or non-seizure (0). It is important to note that the response vector was generated with the information provided by an experienced neurologist that located the seizure onset and offset. Therefore, the classification resulted in a vector composed of 1's and 0's for each channel. The explanations of parameters used can the found below.

A. t-location-scale distribution

The t-location-scale distribution is a statistical model that belongs to the location-scale family formed by translation and rescaling of the Student's t-distribution.

The probability density function (PDF) of a location-scale distribution, is given by

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

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

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

Therefore applying (2) to (1), the probability density function (PDF) of the t-location-scale is obtained:

$$f_{\text{tls}}(x|\mu,\sigma,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{\nu+\left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right]^{-\left(\frac{\nu+1}{2}\right)} \tag{3}$$

where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter, $\nu > 0$ is the shape parameter, and $\Gamma(.)$ is the Gamma function. See [5] for more details.

B. Pearson correlation coefficient

The Pearson correlation coefficient (ζ) is a variation of the basic correlation equation, it is normalized so the outcome lies between ± 1. The modification to the covariance equation for two sliding windows ($\zeta_{x_i x_{(i+1.5)}}$) is as follows:

$$\frac{1}{(\mathcal{W}-1)\sigma^2_{x_i}\sigma^2_{x_{(i+1.5)}}}\sum_{n=1}^{\mathcal{W}}(x_{i_n}-\overline{x_i})(x_{(i+1.5)_n}-\overline{x_{(i+1.5)}}) \tag{4}$$

Where x_i and $x_{i+1.5}$ are the two sliding windows, $\mathcal{W}$ is the length of the window (both windows share the same length), $\sigma^2_{x_i}$ and $\sigma^2_{x_{(i+1.5)}}$ are the variances and $\overline{x_i}$ and $\overline{x_{(i+1.5)}}$ are the means. If the correlation value is equal to $+1$ then the two signals are identical, -1 implies that the signals are the exact opposites and if it is equal to 0 there is no correlation [13].

C. Variance

The variance is a measure of signal variability irrespective of its average given by

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_{(i+1.5)}-\overline{x})^2 \tag{5}$$

where N represent the length of the total signal, $x_{i+1.5}$ is the sliding window and $\overline{x}$ is the mean of that given time laps. Note that (5) is applied to one window as it was moved along the whole EEG signal.

D. 1-nearest neighbor classifier (1NN)

k-Nearest-neighbor fit uses the observations in the binary training sets $\boldsymbol{\theta}_1 = [\mu_1, \sigma_1, \nu_1, \sigma_1^2, \zeta_1]$ for seizure events and $\boldsymbol{\theta}_0 = [\mu_0, \sigma_0, \nu_0, \sigma_0^2, \zeta_0]$ for non-seizure events, closest in input space to x to form Y.

$$\boldsymbol{Y}(x) = \frac{1}{k}\sum_{x_i \in N_k(x)} y_i \tag{6}$$

Where $N_k(x)$ is the neighborhood of x defined by the k closest points x_i in the training sample. The idea is to find the k observations with x_i closest to x in input space, and average their responses. The Euclidean distance metric in a feature space is given by:

$$d_i = ||x_i - x_0|| \tag{7}$$

In other words, in kNN given a query point x_0, the k training points x_i with $i = 1, ..., k$ closest in distance to x_0 are found, and then classified using the majority vote among the k neighbors. If $k = 1$ a 1NN classifier is used and $\boldsymbol{Y}$ is assigned the value y_i of the closest point x_i to x in the training data (based on an Euclidean distance) [14].
The following [14], [15] contain a comprehensive treatment of the mathematical properties of nearest neighbors classifiers.

