

Article

## A Maximum Entropy Approach for Predicting Epileptic Tonic-Clonic Seizure

Maria Teresa Martín <sup>1,2</sup>, Angelo Plastino <sup>2,\*</sup> and Victoria Vampa <sup>3</sup>

<sup>1</sup> Depto. de Matemática, Fac. de Cs. Exactas, Universidad Nacional de La Plata, 1900 La Plata, Argentina; E-Mail: mtmartin@fisica.unlp.edu.ar

<sup>2</sup> Instituto de Física (IFLP-CCT-CONICET), Universidad Nacional de La Plata, C.C. 727, 1900 La Plata, Argentina

<sup>3</sup> Depto. de Cs. Básicas, Fac. de Ingeniería, Universidad Nacional de La Plata, 1900 La Plata, Argentina; E-Mail: victoriavampa@gmail.com

\* Author to whom correspondence should be addressed; E-Mail: plastino@fisica.unlp.edu.ar.

Received: 24 June 2014; in revised form: 16 July 2014 / Accepted: 12 August 2014 /

Published: 18 August 2014

---

**Abstract:** The development of methods for time series analysis and prediction has always been and continues to be an active area of research. In this work, we develop a technique for modelling chaotic time series in parametric fashion. In the case of tonic-clonic epileptic electroencephalographic (EEG) analysis, we show that appropriate information theory tools provide valuable insights into the dynamics of neural activity. Our purpose is to demonstrate the feasibility of the maximum entropy principle to anticipate tonic-clonic seizure in patients with epilepsy.

**Keywords:** maximum entropy; pseudo-inverse approach; tonic-clonic EEG transition

---

### 1. Introduction

The processing of information by the brain is reflected in the dynamical changes of its electrical activity in time and frequency. Therefore, methods capable of describing the qualitative variations of the signals are required, such as the methods developed by Rosso *et al.* using wavelet analysis [1].

From the patient's perspective, one of the most troubling aspects of epilepsy is the unpredictability of the occurrence of seizures. This fact has an impact on safety, patient anxiety and on the management of medications. The availability of a reliable predictor of an upcoming seizure would offer great value. An

algorithm with high predictive accuracy and requiring relatively inexpensive computing resources would be very desirable.

In this paper an information theory approach is developed for modelling chaotic time series in parametric fashion. This technique can be used to predict the future behavior of these series, using just a small amount of data. Our working hypothesis is represented by a set of parameters that allows for good predictive power. The Maximum Entropy Principle (MEP) is used for the estimation of the parameters of the pertinent model, using associated time series data. We developed a method effective in anticipating seizures from the scalp recording. This prediction possibility would be of great importance since it would allow for preventive measures to be taken against tonic-clonic seizures. As the name implies, these seizures combine the characteristics of tonic seizures and clonic seizures. The tonic phase comes first: all muscles stiffen. Air being forced past the vocal cords causes a cry or groan. The person loses consciousness and falls. The tongue or cheek may be bitten, so bloody saliva may come out from the mouth. The person may turn a bit blue in the face. After the tonic phase comes the clonic phase: the arms and usually the legs begin to jerk rapidly and rhythmically, bending and relaxing at the elbows, hips, and knees. After a few minutes, the jerking slows and stops. Bladder or bowel control is sometimes lost as the body relaxes. Consciousness returns slowly, and the person may be drowsy, confused, agitated, or depressed. These seizures generally last 1 to 3 min.

The paper is organized as follows. In Section 2 we discuss how to model time series using information theory. We describe how the appropriate estimation of the model’s parameter values translates itself into a technique for time series prediction. In Section 3 these techniques are applied for the analysis of tonic-clonic electroencephalographic (EEG) time series so as to anticipate the seizure start. Some conclusions are given in Section 4.

## 2. Method

Given a signal  $\mathbf{x}$  from a dynamical system  $D : \mathbb{R}^S \rightarrow \mathbb{R}^S$ , the corresponding time series consists of a sequence of measurements  $\{v(t_n), n = 1, \dots, N\}$  on a system considered to be in a state described by  $\mathbf{x}(t_n) \in \mathbb{R}^S$  at discrete times  $t_n$ , where  $N$  is the length of the time series.

