

Detecting determinism with improved sensitivity in time series: Rank-based nonlinear predictability score

Daniel Naro,¹ Christian Rummel,² Kaspar Schindler,³ and Ralph G. Andrzejak^{1,*}

¹*Universitat Pompeu Fabra, Department of Information and Communication Technologies, Barcelona, Spain*

²*Support Center for Advanced Neuroimaging, University Institute for Diagnostic and Interventional Neuroradiology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland*

³*qEEG group, Department of Neurology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland*

(Received 17 July 2014; published 12 September 2014)

The rank-based nonlinear predictability score was recently introduced as a test for determinism in point processes. We here adapt this measure to time series sampled from time-continuous flows. We use noisy Lorenz signals to compare this approach against a classical amplitude-based nonlinear prediction error. Both measures show an almost identical robustness against Gaussian white noise. In contrast, when the amplitude distribution of the noise has a narrower central peak and heavier tails than the normal distribution, the rank-based nonlinear predictability score outperforms the amplitude-based nonlinear prediction error. For this type of noise, the nonlinear predictability score has a higher sensitivity for deterministic structure in noisy signals. It also yields a higher statistical power in a surrogate test of the null hypothesis of linear stochastic correlated signals. We show the high relevance of this improved performance in an application to electroencephalographic (EEG) recordings from epilepsy patients. Here the nonlinear predictability score again appears of higher sensitivity to nonrandomness. Importantly, it yields an improved contrast between signals recorded from brain areas where the first ictal EEG signal changes were detected (focal EEG signals) versus signals recorded from brain areas that were not involved at seizure onset (nonfocal EEG signals).

DOI: [10.1103/PhysRevE.90.032913](https://doi.org/10.1103/PhysRevE.90.032913)

PACS number(s): 05.45.Tp, 87.19.le, 87.19.xm

I. INTRODUCTION

The ultimate goal of many scientific disciplines is to predict the temporal evolution of some dynamics based on observations of the dynamics' states in the recent past and present. Sometimes, however, one does not aim at an actual prediction but rather at quantifying the degree of a dynamics' predictability. The feature of predictability can be exploited in applications as distinct as, for example, damage assessment in structural engineering [1], channel equalization and timing recovery in telecommunications [2], modeling the dynamical response of granular material to large strain in critical state soil dynamics [3], the automatic identification of cover songs in music information retrieval [4], the understanding of El Niño and La Niña networks in atmospheric science [5], the assessment of heart-rate and arterial pressure regulation in cardiovascular physiology [6], or the characterization of electroencephalographic (EEG) recordings during photic stimulation in clinical neurophysiology [7]. As we detail below, the assessment of predictability from EEG signals can furthermore help to localize the epileptogenic zone in the diagnostics of epilepsy patients.

The framework of nonlinear time series analysis comprises different approaches to assess predictability, including local methods in reconstructed state spaces (e.g., Refs. [8–17]). These are based on the following common basic idea. For smooth deterministic dynamics, similar instantaneous states remain similar in their immediate future. To evaluate this criterion for determinism, these measures quantify the goodness with which the future evolution of reference states can be predicted from the future evolutions of states which are similar to the reference states. The similarity used to select the predicting

states is quantified by spatial proximity in reconstructed state spaces. Accordingly, these states are called nearest neighbors. The future evolution of the reference state and its nearest neighbors can be compared, for example, directly on the signal amplitudes [8–11,15,16] or on directional vectors in the reconstructed state space [13]. Different statistics, such as a normalized root-mean-square error [8,15,16], mean-square error [9,16], median-square error [13], decimated independent individual errors [11], or the correlation coefficient between predicted and actual states [10], can be used to quantify the goodness of these predictions. Further, local methods in state space quantify predictability without defining an explicit predictor, e.g., via the average alignment of directional vectors [12], continuity between exceptionally close neighbors and their images [14], or the mean variances of the nearest-neighbor images [17].

A representative of local methods in state spaces is the locally constant nonlinear prediction error (see Refs. [8,9,15] and references therein). Here the time indices of k nearest neighbors are determined and incremented by a prediction horizon h . The mean of the signal amplitude sampled at these incremented times is used to predict the signal amplitude a prediction horizon after the time of the reference state. The root-mean-square error is used to quantify the goodness of the prediction. We here refer to this classical approach as amplitude-based nonlinear prediction error \mathcal{E} . We use \mathcal{E} as benchmark for the rank-based nonlinear predictability score \mathcal{S} . The principle of \mathcal{S} was recently introduced to detect determinism in point processes [18]. We here adapt \mathcal{S} to be applicable to time series. In contrast to \mathcal{E} , the measure \mathcal{S} is not based on signal amplitudes but rather on ranks in sorted lists of amplitude differences. The normalization of the nonlinear predictability score \mathcal{S} is based on the expected values for completely predictable and completely unpredictable signals.

*ralph.andrzejak@upf.edu

It is well established that predictability is a necessary yet not sufficient criterion for determinism. Predictability in a time series can mean that the dynamics is deterministic, noisy deterministic, or simply a stochastic process with some memory. Accordingly, tests for determinism should be sensitive to predictability in time series. They can, however, never be fully specific for determinism since predictability can also be caused by temporal linear correlation. This limited specificity, which is no peculiarity of tests for determinism, but rather common to many nonlinear signal analysis approaches, can be addressed using the concept of surrogates [19,20]. This concept allows one to test whether the results of some nonlinear predictability measure are consistent with the null hypothesis that the signal was measured from a linear stochastic process whose memory results in temporal correlations. Only in combination with surrogate null hypothesis testing can tests for predictability be used as tests for determinism, as stressed, for example, already in Refs. [11,12].

The aim of this paper is to establish the rank-based nonlinear predictability score \mathcal{S} as a sensitive measure for nonlinear deterministic structure in time series. As a benchmark we use the classical amplitude-based nonlinear prediction error \mathcal{E} . Both measures belong to the family of local methods in state spaces. Therefore, we start by describing what is common to both measures. That is the selection of k nearest neighbors in reconstructed state spaces (Sec. II A) and the reasoning of which constraints are applied in this selection (Sec. II B). In Sec. II C 1 we review the definition of the classical amplitude-based nonlinear prediction error, and in Sec. II C 2 we adapt the rank-based nonlinear predictability score to be applicable to time series. The sensitivity of both measures is at first compared under controlled conditions. For this purpose we use signals from the Lorenz dynamics (Sec. II D 1) which we superimpose with noise. Importantly, apart from Gaussian noise, we use noise with non-normally distributed amplitudes (Sec. II D 2). In Sec. II E we briefly specify the type of surrogates we use and the underlying null hypothesis. After the validation of our approach on signals from models, we analyze real-world signals. For this purpose, we use the *Bern-Barcelona EEG database* (Sec. II F). We close our Methods section by reasoning the choice of parameters (Sec. II G). The Results section is divided into the presentation of findings for the noisy Lorenz signals (Sec. III A) and the EEG data (Sec. III B). The parameter dependence of our results is summarized in Sec. III C.

