ELSEVIER

Clinical Neurophysiology 116 (2005) 569–587

www.elsevier.com/locate/clinph

On the predictability of epileptic seizures

Florian Mormann[a,b,*], Thomas Kreuz[a,c], Christoph Rieke[a,b], Ralph G. Andrzejak[a,c], Alexander Kraskov[c], Peter David[b], Christian E. Elger[a], Klaus Lehnertz[a]

[a]*Department of Epileptology, University of Bonn, Sigmund-Freud-Strasse 25, 53105 Bonn, Germany*
[b]*Helmholtz-Institute for Radiation and Nuclear Physics, University of Bonn, Nußallee 14-16, 53115 Bonn, Germany*
[c]*John von Neumann Institute for Computing, Research Center Jülich, 52425 Jülich, Germany*

Accepted 7 August 2004
Available online 6 January 2005

Abstract

Objective: An important issue in epileptology is the question whether information extracted from the EEG of epilepsy patients can be used for the prediction of seizures. Several studies have claimed evidence for the existence of a pre-seizure state that can be detected using different characterizing measures. In this paper, we evaluate the predictability of seizures by comparing the predictive performance of a variety of univariate and bivariate measures comprising both linear and non-linear approaches.

Methods: We compared 30 measures in terms of their ability to distinguish between the interictal period and the pre-seizure period. After completely analyzing continuous inctracranial multi-channel recordings from five patients lasting over days, we used ROC curves to distinguish between the amplitude distributions of interictal and preictal time profiles calculated for the respective measures. We compared different evaluation schemes including channelwise and seizurewise analysis plus constant and adaptive reference levels. Particular emphasis was placed on statistical validity and significance.

Results: Univariate measures showed statistically significant performance only in a channelwise, seizurewise analysis using an adaptive baseline. Preictal changes for these measures occurred 5–30 min before seizures. Bivariate measures exhibited high performance values reaching statistical significance for a channelwise analysis using a constant baseline. Preictal changes were found at least 240 min before seizures. Linear measures were found to perform similar or better than non-linear measures.

Conclusions: Results provide statistically significant evidence for the existence of a preictal state. Based on our findings, the most promising approach for prospective seizure anticipation could be a combination of bivariate and univariate measures.

Significance: Many measures reported capable of seizure prediction in earlier studies are found to be insignificant in performance, which underlines the need for statistical validation in this field.
© 2004 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

Keywords: Intracranial EEG; Epilepsy; Continuous recordings; Seizure anticipation; Seizure prediction; Statistical significance; Receiver-operating-characteristics

1. Introduction

One of the most disabling aspects in epilepsy is the sudden, unforeseen way in which epileptic seizures strike 'like a bolt from the blue'. Apart from the risk of serious injury, there is often a severe feeling of helplessness that has a strong impact on the everyday life of a patient. It is undisputed that a method capable of predicting the occurrence of seizures would significantly improve the therapeutic possibilities (Elger, 2001) and thereby the quality of life for epilepsy patients. In addition to portable warning systems, automated on-demand application of short-acting drugs as well as electrical stimulation or other intervention strategies could be envisioned.

A question of particular interest is whether apart from clinical prodromi (which are found only in some of the patients (cf. Rajna et al., 1997)) characteristic and

* Corresponding author. Address: Department of Epileptology, University of Bonn, Sigmund-Freud-Strasse 25, 53105 Bonn, Germany. Tel.: +49 228 287 9343; fax: +49 228 287 6294.
E-mail address: fmormann@yahoo.de (F. Mormann).

1388-2457/$30.00 © 2004 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.clinph.2004.08.025

objective features can be extracted from the continuous EEG that are predictive of an impending seizure. Much research has been carried out on this topic, and recent studies have reported certain measures derived from the theory of dynamical systems to be to some extent capable of extracting information from the EEG that allow the detection of a preictal state. For an overview of the literature on seizure prediction see Lehnertz and Litt (this issue).