As demonstrated by Takens in 1981 [2], for  $T \in \mathbb{R}, T > 0$ , there exists a functional form of the type,

$$v(t + T) = F(\mathbf{v}(t)), \tag{1}$$

where

$$\mathbf{v}(t) = [v_1(t), v_2(t), \dots, v_d(t)], \tag{2}$$

and  $v_i(t) = v(t - (i - 1)\Delta)$ , for  $i = 1, \dots, d$ .  $\Delta$  is the time lag and  $d$  is the embedding dimension of the reconstruction.  $T$  represents the *anticipation time* and it is of fundamental importance for a prediction model.

We will consider (as in [3]) a particular representation for the mapping function of Equation (1), expressing it—using Einstein’s summation notation—as an expansion of the form

$$F^*(\mathbf{v}(t)) = a_0 + a_{i_1} v_{i_1}(t) + a_{i_1 i_2} v_{i_1}(t) v_{i_2}(t) + a_{i_1 i_2 i_3} v_{i_1}(t) v_{i_2}(t) v_{i_3}(t) + \dots + a_{i_1 i_2 \dots i_{n_p}} v_{i_1}(t) v_{i_2}(t) \dots v_{i_{n_p}}(t), \tag{3}$$

where  $1 \leq i_k \leq d$  with  $1 \leq k \leq n_p$  and  $n_p$  being an adequately chosen polynomial degree so as to series-expand the mapping  $F^*$ . The number of parameters in Equation (3) corresponding to the terms of degree  $k$  depends on the embedding dimension and can be calculated using combination with repetitions,

$$\binom{d}{k}^* = \frac{(d+k-1)!}{k!(d-1)!} \tag{4}$$

Accordingly, the length of the vector of parameters,  $\mathbf{a}$ , adopts the form

$$N_c = \sum_{k=1}^{n_p} \binom{d}{k}^* \tag{5}$$

The computations are made on the basis of a specific information supply, given by  $M$  points of the series

$$\{\mathbf{v}(t_n), v(t_n + T)\}, \quad n = 1, \dots, M. \tag{6}$$

Given the dataset in Equation (6), the parametric mapping in Equation (3) will be determined by the following condition,

$$F^*(\mathbf{v}(t_n)) = v(t_n + T) \quad n = 1, \dots, M, \tag{7}$$

which can be expressed in matrix form as,

$$W\mathbf{a} = \mathbf{v}_T, \tag{8}$$

where  $W$  is a matrix of size  $M \times N_c$ , whose  $n$ -th row is  $[1, v_{i_1}(t_n), v_{i_1}(t_n)v_{i_2}(t_n), \dots, v_{i_1}(t_n)v_{i_2}(t_n) \dots v_{i_{n_p}}(t_n)]$  (cf. Equation (3)) and  $(\mathbf{v}_T)_n = v(t_n + T)$ , for  $n = 1, \dots, M$ . In this work, we use the maximum entropy principle to characterize important probability distributions. Shannon’s entropy, defined for a discrete random variable, can be extended to situations for which the random variable under consideration is continuous.

In order to infer coefficients that are consistent with the data, we shall assume that each set  $\mathbf{a}$  is realized with probability  $P(\mathbf{a})$ . Thus, a normalized probability distribution over the possible sets  $\mathbf{a}$  is introduced,

$$\int_I P(\mathbf{a}) \, d\mathbf{a} = 1, \tag{9}$$

where  $d\mathbf{a} = da_1 da_2 \dots da_{N_c}$  and  $N_c$  is the number of parameters of the model.

The problem then turns into finding  $P(\mathbf{a})$  subject to the requirement that the associated entropy  $H$  be maximized, since this is the best way of avoiding any bias. The expectation value of  $\mathbf{a}$  is defined by

$$\langle \mathbf{a} \rangle = \int_I P(\mathbf{a}) \mathbf{a} \, d\mathbf{a}. \tag{10}$$

For the continuous random variable  $\mathbf{a}$  with probability density function  $p(\mathbf{a})$  on  $I$ , where  $I = (-\infty, \infty)$ , the entropy is given by

$$H(\mathbf{a}) = - \int_I P(\mathbf{a}) \ln P(\mathbf{a}) \, d\mathbf{a}, \tag{11}$$

whenever it exists, and the relative entropy reads

$$H = - \int_I P(\mathbf{a}) \ln \frac{P(\mathbf{a})}{P_0(\mathbf{a})} d\mathbf{a}, \tag{12}$$

where  $P_0(\mathbf{a})$  is an appropriately chosen a priori distribution [4,5].