III. EXPERIMENTS

A. Dataset

In this section the proposed methodology is evaluated using the Children Hospital Boston database. This dataset consists of 36 bipolar 256Hz EEG recordings from pediatric patients suffering from intractable seizures [16]. In this work 45 recordings chosen from 9 different patients were used. Each recording contains a seizure event, whose onset time has been labeled by an expert neurologist. These annotations

were used to extract a short epoch from each recording that contains the seizure and short time intervals before and after the crisis. To be exact, if the seizure lasts a given time interval Δ than the total length of the signal is 3Δ such that the same given time interval is taken before and after the seizure. In the case that the information after or before the seizure is shorter that Δ than the information available was taken. It is important to note that the patients suffered one seizure in the recording used.

The electrodes where placed according to the International 10-20 system which can be seen in Table I and Figure 1.

TABLE I
DURATION OF THE 45 SEIZURES USED IN THIS STUDY.

Patient	Duration (sec)	5 bipolar channels used per patient
1	90	Fp1-F3, Fp2-F8, F8-T8, T8-P8, FT9-FT10
2	101	Fp1-F3, F8-T8, T8-P8, P8-O2, FT9-FT10
3	64	C4-P4, T8-P8, Fz-Cz, Cz-Pz, FT9-FT10
4	53	F7-T7, P7-O1, T8-P8, P7-T7, FT9-FT10
5	120	F4-C4, C4-P4, Fz-Cz, Cz-Pz, FT10-T8
6	117	P3-O1, F4-C4, T8-P8, P8-O2, FT10-T8
7	86	F7-T7, P7-O1, P3-O1, F4-C4, Fz-Cz
8	143	F7-T7, F4-C4, C4-P4, P7-T7, FT9-FT10
9	64	Fp1-F3, Fp2-F8, F8-T8, T8-P8, FT9-FT10

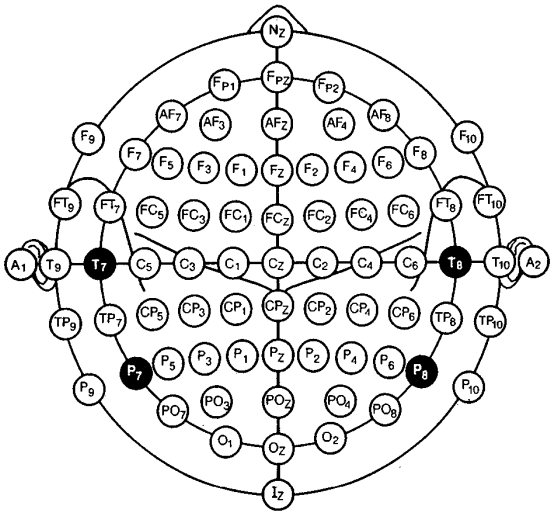

Fig. 1. Location and nomenclature of electrodes, as standardized by the American Electroencephalographic Society, from [17].

Figure 2 displays how the t-location-scale distribution fits the data utilized. On the left hand side, the epileptic signal time domain is found; before, during and after the seizure, with the corresponding histogram on the right hand side. Note how the values of the parameters μ, σ and ν vary in each signal time window.

B. Results and discussion

To check the quality of our seizure onset detection classifier, the set was trained off-line with 40 signals with the 5 features for each vector $\boldsymbol{\theta}_1 = [\mu_1, \sigma_1, \nu_1, \sigma_1^2, \zeta_1]$ for seizure events and $\boldsymbol{\theta}_0 = [\mu_0, \sigma_0, \nu_0, \sigma_0^2, \zeta_0]$ for non-seizure events. The classifier, using Leave-one-out cross-validation, was applied to the training signals, where leave-one-out refers to a patient which has 5 channels. After each training, the remaining 5 bipolar signals (data from