Despite the many publications reporting evidence for the existence of a pre-seizure state, to date no report of a prospective or quasi-prospective prediction of seizures has been published. A major problem with most of the studies presented up to now is that they do not sufficiently (or not at all) investigate the specificity of the described precursors using interictal EEG as control. In addition, many of these studies rely on the use of a posteriori knowledge, e.g. by selecting certain channels out of a large number of channels, or bear the risk of an in-sample over-training of parameters used to calculate measures for the extraction of predictive information. A particular issue that has been neglected by past studies is the need for statistical validation in order to assess the statistical significance of the predictive performance for a given EEG measure (cf. Andrzejak et al., 2003; Mormann et al., 2003b; Winterhalder et al., 2004; Kreuz et al., 2004). Another problem is that up to now no extensive comparison of the performance of different approaches for seizure prediction has been published. Furthermore, there is little experience with continuous long-term-recordings over days, and no study has been reported on comprising different patients from different centers using different pre-surgical evaluation protocols and acquisition systems.

In this paper, we determine the ability of a number of measures to distinguish between the preictal and interictal period using a statistical approach that does not rely on any additional a posteriori knowledge. Retrospectively analyzing continuous multi-day recordings from a group of five patients acquired at different centers, we compare the performance of different univariate and bivariate measures, comprising both linear and non-linear approaches. Applying different smoothing filters to the measures' time profiles and allowing different durations of the preictal period, we evaluate the potential predictive performance of these measures without any antecedent assumptions in the sense of expecting either a preictal increase or a decrease of a measure. We design a number of different evaluation schemes including the analysis of separate channels and seizures and the use of both a constant and an adaptive baseline and pay special attention to issues such as data quality and artifact control as well as statistical validation using seizure time surrogates as a means to assess the statistical significance of an obtained performance value. Based on our results, we discuss the suitability of different measures for seizure prediction.

2. Methods

2.1. Patient characteristics and data acquisition

The analyzed recordings were intracranial multi-day recordings from five patients acquired at different epilepsy centers comprising 46 seizures and a total recording time of 311 h. For detailed information on patient characteristics and data acquisition, refer to Lehnertz and Litt (this issue). Since seizure onset times for the different data sets were supplied by the centers providing the data, we used these times although the criteria applied to determine them may have differed among the centers.

2.2. Characterizing measures of the EEG

For our comparison of different characterizing measures of the EEG we employed a total of 30 different univariate and bivariate measures comprising both linear and non-linear approaches.

As univariate linear measures (12 measures) we used the second, third, and fourth statistical moment of the EEG amplitudes (i.e. the variance σ^2 as a measure for the energy of a signal, the skewness χ, and the kurtosis κ), the relative power of the different spectral bands δ_r, ϑ_r, α_r, β_r, γ_r and the spectral edge frequency f_{50}, the decorrelation time τ_0 as well as the Hjorth mobility HM and Hjorth complexity HC. As univariate non-linear measures (nine measures) we used an effective correlation dimension D^*, the largest Lyapunov exponent L_{max}, the local flow Λ^* as a measure for determinism, the algorithmic complexity AC (an entropy estimate derived from symbolic dynamics), and the loss of recurrence LR as a measure of non-stationarity. In addition we used a surrogate-corrected version of the first four of these measures, termed S–D^*, S–L_{max}, S–Λ^*, and S–AC. As a bivariate linear measure we used a cross-correlation estimate C_{max}. Bivariate non-linear measures (eight measures) comprised different measures for phase synchronization, namely, the mean phase coherence R, the index based on conditional probability λ_{cp}, and the index based on Shannon entropy ρ_{se}, that were applied to phase variables based on both the Hilbert Transform (R^H, λ_{cp}^H, ρ_{se}^H) and the Wavelet Transform (R^W, λ_{cp}^W, ρ_{se}^W), respectively, and two measures for non-linear interdependence, termed S and H. Detailed descriptions and mathematical definitions for all of these measures are given in Appendix A.