II. METHODS

A. Selection of predictor indices

The first steps of analysis are identical for the amplitude-based nonlinear prediction error \mathcal{E} and the rank-based nonlinear prediction score \mathcal{S} . We reconstruct the state space using delay coordinate embeddings and sequentially use all delay vectors as reference states. For each reference state we determine the time indices of its nearest neighbors. The time index of the reference state and the ones of its nearest neighbors are then incremented by a prediction horizon h . We use the signal sampled at the incremented nearest-neighbor indices to make a prediction about the signal at the incremented reference

state index. The two approaches differ only in the form of the prediction and in the way they quantify the goodness of the prediction. As a consequence of the common steps of analysis, both measures have the same parameters. These are the embedding dimension m and time delay τ used for the state space reconstruction, the prediction horizon h , the number of nearest neighbors k , and a decorrelation window length w to exclude temporally close states from the nearest neighbors.

In practice, suppose that a time series x_i with $i = 1, \dots, N$ was measured from some dynamics. To reconstruct the dynamics from the time series by means of delay coordinates, we form embedding vectors as follows [21]:

$$\mathbf{x}_i = (x_i, x_{i-\tau}, \dots, x_{i-(m-1)\tau}) \quad (1)$$

for $i = \eta + 1, \dots, N$, with the embedding window $\eta = (m - 1)\tau$. Each vector \mathbf{x}_i represents the state of the dynamics at the time corresponding to index i . In the next step, we calculate Euclidean interstate distances across all pairings $(i, j = \eta + 1, \dots, N)$ of embedding vectors

$$v_{i,j} = \sqrt{\sum_{d=1}^m (\mathbf{x}_{i,d} - \mathbf{x}_{j,d})^2}, \quad (2)$$

where d indexes the component of the delay vector. For each reference state \mathbf{x}_{i_0} ($i_0 = \eta + 1, \dots, N - h$) these distances are used to determine the indices of its k spatially nearest neighbors: $\{j_{0,r}\}_{(r=1,\dots,k)}$. These are the j indices of the k smallest entries in the set $\{v_{i_0,j}\}_{(j=\eta+1,\dots,N-h;|i_0-j|>w)}$. Importantly, thereby we exclude points that are closer than w in time. This is necessary to ensure that nearest neighbors are taken from neighboring state space trajectory segments rather than from the same segment [22].

As input for the subsequent steps of analysis we use indices of the reference point and its nearest neighbors incremented by the prediction horizon h , i.e., $i_0 + h$ and $j_{0,r} + h$. That explains why we excluded states that are closer to the end of the signal than the prediction horizon h in the ranges specified above. If we would not exclude this range, we would have predictions beyond the end of the signal.

We can now formulate the necessary but not sufficient criterion for determinism given above in terms of our indices. *For smooth deterministic dynamics, similar states at present times identified by i_0 and $j_{0,r}$ should evolve to similar states at future times identified by $i_0 + h$ and $j_{0,r} + h$.*

B. Testing for predictability versus making predictions

In the selection of nearest neighbors we do not impose $j_0 < i_0$. In consequence, the nearest neighbors in general include states from the future of the reference state. Neither do we divide the data into a training and testing set. This is in contrast to some of the techniques introduced and discussed in Refs. [8–10,15]. The reasoning leads us to an important conceptual distinction. We do not aim to make a concrete prediction that the dynamics will be in a certain state at a certain time. Neither is the performance of our approach to be assessed by the goodness of the prediction. Importantly, we do not vary any model parameters, such as the coefficients in a locally linear model, to optimize the prediction. Furthermore, none of the

analysis parameters (i.e., embedding dimension m , time delay τ , prediction horizon h , number of nearest neighbors k , and decorrelation window length w) is optimized. The role of these analysis parameters is further discussed in Secs. II G and III C.

Our aim is to test for predictability. Accordingly, we use as nearest neighbors states that are similar to the reference state, regardless of when the system attained them. We use them regardless of whether they occurred before or after i_0 . The only requirement is that they are temporally separate from i_0 by at least the decorrelation window w . As a test for predictability we then use the measures \mathcal{E} and \mathcal{S} to quantify the degree to which these similar instantaneous states undergo a similar evolution in their immediate future. We assess the performance of these measures in different ways. We determine, for example, the minimal signal-to-noise ratio at which a signal can be distinguished from pure white noise. Another criterion is the power they yield as test statistics in a surrogate null hypothesis test, described in detail in Sec. III E. In particular, we always compare our results against a baseline, for example, pure white noise or the surrogates. This inclusion of a baseline further justifies not imposing $j_0 < i_0$ [15]. If required, the approaches presented below could readily be adapted to include this constraint.

C. Testing for predictability

1. Amplitude-based nonlinear prediction error \mathcal{E}

For the classical nonlinear prediction error (see Refs. [8,9,15] and references therein), the predictability is evaluated on the signal amplitudes. In particular, one considers the amplitudes at the indices of the reference point and its nearest neighbors incremented by the prediction horizon h , i.e., at $i_0 + h$ and $j_{0,r} + h$. We use the average $\langle x_{j_{0,r}+h} \rangle$ taken across $r = 1, \dots, k$ to predict x_{i_0+h} . The nonlinear prediction error \mathcal{E} is then directly defined by the root-mean-square error of this prediction normalized by the standard deviation σ_x of the signal,

$$\mathcal{E} = \frac{\sqrt{\frac{1}{N-\eta-h} \sum_{i_0=\eta}^{N-h} (x_{i_0+h} - \langle x_{j_{0,r}+h} \rangle)^2}}{\sigma_x}. \quad (3)$$

Recalling the definition of the standard deviation, it can be read in this context as the root-mean-square error when the mean is used for the prediction. For predictable signals we get positive \mathcal{E} values close to zero. In contrast, for uncorrelated noise the expected value of \mathcal{E} approaches 1 from above upon increasing k . For the value used here, $k = 5$, we derived numerically that the expected value of \mathcal{E} is approximately 1.096.

2. Rank-based nonlinear prediction score \mathcal{S}

For the rank-based nonlinear prediction score \mathcal{S} , we do not directly evaluate signal amplitudes but rather ranks in sorted lists of amplitude differences. Similarly to Refs. [12,14,17], predictability is quantified without defining an explicit predictor. For this purpose, we at first determine for $i, j = \eta, \dots, N$

$$u_{i,j} = |x_i - x_j|, \quad (4)$$

which is equivalent to Eq. (2) with $m = 1$. We denote by g_{i_0, j_0} the ranks which u_{i_0, j_0} have in a sorted list of all amplitude differences included in the set $\{u_{i_0, j}\}_{j=\eta+1, \dots, N; |i_0-j|>w}$. The

number of differences in this set is denoted by M_{i_0} . For $w < i_0 < N - w + 1$ we have $M_{i_0} = N - 2w - 1$. Below and above this range M_{i_0} increases linearly and reaches $M_{i_0} = N - w - 1$ at $i_0 = 1$ and $i_0 = N$.