This measure resembles discrete entropy in many aspects, but unlike the entropy of a discrete random variable, the entropy for a continuous random variable may be infinitely large, negative, or positive (Ash, 1965 [6]). We characterize, via the maximum entropy principle, various probability distributions, subject to Equation (9) and Equation (8) as constraints, for the expectation  $\langle \mathbf{a} \rangle$  of  $\mathbf{a}$ . The method for solving this constrained optimization problem is to use Lagrange multipliers for each of the operating constraints and maximize the following functional with respect to  $P(\mathbf{a})$ ,

$$J = - \left[ \int_I P(\mathbf{a}) \ln \frac{P(\mathbf{a})}{P_0(\mathbf{a})} d\mathbf{a} + \lambda_0 \left[ \int_I P(\mathbf{a}) d\mathbf{a} - 1 \right] + \lambda^t \int_I [WP(\mathbf{a})\mathbf{a} - \mathbf{v}_T] d\mathbf{a} \right], \tag{13}$$

where  $\lambda_0$  and  $\lambda$  are Lagrange multipliers associated, respectively, with the normalization condition and with the constraints, *i.e.*, Equation (9) and Equation (8).

Taking the functional derivative with respect to  $P(\mathbf{a})$  we get

$$\frac{\partial J}{\partial P(\mathbf{a})} = \ln \left( \frac{P(\mathbf{a})}{P_0(\mathbf{a})} \right) + 1 + \lambda_0 + \lambda^t W\mathbf{a} = 0, \tag{14}$$

which implies that the maximum entropy distribution must have the form

$$P(\mathbf{a}) = \exp \left( - (1 + \lambda_0) \exp(\lambda^t W\mathbf{a}) P_0(\mathbf{a}) \right) \tag{15}$$

If the a priori probability distribution  $P_0(\mathbf{a})$  is chosen to be proportional to  $\exp(-\frac{1}{2}\mathbf{a}^t[\sigma^2]^{-1}\mathbf{a})$ , where  $\sigma^2$  is the covariance matrix, a Gaussian form for the probability distribution  $P(\mathbf{a})$  is obtained, with

$$\langle \mathbf{a} \rangle = -\sigma W^t \lambda \tag{16}$$

Considering Equation (8), the Lagrange multipliers  $\lambda$  can be eliminated:

$$\lambda = \sigma^{-1} (WW^t)^{-1} \mathbf{v}_T, \tag{17}$$

and one can thus write

$$\langle \mathbf{a} \rangle = W^t (WW^t)^{-1} \mathbf{v}_T. \tag{18}$$

The matrix  $W^t (WW^t)^{-1}$  is known as the Moore–Penrose pseudo-inverse of the matrix  $W$  (see [7] and references therein). Consequently, this result shows that the maximum entropy principle coincides with a least square criterion. Once the pertinent parameters are determined, they are used to predict  $M_P$  new series' values,

$$(\widehat{v}(t_n + T))_{n=1, \dots, M_P} = \widehat{W}\mathbf{a}, \tag{19}$$

where  $\widehat{W}$  is a matrix of size  $M_P \times N_c$ .

