35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Estimation of temporal scales of variation in long-term scalp electroencephalograms from epilepsy patients

Catherine Stamoulis, *Member IEEE*, Bernard S. Chang

***Abstract*— Long-term neurophysiological recordings, such as scalp encephalograms (EEG), have been routinely used in studies that aim to characterize dynamic changes in brain activity associated with normal biological processes, such as sleep, but are also becoming increasingly common for clinical evaluation of patients with neurological disorders, such as epilepsy. Analysis of non-stationary recordings from multiple days poses new signal processing challenges, in regard to algorithm efficiency and computational cost, as well as adequate dimensionality data reduction. We compared four approaches for estimating the underlying temporal dynamics of long-term recordings from patients with medically refractory epilepsy: (i) model order selection using the minimum description length principle, (ii) approximate entropy, (iii) mutual information, and (iv) Detrended Fluctuation Analysis (DFA). Individual approaches were found to be sensitive only to specific scales of variation. Approximate entropy and mutual information were sensitive to local dynamics, whereas dynamic model order estimation captured only slowly varying dynamics. DFA was sensitive to multiple temporal scales.**

I. INTRODUCTION

Improvement in technologies for data collection and storage, and emergence of novel algorithms for computationally efficient data analysis, including dimensionality reduction and compressed sensing, have allowed the generation and analysis of very large datasets. Long-term monitoring of physiological processes have provided important insights into human health and disease. For example, long-term monitoring of brain activity during human sleep has provided fundamental information of the neural correlates of sleep, wakefulness and the transition between the two processes [1]. Continuous recordings in patients with neurological disorders, such as epilepsy, are becoming increasingly common for diagnostic purposes and decisions for further clinical care. For example, epilepsy patients often undergo long-term non-invasive electrophysiological studies, in order to assess their candidacy for surgical resection of their seizure focus, and determine if further invasive studies are necessary for precise localization of the epileptogenic tissue.

Analysis of high-dimensional data poses a number of challenges. Dimensionality reduction, particularly in the time domain, and data compression are necessary for reducing the computational cost associated with processing large volumes of data. However, the structure of these data, i.e., the underlying temporal scales of variability are not a priori known, and need to be estimated directly from the data. Furthermore, although some physiological processes are associated with robust and regular patterns of dynamic variability, such as different stages of sleep, abnormal electrophysiological events, such as seizures, may cause irregularly varying neural dynamics, e.g., rapidly varying during ictal epochs versus slowly varying during interictal periods. In order to preserve multi-scale information (in the time domain) encoded in the data, adaptive compression/dimensionality reduction approaches are necessary [2].

A wide range of approaches have been proposed for estimating the underlying temporal structure of electrophysiological signals, such as electroencephalograms (EEG). These range from signal complexity-based approaches to correlation methods and model-order estimation, e.g. [3], [4], [5], [6], among many others. All have advantages and drawbacks, depending on the application and research question. For example, model-order estimation, e.g., for AR modeling, is often used as a measure of data complexity. It relies on the existence of temporal patterns in the data identified from the autocorrelation function. In cases of noisy signals or epochs of increased noise, the low model order may be misleading if data complexity is of interest. For example, white noise, although intuitively a complex signal corresponds to a zeroth order AR process. Nevertheless, if one is only interested in relative changes from an epoch of longer local correlations in time (high-order AR process), to an epoch of lower temporal correlation or even sample independence (lower-order AR process), model-order estimation may be used to detect such transitions. For that purpose, the Minimum Description Length (MDL) principle may be used [9]. In addition, various entropy-based approaches have been proposed, to quantify either transient non-linearities (chaos) in the data, e.g., the Kolmogorov-Sinai entropy, or uncertainty and randomness in e.g., approximate entropy [3]. Entropy-based methods that aim to identify a chaotic structure in the data, make an a priori assumption of at least transient non-linearity. Finally, methods such as Detrended Fluctuation Analysis (DFA) estimate the underlying scale of self-similarity in non-stationary signals, and are useful for detecting scales of long-range, power-law correlations in these signals [7].