For each of these measures, time profiles were calculated from the EEG signals using a moving-window technique. In accordance with previous publications we chose a segment length of 4096 data points with non-overlapping segments. Depending on the sampling rate of the respective data set, the corresponding duration of these segments ranged from 17 to 20.5 s. (For data set A, data were downsampled from 480 to 240 Hz to ensure a segment duration similar to that of the other patients.) This duration can be regarded as a compromise between the required statistical accuracy for

the calculation of a measure and approximate stationarity within a window's length (Blanco et al., 1995; Lopes da Silva, 1987). A fixed number of sample points rather than a fixed duration was chosen to permit the use of a Fast Fourier Transform algorithm for the calculation of certain measures as well as for the surrogate generation.

For every segment from every patient, the value of each of the 30 respective measures was calculated, which resulted in time profiles for every EEG channel (or, in the case of bivariate measures, for every combination of neighboring EEG channels). To ensure the comparability of univariate and bivariate measures in terms of the number of channels or channel combinations, respectively, we restricted the following analyses for the bivariate measures to combinations of neighboring channels only. (For data set A, for which only every other electrode contact was available, combinations of successive channels were used.)

2.3. Statistical evaluation

To investigate whether the pre-seizure period can be distinguished from the interictal period we chose a statistical rather than an algorithmic approach.

The most straightforward concept is to evaluate the hypothesis that a measure capable of discriminating an interictal from an assumed preictal state will attain different values for each state. This hypothesis can be tested by comparing the amplitude distributions of interictal and preictal values of a characterizing measure on a statistical basis. In line with these considerations, a number of studies have reported certain patterns such as local drops or peaks in the profiles of a measure to be of predictive value for an impending seizure. In previous publications, such drops have been characterized by their duration and maximum depth relative to a given reference level (Elger and Lehnertz, 1998; Lehnertz and Elger, 1998). Such a characterization, however, is strongly dependent on the choice of the reference level. In this study, we therefore chose a more robust way for the quantification of drops and peaks in the time profiles of a characterizing measure. After smoothing the profiles with a backward moving average filter of size d we compared the resulting distributions for interictal and preictal values.

Note that this technique is equivalent to applying a moving-window technique to the profiles and calculating the area under the profile in a window of length d. If a window contains a local drop, its area will attain a smaller value, whereas for a local peak it will attain a larger value. The parameter d thus governs the minimum duration of such a drop or peak because events with a duration substantially smaller than d are likely to cancel each other out. Since the profiles can then be characterized by their level only, simple thresholding with a continuously varied threshold can be applied to compare the distributions of drops and peaks in the interictal and the preictal period.

In order to be sensitive to different characteristic widths of drops or peaks, we performed our analysis both with and without smoothing, i.e. for parameter values $d=0$ and 5 min.

A second parameter in our analysis is the presumed duration of the preictal state. Suppose a preictal state were reflected by a change in values of a characterizing measure starting a certain time before seizure onset. Then this effect would be best resolved by defining the preictal period according to this time and then comparing interictal and preictal distributions. Selecting a preictal period of a larger or smaller length would result in an increased number of false negative or false positive classifications, respectively. Since in the literature seizure precursors have been reported to occur on different time scales, it would be ideal to perform a statistical test for every possible duration of the preictal period and to choose the duration with the maximum test performance. For a practical implementation, however, we restricted ourselves to four different durations s of a possible preictal state, selected in accordance with typical anticipation times for different measures described in the literature: $s=5$ min (cf. Duckrow and Spencer, 1992; Rogowski et al., 1981), $s=30$ min (cf. Elger and Lehnertz, 1998; Le Van Quyen et al., 1999, 2000, 2001; Lehnertz and Elger, 1995, 1998; Martinerie et al., 1998; Navarro et al., 2002), $s=120$ min, and $s=240$ min (cf. Iasemidis et al., 2001; Litt et al., 2001; Mormann et al., 2000, 2003a,b).

In order to exclude effects from the postictal period, which can be accompanied by alterations in the EEG, recording periods within 30 min after the onset of a seizure were discarded from the analysis. In cases where the time between two successive seizures was less than $s+30$ min, the maximum amount of data available (i.e. from seizure onset back to the end of the postictal phase of the preceding seizure) was used instead.

