

Measuring randomness by leave-one-out prediction error. Analysis of EEG after painful stimulation

N. Ancona^a, L. Angelini^{b,c,d}, M. De Tommaso^{b,e}, D. Marinazzo^{b,c,d}, L. Nitti^{b,d,f},
M. Pellicoro^{b,c,d}, S. Stramaglia^{b,c,d,*}

^a*Istituto di Studi sui Sistemi Intelligenti per l'Automazione, C.N.R., Bari, Italy*

^b*TIRES-Center of Innovative Technologies for Signal Detection and Processing, University of Bari, Italy*

^c*Dipartimento Interateneo di Fisica, Bari, Italy*

^d*Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Italy*

^e*Dipartimento di Scienze Neurologiche e Psichiatriche, University of Bari, Italy*

^f*Dipartimento di Biochimica Medica, Biologia Medica e Fisica Medica, University of Bari, Italy*

Received 6 July 2005; received in revised form 30 August 2005

Available online 27 October 2005

Abstract

A parametric approach, to measure randomness in time series, is presented. Time series are modelled by a kernel machine performing regularized least squares and the leave-one-out (LOO) error is used to quantify unpredictability. On analyzing simulated data sets, we find that structure in data leads to a minimum of the LOO error as the regularizing parameter is varied. We consider electroencephalographic signals from migraineurs and healthy humans, after painful stimulation and use the proposed approach to detect changes of physiological state and to find differences between the response from patients and healthy subjects. As painful stimulus causes organization of the local activity in the cortex, EEG series become more predictable after stimulation. This phenomenon is less evident in patients: the inadequate cortical response to pain in migraineurs separates patients from controls with a probability close to 0.005.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Randomness; Predictability; EEG analysis

1. Introduction

Quantification of the regularity (orderliness) of physiological signals has been the focus of several approaches in the last few years [1]. Diseased systems typically show enhanced regularity compared to the dynamics of healthy systems [2–4]. More generally, loss of complexity has been proposed as a generic feature of pathologic dynamics [5]. The traditional non-parametric algorithm to measure regularity is based on the evaluation of the Kolmogorov–Sinai (KS) entropy [6], the mean rate of creation of information. This quantity is related to the probability that two sequences, extracted from the time series, which are similar for m points remain similar at the next point: hence, KS entropy measures the unpredictability of the time series. Although

*Corresponding author.

E-mail address: Sebastiano.Stramaglia@ba.infn.it (S. Stramaglia).

effective methods for estimating the entropy of a time series have been developed [1,7–9], the analysis of short and noisy data remains a difficult task. In this paper we describe a parametric approach to measure the predictability of short and noisy time series, where time series are modelled by the regularized least squares (RLS) approach and the leave-one-out (LOO) error [10] is used to quantify predictability. This approach has already been used in Ref. [11] to analyze systolic arterial pressure time series from healthy subjects and chronic heart failure patients undergoing paced respiration. Here we extend and elaborate previous findings showing its applicability in the analysis of EEG activity evoked by CO₂ laser painful stimulation [12,13].

It is worth stressing that the LOO error, like the KS entropy, is expected to grow monotonically with the degree of randomness of the signal; hence, it has no straightforward correspondence with the complexity of the signal, which is associated with *meaningful structural richness* [14]. For example, white noise is highly unpredictable but not structurally complex since at the global level it admits a very simple description. Substantial attention has been focused on defining a measure of complexity that assigns minimum values to both determinist/predictable and uncorrelated signals [15–19], but no consensus has been reached on this issue.

2. Method

The regularized least-squares models for regression have been described in detail in Ref. [11]; here, we briefly recall the main properties of these machines. Let us denote $\{x_i\}_{i=1,\dots,N}$ as a physiological time series, which we assume to be stationary (this assumption is justified in the case of a short length of recording). In the preprocessing stage, the time series is normalized to have zero mean and unit variance. We fix the length of a window m , and for $k = 1-\ell$ (where $\ell = N - m$), we denote $\mathbf{x}_k = (x_{k+m-1}, x_{k+m-2}, \dots, x_k)$ and $y_k = x_{k+m}$; we treat these quantities as ℓ realizations of the stochastic variables \mathbf{x} (input variables) and y (output variable) with an unknown probability distribution $p(\mathbf{x}, y)$. The problem of learning consists in providing an estimator $f: \mathbf{x} \rightarrow y$. In our approach it has the form:

$$y = f(\mathbf{x}) = \sum_{i=1}^{\ell} c_i K(\mathbf{x}_i, \mathbf{x}), \quad (1)$$

where the kernel $K(\mathbf{x}, \mathbf{y})$ is a positive-definite symmetric function, and coefficients \mathbf{c} are given by

$$\mathbf{c} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}, \quad (2)$$

\mathbf{K} being an $\ell \times \ell$ matrix with a generic element $\mathbf{K}_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$. λ is the regularization parameter. Predictor (1) corresponds to a linear predictor in the feature space spanned by the eigenvectors of the integral operator determined by K . Many choices of the kernel function are possible, for example the polynomial kernel of degree p has the form $K(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x} \cdot \mathbf{y})^p$ (the corresponding features are made up of all the powers of \mathbf{x} up to the p th). The radial basis function Gaussian kernel is $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2/2\sigma^2)$ and deals with all the degrees of nonlinearity of \mathbf{x} : the width σ plays a role similar to that played by the regularization parameter, i.e., it must be tuned to avoid overfitting. As a measure of the generalization ability of the trained model, we consider the LOO error ε , i.e., the average empirical square error when the data point, whose error is under consideration, is removed from the training set. It can be calculated as follows:

$$\varepsilon = \frac{1}{\ell} \sum_{i=1}^{\ell} \left(\frac{y_i - \sum_{j=1}^{\ell} K_{ij} c_j}{1 - \mathbf{G}_{ii}} \right)^2, \quad (3)$$

where the matrix \mathbf{G} is equal to $\mathbf{K}(\mathbf{K} + \lambda \mathbf{I})^{-1}$. We conclude this section by noting the interesting properties of the class of models we are dealing with; the most important one is that such models have a high generalization capacity. This means that they are able to predict complex signals when a finite and small number of observations of the signal itself are available. Moreover, the degree of nonlinearity present in the modelling, introduced by the kernel method, may be easily controlled. They allow an easy calculation of the LOO error [10], the quantity that we use to quantify predictability. They possess an invariance property that renders them suitable to evaluate causality between simultaneously acquired signals [20]. Finally, this approach generalizes

the classical autoregressive (AR) approach to time series analysis [21], which is recovered for $p = 1$ polynomial kernel in the limit $\lambda \rightarrow 0$.

3. Results on simulated time series

We first describe the application of our method on simulated white and 1/f noise [22], as well as on time series from a chaotic map (logistic with $A = 3.8$) with additive white noise proportional to parameter η .

Firstly let us consider the role of λ in the polynomial kernel case: the presence of a minimum in the plot ε vs. λ is a signature of the presence of some structure in the data. In Fig. 1 we depict these plots for a typical realization of the chaotic map with $\eta = 0.05$ (up), a realization of 1/f noise (middle) and white noise (down). The length of these series is $N = 100$ points and a polynomial kernel with $p = 2$ is used. We note the presence of a minimum both for the chaotic map and 1/f noise; the minimum is missing in the case of white noise due to the absence of structure.

In Fig. 2 the plots ε vs. σ are reported, for the same realizations of time series as in Fig. 1, in the Gaussian kernel case with $\lambda = 0.01$ fixed. Again we see a minimum both for the chaotic map and 1/f noise, and the absence of the minimum in the case of white noise. The fact that white noise is less predictable than 1/f noise is in agreement with the results by sample entropy [5].

Our simulations show that Gaussian models work better than polynomial ones for high noise levels. For example in Fig. 3 we depict ε vs. η corresponding to $N = 100$ chaotic map series. The continuous line refers to the $p = 2$ polynomial kernel model where, at each value of η , λ is tuned to minimize ε . The dashed line refers to a Gaussian kernel model with $\lambda = 0.001$ and, at each value of η , σ is tuned to minimize ε . Results are averaged over 100 realizations of the noise. At small noise, the polynomial model (which is the exact model in the absence of noise) works better; as the noise increases ($\eta > 0.01$) the performance by the Gaussian model becomes the best. The ratio between the LOO errors from polynomial ε_p and Gaussian ε_g models is depicted in the inset figure.