To test the above given criterion for predicability we again compare the future evolution of the nearest neighbors with the one of the reference state. For this purpose, we determine the mean rank of the amplitude differences between x_{i_0+h} and the k different $x_{j_{0,r}+h}$,

$$R_{i_0} = \frac{1}{k} \sum_{r=1}^k g_{i_0+h, j_{0,r}+h}. \quad (5)$$

For complete predictability, this is the mean of the k lowest ranks, and we get

$$R_{i_0} = \frac{k+1}{2}. \quad (6)$$

We denote this lower boundary, which is independent of i_0 , by

$$R^L = \frac{k+1}{2}. \quad (7)$$

In contrast, for no predictability, nearest neighbors of i_0 have no predictive value h time steps later. Accordingly, $g_{i_0+h, j_{0,r}+h}$ are just k random samples from a uniform distribution on the integers $1, \dots, M_{i_0}$. In consequence, for no predictability, R_{i_0} has an expected value of

$$R_{i_0}^U = \frac{M_{i_0} + 1}{2}. \quad (8)$$

Since $k \ll M_{i_0}$, we here neglect that the k samples are taken without replacement. Using a general normalization principle [23–26], we define the rank-based prediction score as follows [18]:

$$\mathcal{S} = \frac{1}{N - \eta - h} \sum_{i_0=1+\eta}^{N-h} \frac{R_{i_0}^U - R_{i_0}}{R_{i_0}^U - R^L}. \quad (9)$$

For complete predictability we get $R_{i_0} = R^L$, and \mathcal{S} takes its maximal value of 1. In contrast, \mathcal{S} is normally distributed with an expected value of zero for no predictability, since here the expected value of R_{i_0} is $R_{i_0}^U$.

D. Signals

1. Lorenz signals

As exemplary model system we use Lorenz dynamics with standard parameters as follows: $\dot{x}(t) = 10(y(t) - x(t))$, $\dot{y}(t) = 28x(t) - y(t) - x(t)z(t)$, $\dot{z}(t) = x(t)y(t) - \frac{8}{3}z(t)$. We integrated this set of differential equations using a fourth-order Runge-Kutta algorithm with step size of 0.005 time units. The integrations were started at random initial conditions and preiterations were used to let transients die out. As raw signal we used $x_n = x(t_n)$ for $t_n = n\Delta t$ with $n = 1, \dots, N = 2560$ and $\Delta t = 0.03$ time units.

2. Noise

The raw Lorenz signals x_n were distorted with white additive measurement noise. Apart from Gaussian noise, we used noise with distributions having narrower central peaks

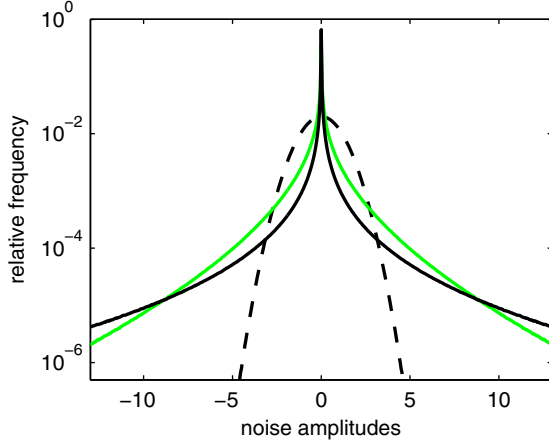


FIG. 1. (Color online) Relative frequencies of the noise amplitudes $\xi_{l,n}$ for different l determined in bins of 0.05. We use a logarithmic scale for the ordinate to better resolve characteristics of the distribution tails. The abscissa is truncated, and nonzero relative frequencies are obtained for amplitudes towards $\pm\infty$. $l = 1$. Black dashed line: $l = 1$; Green (light gray) solid line: $l = 3$; Black solid line: $l = 5$.

and heavier tails than the normal distribution. To generate the noise, we started with zero-mean unit-variance Gaussian white noise ζ_n . The amplitude distribution was then altered by applying $\xi_{l,n} = (\zeta_n)^l$ for $l = 1, 3, 5$. To illustrate the effect of l , we display the amplitude distributions in Fig. 1. All distributions are normalized to unit variance. Evidently, for $l = 1$ the original Gaussian noise is maintained. With increasing l , the values at the center of the peak are concentrated closer to zero. In contrast, values in the tails are driven away from the center towards even higher values. In consequence, for $l = 3$ and even more so for $l = 5$, the noise consists of many small values and some very high values. We vary the signal-to-noise ratio,

$$q = \sigma_{x_n}^2 / \sigma_{\xi_{l,n}}^2, \quad (10)$$

from 10^{-6} to 10^6 sampled evenly on a logarithmic scale. We furthermore include the cases $q = 0$ (noise-only) and $q \rightarrow \infty$ (noise-free). For each signal-to-noise ratio and all l values, we generate 250 independent realizations of the noisy signal.

E. Surrogates

Suppose we analyzed an experimental signal and obtained some low value of the nonlinear prediction error \mathcal{E} and some high value of the nonlinear predictability score \mathcal{S} . As stated in the Introduction, this result by itself does not provide sufficient evidence that the signal was measured from a deterministic dynamics. Likewise, the dynamics could be linear stochastic and the predictability which \mathcal{E} and \mathcal{S} detect in the signal could reflect its autocorrelation caused by the memory of the process. Even for memoryless linear stochastic processes, autocorrelation and thereby predictability can be caused simply by low-pass filtering the signal. Depending on the parameters, in particular m , τ , and h , this autocorrelation can lead to low values of \mathcal{E} and high values of \mathcal{S} . To test the null hypothesis that the signal was measured from a linear stochastic correlated process, we can use surrogate signals.

These surrogate signals are generated by randomizing the original signal. In our case, this randomization is constrained such that the surrogates have the same autocorrelation and amplitude distribution as the original signal [27]. Any potential nonlinear deterministic structure or nonstationary features of the original signal are destroyed by the randomization. The complete null hypothesis tested by this particular type of surrogates is: ‘The dynamics is a stationary linear stochastic correlated Gaussian process. The measurement function by which the signal was derived from the dynamics is invertible but potentially nonlinear. The autocorrelation, mean, and variance of the underlying Gaussian process are such that the measurement results in the autocorrelation, and amplitude distribution of the observed time series.’ To test this null hypothesis we generate a set of surrogates from independent randomizations of the signal and calculate \mathcal{E} and \mathcal{S} from them. In this way, we can estimate the mean values and variances of \mathcal{E} and \mathcal{S} which we would expect if this null hypothesis was true and given the signal’s amplitude distribution and autocorrelation. If the result for the original signal deviates significantly from the distribution of results obtained from the surrogates, the null hypothesis is rejected.