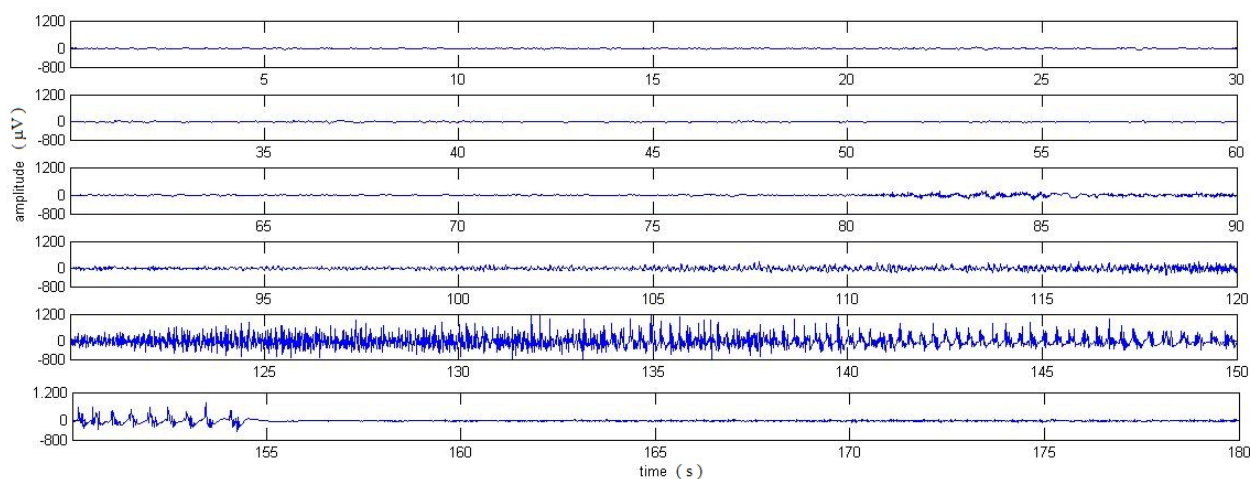
### 3. Application to EEG Time Series

#### 3.1. Clinical Data and Experimental Setup

A scalp EEG signal is essentially a non-stationary time series that presents artefacts mainly due to eye movements, muscle activities (or electromyogram), among others. Artefacts related to muscle contractions are especially troublesome in the case of tonic-clonic epileptic seizures, where they reach very high amplitudes that contaminate the whole seizure recording.

In [8], Rosso, Martín and Plastino presented the analysis of twenty tonic-clonic secondary generalized epileptic records pertaining to eight patients showing the self-organization in brain electrical activity. The patient group consisted of four males and four females with a diagnosis of pharmaco-resistant epilepsy and no other accompanying disorders. From this data we chose one of the three seizures recorded for a female (39 years old) with epileptic focus in the left temporal lobe to be analyzed with the method we propose in this work. Scalp electrodes with bimastoideal reference were applied following the 10–20 international system. In Figure 1 we present the scalp EEG signal recorded in a central right location (C4 channel) [8]. This electrode was chosen after visual inspection of the EEG records, as the one with the minimum amount of artefacts. The signal was digitized at 409.6 Hz through a 12 bit A/D converter and filtered with an antialiasing eight pole lowpass Bessel filter, with a cut-off frequency of 50 Hz. Then, the signal was digitally filtered with a 1–50 Hz bandwidth Butterworth filter and stored, after decimation, at 102.4 Hz, in a PC hard drive. Recordings were performed under video control in order to have an accurate determination of the stages of the seizure. The different stages of the EEG signals (preseizure, seizure start, tonic phase, clonic phase, seizure end, and postseizure) were determined by a team of physicians. The preseizure state is the hypothetical state in which processes leading to the seizure onset start.

**Figure 1.** Scalp EEG signal for a generalized epileptic tonic-clonic seizure. Preseizure stage (0–81 s), seizure start (81 s), tonic-clonic phase (82–155 s), seizure end (155 s), and postseizure (155–180 s).



According to physician diagnosis, the epileptic seizure starts at about to 81 s with a “discharge” of slow waves superposed by fast ones with lower amplitude. This discharge lasts approximately 8 s and

has a mean amplitude of  $100 \mu\text{V}$ . Afterwards, the seizure spreads, making the analysis of the EEG more complicated, due to muscle artefacts; however, it is possible to establish the beginning of the clonic phase at around 125 s, and the end of the seizure at 155 s, where one encounters an abrupt decay of the signal's amplitude.