This study aimed to estimate the underlying temporal dynamics of long-term scalp EEG recordings from patients with epilepsy, and thus identify relevant scales of data segmentation and analysis. Four methods were compared, model-order

This work was supported by the department of Radiology, Children's Hospital Boston (CS), the Harvard Clinical and Translational Science Center (CS) (NIH Award #UL1 RR 025758), and NIH grant R01 NS073601 (BC)

C. Stamoulis is with the Departments of Radiology and Neurology, Boston Children's Hospital and Harvard Medical School, Boston, MA 02115 USA caterina@mit.edu, caterina.stamoulis@childrens.harvard.edu

B.S. Chang is with the Department of Neurology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston MA 02215

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

selection using the MDL principle, approximate entropy, mutual information and DFA. As we were solely interested in the estimation of these time scales, and thus breakpoints between intervals of increased randomness or higher local correlation in the time domain, rather than characterization of the data complexity, these methods were selected based on their limited number of a priori assumptions on the data characteristics. Note that model order selection, approximate entropy and DFA were estimated from individual signals, whereas mutual information was estimated from signal pairs.

II. METHODS

A. EEG data

1) Data collection: All continuous EEGs were recorded in the Clinical Neurophysiology Laboratory of the Comprehensive Epilepsy Center at Beth Israel Deaconess Medical Center. A standard international 10-20 clinical EEG system was used. All recordings were part of inpatient, long-term electrophysiology studies for patient evaluation and monitoring, and typically spanned several days. Signals were sampled at 500 samples/s and were referenced to a common average reference.

2) EEG preprocessing: Power-line noise was attenuated with a stopband filterbank, centered at the 60 Hz harmonics of the noise, in the range 60-250 Hz, with a 1 Hz bandwidth for center frequencies $\leq$ 150 and a 1.5 Hz bandwidth for center frequencies $>$ 150 Hz. Third order elliptical filters (20 dB attenuation in the stopband, 0.5 dB ripple in the passband) were used. Signals were filtered in both forward and reverse directions to eliminate potential phase distortions due to the non-linear phase of the filter.

3) Patient details: Table I summarizes the patient demographics and data details. All patients included in this analysis had been diagnosed with medically refractory epilepsy. Long-term EEGs used for temporal scale estimation in this study, were previously analyzed in [8] for different purposes.

B. Estimation Methods

1) Model Order Selection: Consider the real-valued vector $y(t)$ of length N, and the corresponding real-valued n-dimensional parameter vector θ. A model order based on the MDL principle is selected as follows: For a set of candidate models $\mathscr{H}^1$,... $\mathscr{H}^n$ we choose a model that minimizes $L(\mathscr{H}|y)+L(\mathscr{H})$, where $\mathscr{H} \in \mathscr{H}^1 \cup \mathscr{H}^2$.... Each hypothesis $\mathscr{H}$ may be thought of as a probability distribution. Code-length L may be expressed as $L = -\log P(y^n|\mathscr{H})$. More details on the MDL principle may be found in [9].

2) Mutual Information: For variables X and Y, with respective marginal distributions $p_X(x)$ and $p_Y(y)$, and joint distribution $p_{XY}(x,y)$, their mutual information $I(X;Y)$ measures the amount of information Y contains about X.

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p_{XY}(x,y) log_2 \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)} = \quad (1)$$
$$H(X)+H(Y)-H(X,Y)$$

where $H(\cdot)$ is the entropy of the random variable and $H(X,Y)$ the joint entropy between X and Y.

3) Approximate Entropy: For a real-valued vector $y(t)$ of length N, positive real number r and fixed positive integer m, one may construct a sequence of vectors $x(1),x(2),...x(n-m+1)$, defined as x(i) = [y(i), y(i+1),...., y(i+m-1)]/. Then, for each i $1 \leq i \leq N-m+1$ the correlation dimension is given by:

$$C_i^m(r) = j : \frac{d[x(i),x(j)] \leq r}{N-m+1} \quad (2)$$

where d is the maximum absolute distance between $x(i)$ and $x(j)$. We can then define, the measure Φ, such that:

$$\Phi^m(r)\frac{1}{(N-m+1)}\sum_{i=1}^{N-m+1} \log C_i^m(r) \quad (3)$$

Approximate entropy is defined as:

$$ApEn(m,r) = lim_{N\to\infty}[\Phi^m(r) = \Phi^{m+1}(r)] \quad (4)$$

4) Detrended Fluctuation Analysis: Signals are first integrated over a typically small data window. Integrated signals are segmented using a variable segment length, corresponding to potentially distinct scales in the data. In each segment, the data trend is estimated by fitting a least-squares line through the data. The fluctuation $F(k)$ at each segment k is defined as:

$$F(k) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}[y(n)-y_k(n)]^2} \quad (5)$$

where N is the signal length and $y_k(n)$ the local trend in segment k. When the logarithm of $F(k)$ is plotted as a function of the logarithm of the scale (window in which the local trend was estimated), the slope of the best fitted line through these data represents the scaling exponent α. The value of α reflects the existence, range and type of correlation in the data. For uncorrelated data, $\alpha \sim 0.5$. For long-range correlations $0.5 < \alpha < 1$, and for non-power law correlations, $\alpha > 1$.