Evidently, the power of this statistical surrogate null hypothesis test depends on the sensitivity of our measures \mathcal{E} and \mathcal{S} . The better the sensitivity for deterministic structure, the higher the probability that the null hypothesis is correctly rejected if it is false. If the signal was measured from a deterministic dynamics but is strongly contaminated by noise, then the null hypothesis is false. It will, however, depend on the measure’s sensitivity if it is capable of detecting the potentially subtle differences between the noisy signal and the surrogates. Accordingly, we can assess the sensitivity of \mathcal{E} and \mathcal{S} by quantifying the statistical power of a surrogate null hypothesis test based on these measures in dependence on the noise amplitude.

F. EEG signals

We analyze the Bern-Barcelona EEG database which is available in Ref. [28] and which we published along with our previous work [29]. This anonymized database comprises intracranial EEG recordings from five epilepsy patients, which were performed prior to and independently from the study presented in Ref. [29] and the present study as part of the epilepsy diagnostics in these patients. We refer to Sec. II A of Ref. [29] for details on this diagnostics and recording techniques.

The database comprises 3750 focal signal pairs and 3750 nonfocal signal pairs. Focal signals were recorded from brain areas where the first ictal EEG signal changes were detected as judged by expert visual inspection by at least two neurologists who are also board-certified electroencephalographers. Nonfocal signals were recorded from brain areas that were not involved at seizure onset. Importantly, also this delineation of brain areas where the first ictal EEG signal changes were detected was done prior to and independently from our study as part of the epilepsy diagnostics. Both focal and nonfocal signals include only seizure-free EEG activity. Individual signal pairs were recorded simultaneously at neighboring contacts of intracranial multicontact electrodes. The entirety

of the signal pairs were selected at random from continuous recordings of the seizure-free interval of the patients. We refer to Sec. II B of Ref. [29] for details on this selection. Since, in contrast to Ref. [29], our analysis is univariate, we only take the first signals from the signal pairs, referred to as x signals in Ref. [29].

The signals of the Bern-Barcelona EEG database have a length of 20 s sampled at 512 Hz. We followed Ref. [29] and low-pass filtered and down-sampled the signals prior to the analysis. For this purpose, we used an eighth-order Butterworth filter with a cut-off frequency of 40 Hz and subsequently down-sampled the signals by a factor of 4, resulting in a signal length of $N = 2560$ and a sampling time of 7.8 ms.

The EEG recordings analyzed here were previously studied in Ref. [29]. We provided these recordings, the source code, and detailed results of this previous study in Ref. [28]. This includes the source code to calculate \mathcal{E} and to generate surrogates. Results obtained for small subsets of the EEG signals were published by others in Refs. [30,31]. We now extend these resources by source code that allows for the simultaneous calculation of \mathcal{S} and \mathcal{E} , as well as by the detailed results obtained in the present study for the EEG recordings.

G. Choice of parameters and connection to previous work

In the first analysis of the Bern-Barcelona EEG database, we analyzed these signals with regard to their nonrandomness, nonlinear dependence, and nonstationarity [29]. The test for nonrandomness was based on the nonlinear prediction error \mathcal{E} as a test statistics in combination with the surrogates also used here. Accordingly, \mathcal{E} results obtained in our previous

study [29] for the EEG can be used as a benchmark for \mathcal{S} in our present study. In the previous study detailed results were shown for the following parameters: embedding dimension $m = 8$, time delay $\tau = 4$ sampling times, $k = 5$ nearest neighbors, prediction horizon $h = 4$ sampling times, and decorrelation window $w = 19$ sampling times. We therefore use the same parameter setting here. We use it for the EEG signals and also for the Lorenz signals. These are reasonable parameter values for both types of signals but certainly not the optimal ones. However, our aim is not to find the optimal parameters for the individual signal types and measures. Rather, we want to compare the measures \mathcal{E} and \mathcal{S} and avoid any in-sample optimization of the parameters. That is the reason why we use the values already used in Ref. [29] and discuss only briefly the influence of these parameters in Sec. III C. With the same reasoning, the preprocessing steps for the EEG signals are as in Ref. [29]. This includes the use of $N = 2560$ samples for the analysis. Accordingly, we also use the same number of samples for the noisy Lorenz signals.

III. RESULTS

A. Noisy Lorenz signals

Figure 2 shows how the noise amplitude distribution parameter l influences the noise robustness of the nonlinear prediction error \mathcal{E} and the nonlinear predictability score \mathcal{S} . At the noise-only case ($q = 0$) and very small q values, both measures show results expected for unpredictable signals; the mean values $\langle \mathcal{E} \rangle$ and $\langle \mathcal{S} \rangle$ obtained across the 250 independent signal realizations are statistically consistent with 1.096 and 0, respectively. On increasing q , $\langle \mathcal{E} \rangle$ decreases and $\langle \mathcal{S} \rangle$ increases, reflecting the increased predictability. This predictability is