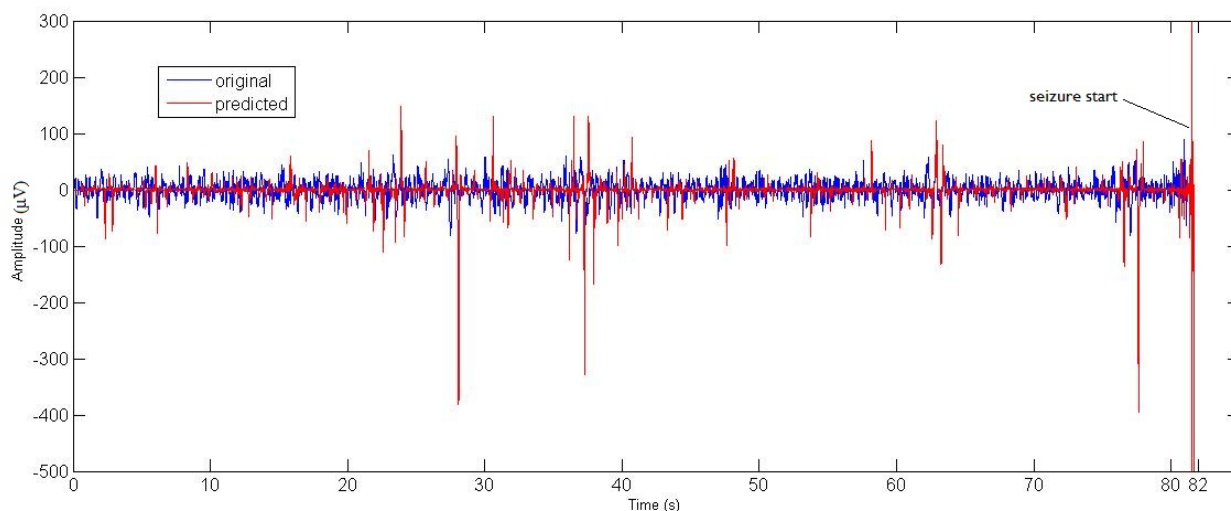
One of our goals was to model the EEG signals at the pre-seizure stage finding the corresponding parameters, in order to be able to predict a seizure's start.

### 3.2. Maximum Entropy Prediction

The method described in Section 2 was applied to the EEG data corresponding to the pre-seizure stage. The data were subdivided into two parts. One of them was employed in order to adjust the parameters of the model, while the other served to check upon the subsequent predictive power. We include into the model polynomials up to cubic degree,  $n_p = 3$ . The frequency sample is  $\Delta = 9.765 \times 10^{-3}$  Hz. We considered 3000 series values, which corresponds to approximately 30 s, to obtain the parameter vector  $\mathbf{a}$ . Using a  $d = 2$  embedding dimension, the number of estimated parameters is  $N_c = 14$ . We consider an anticipation time  $T \approx 1$  s, corresponding to 60 samples.

In Figure 2, 81 seconds that encompass the pre-seizure stage taken from the original EEG series are displayed together with the MEP prediction. A zoom of the seizure start is presented in Figure 3 to show the prediction clearly. From  $t = 81$  s to  $t = 81.6$  s only predicted values are shown (prediction zone). It is important to note that the value of the anticipation time  $T$  allows one to predict the seizure start ( $t_s = 81.5$  s) one second before it happens.

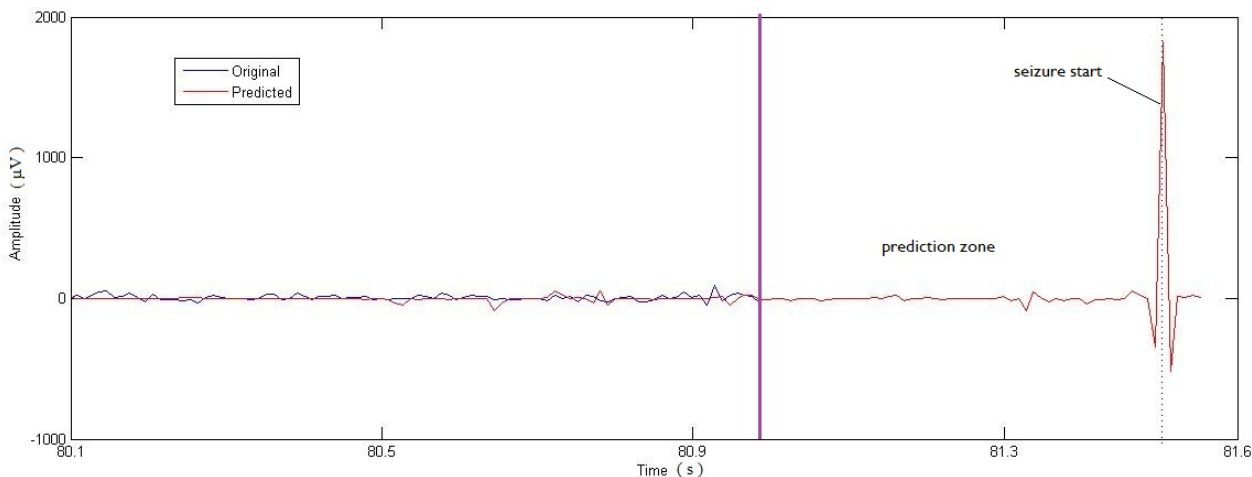
**Figure 2.** Pre-seizure stage and seizure start. Original data (0–81 s) and MEP prediction (0–82 s).