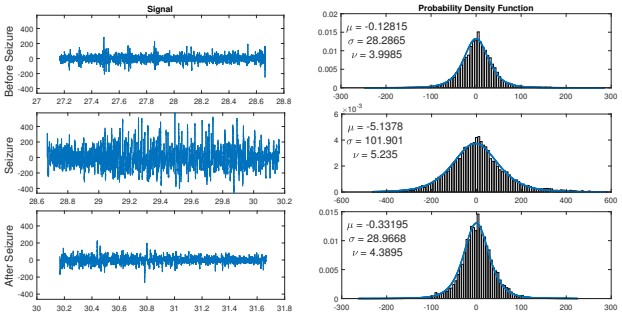

Fig. 2. Signal time domain example show the evolution of the signal through time in seconds (before, during and after the seizure) fitted by the t-location scale distribution and the value of the three parameters for each time window. Note the high amplitude in the time domain during the seizure and the different values for the parameters μ, σ and ν.

one patient), which where not included in the training but the seizure onset and offset were known, where predicted. This prediction was done 9 times in order to predict the 5 channels corresponding to each patient, resulting in a total of 45 predictions. The percentage of correct classifications was analyzed, in terms of sensitivity, specificity and accuracy. The values obtained are; 91% sensitivity (True positive rate), 95% specificity (True negative rate) and 95% accuracy for seizure on set detection in epilepsy signals. For each iteration, the prediction speed was 230000 obs/sec and training time was 4.4014 sec for 12480 observations.

The Figures 3 to 5 show all the scatter plots from the parameters estimated in pairs in the leave-one-out cross validation. The objective is to find parameters that distinguish seizure events (red) and non-seizure events (blue) and furthermore permit onset detection in epileptic EEG signals.

In Figure 3, for t-location-scale parameters, both seizure events and non-seizure events have a range of values that allow them to be differentiated in the combinations of σ, ν and μ. For μ vs. σ, a cluster of points can be found for non seizure events where σ takes relatively low values and μ tends to zero. For the seizure, the points are spread out and separated from the blue cluster, σ takes much higher values and μ belongs to a wider range of negative and positive values. In the case of ν and σ, σ allows to differentiate seizure and non seizure, as σ has low values for blue and high values for red, while ν has a relatively similar nature for both cases.

Figure 4 show the relationship between t-location-scale parameters and σ^2. For μ vs. σ^2 it can be observed that σ^2 is set at high numerical values for seizure events with respect to μ and that non-seizure events are concentrated in a small range near zero. With respect to ν vs. σ^2, σ^2 is able to differentiate non-seizure events from seizure events as it takes high values for the latter and low values for the former. For σ and σ^2 non-seizure events are set at low values of both parameters (cluster), while seizure events are set at high values of σ^2 with respect to σ.

In Figure 5 t-location-scale parameter vs ζ can be analyzed. It is visible that ζ, in all the cases, acquires numerical values around zero for seizure events. The subfigure that demonstrates the greatest segmentation between red and blue is ζ vs σ, for seizures σ values are high and ζ tends to zero.

As for non seizures, σ values are low and ζ can be found around a wider numerical range.

The research reflected in this paper using a feature vector $\boldsymbol{\theta} = [\mu, \sigma, \nu, \sigma^2, \zeta]$ in 45 epileptic signals for the classes seizure and non-seizure, suggests that the proposed methodology based on the t-location-scale distribution coupled with the variance and the Pearson correlation coefficient and the 1NN-based classifier, is potentially useful for seizure onset detection in epileptic EEG signals. The signals studied have a delay on average of 4.5 seconds caused by the length of the empirical windows and their overlap. The latter is an acceptable time, from a clinical point of view, in automatic detection systems in EEG extracranial signals [18].

IV. Conclusion

This work presents a new method based on parameters of the t-location-scale distribution coupled with the variance and the Pearson correlation coefficient as features for onset seizure detection in epileptic EEG signals with a delay of 4.5 seconds in average. The performance of the proposed method was evaluated on a real dataset containing 45 epileptic signals achieving an accuracy of 95% through the 1NN-based classifier, which suggests that the proposed methodology is potentially useful for seizure onset detection in epileptic EEG signals. Even though the values of specificity, sensitivity and accuracy are not optimum, the user interface of the algorithm provides clear visual detection of the onset, given that it is handled by an expert on the matter.