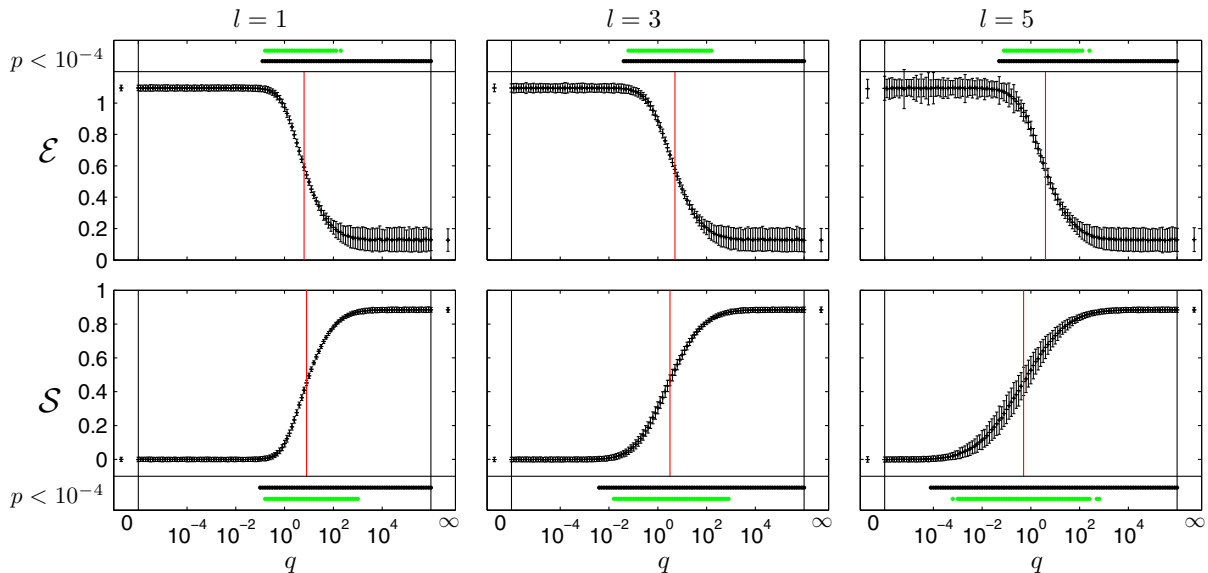


FIG. 2. (Color online) Results for the noisy Lorenz signals in dependence on the signal-to-noise ratio q and amplitude distribution parameter l . Error bars show the distribution means, $\langle \mathcal{E} \rangle$ and $\langle \mathcal{S} \rangle$, \pm two standard deviations obtained for the 250 independent realizations of the signals. All simulations are independent across different q and l values. The leftmost and rightmost distributions correspond to the noise-only ($q = 0$) and noise-free ($q \rightarrow \infty$) case, respectively. The vertical red lines mark the minimal q value for which the distributions' mean values pass the midpoint between the mean values obtained for $q = 0$ and $q \rightarrow \infty$. Black dots indicate those q for which the distributions significantly differ from the ones obtained for $q = 0$ (Wilcoxon rank-sum test rejected with $p < 10^{-4}$). Green (light gray) dots indicate those q for which the distributions significantly differ from the ones obtained for next lower q value (Wilcoxon rank-sum test rejected with $p < 10^{-4}$).

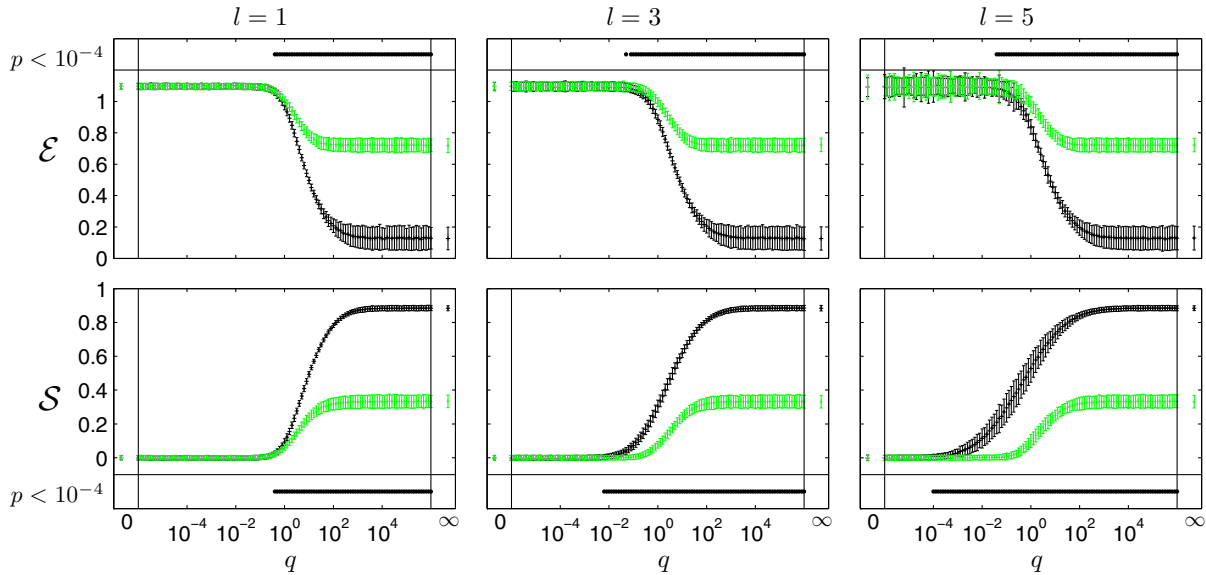


FIG. 3. (Color online) Results for the noisy Lorenz signals and corresponding surrogates in dependence on the signal-noise-ratio q and the amplitude distribution parameter l . Black: Distributions of \mathcal{E} and \mathcal{S} obtained for 250 independent realizations of the noisy Lorenz signals (already shown in Fig. 2). Green (light gray): distributions of \mathcal{E} and \mathcal{S} obtained for 250 surrogates. Here one surrogate was generated for each of the 250 independent noisy Lorenz signals. Error bars again correspond to the means ± 2 standard deviations. In contrast to Fig. 2, black dots now indicate those q for which the distribution obtained for the original signals significantly differ from the one obtained for the surrogates (Wilcoxon rank-sum test rejected with $p < 10^{-4}$).

caused by the increased proportion of the deterministic Lorenz signal. Towards high q and the noise-free case ($q \rightarrow \infty$), both measures level off. However, they do not reach values expected for completely predictable signals, which are $\langle \mathcal{E} \rangle = 0$ and $\langle \mathcal{S} \rangle = 1$. This is due to the finite length and complex, nonperiodic nature of the Lorenz signals.

Between the two limits $q = 0$ and $q \rightarrow \infty$, we can assess the sensitivity of our measures in different ways. At first we ask how much signal the measures need to detect it from the noise. We can answer this question by determining the minimal signal-to-noise ratio q at which the measures' distribution significantly differ from their distributions at the noise-only limit ($q = 0$) (see black dots in Fig. 2). Furthermore, we compare the ranges of the signal-to-noise ratio for which a gradual increase of the signal-to-noise ratio results in a significant decrease of \mathcal{E} and significant increase of \mathcal{S} [see green (light gray) dots in Fig. 2]. Finally, we determine the q value for which $\langle \mathcal{E} \rangle$ and $\langle \mathcal{S} \rangle$ reach the midpoint between the results obtained for $q = 0$ and $q \rightarrow \infty$ (see red vertical lines in Fig. 2).