We analyzed the predictive power of the method for different anticipation time values. This is illustrated in Figures 4, 5 and 6, for 60, 30 and 5 samples, respectively. The best prediction was obtained for 5 samples ( $T \approx 0.1$  s), too short an interval for adopting preventive measures.

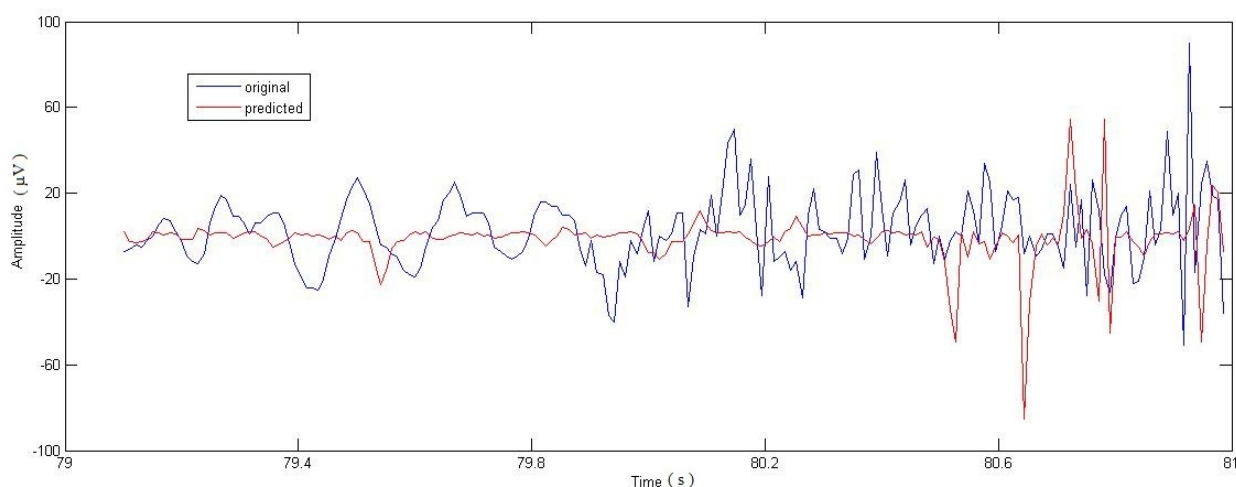
We also investigated the effect of varying the embedding dimension  $d$ . In Figure 7, results obtained with  $d = 3$  are presented. Making pertinent comparisons, it can be noted that the predictive power becomes greater than that for  $d = 2$  (Figure 3), but the concomitant number of parameters and consequently the computational cost also increase. For this reason  $d = 2$  seems to be the most adequate choice.