Finally, in Fig. 4 we show how the LOO error depends on the length of the sampled series. Segments of chaotic map series, with $\eta = 0.05$ and varying length, are modelled by $p = 2$ polynomial kernel (upper curve): at each value of N , λ is tuned to minimize ε . The lower curve refers to a Gaussian kernel with $\lambda = 0.001$ and, at each value of N , σ is tuned to minimize ε . Results are averaged over 500 realizations of the noise. The plots show the robustness of index ε : the LOO error at $N = 50$ differs by only a few percent from the asymptotic value.

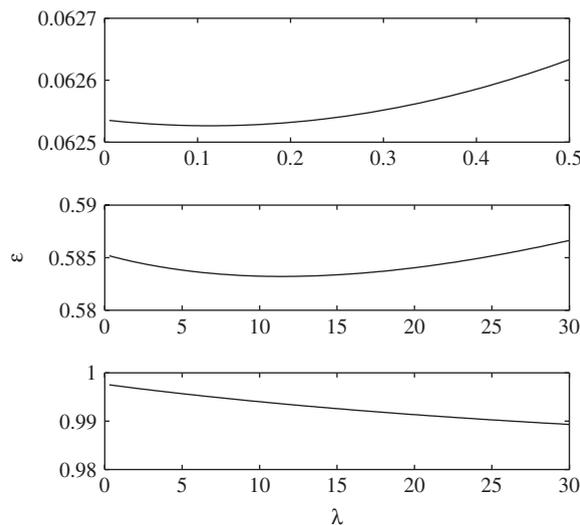


Fig. 1. The LOO error ε is plotted versus λ for a typical realization of the chaotic map with $\eta = 0.05$ (up), a realization of 1/f noise (middle) and white noise (down). Here $N = 100$ and a polynomial kernel with $p = 2$ is used.

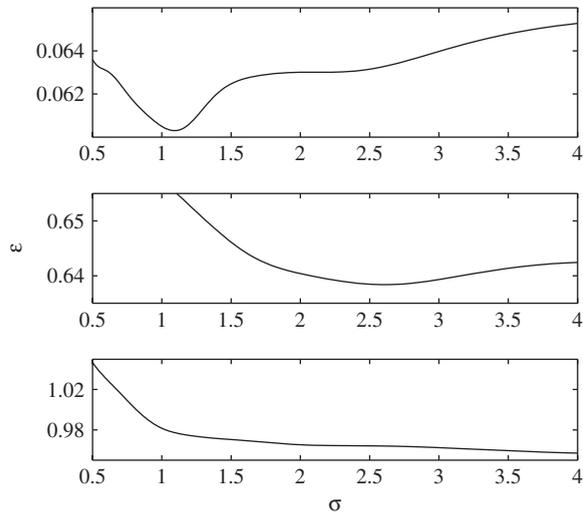


Fig. 2. ε is plotted versus σ for the same realizations of time series as in Fig. 1, using a Gaussian kernel with $\lambda = 0.01$.

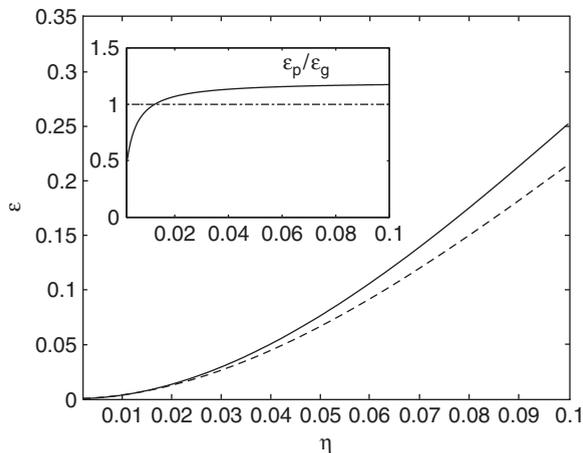


Fig. 3. ε is plotted versus η in the case of a chaotic map series. $N = 100$; the continuous line refers to the $p = 2$ polynomial kernel model and the dashed line to the Gaussian kernel model with $\lambda = 0.001$.