We see that for normally distributed noise ($l = 1$) the sensitivity of \mathcal{E} and \mathcal{S} is almost identical. For increasing l , the performance of both measures increases. This improvement is more pronounced for \mathcal{S} , so for $l = 5$ its performance is clearly higher than the one of \mathcal{E} . At this l , \mathcal{S} detects the nonrandom structures at signal-to-noise ratios q that are more than two orders of magnitude smaller than the ones required by \mathcal{E} . Furthermore, the variability of \mathcal{E} across independent signal realizations is increased at high l and low q . These effects can be explained in the following way. For higher l , the proportion of very small noise values is higher (Fig. 1). Therefore, the overall impact of the noise is smaller. On the other hand, for higher l the noise has some very high

values. If these high amplitudes contribute in Eq. (3), either as erroneously predicted or erroneously predicting values, \mathcal{E} is increased substantially towards its limit for unpredictable signals. For \mathcal{S} , in contrast, the nuisance caused by these high values is limited since ranks are used instead of amplitudes. R_{i_0} in Eq. (5) has an upper bound for erroneous predictions, namely the mean of the k highest ranks. In consequence, the cost of a wrong prediction is limited for the rank-based nonlinear predictability score \mathcal{S} . Therefore, it outperforms the amplitude-based nonlinear prediction error \mathcal{E} for noise, having a narrower peak and heavier tails than the normal distribution.

We now contrast the results obtained for the original noisy Lorenz signals with those for surrogates (Fig. 3). In dependence on the signal-to-noise ratio q , the surrogate results have the same characteristics as the original results, an inverse S-shaped curve for \mathcal{E} and an S-shaped curve for \mathcal{S} . At the noise-only limit ($q = 0$) and very small q values, the results for the original noisy Lorenz signals are statistically indistinguishable from those for the surrogates. Both for the original signals and surrogates results coincide with those expected for unpredictable signals, $\langle \mathcal{E} \rangle$ and $\langle \mathcal{S} \rangle$ are statistically consistent with 1.096 and 0, respectively. On increasing q , $\langle \mathcal{E} \rangle$ decreases and $\langle \mathcal{S} \rangle$ increases, not only for the original signals but also for the surrogates. For the latter this reflects the increased predictability caused by the increased autocorrelation of the noisy signal. Recall that the surrogates have the same autocorrelation as the original signals and take into account that the autocorrelation strength increases for a decreased proportion of the white, i.e., temporally uncorrelated, noise.

Towards high q and the noise-free case ($q \rightarrow \infty$) the surrogate results also level off. However, they converge to mean values that are clearly distinct from those of the original

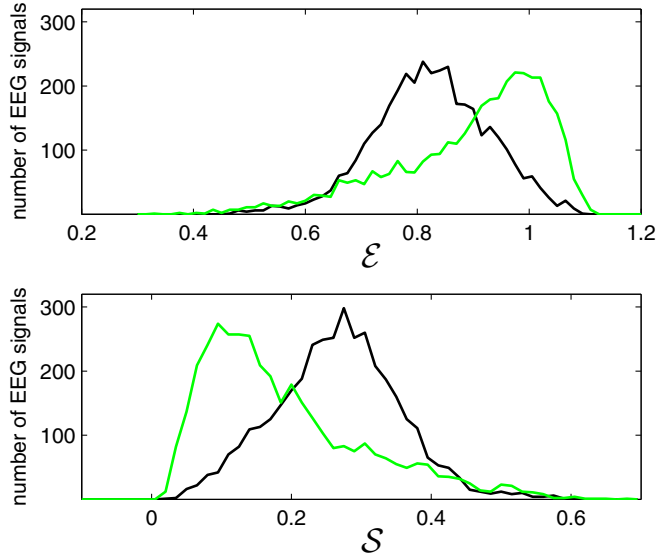


FIG. 4. (Color online) Comparison of \mathcal{E} and \mathcal{S} values for focal and nonfocal EEG signals. Histograms across the 3750 focal (black) and 3750 nonfocal [green (light gray)] EEG signals. The bin size is 0.015.

signals. This is due to the stochastic nature of the surrogates and the power of the measures to distinguish them from the original deterministic signals. The minimal q value at which the original distributions significantly differ from the corresponding surrogate distributions vary across l and for \mathcal{E} versus \mathcal{S} . In Fig. 3 we compare the distributions obtained for the 250 independent realizations of the original signals versus distributions of 250 corresponding individual surrogates. We observe that the power of the surrogate statistical test increases with increasing l . Again this improvement is substantially higher for the rank-based nonlinear predictability score \mathcal{S} as compared to the amplitude-based nonlinear prediction error \mathcal{E} . We get analogous results (not shown) when we determine the fraction of the 250 independent signal realizations for which the null hypothesis test was rejected based on a set of 19 surrogates generated for each individual signal realization. In consequence, the nonlinear predictability score \mathcal{S} again outperforms the nonlinear prediction error \mathcal{E} for high l . Here this is evidenced by a higher statistical power in a surrogate test of the null hypothesis of a correlated linear stochastic signal.

B. EEG signals

We now consider the distributions of \mathcal{E} and \mathcal{S} obtained across all 3750 focal and 3750 nonfocal EEG signals (Fig. 4). For both measures the focal and nonfocal distributions are clearly distinct. Recall that higher predictability is reflected in lower \mathcal{E} and higher \mathcal{S} values. Accordingly, for both \mathcal{E} and \mathcal{S} , the focal EEG signals are on average more predictable than the nonfocal EEG signals. To quantify this contrast between focal and nonfocal signals we determine the probability p that a random sample from the focal EEG signals appears more predictable than a random sample from the nonfocal EEG signals. This is done by evaluating the distributions of \mathcal{S} and \mathcal{E} numerically for both measures separately, and we get $p = 0.728$ for \mathcal{S} and $p = 0.696$ for \mathcal{E} . Similarly, this

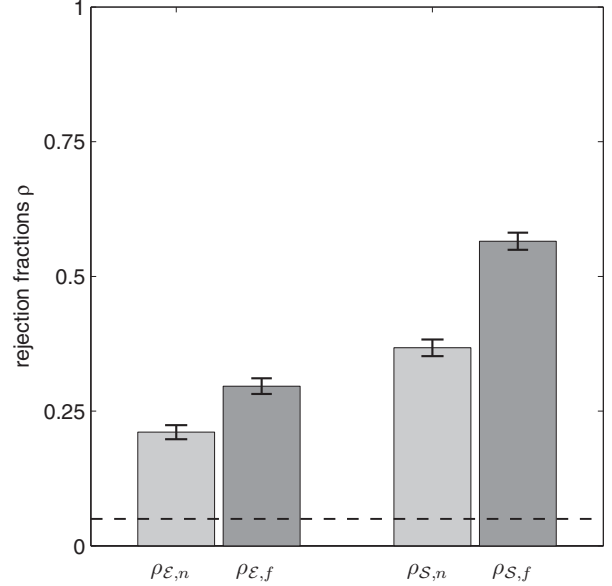


FIG. 5. Comparison of rejection fractions ρ of the surrogate null hypothesis test obtained by \mathcal{E} and \mathcal{S} for focal and nonfocal EEG signals. The percentages are with regard to the 3750 focal and 3750 nonfocal EEG signals. The dashed line indicates the chance level. The error bars illustrate 95% confidence intervals determined by $[\rho - 1.96\sqrt{\rho(1-\rho)/3750}, \rho + 1.96\sqrt{\rho(1-\rho)/3750}]$, see Ref. [29]. (Note that the \mathcal{E} results are a reproduction of the corresponding nonrandomness test results shown as part of Fig. 2(a) of Ref. [29]. There are slight differences, smaller than the confidence intervals, between the rejection percentages depicted here and those of Ref. [29]. They are caused by the random component of statistical type-I errors of the surrogate test across all signals.)

contrast can be assessed using the Z values of a Wilcoxon rank-sum test, which give $Z = 34.1$ for \mathcal{S} and $Z = 29.3$ for \mathcal{E} . If results for the focal and nonfocal signals were statistically indistinguishable, we would obtain approximately $p = 0.5$ and $Z = 0$. Hence, the nonlinear predictability score \mathcal{S} yields a higher discriminative power between the focal and nonfocal EEG signals than the nonlinear prediction error \mathcal{E} .

