

Article

Selection of Embedding Dimension and Delay Time in Phase Space Reconstruction via Symbolic Dynamics

Mariano Matilla-García ^{1,*}, Isidro Morales ², Jose Miguel Rodríguez ³ and Manuel Ruiz Marín ³ 

¹ Facultad de Economicas y Empresariales, Universidad Nacional de Educación a Distancia (UNED), 28050 Madrid, Spain

² Telefónica, 28040 Madrid, Spain; isildur20022003@yahoo.es

³ Departamento Metodos Cuantitativos, Ciencias Juridicas y Lenguas Modernas, Universidad Politecnica de Cartagena, 30201 Cartagena, Spain; Jose.Rodriguez@upct.es (J.M.R.); manuel.ruiz@upct.es (M.R.M.)

* Correspondence: mmatilla@cee.uned.es

Abstract: The modeling and prediction of chaotic time series require proper reconstruction of the state space from the available data in order to successfully estimate invariant properties of the embedded attractor. Thus, one must choose appropriate time delay τ^* and embedding dimension p for phase space reconstruction. The value of τ^* can be estimated from the Mutual Information, but this method is rather cumbersome computationally. Additionally, some researchers have recommended that τ^* should be chosen to be dependent on the embedding dimension p by means of an appropriate value for the time delay $\tau_w = (p - 1)\tau^*$, which is the optimal time delay for independence of the time series. The C-C method, based on Correlation Integral, is a method simpler than Mutual Information and has been proposed to select optimally τ_w and τ^* . In this paper, we suggest a simple method for estimating τ^* and τ_w based on symbolic analysis and symbolic entropy. As in the C-C method, τ^* is estimated as the first local optimal time delay and τ_w as the time delay for independence of the time series. The method is applied to several chaotic time series that are the base of comparison for several techniques. The numerical simulations for these systems verify that the proposed symbolic-based method is useful for practitioners and, according to the studied models, has a better performance than the C-C method for the choice of the time delay and embedding dimension. In addition, the method is applied to EEG data in order to study and compare some dynamic characteristics of brain activity under epileptic episodes

Keywords: symbolic analysis; symbolic entropy; delay time selection; dynamic reconstruction



Citation: Matilla-García, M.; Morales, I.; Rodríguez, J.M.; Ruiz Marín, M. Selection of Embedding Dimension and Delay Time in Phase Space Reconstruction via Symbolic Dynamics. *Entropy* **2021**, *23*, 221. <https://doi.org/10.3390/e23020221>

Academic Editor: Christian H. Weiss

Received: 5 January 2021

Accepted: 8 February 2021

Published: 11 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

For the theory of state space reconstruction suggested by Packard, Takens et al. [1,2] is the base for data-driven analysis and prediction of chaotic systems. It can be proved through Taken's theorem [2] that the strange attractor of the chaotic systems could be properly recovered from only one projection of the dynamic system. The fundamental theorem of reconstruction of Takens establishes a sufficient condition (but not necessary) given by $p \geq 2d + 1$, where d is the fractal dimension of the underlying chaotic attractor, and p stands for the embedding dimension used for phase space reconstruction. Nevertheless, no condition is given regarding the time delay.

A popular method for state space reconstruction is the method of delays. It consists of embedding the observed scalar time series $\{X_t\}_{t \in I}$ in one p -dimensional space $X_p^\tau(t) = (X_t, X_{t+\tau}, \dots, X_{t+(p-1)\tau})$ for $t \in I$, where τ is the time delay for the reconstruction, p is the embedding dimension, and I is a set of time indexes of cardinality T . Notice that the number of points inserted in the p -dimensional space is $M = T - (p - 1)\tau$ and all dynamic properties, such as dependencies, periodicity, and complexity changes, can be extracted from it. That is, there is a differentiable homomorphism from the orbits of the chaotic attractor in the reconstructed space \mathbb{R}^p to the original system.

The selection of the parameters p and τ^* is a challenge. An improper choice can result in a spurious indication of a nonlinear complex structure when the system is linear. Albeit specialized literature provides different methods to select the parameters for state space reconstruction, none of them that are superior to the remaining ones in all aspects. In general, the optimal strategy for parameter selection will depend on the time series and a complexity measure (e.g., Lyapunov Exponents or Correlation Dimension). There are two different approaches to the selection of the parameters p and τ^* . The first approach considers that p and τ^* are selected independently from each other. For example, the G-P algorithm for the selection of p proposed by Albano et al. [3] and different proposals for the selection of the time delay τ^* based on Mutual Information [4], autocorrelation and high-order correlations [5], filling factor [6], wavering product [7], average displacement AD [8], and multiple autocorrelation [9]. The second approach considers that the parameters p and τ^* are closely related when the time series under consideration is noisy and has finite length. A great number of experiments indicate that p and τ^* are related with the time delay for independence of the time series through $\tau_w = (p - 1)\tau^*$. Therefore, a bad selection of the parameters will directly impact the equivalence between the original system and the reconstructed phase space. Thus, some authors are in favor of jointly selecting p and τ^* as, for example, the small-window solution [10], C-C method [11], and automated embedding [12].

Many researchers consider that the second approach (joint selection) is more reasonable than the first one (independent selection) in the engineering practice. They consider that the estimation of mutual information is rather cumbersome computationally, whereas the autocorrelation function only accounts for the linear dependence and therefore does not properly treat the presence of nonlinearities. On the other hand, the C-C method suggested by Kim et al. [11] is the most popular, which provides the delay time τ^* and embedding dimension p simultaneously by using correlation integral, and it has the advantage of low complexity and robustness in finite samples [13].

In the present paper, we propose a new method for selection p and τ^* based on symbolic dynamics and Information Theory. Symbolic Dynamics studies dynamical systems on the basis of the symbol sequences obtained for a suitable partition of the state space. The basic idea behind symbolic dynamics is to divide the phase space into finite number of regions and label each region by an alphabetical letter. In this regard, symbolic dynamics is a coarse-grained description of dynamics. Even though coarse-grained methods lose a certain amount of detailed information, some essential features of the dynamics may be kept, e.g., periodicities and dependencies, among others. Symbolic dynamics has been used for investigation of nonlinear dynamical systems (central references for the interested reader are [14–18]; for an overview, see Hao and Zheng [19]). In general terms, there is a broad agreement in that symbolization can increase the efficiency of finding and even quantifying information for characterizing and recognizing temporal patterns (see [20] for a review on experimental data). The process of symbolizing a time series is based upon the method of delay time coordinates, introduced by Takens, in order to carry out the phase space reconstruction.

Since the methods of state space reconstruction are based to some extent on detection of delays for which there is some sort of dependence (linear or nonlinear), and Symbolic Dynamics has been used as a statistical tool to detect the presence of dependence in time series [21]; symbolic dynamics is a suitable tool to select the optimal state space reconstruction parameters of chaotic time series.