Note that the introduced approach is solely based on changes of the level of a characterizing measure and is not directly sensitive to trends as a predictive feature. However, our approach could easily be expanded to the detection of trends by simply taking the derivative of the profiles before determining their amplitude distributions. This would, however, double the amount of results and thus for the sake of structural clarity is left to further studies.

Another important aspect is that EEG epochs with a duration of minutes to hours (as for the presumed preictal periods) or even days (as for the interictal periods) invariably undergo dynamical changes and thus present non-stationary properties. Among the factors that could potentially influence the present analyses are changing antiepileptic drug levels during medical tapering (Lehnertz and Elger, 1997), changes in vigilance states and other circadian fluctuations (cf. Kreuz et al., 2004). The quantification of their influence, however, requires much clinical information and will be subjected to further research.

2.4. Discriminating the preictal from the interictal period

One of the best-known approaches for the discrimination of two amplitude distributions is the so-called receiver-operating-characteristics (ROC) which provides a good indication for the overall separability in terms of sensitivity and specificity of a characterizing measure.

With ROC statistics, a threshold for amplitude values of a measure is continuously varied, and the sensitivity (ratio of true positive classifications to total number of positive classifications) of the discrimination obtained for this threshold is plotted against 1 minus the corresponding specificity (ratio of true negative classifications to total number of negative classifications). The resulting curve is termed ROC curve.

The definitions of sensitivity and specificity can either be based on the hypothesis that values from the second amplitude distribution are generally lower than those from the first distribution or on the opposite hypothesis.

In the case of the first hypothesis (*preictal decrease*, '↓'), the terms 'positive' and 'negative' relate to whether an amplitude value is below or above the threshold, respectively, while the characterization 'true' or 'false' indicates whether values below the threshold belong to the second distribution (i.e. to the preictal period) and values above the threshold belong to the first distribution (i.e. to the interictal period) or not. If the second hypothesis (*preictal increase*, '↑') is chosen as the ROC hypothesis, definitions must be adjusted accordingly.

The area under the ROC curve can be used to quantify the degree to which the two distributions can be distinguished (cf. Fig. 1). For identical distributions this area is 0.5, while for distributions that are completely non-overlapping, values of 0 or 1 are attained, depending on the ROC hypothesis used for the definition of sensitivity and specificity. The capability of a measure to distinguish between the interictal and preictal period, i.e. its potential predictive performance, can thus be quantified by the area under the ROC curve where in case of the first hypothesis an area greater than 0.5 corresponds to preictally decreased values as compared to the interictal values and vice versa.

For a better comparability of the different measures, we always determined ROC values for both hypotheses (preictal *decrease* and *increase*, respectively) and selected the larger one thus achieving a performance value that is always ≥ 0.5 by construction. Furthermore, the analysis parameters d and s were chosen to yield maximum performance values.

2.5. Evaluation schemes

For each of the characterizing measures and every patient, four different evaluation schemes were designed.

Scheme #1. To account for possible global effects, in a first approach we compared the distribution of all interictal values with the distribution of all preictal values from all channels (or, in the case of bivariate measures, from all