III. RESULTS

We estimated both approximate entropy and mutual information using a 10-s sliding window, and DFA using a 4-sec integration window. Although this a priori data segmentation choice may be computationally inefficient, and as previously mentioned the goal of the estimation of temporal scales is to avoid a priori decisions on data segmentation, this preliminary study aimed to identify the underlying dynamic scales of the epilepsy data, and compare the four measures of uncertainty and complexity/correlation. Figures 1-3 show examples of estimated parameters in two 2-hr segments, from two patients. Mutual information (shown in red) is superimposed to approximate entropy (black) for different channels. In Figure 1 mutual information between channels Fp1 and F7 (in close proximity to each other), Fp1 and T3, Fp1 and Fp2 (across hemispheres) and Fp1 and O2 (at long distances from each other) are shown. Corresponding entropies for each of these channels (F7, T3, Fp2, 02) are superimposed. For both patients, these 2-hr intervals correspond to non-ictal epochs. Approximate entropy measures

TABLE I

PATIENT CLINICAL DEMOGRAPHICS.

Patient #	Age (yrs)	Etiology	Seizure focus	# Recording hrs	# Seizures
1	47	Cryptogenic	L temporal	63.5	3
2	27	Head injury	L temporal	193.5	7
3	24	Cryptogenic	L/R/simult-bilat. temporal	40	4
4	23	Brain malformation	L temporal	48	1
5	27	Cryptogenic	R/L temporal	137.5	6

the unpredictability (randomness) of the data, and thus small values of ApEn imply that the data are predictable. In case of coordinated signals, and thus individually predictable and jointly synchronized, mutual information is expected to be high. So, in comparing the two measures, high mutual information may be accompanied with low ApEn. Indeed, this was often observed locally in time (order of a few min or less), but anti-correlations between MI and ApEn were not clear at longer time scales. The estimated optimum model order varied at longer time scales, though only in a few channels. In many channels it was either constant, for long intervals, or zero.

We also compared these scales to the scaling exponents estimated using DFA. The spatio-temporal distribution of these exponents, for the data example in Figure 1, is shown in Figure 2. Intervals of power-law correlations ($0.5 < \alpha < 1$) were followed by intervals of non-power-law correlations ($\alpha > 1$) and even intervals of non-correlation ($\alpha < 0.5$). The scales corresponding to correlations that did not follow a power-law were approximately time-locked to the maxima of the estimated model-order trajectories, as well as maxima in MI, e.g., third panel in Figure 1. The temporal variability of estimated exponents was consistent across EEG channels.

Finally, we also examined the temporal scales of variation in the entire recording for each patient (see Table 1). Figure 4 shows estimated parameters for patient 1, during an interval of 11.7 hours. Mutual information estimated detailed structure at localized scales (peaks in the mutual information signal), but the overall envelope of this signal was also informative in that it revealed another scale of slower variation. Approximate entropy varied non-specifically and was difficult to interpret, suggesting that it may be a better measure for estimating local dynamic scales. The model order also varied slowly over time, at the same approximate rate as the mean envelope of mutual information. The scaling exponent estimated using DFA varied similarly to MI, although intervals where the exponent was $\leq$0.5 were not clearly correlated with any of the scales estimated with other approaches.