Thus, we will select p and τ^* by translating the problem into symbolic dynamics and then we use an entropy measure associated with the symbols space (symbolic entropy) as a tool for the parameter selection. On the one hand, we have compared the performance of the proposed method with other available methods. Results seem to be in favor of this proposal. On the other hand, and from an empirical point of view, we have applied it to EEG data, which allows for understanding some dynamic characteristics of brain activity under epileptic episodes.

The rest of the paper is structured as follows: in Section 2, we introduce the basic concepts of symbolic analysis, and we also provide a symbolization procedure that works for estimation of the parameters for Phase Space reconstruction. In Section 3, we show the performance of the symbolic method to estimate phase space reconstruction parameters, and we compare it with the well known Mutual Information based methods and C-C method. In Section 4, the new techniques presented in this paper are applied to a real EEG database obtained from the University of Bonn and studied well for understanding epileptic phenomena. Finally, Section 5 presents conclusions.

2. Definitions and Symbolization Procedure

In this section, we will introduce some definitions and basic notations referred to symbolic dynamics.

Let $\{X_t\}_{t \in I}$ be a real valued time series. We will use symbolic analysis to study the state space reconstruction parameters associated with it. Symbolic analysis in our context is a coarse-grained approach to the study of time series, consisting of embedding the time series in a p -dimensional space, then constructing a partition of this p -dimensional space and labeling each set of the partition with a symbol. Therefore, all the p -dimensional vectors belonging to the same set of the partition are labelled with the same symbol. Afterwards, with information theory based measures, we will study the distribution of the symbols that will help us in the estimation of the parameters for state space reconstruction.

More concretely, in mathematical terms, given a positive integer $p \geq 2$, and a time delay τ , we consider that the time series is embedded in an p -dimensional space as follows:

$$X_p^\tau(t) = (X_t, X_{t+\tau}, \dots, X_{t+(p-1)\tau}). \tag{1}$$

The parameter p is usually known as embedding dimension and $X_p^\tau(t)$ p, τ -history.

Next, given a positive real number ϵ and in order to provide a partition of \mathbb{R}^p , we define for any element $v = (v_1, v_2, \dots, v_p) \in \mathbb{R}^p$ the following indicator function:

$$\delta_{ij}(v) = \begin{cases} 1 & \text{if } |v_i|, |v_j| < \epsilon \text{ or } |v_i|, |v_j| \geq \epsilon \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

That is, $\delta_{ij}(v) = 1$ if and only if its entries v_i and v_j satisfy that $|v_i|$ and $|v_j|$ are both either smaller or greater than ϵ . Let the set Γ_p be the set of cardinality 2^{p-1} formed by the vectors of length $p - 1$ with entries in the set $\{0, 1\}$. Then, we can define a map $f_\epsilon : \mathbb{R}^p \rightarrow \Gamma_p$ defined as

$$f_\epsilon(v = (v_1, v_2, \dots, v_p)) = (\delta_{12}(v), \delta_{13}(v), \dots, \delta_{1p}(v)).$$

The map f_ϵ defines an equivalence relation in \mathbb{R}^p such that $v_1 \sim v_2$ if and only if $f_\epsilon(v_1) = f_\epsilon(v_2)$. Therefore, this equivalence relation provides a partition of \mathbb{R}^p in 2^{p-1} disjoint sets. Each of these sets is labeled with an element of Γ_p . The elements in Γ_p are called symbols and f_ϵ symbolization map. In general, if $\pi \in \Gamma_p$ is a symbol and $v \in \mathbb{R}^p$ is such that $f_\epsilon(v) = \pi$, then we will say that v is of π -type.

Next, we are interested in the application of the symbolization map f_ϵ to the p, τ -history $X_p^\tau(t) = (X_t, X_{t+\tau}, \dots, X_{t+(p-1)\tau})$. Notice that $f_\epsilon(X_p^\tau(t))$ is a vector (symbol) whose i -th entry provides information on whether $|X_t|$ and $|X_{t+i}|$ are both either smaller or greater than ϵ . Then, we want to extract information on the dynamics of the time series $\{X_t\}_{t \in I}$ by using information theory based measures on its associated symbols distribution $\left\{ f_\epsilon(X_p^\tau(t)) \right\}_{t=1}^{T-(p-1)\tau}$. More concretely, we can estimate the probability of a symbol $\pi \in \Gamma_p$ as

$$p_\pi = \frac{\# \{X_p^\tau(t) \text{ of } \pi - \text{type}\}}{T - (p - 1)\tau} \tag{3}$$

Now, under this setting, given a time delay τ and embedding dimension $p \geq 2$, we can define the symbolic entropy of a time process $\{X_t\}_{t \in I}$. This entropy is defined as the Shannon's entropy [22] of the 2^{p-1} distinct symbols as follows:

$$h(p, \tau) = - \sum_{\pi \in \Gamma_m} p_\pi \ln(p_\pi). \quad (4)$$

Symbolic entropy $h(p, \tau)$ is the information contained in comparing p, τ -histories generated by the time process. Notice that $0 \leq h(p, \tau) \leq \ln(n)$, where the lower bound is attained when only one symbol occurs, and the upper bound for a completely random system (i.i.d temporal sequence) where all possible symbols appear with the same probability.

Then, if $\tau = \tau^*$ is an optimal time delay, for a positive integer k , the dependence between X_t and $X_{t+k\tau^*}$ vanishes, and hence the symbolic entropy associated with the time series $\{X_t\}$ should be maximum. Therefore, in order to select the optimal time delay τ^* , we select the first τ satisfying

$$\tau^* = \arg \max_{\tau} \{h(p, \tau)\} \quad (5)$$

With respect to the optimal embedding window $\tau_w = (p-1)\tau^*$, this can be associated with the mean orbital period P_w of low-dimensional chaotic systems that shows pseudo-periodicity. That is, P_w can be considered the time dependence of the chaotic time series. Although the chaotic systems oscillate without periodicity, low dimensional chaotic systems show pseudo-periodicity. The mean orbital period could naturally be associated with the mean time between two consecutive visits to a Poincare section [23]. For the time series with mean orbital period P_w , all points at a time multiple of P_w are in the same Poincare section in phase space. Therefore, a local minimum of symbolic entropy $h(p, \tau)$ is reached for $\tau = kP_w$ and thus

$$\tau_w = \arg \min_{\tau} \{h(p, \tau)\}. \quad (6)$$

To finish this section, we are going to illustrate the symbolization procedure with an easy example. Let $\{X_t\}_{t \in I}$ be the following finite time series:

$$\{X_1 = 2; X_2 = -7; X_3 = -12; X_4 = 5; X_5 = -1; X_6 = 9; X_7 = 14\} \quad (7)$$

and assume that $\epsilon = 3$, $\tau = 1$ and $p = 3$. Then, the symbols' set remains as

$$\Gamma_3 = \{(0,0); (0,1); (1,0); (1,1)\}.$$

Under this setting, we can construct the following five p, τ -histories: $X_3^1(1) = (1, -7, -12)$; $X_3^1(2) = (-7, -12, 5)$; $X_3^1(3) = (12, 5, -1)$; $X_3^1(4) = (5, -1, 9)$; and $X_3^1(5) = (-1, 9, 14)$. Then, the symbolization map f_ϵ associate each p, τ -history to a symbol. Concretely, $f_3(X_3^1(1) = (1, -7, -12)) = (0,0)$ because the first entry of the m-history, 1 is in absolute value smaller than $\epsilon = 3$ while the second and the third are both greater than $\epsilon = 3$, and hence the agreement indicator that defines the symbolization map takes the value 0. Similarly, we find that $X_3^1(2)$ is of (1,1)-type, $X_3^1(3)$ is of (1,0)-type, $X_3^1(4)$ is of (0,1)-type, and $X_3^1(5)$ is of (0,0)-type. Thus, we can estimate the symbols distribution by its relative frequency $p((0,0)) = \frac{2}{5}$, $p((0,1)) = p((1,0)) = p((1,1)) = \frac{1}{5}$ and the entropy associated with them $h(p, \tau) = h(3, 1) = -\frac{2}{5} \log(\frac{2}{5}) - \frac{1}{5} \log(\frac{1}{5}) - \frac{1}{5} \log(\frac{1}{5}) - \frac{1}{5} \log(\frac{1}{5}) = 1.3322$.

Selection of p and ϵ for Finite Sample Sizes

When determining the parameters of phase space reconstruction of a finite chaotic time series by using the symbolic entropy, one needs to select in advance the values of p and ϵ . In addition, sample size T also plays an important role. In [21], some general criteria are recommended to select the embedding dimension p and sample size T in order to compute the symbolic entropy. First, the sample size T should be as larger than the number

of symbols 2^{p-1} of the symbolization map f_ϵ . Second, from a statistical point of view, data sets must contain at least five times the number of possible events or symbols. Thus, the embedding dimension will be the largest positive integer p that satisfies $5 \cdot 2^{p-1} \leq T$.

To select ϵ , we propose to use a data driven method which is based on symbolic entropy. Particularly, we partially rely on the methodology described in [24] based on the construction of peak detection functions (FPs). The selected ϵ will be the largest ϵ that locally maximizes the absolute value of a pick function $FP(i, x_i)$, where FP associates values to the symbolic entropy of a time series. More concretely, define

$$FP^+(k, \tau) = \frac{\max\{h_\tau - h_{\tau-1}, h_\tau - h_{\tau-2}, \dots, h_\tau - h_{\tau-k}\} + \max\{h_\tau - h_{\tau+1}, h_\tau - h_{\tau+2}, \dots, h_\tau - h_{\tau+k}\}}{2} \quad (8)$$

and

$$FP^-(k, \tau) = \frac{\min\{h_\tau - h_{\tau-1}, h_\tau - h_{\tau-2}, \dots, h_\tau - h_{\tau-k}\} + \min\{h_\tau - h_{\tau+1}, h_\tau - h_{\tau+2}, \dots, h_\tau - h_{\tau+k}\}}{2} \quad (9)$$

where $h_{\tau+l} = h(p, \tau + l)$ for $l = 0, 1, \dots, k$. The functions $FP^+(k, \tau)$ (respectively $FP^-(k, \tau)$) allows for selecting the time delay τ for which value $h(p, \tau)$ is maximum (respectively minimum) in the neighborhood of $(\tau - k, \tau + k)$. As stated in [24], values of k in the range $[3, 5]$ are usually suitable. Notice that, by construction, $0 < \epsilon < \max\{X_t\}$. Then, the selected parameter, namely ϵ^* , will be the one in the interval $(0, \max\{X_t\})$ that satisfies

$$\epsilon^* = \max_{\epsilon} \left\{ \max_{\tau} \{FP^+(k, \tau)\}, \min_{\tau} \{|FP^-(k, \tau)|\} \right\} \quad (10)$$

3. Simulation Analysis

The following examples illustrate the performance of the proposed symbolic method when estimating the parameters time delay τ^* and embedding dimension p for phase space reconstruction of a chaotic time series. The aim of this set of simulations is, firstly, to empirically evaluate the performance of the new symbolic procedure to select the “correct” parameters. Secondly, we aim to compare with the symbolic method with other competitive available methods that have been commented in the introductory section and that are fully documented in the bibliographical references of this paper.

To this end, we extract univariate time series $\{X_t\}$ of length $T = 3000$ from five chaotic systems that have been extensively studied. In all cases, we set the embedded time series in a six-dimensional space that is $p = 6$. To evaluate the performance of the novel symbolic method, we compare the performances of the new method with other available selection methods: the C-C method (C-C), the Nearest Neighbor method, and the method based on the first minimum of the autocorrelation function (FAC) selection parameters of these last two methods are based on the Mutual Information (MI) criteria.

Scientific literature has shown that the C-C method has a good performance when used for selecting time delays and embedding dimensions. Thus, the C-C method can be thought of as a natural competitor and therefore it is worth comparing the performance against it. For this reason, we will compare results for several well-known dynamic systems. In order to compare and evaluate the performance of each method, we will use the selected parameters of each method for reconstructing the attractor and estimating two complexity measures of each system. These two measures are theoretically known for each of the three systems, and therefore they are used as a base of comparison. A final user will prefer using reconstruction parameters that lead to estimations that are as close as possible to the theoretical ones.