These differences in predictability can indicate that the focal EEG has a higher degree of nonlinear deterministic structure than the nonfocal EEG. However, as indicated above, an alternative or complementary explanation is that the focal EEG has stronger autocorrelation. We again address this ambiguity by a surrogate null hypothesis test. For each of the 7500 EEG signals we generated 19 surrogates. For \mathcal{E} , we rejected the null hypothesis if the result of the original EEG signal was lower, i.e., the original signal was more predictable, than the ones of all 19 surrogates. Analogously, for \mathcal{S} the original result had to be above the ones of all 19 surrogates to reject the null hypothesis. Since we use 19 surrogates, this test has a chance level of 0.05 for individual EEG signals.

We denote the rejection fractions for the focal and nonfocal EEG signals for \mathcal{E} by $\rho_{\mathcal{E},f}$ and $\rho_{\mathcal{E},n}$, respectively, and for \mathcal{S} by $\rho_{\mathcal{S},f}$ and $\rho_{\mathcal{S},n}$, respectively. Figure 5 shows that all four rejection fractions exceed the chance level clearly. Importantly, we get substantially more rejections for \mathcal{S} . In addition, for both measures more rejections are found for

the focal EEG signals as compared to the nonfocal signals. This contrast is higher for \mathcal{S} : $\rho_{\mathcal{S},f} - \rho_{\mathcal{S},n} = 0.198$ versus $\rho_{\mathcal{E},f} - \rho_{\mathcal{E},n} = 0.085$. Hence, for the nonlinear predictability score \mathcal{S} , the rejection percentage of the null hypothesis of linear correlated signals is overall higher than for the nonlinear prediction error \mathcal{E} . Furthermore, also with regard to these rejection percentages, the contrast between the focal and nonfocal signal is higher for \mathcal{S} than for \mathcal{E} .

C. Parameter dependence

As reasoned in Sec. II G, we here used the same values for the parameters as in the main analysis of Ref. [29]. We furthermore explored the parameter range used in Appendix A of Ref. [29] ($m = [2,4,6,8,10,12]$, $\tau = [1,2,4,8]$ sampling times, $k = [1,5,10]$, $h = [1,2,4,8,16,32]$ sampling times). For the sake of conciseness we do not show these results here. We report however, that all conclusions given here with regard to the advantage of \mathcal{S} over \mathcal{E} hold true across this range of parameters for both the noisy Lorenz signals as well as for the EEG signals. Furthermore, since we continue to provide all EEG signals and source codes in the public domain, the scientific community can readily compare \mathcal{E} and \mathcal{S} for whatever set of parameters it might consider optimal.

IV. DISCUSSION

We compared the sensitivity of the amplitude-based nonlinear prediction error \mathcal{E} and the rank-based nonlinear predictability score \mathcal{S} for deterministic structure in time series. In the first part, we worked under controlled conditions by analyzing signals from the Lorenz dynamics superimposed with different types of noise. We determined the minimal signal-to-noise ratio q at which the noisy signal can be distinguished from the noise-only case ($q \rightarrow \infty$) and furthermore compared how quickly the measures rise with further increasing q . For Gaussian white noise the sensitivity of both measures is almost identical. However, the nonlinear predictability score \mathcal{S} is more robust against noise with a narrower peak and heavier tail than the normal distribution. We then compared the statistical power which the measures yield when used as test statistics in a surrogate test of the null hypothesis that the signal was measured from a correlated linear stochastic process. Again the nonlinear predictability score \mathcal{S} outperforms the nonlinear prediction error \mathcal{E} for noise with narrow peaks and heavy tails.

One might object that this particular type of noise is somewhat *ad hoc* to stress the advantages of \mathcal{S} over \mathcal{E} . However, we provide an application to real-world signals for which our results are in close analogy to the ones obtained under controlled conditions for model signals. As real-world signals we used the Bern-Barcelona EEG database. This EEG database was composed and published before we carried out the analysis for the present study. Accordingly, any sampling bias towards favoring the advantage of \mathcal{S} over \mathcal{E} by the selection of individual EEG signals can be ruled out.

For the analysis of the EEG signals we again use a two-step evaluation procedure. We at first consider the measures' values as such without the use of surrogates. We then use the measures as test statistics in the surrogate null hypothesis test. As

for the first step, we observe that the focal EEG signals are on average deemed by both measures more predictable than the nonfocal signals. With regard to this feature, the nonlinear predictability score \mathcal{S} yields a higher discriminative power between the focal and nonfocal EEG signals. The main difference in the measures' sensitivity in application to the EEG signals, however, becomes evident in the surrogate test. The nonlinear predictability score \mathcal{S} has a substantially higher rejection percentage than the nonlinear prediction error \mathcal{E} . Importantly, for both measures more rejections are found for the focal EEG signals as compared to the nonfocal signals. Finally, also with regard to these rejection percentages, the contrast between the focal and nonfocal signal is higher for \mathcal{S} than for \mathcal{E} .

The fact that in a study of seizure-free EEG recordings we obtain more rejections of the surrogate null hypothesis for focal EEG signals as compared to nonfocal EEG signals is consistent with previous studies [17,25,26,29,32–39]. An important confounding variable in such results derived from surrogate tests is nonstationarity. Since stationarity is included in the surrogate null hypothesis, nonstationarity can cause a rejection of the surrogate test. To address this ambiguity, we included a stationarity test in our previous work [29]. A main result was that we found stronger indications of nonstationarity for nonfocal signals as compared to the focal signals. Furthermore, the contrast between the focal and nonfocal signals obtained from the \mathcal{E} -based randomness test of Ref. [29] was further enhanced when applied only to stationary EEG signals. The fact that the present study and our previous work are based on the same database of EEG signals allows us to link the \mathcal{S} results obtained here to those of the stationarity test included in our previous work. We do not include detailed results here but report that all conclusions drawn in Ref. [29] with regard to the effect of nonstationarity on the \mathcal{E} -based surrogate test hold for the \mathcal{S} -based surrogate test as well. In particular, we find that the contrast between the focal and nonfocal signals is enhanced when applied only to stationary EEG signals.