**Figure 3.** Zoom from Figure 2. Prediction of seizure start ( $d = 2$ ).

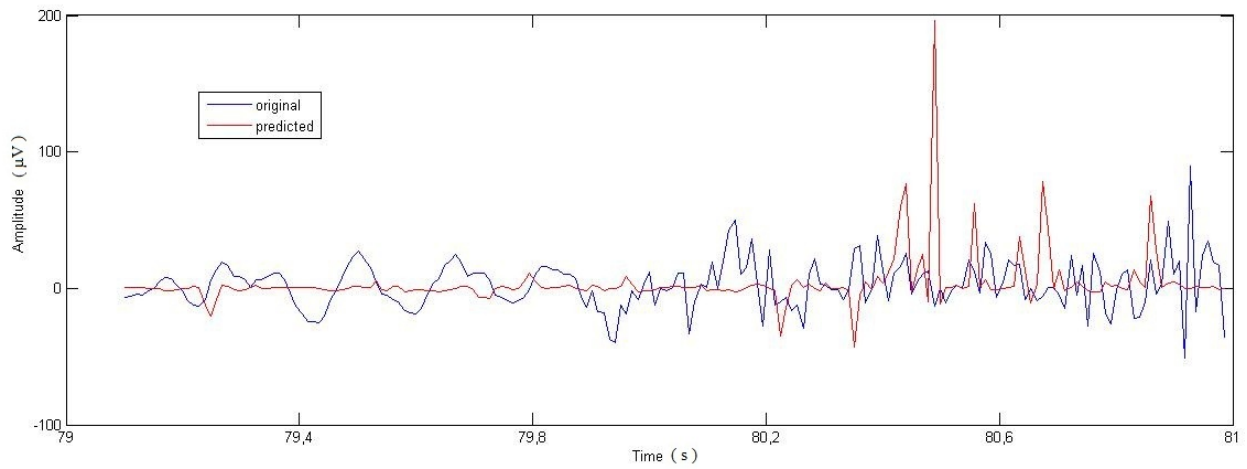


One critical point is the possible distortion due to spatial propagation of the seizure, since only data from the C4 electrode were analyzed and the source of the seizure was located in the left temporal location. For this reason, we also analyzed the T3 and T4 electrodes to compare with the C4 electrode results. Similar values for the time of the seizure start were obtained. As expected, due to its location, in the T3 electrode the seizure start was detected earlier, at  $t = 80.7$  s.

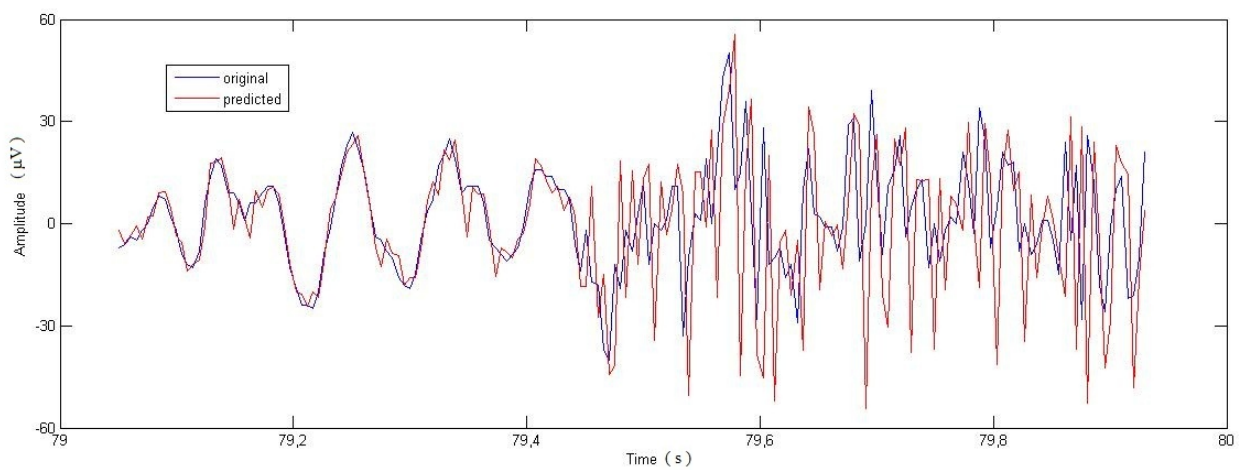
**Figure 4.** Seizure start (79–81 s).  $T \approx 1$  s.