Accordingly, we will use the following systems to conduct the comparisons:

- Lorenz system [25]:

$$\begin{aligned} \dot{x} &= -a(x - y) \\ \dot{y} &= -xz + cx - y \\ \dot{z} &= xy - bz \end{aligned} \quad (11)$$

The time series was obtained by projecting the x -coordinate of the system defined by the parameters $a = 16$, $c = 45.92$, $b = 4$, with an integral step 0.01, and initial conditions $x_0 = -1$, $y_0 = 0$ $z_0 = 1$. The computed optimal ratio is $\epsilon^* = 1.2\sigma_x$, where σ_x is the standard deviation of the chaotic time series under consideration. For this optimal ratio, Figure 1 illustrates the series of the normalized symbolic entropy $h(6, \tau)/6$ as a function of the time delay τ for the Lorenz system. Clearly, we observe that the first local maximum is attained at $\tau^* = 12$ and the minimum at $\tau_w = 46$. Then, an estimated value of embedding dimension p can be computed by solving $\tau_w = (p - 1)\tau^*$ obtaining an approximate value of $p = 5$. For the Mutual Information method, the optimal time delay was $\tau^* = 11$, while, for the C-C method, the estimated parameters were $\tau^* = 10$ $\tau_w = 100$ and $p = 11$. Notice that the optimal time delay τ^* estimated by the three methods are quite close to each other while the estimated time delay window τ_w strongly differs between C-C and symbolic methods:

- Rossler system [26]:

$$\begin{aligned}\dot{x} &= -y - z \\ \dot{y} &= x + dy \\ \dot{z} &= z(x - f) + e\end{aligned}\quad (12)$$

The time series was obtained by projecting the x -coordinate of the system defined by the parameters $d = 0.15$, $e = 0.2$ and $f = 10$, with an integral step 0.05, and initial conditions $x_0 = -1$, $y_0 = 0$ and $z_0 = 1$. The computed optimal ratio is $\epsilon^* = 0.4\sigma_x$, where σ_x is the standard deviation of the chaotic time series under consideration. Figure 2 shows the normalized symbolic entropy $h(6, \tau)/6$ as a function of the time delay τ for the Rossler system. It can be seen that the selected parameters by the symbolic method are $\tau^* = 18$ and $\tau_w = 121$, and consequently the estimated value for p is 8. The estimated time delay for Mutual Information method is $\tau^* = 20$. For the C-C method, the estimated parameters are $\tau^* = 17$ and $\tau_w = 191$. Again, the optimal time delay τ^* estimated by the three methods is very similar while the time delay window τ_w estimated by the C-C method is much different than the one estimated with symbolic entropy.

- Duffing System [27]:

$$\begin{aligned}\dot{x} &= y \\ \dot{y} &= -gy - kx(1 + x^2) + l\cos z \\ \dot{z} &= v\end{aligned}\quad (13)$$

The time series was obtained by projecting the x -coordinate of the system defined by the parameters $g = 0.05$, $k = 0.25$, $l = 7.5$ and $v = 1$, with an integral step 0.05, and initial conditions $x_0 = -1$, $y_0 = 0$ $z_0 = 1$. The computed optimal ratio is $\epsilon^* = 0.275\sigma_x$, where σ_x is the standard deviation of the chaotic time series under consideration. Figure 3 shows the normalized symbolic entropy $h(6, \tau)/6$ as a function of the time delay τ for the Duffing system. The estimated optimal time delay and time delay window with the symbolic method are $\tau^* = 14$ and $\tau_w = 126$, respectively. Then, the estimated embedding dimension is $p = 10$. As in the previous examples, the estimated time delay for the Mutual Information method ($\tau^* = 12$) and for the C-C method ($\tau^* = 12$) are fairly close to the one estimated by symbolic method. Again, the time delay window estimated based on the C-C method $\tau_w = 161$ is far from the one estimated with symbolic method.

These first three models are well-known and well-studied and have served as a base of comparison of new techniques for similar aims as this paper. In order to complete this analysis, we have also considered the next two models that we refer to as the Mackey–Glass model and Chen model:

- Mackey-Glass system [28]:

$$\frac{dx}{dt} = \frac{ax(t-\tau)}{1+x^c(t-\tau)} - bx \tag{14}$$

The time series was obtained by fixing parameters $a = 0.2, b = 0.1, c = 10, y = 17$, with initial conditions $x(t < 0) = 0$ y $x(t = 0) = 1.2$. The first 2000 iterations were discarded. The computed optimal ratio is $\epsilon^* = 0.79\sigma_x$, where σ_x is the standard deviation of the chaotic time series under consideration. The estimated optimal time delay and time delay window with the symbolic method are $\tau^* = 13$ and $\tau_w = 49$, respectively. Figure 4 shows the normalized symbolic entropy $h(6, \tau)/6$ as a function of the time delay τ for the Mackey-Glass system. Then, the estimated embedding dimension is $p = 5$. In this case, the estimated time delay for the Mutual Information method ($\tau^* = 12$) and for the C-C method ($\tau^* = 14$) are fairly close to the one estimated by symbolic method ($\tau = 13$). Again, the time delay window estimated based on the C-C method $\tau_w = 166$ is far from the one estimated with symbolic method ($\tau_w = 49$):

- Chen system [29]:

$$\begin{aligned} \dot{x} &= a(y - x) \\ \dot{y} &= (c - a)x + cy - xz \\ \dot{z} &= xy - bz \end{aligned}$$

The time series was obtained by projecting the x -coordinate of the system defined by the parameters $a = 35, b = 3, c = 28$, with an integral step 0.01, and initial conditions $x_0 = -1, y_0 = 0$ y $z_0 = 1$. The first 2000 iterations were discarded. The computed optimal ratio is $\epsilon^* = 0.89\sigma_x$, where σ_x is the standard deviation of the chaotic time series under consideration. The estimated optimal time delay and time delay window with the symbolic method are $\tau^* = 11$ and $\tau_w = 60$, respectively. Figure 5 shows the normalized symbolic entropy $h(6, \tau)/6$ as a function of the time delay τ for the Chen system. Then, the estimated embedding dimension is $p = 6$. As in the previous examples, the estimated time delay for the Mutual Information method ($\tau^* = 10$) and for the C-C method ($\tau^* = 9$) are fairly close to the one estimated by symbolic method. Again, the time delay window estimated based on the C-C method $\tau_w = 104$ is far from the one estimated with symbolic method ($\tau_w = 60$).

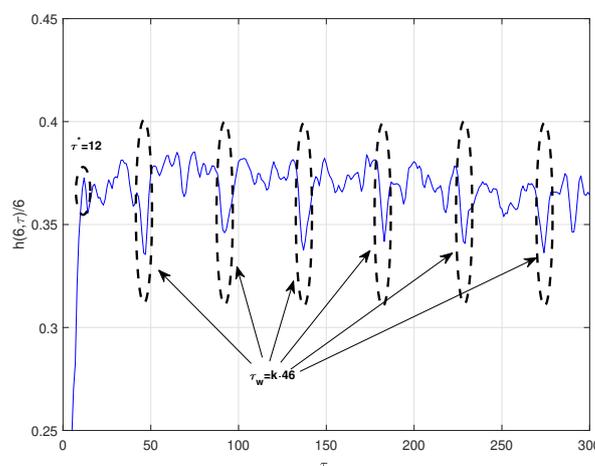


Figure 1. Normalized symbolic entropy $h(6, \tau)/6$ for $\epsilon = 1.2\sigma_x$ of Lorenz system.