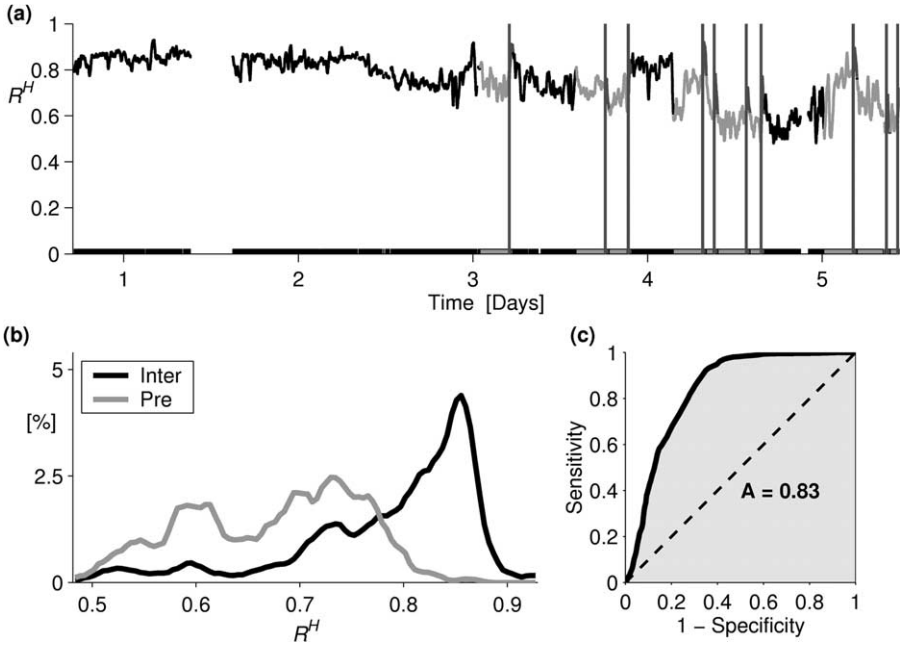

Fig. 1. Illustration of the ROC statistics using the second evaluation scheme. Here, the Hilbert mean phase coherence R^{H} over time is displayed for one channel combination (TR8 and TR9) of data set B (upper part) along with the distributions of values from the preictal and interictal periods from this channel combination (lower part left) and the ROC curve calculated from these distributions (lower part right). For this example, a 5 min smoothing filter as well as a preictal duration of 240 min were chosen, and the hypothesis of a preictal decrease was used for ROC-testing. Note that the parameters and the channel combination used in this example were chosen to yield the optimum performance value for this patient and measure. Discontinuities in the time profile are due to recording gaps.

neighboring channel combinations) and all seizures of a patient (*all seizures*, *all channels*). For this scheme, we expected high performance values only for measures exhibiting a preictal increase or decrease uniformly and on a similar level in all channels.

Scheme #2. In most studies, seizure precursors have been reported to occur only in certain distinct channels. We therefore performed our analysis separately for each channel, turning our attention particularly to the channel with the best performance (cf. Fig. 1). That way we expected to find an increased performance particularly for those measures for which seizure precursors occur constantly in the same channel(s) and on a similar level for all the seizures of a patient (*all seizures, each channel separately*).

Scheme #3. Since the predictive performance of a measure may be of practical value even though it may not detect *every* seizure of a patient, we considered it useful to evaluate the predictive performance separately for each seizure and evaluate the distribution of performance values for different seizures of a patient. This means that for every channel the distribution of preictal values for each seizure is tested separately against the distribution of the entire interictal values of that channel (*each seizure separately*, *each channel separately*, *constant baseline*).

Scheme #4. Using the entire distribution of all the interictal periods recorded over days corresponds to a non-adaptive, constant reference level. In order to account for possible slow changes in dynamics resulting in slow baseline shifts, we designed another evaluation scheme in which for every channel the preictal distribution of each seizure is tested only against the interictal period preceding this seizure's preictal period (*each seizure separately*, *each channel separately*, *adaptive baseline*). A practical problem with this scheme is the clustering of seizures observed in some of the patients. Since a comparison of distributions can only be performed in cases where the interval between two successive seizures is longer than $s + 35$ min (30 min for the postictal period and 5 min as minimum length for the interictal distribution), some of the total of 46 seizures had to be discarded, particularly for large values of the parameter s. The number of discarded seizures was 0 for $s = 5$ min, 4 for $s = 30$ min, 21 for $s = 120$ min, and 31 for $s = 240$ min.

Note that other analysis schemes such as *best channel per seizure* are conceivable. However, in terms of clinical practicability and prospective implementation, we consider such an approach less promising.

2.6. Data quality and artifact control

Any recording of a biological signal bears a certain risk of containing artifacts. In our study, artifacts can be caused by the patient (e.g. eye movements, muscle artifacts) or by the recording system (e.g. power line interferences due to insufficient shielding or insulation, data clipping due to insufficient input level adjustment). In intracranial recordings, patient artifacts are generally much less frequent than in scalp recordings. Patient artifacts caused by an external reference electrode (e.g. ear lobe) can be minimized by using a common average reference instead.