4. Analysis of EEG signals from migraineurs after painful stimulation

In this section we describe the application of our approach to analysis of EEG signals from patients affected by migraine. Ten patients suffering from migraine without aura [23] were submitted to CO₂ laser stimulation of the skin of the dorsum of the right-hand and the right supraorbital zone, during the non-symptomatic phase. The pain stimulus was laser pulses (wavelength 10.6 m) generated by a CO₂ laser (Neurolas, Electronic Engineering, Florence, Italy; www.elengroup.com). The beam diameter was 2.5 mm, and the duration of the stimulus pulse was 20 ms. We used a fixed-intensity set at 7.5 W for both the hand and the supraorbital zone. For each site of stimulation, 40 laser stimuli were delivered; the first 20 stimuli were delivered without any warning signal, the second series was preceded by a flash, after which patients are instructed to wait for the painful stimulation. The inter-stimulus-interval (ISI) was at least 10 s. Ten sex- and age-matched healthy subjects were also submitted to the experimental procedure. EEG recordings were obtained using a MICROMED (Mogliano Veneto, Italy) System 98 apparatus. Electrodes were positioned according to the

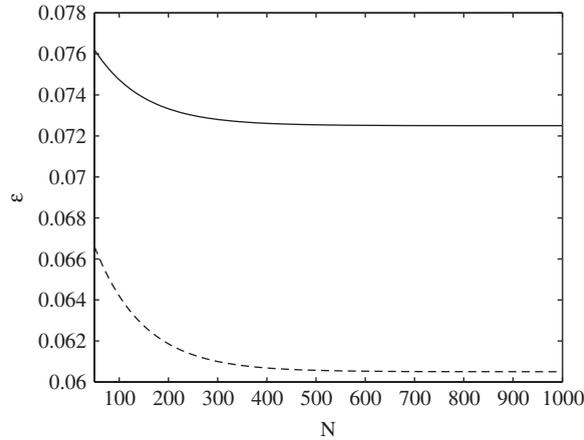


Fig. 4. ε is plotted versus the length of the time series N . Segments of noisy chaotic map with $\eta = 0.05$, with varying length, are modelled by $p = 2$ polynomial kernel (upper curve). The lower curve refers to the Gaussian kernel with $\lambda = 0.001$.

international 10–20 system, at Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, C4, T4, T5, P3, Pz, P4, T6, O1 and O2. The reference electrode were positioned at the linked earlobes (A1–A2), with the ground electrode placed over the nasion. The EEG signals were sampled at a rate of 256 Hz, and filtered at 0.1–60 Hz. Eye movements were monitored by a pair of electrodes placed at the outer canthi of both eyes in order to show ocular artefacts: EEG tracks containing artefacts were rejected by visual inspection.

All EEG signals were filtered in the traditional bands: alpha (8–12.5 Hz), beta (13–30 Hz), etc. In order to test to what extent the brain rhythms are reset after the stimulus, we investigated the cross-correlation of filtered time series from pairs of electrodes. In the case of M stimuli, for electrodes x and y , the cross-trial cross-correlation [24] is defined as

$$C_{xy}(t) = \frac{\sum_{j=1}^M x(t + \tau_j)y(t + \tau_j)}{\sqrt{\sum_{j=1}^M x^2(t + \tau_j)}\sqrt{\sum_{j=1}^M y^2(t + \tau_j)}}, \quad (4)$$

where τ_j is the time at which the j -th stimulus is delivered. The quantity C varies in $[-1, 1]$ and it assumes large values in the presence of a strong reset triggered by the stimuli. The evaluation of this parameter revealed a strong reset of the beta rhythms (12.5–30 Hz) after the painful stimuli, in accordance with [25]: therefore the following analysis will refer to the EEG signals filtered in the beta band. We found a significant enhancement of the cross-correlation, after the stimulus, for the three electrodes FZ, F3 and F4. The same effect was found for the other two groups of electrodes C3, CZ, C4 and P3, PZ, P4. The effect, due to stimulus, persists for a relatively long interval of time (more than 1 s). In Fig. 5 C is plotted versus time for the pair FZ-F3 and for the pair O1-O2, for a typical patient. The dotted vertical line on 0 represents the time of delivery of the stimulus. For the pair FZ-F3, after the reaction time the value of C increases and becomes close to one, going back to pre-stimulus values after more than one second. This does not occur for the pair O1–O2, that are close to each other, and thus reasonably inter-correlated, but this correlation does not increase after the stimulus. For the same patient, in Fig. 6 we report a map representing the cross-correlation (averaged over an interval of 1 s after the stimulus) between all pairs of electrodes. The presence of the three groups of correlated electrodes corresponds to three almost white 3×3 squares along the diagonal. The same qualitative pattern was found in all the remaining patients and in all the controls. From these results we state, in agreement with [13], that central, parietal and frontal electrodes represent the cortex areas most sensitive to painful stimuli, and therefore in the subsequent analysis we will concentrate on these 9 electrodes.