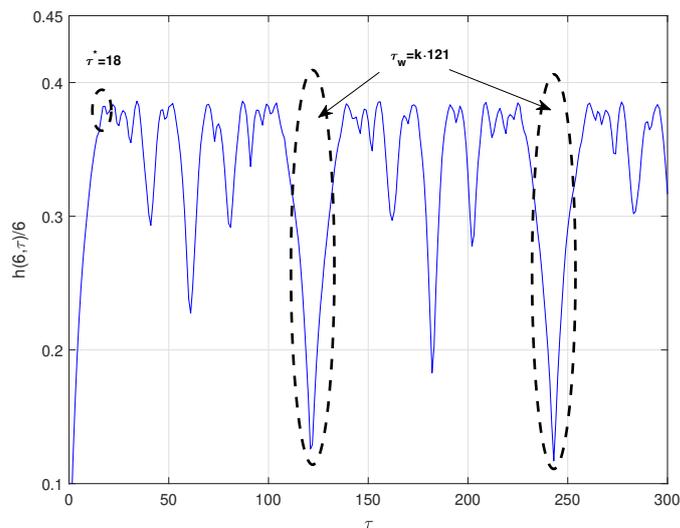


Figure 2. Normalized symbolic entropy $h(6, \tau)/6$ for $\epsilon = 0.4\sigma_x$ of Rossler system.

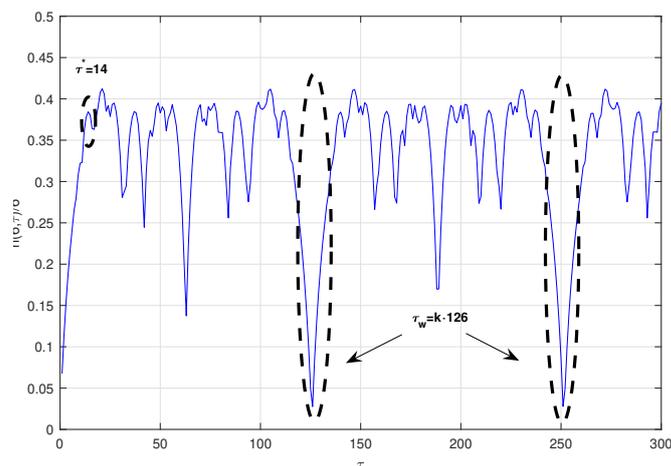


Figure 3. Normalized symbolic entropy $h(6, \tau)/6$ for $\epsilon = 0.275\sigma_x$ of the Duffing system.

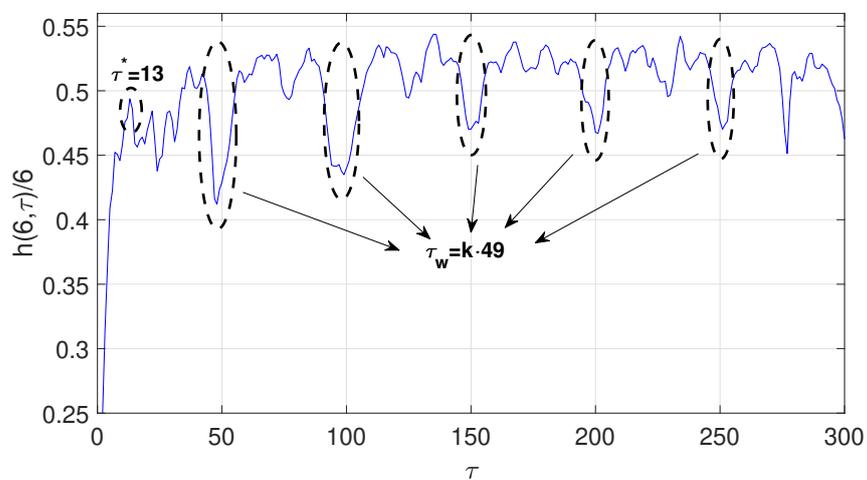


Figure 4. $h(6, \tau)/6$ for $\epsilon = 0.79\sigma_x$ of the Mackey–Glass system.

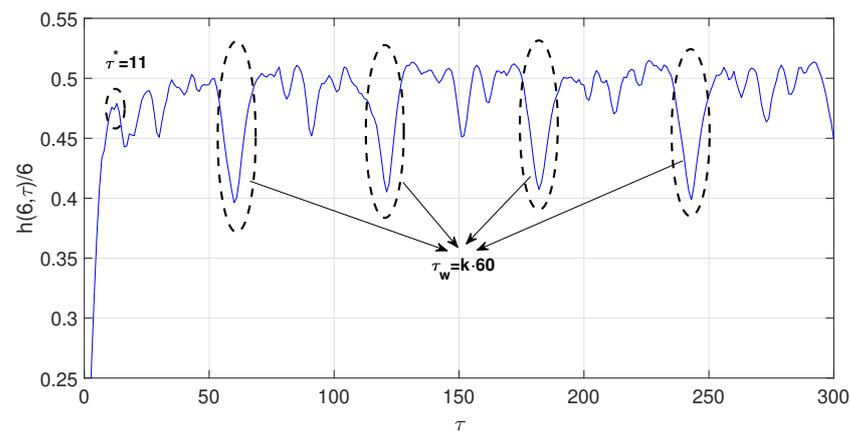


Figure 5. Normalized symbolic entropy $h(6, \tau)/6$ for $\epsilon = 0.89\sigma_x$ of the Chen system.

Table 1 summarizes for each method the estimated parameters for phase space reconstruction of the five systems. Bold is reserved for the results obtained with the new selecting method.

Table 1. Estimated parameters for phase space reconstruction for the studied systems.

System	MI Method		C-C Method		Symbolic Method			NN		FAC	
	τ^*	τ^*	τ_w	p	τ^*	τ_w	p	τ^*	p	τ^*	p
Lorenz	11	10	100	11	12	46	5	11	3	11	3
Rosler	20	17	191	12	18	121	8	20	4	20	4
Duffing	12	14	161	12	14	126	10	12	2	12	2
Mc-Glass	12	14	166	13	13	49	5	12	4	12	4
Chen	10	9	104	27	11	60	6	10	4	10	4

In order to check whether the symbolic method is reliable when estimating the parameters for phase space reconstruction, τ^* , τ_w , and p , we will compute, based on this estimation, two complexity measures for each one of the systems that needs these parameters for its computation. These complexity measures are the largest Lyapunov exponent LLE [30], which is a measure of the complexity of the time process, and the Correlation Dimension \mathcal{D} [31], which is a measure of the dimension of the space occupied by the chaotic attractor. For the computation of these two geometric invariants, the time delay τ^* and embedding dimension p are essential parameters, and a bad selection of them would produce a big bias in LLE and \mathcal{D} . The largest Lyapunov exponent LLE for the five systems have been computed in [27,32–35] and the Correlation Dimension \mathcal{D} in [32,33,36,37]. Furthermore, we have completed the study by increasing the sample size to 10,000 observations. Tables 2 and 3 show the values of LLE and \mathcal{D} based on the values of the estimated parameters τ^* , τ_w and p with symbolic and C – C methods together with the reference values, respectively.