For the analyses performed in this study, we transformed the EEG from data set E using a new referencing scheme and thereby suppressing the very prominent ECG and movement artifacts found in this recording. For this purpose we defined four different groups of contacts (group 1: LFDLN, LSTN, LDTLA; group 2: LDTLH, LDTLP; group 3: RFDLN, RSTN, RPTN contacts 1–4 with contacts 1–3 inverted; group 4: RPTN contact 5, RDTLA, RDTLH, RDTLP) and had every contact referenced against the common average of its respective group.

As for technical artifacts, we first checked all recordings for peaks in the power spectrum caused by the line voltage (60 Hz for recordings performed in the US and 50 Hz for European recordings). Here, we found a rather prominent 60-Hz-peak in data set D.

Furthermore, we scanned every analysis window for the existence of plateaus (i.e. consecutive sampling points of identical value) in order to identify recording dropouts and data clippings. Any analysis window with a plateau of more than 40 data points or with more than 1000 data points contained in different plateaus were declared as artifacts. If the number of artifact windows thus identified was more than 5% of all windows for a given channel, this entire channel was discarded from the analysis. The channels thus discarded were TR01, TR06, TR07 for data set B, LTD1, LTD2 for data set C, LDTLH3–LDTLH7, LDTLP1–LDTLP8 for data set E (using new references) as well as the corresponding channel combinations.

For the time profiles of the remaining channels, values corresponding to artifact windows were set to the median of the respective amplitude distributions (i.e. interictal or preictal) in accordance with the respective evaluation scheme.

2.7. Seizure times surrogates as a test for statistical validity

Following Andrzejak et al. (2003), let us now consider the following null hypothesis: 'The transition from the interictal to the ictal state is an abrupt phenomenon. An intermediate preictal state does not exist.' Despite the fact that in this case no information predictive of an impending seizure could be extracted from the EEG, many of our characterizing measures would probably still render a performance value (as quantified by the area under the ROC curve) different from 0.5. This is due to the fact that time series such as the profiles of our characterizing measures that were obtained from recorded data usually contain fluctuations and therefore in principle have a non-zero probability of attaining any course within their range of definition. Even though this probability may be very small for a single time profile, the number of different variables (different measures, different

patients, different channels, different seizures, different parameters, different ROC hypotheses) included in this study leads to an increased probability for finding a combination of these variables which leads to results that may appear significant without actually being so. For our particular evaluation this means that the optimization over parameters s and d and the applied ROC hypotheses as well as the best channel selection applied in the last three evaluation schemes is likely to decrease the actual statistical significance of the results. Concerning our analysis this means that there is no guarantee that any observed effects are indeed due to different characteristics of the interictal and preictal state, but instead could be caused, at least in part, by random fluctuations (cf. Mormann et al., 2003b) selected by the optimization process. The influence of the different degrees of freedom on the statistical significance of the obtained results is difficult if not impossible to estimate on a theoretical basis. Hence it is impossible to decide whether a given ROC performance value obtained from our data indicates the existence of a preictal state or whether it is consistent with the null hypothesis stated above. We therefore applied a surrogate test to evaluate the null hypothesis of finding the same results just due to random fluctuations contained in the data (cf. Andrzejak et al., 2003).

For this purpose we designed a total of 19 different surrogate sets of randomized seizure onset times by randomly permutating the intervals between different seizures including the interval between the first seizure and the beginning of the recording. We then once again applied the different evaluation schemes to the original time profiles of all of our characterizing measures using now instead of the original seizure onset times one of the surrogates. This process was carried out separately for each of the 19 seizure times surrogates. By proceeding this way we ensured that the a priori significance of every performed ROC test, which is determined by the sizes of the two distributions, is the same for the surrogates as for the original. Note that each of the seizure times surrogates was applied to the original unprocessed profiles with the only constraint that none of the surrogate seizure times were allowed to coincide with an original onset time.