In Fig. 7 we present the graphs of the stimulus-averaged signals event-related potentials (ERP), at the FZ electrode, for a typical control and a typical patient, in the case of complete signals (upper part of the figure) and beta-filtered signals (lower part of the figure). We added an offset (equal to $150 \mu\text{V}$ in the case of the complete signal, and $20 \mu\text{V}$ in the case of beta-filtered signal) to the curves corresponding to the patient, so as

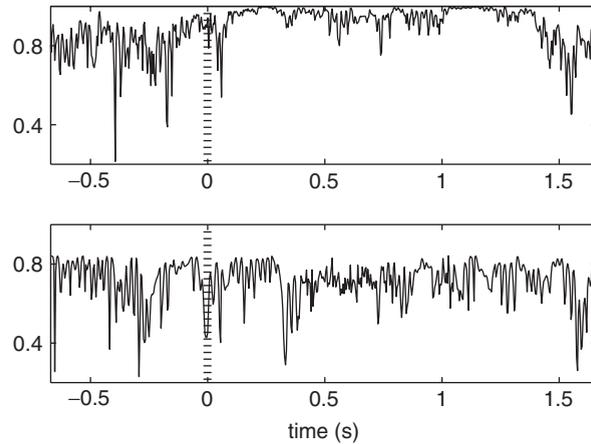


Fig. 5. Cross-trial cross-correlation C is plotted versus time for the pair FZ-F3 (up) and for the pair O1-O2 (down), for a typical patient.

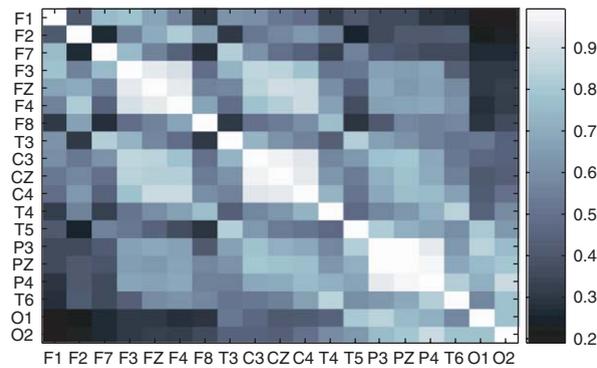


Fig. 6. The map (with gray-level scale) representing the cross-trial cross-correlation (averaged over the interval of 1 s after the stimulus) between all pairs of electrodes, for the same patient as in Fig. 5.

to make the two curves distinguishable. No significant differences were found in the ERPs from patients and controls.

The basic idea in our approach is that different physiological states may be characterized in terms of the predictability of time series. Segments of EEG series are modelled using a Gaussian kernel with $m = 32$, $\sigma = 6$ and a regularization parameter $\lambda = 0.01$. The value of m ensures that all the oscillations in the beta band contribute to the input pattern; σ and λ are fixed once for all subjects, by minimizing the average, LOO error on a subset made of an equal number of control and migraineurs time series. For each patient and for each electrode, we extract a time series after each stimulus, starting from $\tau_j + \tau_r$, where τ_j is the instant when the stimulus is delivered and τ_r is a reaction time, which in previous studies has been found to be equal to 0.25 and 0.35 s for stimulus delivered on the eye and the hand, respectively. The same is done after the flash used as a warning. The duration of these series is 1 s (256 points), corresponding to the average post-stimulus correlation period. Furthermore, for every post-stimulus series, a time series of equal length is extracted from the EEG signal before the stimulus, in order to have a control sample. For each of these series we calculated the LOO error. Finally, averaging on stimuli and on the nine channels under consideration, we obtained eight values of the LOO error for each subject: basal (prior to hand stimulation); basal (prior to eye stimulation); after the flash preceding hand stimulation; after the flash preceding eye stimulation; after hand stimulation with and without warning; and after eye stimulation with and without warning.