Table 2. Largest Lyapunov exponent LLE based on the estimation of the phase space parameters τ^* , τ_w and p if C-C method and symbolic method, together with the reference true value. Values in parentheses report estimated LLE for series of 10,000 observations.

	LLE Reference Value	Estimated LLE C-C Method	Estimated LLE Symbolic Method	Estimated LLE NN	Estimated LLE FAC
Lorenz	1.500	0.940 (1.667)	1.50 (1.670)	1.438 (1.659)	0.771 (0.742)
Rosler	0.090	0.095 (0.066)	0.09 (0.080)	0.061 (0.079)	1.108 (0.068)
Duffing	0.183	0.168 (0.021)	0.184 (0.200)	0.014 (0.215)	0.019 (0.02)
Mc-Glass	0.007	0.006 (0.008)	0.007 (0.007)	0.006 (0.009)	0.007 (0.008)
Chen	1.997	1.852 (1.773)	1.982 (1.852)	2.483 (2.160)	1.359 (1.120)

Table 3. Correlation dimension \mathcal{D} based on the estimation of the phase space parameters τ^* , τ_w , and p of the C-C method and symbolic method, together with the reference true value. Values in parentheses report estimated D for series of 10,000 observations.

	\mathcal{D} Reference Value	Estimated \mathcal{D} C-C Method	Estimated \mathcal{D} Symbolic Method	Estimated \mathcal{D} NN	Estimated \mathcal{D} FAC
Lorenz	2.06	1.93 (2.41)	2.01 (2.02)	1.76 (2.23)	2.02 (3.63)
Rosler	2.01	2.13 (2.31)	2.09 (2.01)	1.79 (1.77)	1.92 (1.49)
Duffing	2.23	2.10 (2.56)	2.21 (2.37)	1.13 (1.16)	1.51 (1.47)
Mc-Glass	2.10	2.83 (2.20)	2.12 (2.11)	1.94 (1.80)	2.01 (1.97)
Chen	2.16	2.19 (3.67)	2.13 (2.12)	2.05 (2.26)	2.36 (3.49)

We can observe the estimated values of the largest Lyapunov exponent and Correlation dimension based on the Symbolic method in Tables 2 and 3, respectively. Importantly, these symbolic-based estimations are very close to their reference (theoretical) values, regardless of the sample size, which suggests the good behavior of the new method for reconstruction of the dynamics of the system. On the other hand, we were wondering if the symbolic method is competitive with its main competitor, namely, the C-C method. In this regard, we can devise that the estimated values for the Lyapunov Exponents are clearly in favor of the Symbolic method as the estimation is closer to the theoretical reference value than in the case of the C-C method. This is true for the five systems. Similar conclusions can be obtained from the results regarding correlation dimension: the symbolic-based estimated dimensions are closer to the true value than C-C estimation, regardless the studied system. On the other hand, methods based on nearest neighbors and autocorrelation function are reported. Results show the symbolic based method also has better empirical behavior. All of these results could be explained by a wrong selection of delay time window τ_w by the C-C method as stated in [23,38–40].

4. EEG Dynamics under Epileptic Activity

The Electroencephalogram (EEG) is a spontaneous bioelectricity activity that is produced by the central nervous system. Therefore, EEG can be understood as a representative signal containing information about the activity of the brain. Currently, EEG is widely used in clinic and neuralelectricity physiological research. The shape of the waves may contain useful information about the state of the brain. EEG does include abundant information about the state and change of the neural system.

The dynamics of brain activity is considered to be of a nonlinear nature. Accordingly, EEG signals are studied by means of nonlinear dynamic tools. Indeed, a large body of studies have reported that the EEG was derived from chaotic systems [41–44].

In this section of the paper, we apply the symbolic-based approach for reconstruction of dynamics generated by empirical EEG recording from a public dataset by the University of Bonn [41]. Epilepsy is characterized by recurring seizures in which abnormal electrical activity in the brain causes the loss of consciousness or a whole body convulsion. From this point of view, our results will contribute to the empirical analysis of role on nonlinear dynamics in epileptology. The Bonn University EEG database is comprised of five types of EEG signals (EEG recordings from healthy volunteer with eyes open and closed, epilepsy patients in the epileptogenic zone during a seizure-free interval and in an opposite brain zone, and epilepsy patients during epileptic seizures) were studied.

To conduct this empirical analysis, we firstly use Theiler's method of surrogate data to distinguish between linearity and nonlinearity. To do so, the null hypothesis of linearity is tested against nonlinearity [45]. Chaos cannot come from a linear signal. Secondly, we test for chaoticity against pure stochasticity. Linear signals are expected to be of stochastic nature while nonlinear signals can come from either a stochastic process or a pure chaotic one. The statistical test for chaos [46] tests the null hypothesis of chaos versus the alternative of stochastic process. We also estimate correlation dimension using Theiler's approach in order to exclude time correlated states in the correlation integral estimation [47].

Table 4. Complexity measures for EEG data sets.

Brain Status	τ^*	τ_w	p	Non Linearity Test	Chaos vs. Stochastic	Estimated \mathcal{D}
Healthy with open eyes	13	98	9	NL ($pval = 0.03$)	Stochastic($pval < 0.001$)	5.56
Healthy with closed eyes	9	109	13	L ($pval = 0.64$)	-	7.98
Seizure free non-epileptogenic zone	9	120	14	L ($pval = 0.76$)	-	4.99
Seizure free epileptogenic zone	18	108	7	NL ($pval < 0.001$)	Chaos($pval = 1$)	3.48
Seizure activity	11	65	7	NL ($pval < 0.001$)	Chaos($pval = 1$)	3.69

Table 4 collects the outcomes of all procedures. Results firstly indicate that the brain's activity is of a nonlinear nature for a healthy person with open eyes and for records of epileptic person regardless if s/he is under seizure activity or not, whenever measurement is done in the epileptogenic zone. The test for chaos applied to nonlinear signals helps to conclude that only the nonlinear dynamics found for epileptic patients are statistically compatible with chaotic dynamics, while the dynamics are nonlinear stochastic for a healthy person with open eyes. Finally, the estimated correlation dimensions show how (correlation) dimension is reduced as the process moves from stochastic to chaotic, as expected.