In summary, we have shown under controlled conditions that the rank-based nonlinear predictability score \mathcal{S} is of higher sensitivity to deterministic structure in noisy time series than the classical amplitude-based nonlinear prediction error \mathcal{E} . In an application to intracranial EEG recordings we have evidenced that this increased sensitivity for nonrandomness translates to an improved contrast between focal and nonfocal EEG signals. Such improvements can be decisive towards establishing nonlinear signal analysis measures in an automated quantitative evaluation of the EEG. In addition, from a clinical point of view it will be important to test if the method presented here allows one to differentiate focal from nonfocal EEG signals not only on the group-level but also for individual patients.

ACKNOWLEDGMENTS

R.G.A. acknowledges Grant No. FIS-2010-18204 of the Spanish Ministry of Education and Science and funding from the Volkswagen Foundation. K.S. and C.R. are grateful for support by the Swiss National Science Foundation (Projects No. SNF 320030-122010 and No. 33CM30-124089).

- [1] S. Torkamani, E. A. Butcher, M. D. Todd, and G. Park, *Mech. Syst. Signal Pr.* **29**, 457 (2012).
- [2] Z. Zhu and H. Leung, in *IEEE International Symposium on Circuits and Systems (ISCAS), 2012* (IEEE, Piscataway, NJ, 2012), pp. 624–627.
- [3] M. Small, D. M. Walker, A. Tordesillas, and C. K. Tse, *Chaos* **23**, 013113 (2013).
- [4] J. Serra, H. Kantz, X. Serra, and R. G. Andrzejak, *IEEE Trans. Audio Speech* **20**, 514 (2012).
- [5] A. A. Tsonis and K. L. Swanson, *Phys. Rev. Lett.* **100**, 228502 (2008).
- [6] G. Nollo, L. Faes, R. Antolini, and A. Porta, *Phil. Trans. R. Soc. A* **367**, 1423 (2009).
- [7] S. Erla, L. Faes, E. Tranquillini, D. Orrico, and G. Nollo, *Med. Eng. Phys.* **33**, 504 (2011).
- [8] J. D. Farmer and J. J. Sidorowich, *Phys. Rev. Lett.* **59**, 845 (1987).
- [9] M. Casdagli, *Phys. D* **35**, 335 (1989).
- [10] G. Sugihara and R. May, *Nature* **344**, 734 (1990).
- [11] M. B. Kennel and S. Isabelle, *Phys. Rev. A* **46**, 3111 (1992).
- [12] D. T. Kaplan and L. Glass, *Phys. Rev. Lett.* **68**, 427 (1992).
- [13] R. Wayland, D. Bromley, D. Pickett, and A. Passamante, *Phys. Rev. Lett.* **70**, 580 (1993).
- [14] D. T. Kaplan, *Physica D* **73**, 38 (1994).
- [15] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, UK, 1997).
- [16] D. Kugiumtzis, O. Lingjærde, and N. Christophersen, *Physica D* **112**, 344 (1998).
- [17] T. Gautama, D. P. Mandic, and M. M. Van Hulle, *Phys. Rev. E* **67**, 046204 (2003).
- [18] R. G. Andrzejak, F. Mormann, and T. Kreuz (unpublished).
- [19] D. Kaplan and R. Cohen, *Circ. Res.* **67**, 886 (1990).
- [20] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, *Physica D* **58**, 77 (1992).
- [21] F. Takens, in *Dynamical Systems and Turbulence*, Vol. 898 of Lecture Notes in Mathematics, edited by D. A. Rand and L.-S. Young (Springer-Verlag, Berlin, 1981), pp. 366–381.
- [22] J. Theiler, *Phys. Rev. A* **34**, 2427 (1986).
- [23] R. G. Andrzejak, A. Kraskov, H. Stögbauer, F. Mormann, and T. Kreuz, *Phys. Rev. E* **68**, 066202 (2003).
- [24] D. Chicharro and R. G. Andrzejak, *Phys. Rev. E* **80**, 026217 (2009).
- [25] C. Rummel, M. Müller, G. Baier, F. Amor, and K. Schindler, *J. Neurosci. Meth.* **191**, 94 (2010).
- [26] C. Rummel, E. Abela, M. Müller, M. Hauf, O. Scheidegger, R. Wiest, and K. Schindler, *Phys. Rev. E* **83**, 066215 (2011).
- [27] T. Schreiber and A. Schmitz, *Phys. Rev. Lett.* **77**, 635 (1996).
- [28] R. G. Andrzejak, K. Schindler, and C. Rummel, *Bern-Barcelona EEG Database*, <http://ntsa.upf.edu/downloads>.
- [29] R. G. Andrzejak, K. Schindler, and C. Rummel, *Phys. Rev. E* **86**, 046206 (2012).
- [30] G. Zhu, Y. Li, P. P. Wen, S. Wang, and M. Xi, in *AIP Conference Proceedings* (AIP, New York, 2013), Vol. 1559, p. 31.
- [31] R. Sharma, R. B. Pachori, and S. Gautam, in *International Conference on Medical Biometrics (ICMB), 2014* (IEEE, Piscataway, NJ, 2014), p. 135.
- [32] M. Casdagli, L. Iasemidis, J. Sackellares, S. Roper, R. Gilmore, and R. Savit, *Physica D* **99**, 381 (1996).
- [33] M. Casdagli, L. Iasemides, R. Savit, R. Gilmore, S. Roper, and J. Sackellares, *Electroenceph. Clin. Neurophysiol.* **102**, 98 (1997).
- [34] R. G. Andrzejak, G. Widman, K. Lehnertz, P. David, and C. E. Elger, *Epilepsy Res.* **44**, 129 (2001).
- [35] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, *Phys. Rev. E* **64**, 061907 (2001).
- [36] R. G. Andrzejak, F. Mormann, G. Widman, T. Kreuz, C. E. Elger, and K. Lehnertz, *Epilepsy Res.* **69**, 30 (2006).
- [37] C. Wilke, W. van Drongelen, M. Kohrman, and B. He, *Epilepsia* **51**, 564 (2010).
- [38] R. G. Andrzejak, D. Chicharro, K. Lehnertz, and F. Mormann, *Phys. Rev. E* **83**, 046203 (2011).
- [39] R. G. Andrzejak, O. David, V. Gnatkovsky, F. Wendling, F. Bartolomei, S. Francione, P. Kahane, K. Schindler, and M. de Curtis, Brain topography, doi:10.1007/s10548-014-0380-8.