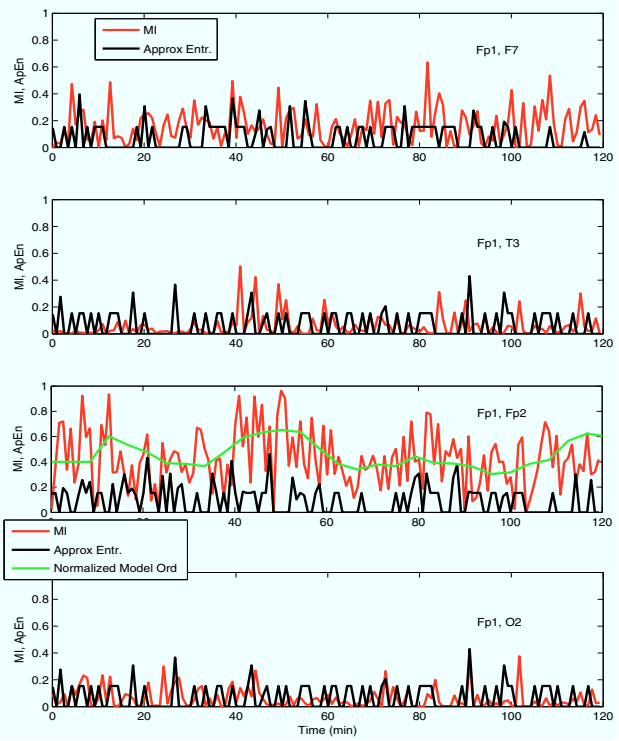

Fig. 1. Mutual information (MI, red), approximate entropy (ApEn, black) and model order (M, green) as a function of time for 4 pairs of signals (Fp1, F7), (Fp1, T3), (Fp1, Fp2), (Fp1, O2) - for MI. Model order was normalized so that it can be superimposed to the other plots. This segment is from patient 2, and corresponds to a nonictal epoch.

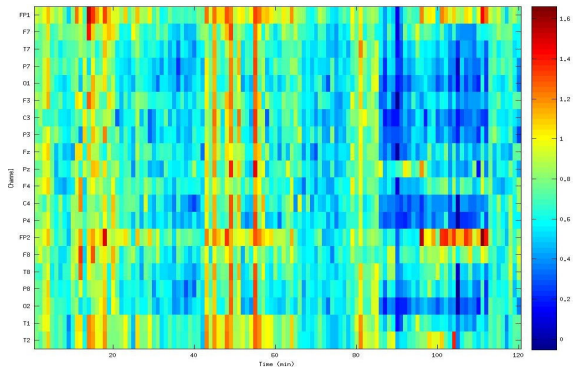

Fig. 2. Spatio-temporal distribution of scaling exponents estimated from all EEG signals using DFA. The colorbar corresponds to exponent values.

IV. DISCUSSION

We have investigated four approaches for estimating the data structure and complexity in long-term EEGs from patients with epilepsy, for data dimensionality reduction and computationally efficient analysis. Approximate entropy was the least sensitive to long-range dynamic variability, i.e., of the order of hours. Mutual information also identified local scales. Evidently a longer analysis window for estimation

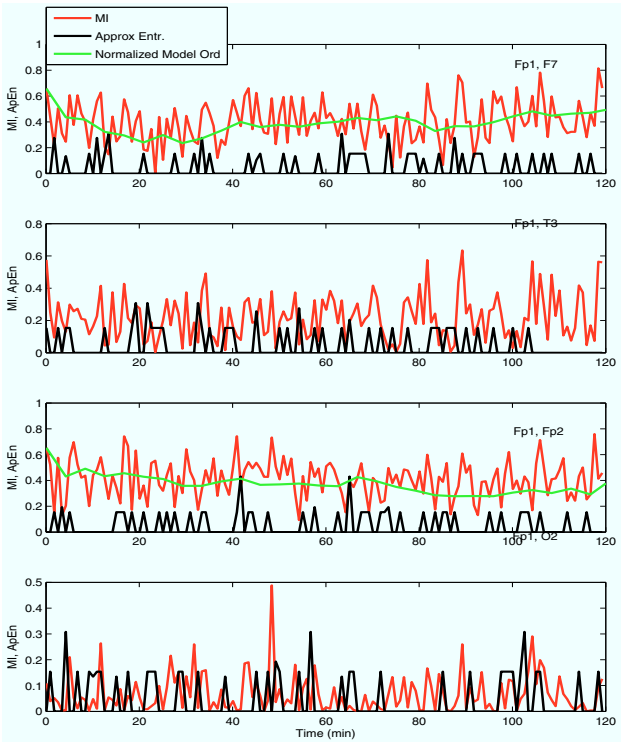

Fig. 3. Mutual information (MI, red), approximate entropy (ApEn, black) and model order (M, green) as a function of time for 4 pairs of signals (Fp1, F7), (Fp1, T3), (Fp1, Fp2), (Fp1, O2) - for MI. Model order was normalized so that it can be superimposed to the other plots. This segment is from patient 3, and corresponds to a nonictal epoch.