These results support the nonlinear deterministic structure of brain dynamics related to epileptic activity as earlier reported in [48,49]. Our estimates of correlation dimensions are in line with other previous studies [50] on the same dataset, although with different parameter configurations. Thus, the conclusion in this regard is that epileptic seizures are emergent states with reduced dimensionality compared to non-epileptic activity. This is in line with the clinical common knowledge that establishes that healthy systems evolve with time and their adaptive capability is higher, resulting in higher complexity. On the other hand, the alternations in the structural components and/or decreased functional capability of the subsystem cause dysfunction in the regularity mechanism of the overall system, which results in the loss of complexity, as indicated in [51,52].

5. Conclusions

In this paper, we have introduced a new method based in Symbolic Dynamics, for the estimation of the phase space reconstruction parameters τ^* , τ_w and p . In the simulation analysis, we applied the Symbolic method to choose the phase space reconstruction parameters from the time series generated from several dynamical models that have been well-studied and used for evaluating the ability of different reconstruction methodologies. The values found for τ^* agree well with those found for the mutual information and the C-C method. The values found for τ_w do not agree with the values estimated by the C-C method. For this reason, in order to clarify which method for selecting phase space reconstruction parameters is more reliable, we use them in the computation of two complexity measures, namely largest Lyapunov exponent (LLE) and Correlation dimension (\mathcal{D}). Results indicate that the parameters estimated by the Symbolic method produces a closer approach to reference (theoretical) values of LLE and \mathcal{D} than the C-C method. Finally, the proposed method is used to study the dynamics of brain activity under epilepsy by means of real EEG signals. The empirical results hint that epileptic patients show chaotic dynamics in EEG signals. Furthermore, our results are statistically significant and therefore hint the potential of symbolic based tools in distinguishing healthy and epileptic subjects.

Author Contributions: Conceptualization, M.M.-G. and M.R.M.; investigation, I.M. and J.M.R.; methodology, M.M.-G. and M.R.M.; software, I.M.; writing—original draft, I.M.; Writing—review & editing, M.M.-G., I.M., J.M.R. and M.R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This was partially funded by Ministerio de Ciencia e Innovación under Grant PID2019-107192GB-I00 and under Grant PID2019-107800GB-I00/AEI/10.13039/501100011033. This study is part of the collaborative activities carried out under the program Groups of Excellence of the region

of Murcia, the Fundacion Seneca, Science and Technology Agency of the region of Murcia project 19884/GERM/15.

Acknowledgments: Mariano Matilla-García is grateful for the financial support of Ministerio de Ciencia e Innovación under Grant PID2019-107192GB-I00. Manuel Ruiz is grateful for the financial support of Ministerio de Ciencia e Innovación of Spain under Grant PID2019-107800GB-I00/AEI/10.13039/501100011033. This study is part of the collaborative activities carried out under the program Groups of Excellence of the region of Murcia, the Fundación Séneca, Science and Technology Agency of the region of Murcia project 19884/GERM/15.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Packard, N.H.; Crutchfield, J.P.; Farmer, J.D.; Shaw, R.S. Geometry from a time series. *Phys. Rev. Lett.* **1980**, *45*, 712. [CrossRef]
- Takens, F. Detecting Strange Attractors in Turbulence. In *Dynamical Systems and Turbulence*; Springer:Berlin/Heidelberg, Germany, 1981.
- Albano, A.M.; Muench, J.; Schwartz, C.; Mees, A.I.; Rapp, P.E. Singular-value decomposition and the Grassberger-Procaccia algorithm. *Phys. Rev. A* **1988**, *38*, 3017. [CrossRef]
- Fraser, A.M.; Swinney, H.L. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **1986**, *33*, 1134. [CrossRef] [PubMed]
- Albano, A.M.; Passamante, A.; Farrell, M.E. Using high-order correlations to define an embedding window. *Physica D* **1991**, *54*, 85–97. [CrossRef]
- Pfister Buzug, T. Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global static and local dynamical behavior of strange attractors. *Phys. Rev. A* **1992**, *45*, 7073–7084. [CrossRef]
- Pfister Buzug, T. Comparison of algorithms calculating optimal embedding parameters for delay time coordinates. *Physica D* **1992**, *58*, 127–137. [CrossRef]
- Rossenstein, M.T.; Colins, J.J.; de Luca, C.J. Reconstruction expansion as a geometry-based framework for choosing proper delay times. *Physica D* **1994**, *73*, 82–98. [CrossRef]
- Lin, J.; Wang, Y.; Huang, Z.; Shen, Z. Selection of proper time-delay in phase space reconstruction of speech signals. *Signal Process* **1999**, *15*, 220–225.
- Kugiumtzis, D. State space reconstruction parameters in the analysis of chaotic time series—The role of the time window length. *Physica D* **1996**, *95*, 13–28 [CrossRef]
- Kim, H.S.; Eykholt, R.; Salas, J.D. Nonlinear dynamics, delay times, and embedding windows. *Physica D* **1999**, *127*, 48–60. [CrossRef]
- Otani, M.; Jones, A. Automated Embedding and Creep Phenomenon in Chaotic Time Series. 2000. Available online: <http://users.cs.cf.ac.uk/O.F.Rana/Antonia.J.Jones/UnpublishedPapers/Creep.pdf> (accessed on 5 January 2021).
- Lu, J.H.; Lu, J.A.; Chen, S.H. *Analysis and Application of Chaotic Time Series*; Wuhan University Press: Wuhan, China, 2002.
- Matilla-Garcia, M.; Ruiz, M. Detection of nonlinear structure in time series. *Econ. Lett.* **2009**, *105*, 1–6. [CrossRef]
- Alves, J.F.; Ramos, J.S. Kneading theory: A functorial approach. *Comm. Math. Phys.* **1999**, *204*, 89–114. [CrossRef]
- Milnor, J.; Thurston, W. On iterated maps of the interval. In *Dynamical Systems*; Springer: Berlin/Heidelberg, Germany, 1988; Volume 1342, pp. 465–563.
- Rocha, J.L.; Ramos, J.S. Weighted kneading theory of one-dimensional maps with a hole. *Int. J. Math. Math. Sci.* **2004**, *2004*, 2019–2038. [CrossRef]
- Alves, J.F.; Ramos, J.S. Kneading theory for tree maps. *Ergod. Theory Dyn. Syst.* **2004**, *24*, 957–985. [CrossRef]
- Hao, B.; Zheng, W. *Applied Symbolic Dynamics and Chaos*; World Scientific: Singapore, 2003.
- Daw, C.S.; Finney, C.E.A.; Tracy, E.R. A review of symbolic analysis of experimental data. *Rev. Sci. Instrum.* **2003**, *74*, 915–930. [CrossRef]
- Matilla-Garcia, M.; Ruiz, M. A non-parametric independence test. *J. Econom.* **2008**, *144*, 39–155. [CrossRef]
- Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Champaign, IL, USA, 1949.
- Yibin, L.; Yong, S.; Caihong, L. Selection of parameters for Phase Space reconstruction of chaotic time series. In Proceedings of the 2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), Changsha, China, 23–26 September 2010; pp. 30–33.
- Palshikar, G. Simple algorithms for peak detection in time-series. In Proceedings of the first IIMA International Conference on Advanced Data Analysis, Business Analytics and Intelligence (ICADABAI 2009), Ahmedabad, India, 6–7 June 2009.
- Lorenz, E.N. Deterministic nonperiodic flow. *Atoms J. Sci.* **1963**, *20*, 130. [CrossRef]
- Rossler, O.E. An equation for continuous chaos. *Phys. Lett. A* **1976**, *57*, 397. [CrossRef]
- Zeni, R.A.; Gallas, J.A.C. Lyapunov Exponents for a Duffing Oscillator. *Physica D* **1995**, *89*, 79–85. [CrossRef]
- Mackey, M.; Glass, L. Oscillation and chaos in physiological control systems. *Science* **1977**, *197*, 287–289. [CrossRef]
- Lu, J.; Chen, G. A new chaotic attractor coine. *Int. J. Bifurc. Chaos* **2002**, *12*, 659–661. [CrossRef]