Perspective for future work include; determining the optimal window length, improving the specificity, sensitivity and accuracy, predict the seizure onset detection in new signals without ground truth, estimate the weights of the 5 features parameters used in order to localize a possible spread of the seizure and relate them to the principal components analysis (PCA) of the EEG raw.

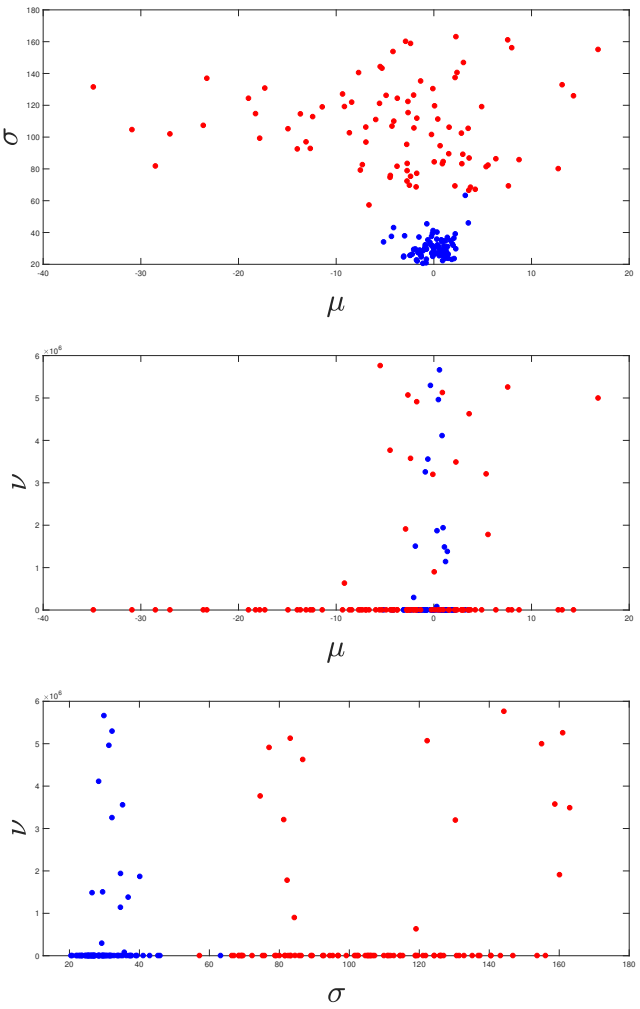

Fig. 3. Scatter plots example for t-location-scale parameters: location (μ), shape (ν) and a non-negative scale (σ). Both seizure events (red circles) and non-seizure events (blue circles) have a range of values that allow them to be differentiated clearly.

Acknowledgements

Part of this work was supported by the I+D2017 grant from Instituto Tecnológico de Buenos Aires.

References

[1] J. N. Parker and P. M. Parker, *The Official Patient's Sourcebook on Seizures and Epilepsy*. The MIT Press, 2003.

[2] W. H. Smithson and M. C. Walker, *ABC of Epilepsy*. BMJ Books, 2012.

[3] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. E. Jr., L. Forsgren, J. A. French, M. Glynn, D. C. Hesdorffer, B. Lee, G. W. Mathern, S. L. Moshe, E. Perucca, I. E. Scheffer, T. Tomson, M. Watanabe, and S. Wiebe, "ILAE official report: A practical clinical definition of Epilepsy," *Epilepsy*, vol. 55, no. 4, pp. 475–482, 2014.

[4] H. O. Luders, *Textbook of Epilepsy Surgery*. CRC Press, 2008.

[5] A. Quintero-Rincón, J. Prendes, V. Muro, and C. D'Giano, "Study on spike-and-wave detection in epileptic signals using t-location-scale distribution and the k-nearest neighbors classifier." *IEEE URUCON Congress on Electronics, Electrical Engineering and Computing*, vol. 2017, pp. 1–4, 2017.