Different physiological states. For control subjects, the mean value of the LOO error decreases after the flash: the hypothesis that two independent samples, ten LOO errors before and after the flash, come from

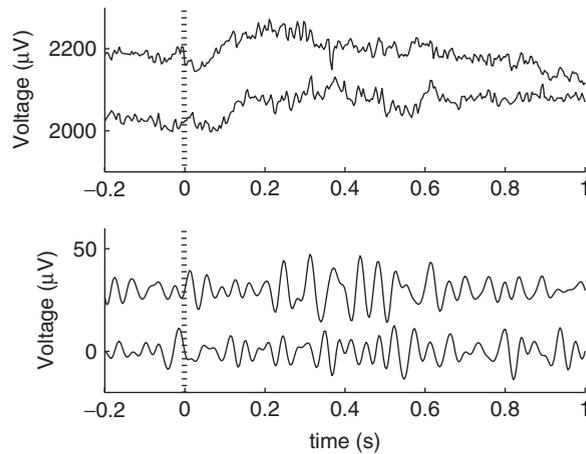


Fig. 7. The graphs of event-related potentials, at the FZ electrode, are depicted for a typical control and a typical patient. The upper part of the figure refers to the complete signals, whereas the lower part refers to beta-filtered signals. An offset (equal, respectively to 150 and 20 μV) is added to the curves corresponding to the patient.

Table 1
Mean values of LOO error for patients (in the parentheses: controls)

	Hand no flash	Hand flash	Eye no flash	Eye flash
Basal	0.0399(0.0344)		0.0390(0.0349)	
After flash		0.0377(0.0274)		0.0348(0.0256)
After stimuli	0.0326(0.0257)	0.0309(0.0244)	0.0323(0.0242)	0.0323(0.0261)

distributions with equal medians has probability (evaluated by the Wilcoxon rank sum test) 0.0379 both for stimulus on the hand and for the eye. After the painful stimulus, without flash, there is a major increase in the predictability of the series, and the values of the LOO error separate the two states (basal and after stimulus) with a p -value of 0.0012 for stimulus on the hand and 0.0023 for stimulus on the eye. When the stimulus is delivered after the flash, the predictability increases even further, and in this case we find p -values of 0.0006 for both the hand and the eye.

For migraine patients we observe the same trend, but in this case the decrease of the LOO error is not significant after the flash. After the stimulus migraine patients display a significant increase in predictability only when the painful stimulus is delivered after the flash, with p -values of 0.0070 for the hand and 0.0379 for the eye. The mean values of the LOO error in basal conditions, after the flash, and after the painful stimulus are reported in Table 1.

We also evaluate the sample entropy [1] and the integrated spectral power in beta band in the same EEG segments described before, and unsuccessfully look for indicators, based on these quantities, able to significantly detect changes in physiological state: only LOO error provides statically meaningful results.

Discrimination between migraine patients and controls. LOO errors corresponding to the states of basal acquisition, and after the warning, do not discriminate between control subjects and patients. After the painful stimulus (without warning) the predictability of the series changes differently in controls and patients, leading to a separation with p -values of 0.0046 for stimulus on the hand and 0.0068 for stimulus on the eye. When the stimulus was preceded by the flash the p -values were 0.0046 for both hand and eye.

Discussion. The increase of EEG predictability after painful stimulation is a sign of cortical reactivity to external conditions: the cortex may reduce the degree of randomness in order to adequately receive and elaborate the novel painful stimulus. Of course in the condition of warned stimulation, the cortex appears more prone to receiving the pain signal. In migraine patients the tendency towards a less pronounced increase in predictability after the laser stimulus may be explained by an inadequate cortical response to pain, which

becomes more efficacious only after the stimulus warning. The altered cortical reaction to pain may cause an abnormal recruitment of pain-modulation-system, possibly subtending the onset and the persistence of migraine attack [12]. The pathophysiological consequences of this findings will be discussed more widely elsewhere.

5. Conclusions

We have presented an approach to time series modelling which generalizes the classical autoregressive approach in two ways: (i) the model is regularized and (ii) nonlinearity is introduced. This allows to measure randomness in a physiological time series by means of the LOO error. We have shown application of the method to simulated time series and to the analysis of electroencephalographic signals from healthy subjects and migraineurs, after painful stimulation. The analysis of simulated data has shown that the presence of structure in data is connected to the presence of a minimum of the LOO error as the regularizing parameter is varied. Concerning the physiological application, use of the LOO error allowed to discover a new important phenomenon: migraineurs show an inadequate response, in terms of reset-induced increase of predictability, to painful stimulation.