30. Kaplan, J.; Yorke, J. Chaotic Behavior of Multidimensional Difference Equations. In *Functional Differential Equations and Approximation of Fixed Points*; Peitgen, H.-O., Walther, H.-O., Eds.; Springer: Berlin/Heidelberg, Germany, 1979.
31. Grassberger, P.; Procaccia, I. Characterization of strange attractors. *Phys. Rev. Lett.* **1983**, *50*, 345. [[CrossRef](#)]
32. Ahmed, I.; Mu, C.; Zhang, F. A New Chaotic Attractor with Quadratic Exponential nonlinear Term from Chen's Attractor. *Int. J. Anal. Appl.* **2014**, *5*, 27–32.
33. Maino, D.G.; Uzal, L.; Granito, P. Predicción de Sistemas Dinámicos con Redes Neuronales Profundas. In Proceedings of the 15th Argentine Symposium on Artificial Intelligence, Panamá, Argentina, 25–31 July 2014.
34. Rosenstein, M.T.; Collins, J.J.; De Luca, C.J. A practical method for calculating largest Lyapunov exponents from small data sets. *Phys. D Nonlinear Phenom.* **1992**, *65*, 117–134. [[CrossRef](#)]
35. Tang, L.; Liang, J. C-C Method to Phase Space Reconstruction Based on Multivariate Time Series. In Proceedings of the 2011 2nd International Conference on Intelligent Control and Information Processing, Harbin, China, 25–28 July 2011.
36. Sprott, J.C.; Rowlands, G. Improved correlation dimension calculation. *Int. J. Bifurc. Chaos* **2001**, *11*, 1865–1880. [[CrossRef](#)]
37. Haniyas, M.P.; Tombras, G.S.; Nistazakis, H.E. *Optoelectronic Chaotic Circuits*; IntechOpen: London, UK, 2011; pp. 631–651.
38. Cai, W.; Qin, Y.Q.; Yang, B.-R. Selection of Delay Time Window and Delay Time in Phase Space Reconstruction. In Proceedings of the 2007 International Conference on Computational Intelligence and Security (CIS 2007), Harbin, China, 15–19 December 2007; pp. 526–530.
39. Cai, W.; Qin, Y.Q.; Yang, B.-R. Determination of Phase-Space Reconstruction Parameters of chaotic time series. *Kybernetika* **2008**, *44*, 557–570.
40. Xiu, C.B.; Liu, X.D.; Zhang, Y.H. Selection of embedding dimension and delay time in the phase space reconstruction. *Trans. Beijing Inst. Technol.* **2003**, *23*, 219–224.
41. Andrzejak, R.G.; Lehnertz, K.; Rieke, C.; Mormann, F.; David, P.; Elger, C.E. Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E* **2001**, *64*, 061907. [[CrossRef](#)]
42. Babloyantz, A.; Destexhe, A. Low-dimensional chaos in an instance of epilepsy. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 3513–3517. [[CrossRef](#)]
43. Pritchard, W.S.; Duke, D.W. Measuring chaos in the brain: A tutorial review of EEG dimension estimation. *Brain Cogn.* **1995**, *27*, 353–397. [[CrossRef](#)] [[PubMed](#)]
44. Elger, C.E.; Widman, G.; Andrzejak, R.; Arnhold, J.; David, P.; Lehnertz, K. Nonlinear EEG analysis and its potential role in epileptology. *Epilepsia* **2000**, *41*, 34–38. [[CrossRef](#)]
45. Theiler, J.; Eubank, S.; Longtin, A.; Galdrikian, B. Testing for Nonlinearity in Time Series: The Method of Surrogate Data. *Physica D* **1992**, *58*, 77–94. [[CrossRef](#)]
46. BenSaida, A. A practical test for noisy chaotic dynamics. *SoftwareX* **2015**, *3*, 1–5. [[CrossRef](#)]
47. Theiler, J. Some comments on the correlations dimensions of $1/f$ a noise. *Phys. Lett. A* **1991**, *155*, 480–493. [[CrossRef](#)]
48. Casdagli, M.C.; Iasemidis, L.D.; Savit, R.S.; Gilmore, R.L.; Roper, S.; Sackellares, J.C. Nonlinearity in invasive EEG recordings from patients with temporal lobe epilepsy. *Electroencephalogr. Clin. Neurophysiol.* **1997**, *102*, 98. [[CrossRef](#)]
49. Pijn, J.P.M.; Velis, D.N.; van der Heyden, M.J.; DeGoedde, J.; van Veelen, C.; Lopes da Silva, F.H. Nonlinear dynamics of epileptic seizures on basis of intracranial EEG recordings. *Brain Topogr.* **1997**, *9*, 249. [[CrossRef](#)]
50. Kannathal, N.; Acharya, U.R.; Lim, C.M.; Sadasivan, P.K. Characterization of EEG—A comparative study. *Comput. Methods Programs Biomed.* **2005**, *80*, 17–23. [[CrossRef](#)]
51. Yao, W.; Wang, J. Multi-scale symbolic transfer entropy analysis of EEG. *Phys. A Stat. Mech. Its Appl.* **2017**, *484*, 276–281. [[CrossRef](#)]
52. Hussain, L.; Aziz, W.; Alowibdi, J.S.; Habib, N.; Rafique, M.; Saeed, S.; Kazmi, S.Z.H. Symbolic time series analysis of electroencephalographic (EEG) epileptic seizure and brain dynamics with eye-open and eye-closed subjects during resting states. *J. Physiol. Anthropol.* **2017**, *36*, 21. [[CrossRef](#)]