Using a z-score test, the significance level P for a rejection of the null hypothesis can be determined with $\alpha = 0.05$ being the nominal size of this test. If the performance of a measure for the real seizure onset times exceeds the maximum performance for the surrogates, the results can be assumed to be statistically significant at a level of $P = 0.05$; if no more than one of the surrogates perform better than the original, the significance level is $P = 0.10$, etc.

Note that for data set A, the test for statistical validity could not be performed since with the given constraint no more than two possible seizure time surrogates could be constructed from the three seizures of this patient.

3. Results

The statistical evaluation based on ROC curves is shown in Fig. 1. In this example, the time profile of a measure for phase synchronization (mean phase coherence based on Hilbert Transform R^{H}) for one channel combination of a patient (data set B) is plotted in the upper row of the figure. A smoothing filter of $d = 5$ min was applied and the duration of the preictal period was set to $s = 240$ min in this example. The corresponding amplitude distributions of R^{H} for both the preictal and the interictal periods are displayed in the lower row along with the ROC-curve derived from these distributions under the ROC hypothesis of a preictal decrease. The performance value as quantified by the area under the ROC curve is 0.83 which indicates a fair degree of separability of the two distributions. For this time profile, a fixed threshold value of 0.75, for instance, would render a classification of the preictal and interictal periods at a sensitivity and specificity of 75% each, a threshold value of, e.g. 0.8 would yield a sensitivity of 95% and a specificity of 60%.

For the sake of a better legibility of the following figures, results for measures that were found to be highly correlated to other measures are not displayed. Care was taken, however, that any results described or conclusions drawn are valid for all the measures analyzed.

The performance of each measure was calculated using parameters d and s and a ROC hypothesis that yielded an optimal mean performance of this measure over four of the five patients, i.e. each measure was ‘allowed to pick its optimum parameter setting’. Patient A was excluded from the mean value because, as explained above, no seizure times surrogates and therefore no significance value could be determined for this patient. For completeness, performance values for this patient are displayed along with those of the other patients using the same parameters as for the other patients.

Results for evaluation Scheme #1 (*all seizures*, *all channels*) for the different measures and patients are shown in Fig. 2. Displayed for each measure are the performance values for all five patients as well as the mean value calculated from four of these patients (excluding data set A). Since due to the use of the optimum ROC hypothesis (preictal decrease or increase) mean performance values are by construction equal to or greater than 0.5, bars representing these values are displayed starting at the 0.5 level. Statistically significant values are displayed as black bars, others as gray bars. Optimal values for the parameters d and s as well as the underlying ROC hypothesis are displayed at the bottom of Fig. 2 for each measure. In addition, the significance values as determined from the seizure times surrogate test are displayed at the top of the figure. For better legibility, only P-values smaller than or equal to 0.10 are displayed.

The area under the ROC curves did not deviate from 0.5 by more than 0.06 for any of the examined measures