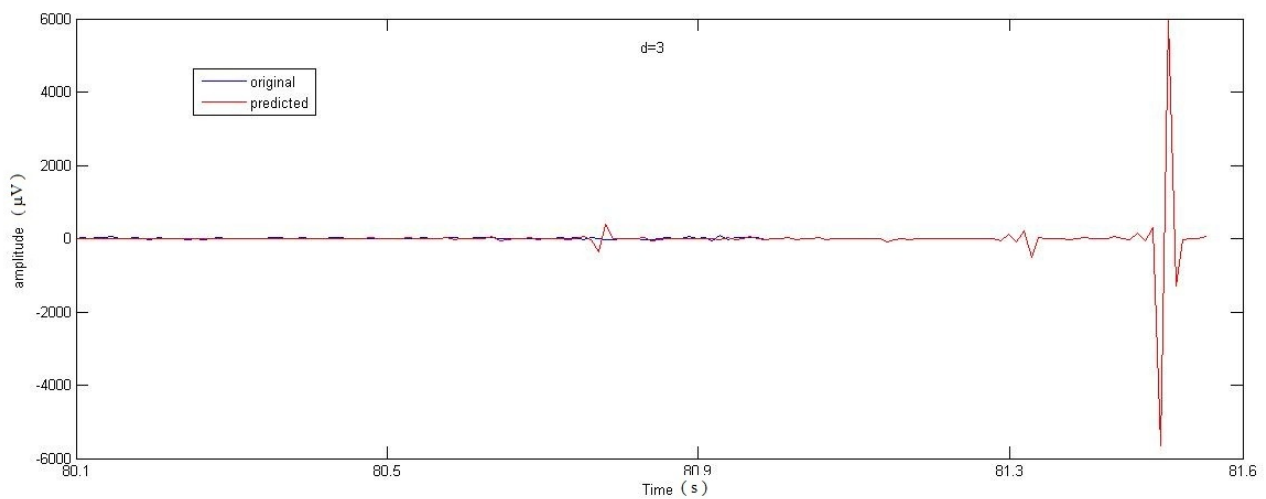
**Figure 5.** Seizure start (79–81 s).  $T \approx 0.5$  s.



**Figure 6.** Seizure start (79–81 s).  $T \approx 0.1$  s.



**Figure 7.** Prediction of seizure start. Embedding dimension  $d = 3$ .





#### 4. Conclusions

We have devised an EEG analysis tool for prediction purposes. The pertinent technique models chaotic time series in parametric fashion. In the case of tonic-clonic epileptic EEG data, we have shown that appropriate information theory tools, associated with the maximum entropy principle, provide valuable insights into the dynamics of neural activity and allow one to anticipate the occurrence of the tonic-clonic transition. Successful detection and anticipation of the associated seizure start was thus obtained with the method we have here proposed. Although it would be desirable to have longer anticipation time than the ones here achieved, our finding seems encouraging and justifies further exploration.

#### Acknowledgments

The authors thank Osvaldo Anibal Rosso for his helpful suggestions.

#### Conflicts of Interest

The authors declare no conflicts of interest.

#### References

1. Rosso, O.A.; Martín, M.T.; Plastino, A. Brain electrical activity analysis using wavelet-based informational tools (II): Tsallis non-extensivity and complexity measures. *Physica A* **2003**, *320*, 497–511.
2. Takens, F. Dynamical Systems and turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*, Proceedings of a Symposium Held at the University of Warwick 1979/80, Coventry, UK, 1979; Rand, D., Young, L.-S., Eds.; Lecture Notes in Mathematics, Volume 898; Springer: Berlin/Heidelberg, Germany, 1981.
3. Diambra, L.; Plastino, A. Modelling time series using information theory. *Phys. Lett. A* **1996**, *216*, 278–282.
4. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.
5. Jaynes, E.T. Information Theory and Statistical Nechanics II. *Phys. Rev.* **1957**, *108*, 171–190.
6. Mallat, S. *A Wavelet Tour of Signal Processing*, 2nd ed.; Academic Press: San Diego, CA, USA, 1999.
7. Martín, M.T.; Plastino, A.; Vampa, V.; Judge, G. A parametric information-theory model for predictions in time series. *Physica A* **2014**, *405*, 63–69.
8. Rosso, O.A.; Martín, M.T.; Plastino, A. Evidence of self-organization in brain electrical activity using wavelet-based informational tools. *Physica A* **2005**, *347*, 444–464.