References

- [1] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, *Am. J. Physiol.* 278 (2000) H2039–H2049.
- [2] A.L. Goldberger, C.K. Peng, L.A. Lipsitz, What is physiological complexity and how does it change with aging and disease?, *Neurobiol. Aging* 23 (2002) 23–26.
- [3] O.A. Rosso, M.L. Mairal, Characterization of time dynamical evolution of electroencephalographic records, *Physica A* 312 (2002) 469–504.
- [4] O.A. Rosso, W. Hyslop, R. Gerlach, R.L.L. Smith, J.A.P. Rostas, M. Hunter, Quantitative EEG analysis of the maturational changes associated with childhood absence epilepsy, *Physica A* 356 (2005) 184–189.
- [5] M. Costa, A.L. Goldberger, C.K. Peng, Multiscale entropy analysis of complex physiological time series, *Phys. Rev. Lett.* 89 (2002) 68102–68105.
- [6] J.P. Eckmann, D. Ruelle, Ergodic theory of chaos and strange attractors, *Rev. Mod. Phys.* 57 (1985) 617–656.
- [7] P. Grassberger, I. Procaccia, Estimation of the Kolmogorov entropy from a chaotic signal, *Phys. Rev. A* 28 (1983) 2591–2593.
- [8] S.M. Pincus, Assessing serial irregularity and its implications for health, *Ann. N.Y. Acad. Sci.* 954 (2001) 245–267.
- [9] O.A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, E. Basar, M. Schürmann, Wavelet entropy: a new tool for the analysis of short duration brain electrical signals, *J. Neurosci. Methods* 105 (2001) 65–75.
- [10] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- [11] N. Ancona, R. Maestri, D. Marinazzo, L. Nitti, M. Pellicoro, G.D. Pinna, S. Stramaglia, Leave-one-out prediction error of systolic arterial pressure time series under paced breathing, *Physiol. Meas.* 26 (2005) 363–372.
- [12] G. Pfurtscheller, F.H. Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles, *Clin. Neurophysiol.* 110 (11) (1999) 1842–1857.
- [13] R. Kakigi, K. Inui, D.T. Tran, Y. Qiu, X. Wang, S. Watanabe, M. Hoshiyama, Human brain processing and central mechanisms of pain as observed by electro- and magneto-encephalography, *J. Chin. Med. Assoc.* 67 (8) (2004) 377–386.
- [14] P. Grassberger, in: H. Atmanspacher, H. Scheingraber (Eds.), *Information Dynamics*, Plenum, NY, 1991, p. 15.
- [15] B.Y. Yaneer, *Dynamics of Complex Systems*, Addison Wesley, Reading, Massachusetts, 1997.
- [16] D.P. Feldman, J. Crutchfield, Measures of statistical complexity: why?, *Phys. Lett. A* 238 (1998) 244–252.
- [17] R. Lopez-Ruiz, H.L. Mancini, X. Calbet, A statistical measure of complexity, *Phys. Lett. A* 209 (1995) 321–326.
- [18] J.S. Shiner, M. Davison, P.T. Landsberg, Simple measure for complexity, *Phys. Rev. E* 59 (1999) 1459–1464.
- [19] P.W. Lamberti, M.T. Martin, A. Plastino, O.A. Rosso, Intensive entropic non-triviality measure, *Physica A* 334 (2004) 119–131.
- [20] N. Ancona, S. Stramaglia, An invariance property of kernel based predictors, cond-mat/0502511, *Neural Comput.*, 2005, in press.
- [21] H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, 1997.
- [22] 1/f noise is generated as follows: we start with uniformly distributed white noise, calculate the fast fourier transform (FFT), and after imposing a 1/f distribution on the power spectrum, we calculate the inverse FFT.
- [23] Headache Classification Subcommittee of the International Headache Society. The International Classification of Headache Disorders, second ed. Cephalalgia, vol. 24, Supplement 1, 2004, pp. 1–159.
- [24] G.D. Dawson, Cerebral responses to nerve stimulation in man, *Br. Med. Bull.* 6 (1950) 326–329.
- [25] S. Ohara, N.E. Crone, N. Weiss, F.A. Lenz, Attention to a painful cutaneous laser stimulation modulates electrocorticographic event-related desynchronization in humans, *Clin. Neurophysiol.* 115 (2004) 1641–1652.