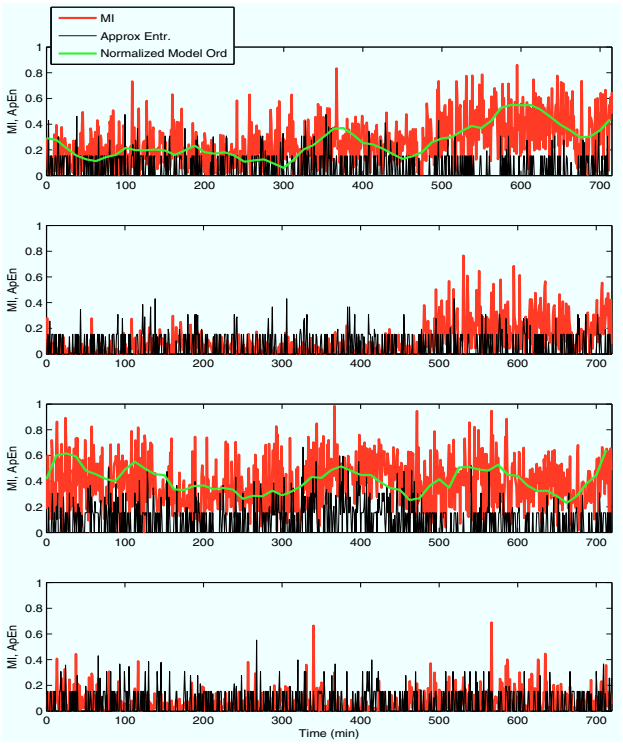

Fig. 4. Mutual information (MI, red), approximate entropy (ApEn, black) and model order (M, green) as a function of time for 4 pairs of signals (Fp1, F7), (Fp1, T3), (Fp1, Fp2), (Fp1, O2) - for MI. Model order was normalized so that it can be superimposed to the other plots. This segment is from patient 1, and corresponds to a $\sim$11.7-hr long nonictal epoch.

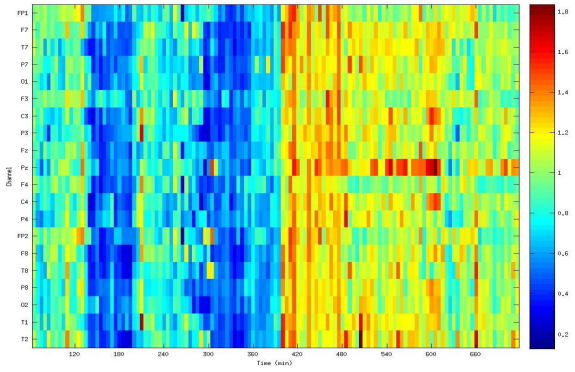

Fig. 5. Spatio-temporal distribution of scaling exponents estimated from all EEG signals using DFA, for the same segment as in Figure 4.

of the necessary marginal and joint pdfs could have been used, potentially leading to the estimation of less variability in the data. Optimum model order, selected according to the MDL principle, only captured slow dynamics, of the order of 30 min to hours. DFA estimated both local and longer-range scales, which in some cases were time-locked to scales estimated by other approaches. This preliminary analysis suggests that multiple measures, which are sensitive to different scales of variation, may be necessary for estimating several relevant scales in long-term recordings. Evidently for these estimations to be meaningful, appropriate thresholds need to be selected. Also, a common window of analysis across channels may be selected by estimating these complexity measures individually at each channel, and then selecting the shortest interval among the estimates. The sensitivity/specificity of these parameters to the underlying structure of the data also need to be assessed via simulation.

REFERENCES

[1] M. Hirshkowitz, Normal human sleep: an overview, *Med Clin N Am*, 88:551-565, 2004.

[2] J. Haupt, R. Nowak, Adaptive sensing for sparse recovery, in Compressed Sensing, Y.C. Eldar and G. Kutyiniok, editors, Cambridge University Press, 2012.

[3] A. Rezek, J. Roberts, Stochastic complexity measures fpr physiological signal analysis, *IEEE Trans Biomed Eng*, 45(9):1186-1191, 1998.

[4] K. Schindler, H. Leung, C.E. Elger, K. Lehnertz, Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG, *Brain*, 130(Pt 1):65-77, 2007.

[5] J.R. Williamson, D.W. Bliss, D.W. Browne, Epileptic seizure prediction using the spatiotemporal correlation structure of intracranial EEG, *Acoustics Speech and Signal Processing (ICASSP)*, 665-668, 2011.

[6] Celka, P., Colditz, P., A computer-aided detection of EEG seizures in infants: a singular spectrum approach and performance comparison, *IEEE Trans Biomed Eng*, 49(5):455-462, 2002.

[7] Peng C-K, Havlin S, Stanley HE, Goldberger AL. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, *Chaos* 5:82-87, 1995.

[8] C. Stamoulis, B.S. Chang, Multiscale information for network characterization in epilepsy, *IEEE EMBC Proc Eng Med Biol*, 5908-5911, 2011.

[9] P. Grunwald, A tutorial introduction to the Minimum Description Length principle, in Advances in Minimum Description Length: Theory and Applications, P. Grunwald, J. Myung, M.A., Pitt, MIT Press, 2004.