[6] K. Shazadi, S. Petrovski, A. Roten, H. Miller, R. Huggins, M. Brodie, M. Pirmohamed, M. Johnson, A. Marson, T. O'Brien, and G. Sills, "Validation of a multigenic model to predict seizure control in newly treated epilepsy," *Epilepsy Research*, vol. 108, no. 10, pp. 1797–1805, 2014.

[7] T. S. Kumar, V. Kanhanga, and R. B. Pachori, "Classification of seizure and seizure-free EEG signals using local binary patterns," *Biomedical Signal Processing and Control*, vol. 15, pp. 33–40, 2015.

[8] U. Acharya, S. L. Oh, Y. Hagiwara, J. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers in Biology and Medicine*, vol. 17, pp. 1–9, 2017.

[9] J.-D. Zhu, C.-F. Lin, S.-H. Chang, J.-H. Wang, T.-I. Peng, and Y.-Y. Chien, "Analysis of spike waves in epilepsy using hilbert-huang transform," *Journal of Medical Systems*, no. 1, p. 170, 2015.

[10] M. Muja, and D.-G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," *VISAPP International Conference on Computer Vision Theory and Applications*, pp. 331–340, 2009.

[11] Z. Haneef, S. Chiang, H. J. Yeh, J. Jerome Engel, and J. M. Stern, "Functional connectivity homogeneity correlates with duration of temporal lobe epilepsy," *Epilepsy Behavior*, vol. 46, pp. 227–233, 2015.

[12] M. Paldino, Z. Chu, M. Chapieski, F. Golriz, and W. Zhang, "Repeatability of graph theoretical metrics derived from resting-state functional networks in paediatric epilepsy patients," *BioMed Research International*, vol. 90, no. 1074:20160656, 2017.

[13] J. L.Semmlow and B. Griffe, *Biosignal and Medical Image Processing*. CRC Press, 2014.

[14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining Inference and Prediction*. Springer, 2016.

[15] E. Alpaydin, *Introduction to Machine Learning*. The MIT Press, 2014.

[16] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, and H. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation 101(23):*, vol. 101, no. 23, pp. e215–e220, 2000.

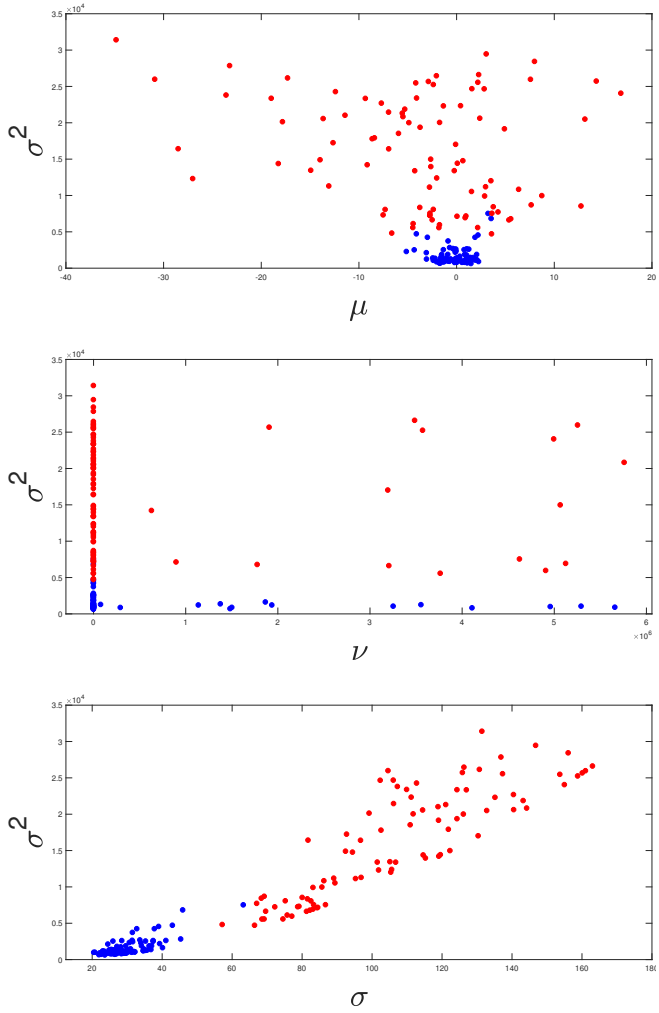

Fig. 4. Scatter plots example for t-location-scale parameters (location (μ), shape (ν) and a non-negative scale (σ)) and variance (σ^2). σ^2 is set at high values for seizure events (red circles). For non-seizure events (blue circles) μ, σ and ν are set low values with respect to σ^2.

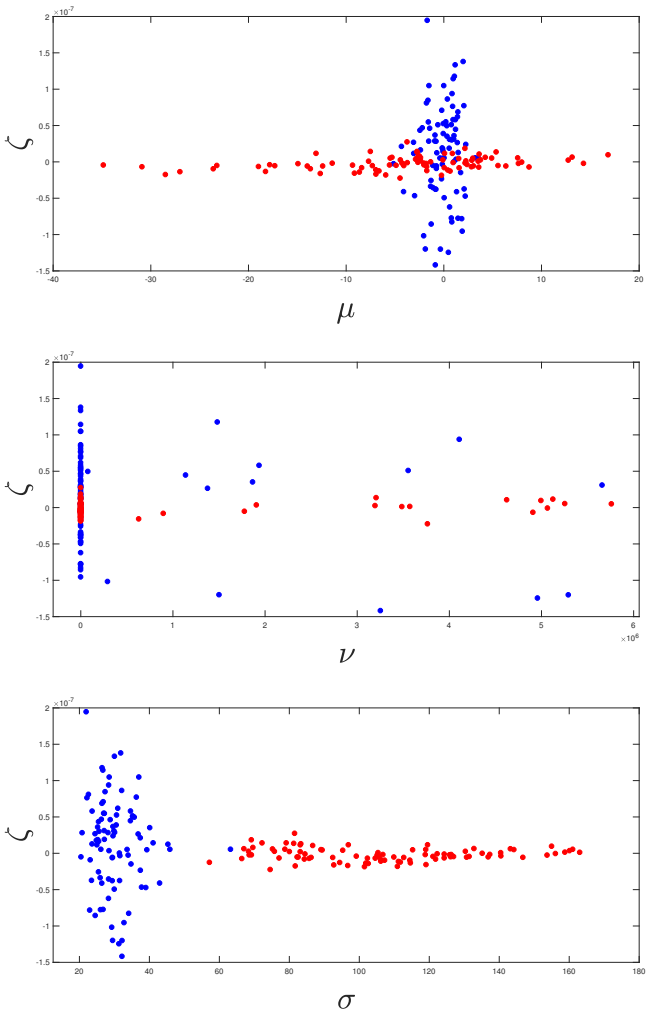

Fig. 5. Scatter plots example for t-location-scale parameters (location (μ), shape (ν) and a non-negative scale (σ)) and Pearson correlation coefficient (ζ). ζ is concentrated in the center around zero for seizure events (red circles), while μ, σ and ν are dispersing outside the center for non-seizure events (blue circles).

[17] F. Sharbrough, G. Chatrian, R. Lesser, H. Luders, M. Nuwer, and T. Picton, "American electroencephalographic society guidelines for nomenclature, standard electrode position," *Journal of clinical Neurophysiology*, vol. 8, no. 2, pp. 200–202, 1991.

[18] S. Nasehi, H. Pourghassem, " A novel fast epileptic seizure onset detection algorithm using general tensor discriminant analysis," *Journal of Clinical Neurophysiology* vol. 30, no. 4 362–370, 2013.