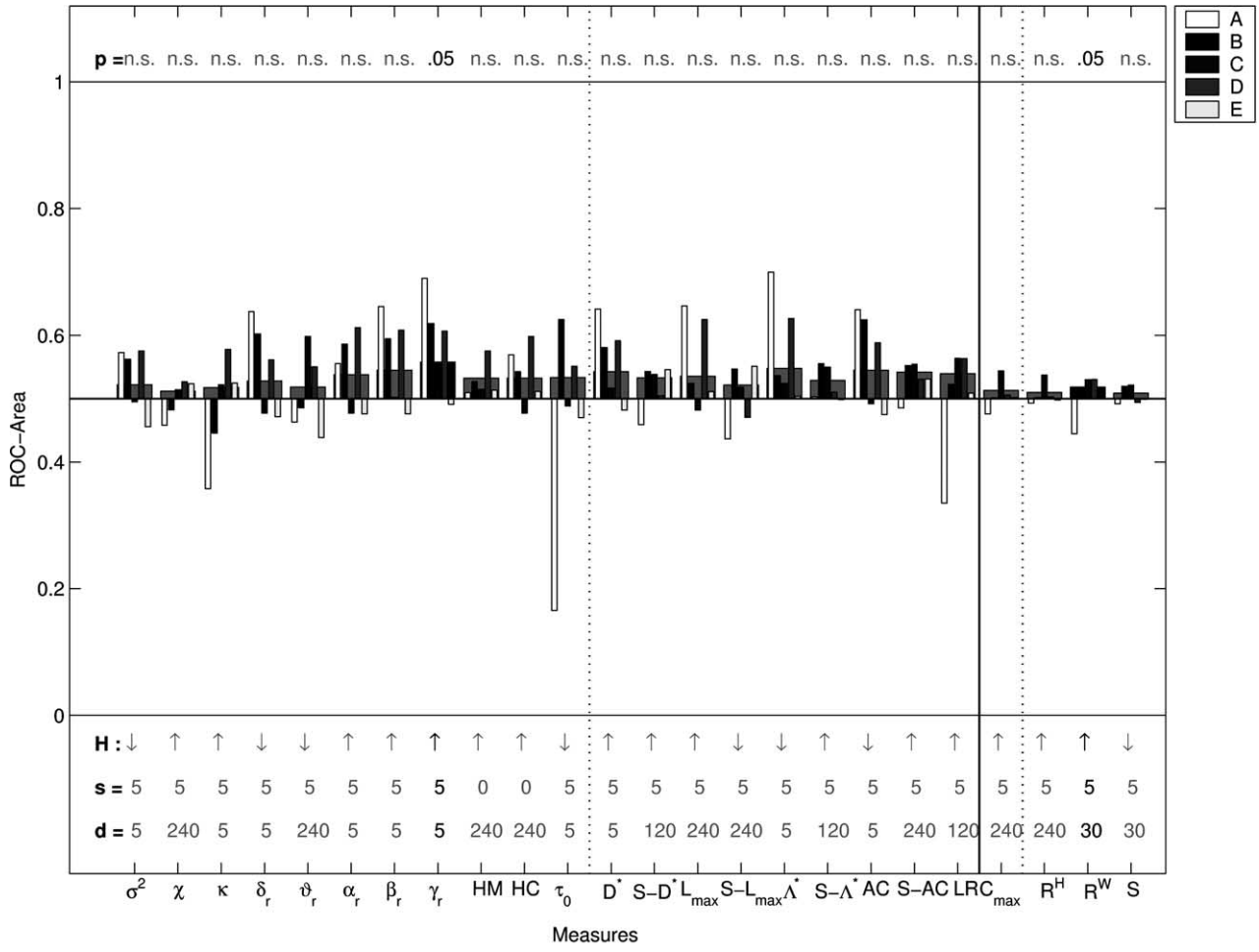

Fig. 2. Performance values for different measures and patients according to evaluation Scheme #1 (*all seizures*, *all channels*) as well as mean values averaged over patients (represented by the black and gray bars in the background) for each measure (omitting data set A). Parameters and ROC hypotheses were chosen to yield a maximum mean value for each measure. At the bottom of the figure, parameters *d* and *s* as well as the ROC hypotheses ('↓' for preictal decrease, '↑' for preictal increase) for each measure are displayed; at the top significance levels of the average performance values are displayed with n.s. (not significant) denoting *P*-values greater than 0.10. Displayed EEG measures comprise variance σ^2, skewness χ, kurtosis κ, spectral bands δ_r, ϑ_r, α_r, β_r, γ_r, Hjorth mobility HM, Hjorth complexity HC, decorrelation time τ_0, correlation dimension D^*, largest Lyapunov exponent L_{max}, local flow Λ^*, algorithmic complexity AC, surrogate-corrected versions of the latter 4 measures S–D^*, S–L_{max}, S–Λ^*, S–AC, loss of recurrence LR, cross-correlation C_{max}, Hilbert mean phase coherence R^H, Wavelet mean phase coherence R^W, non-linear interdependence S.

and patients, which means that none of the measures were able to clearly discriminate the preictal from the interictal period in this evaluation scheme. Only for two measures, namely, for the wavelet mean phase coherence R^W and the relative gamma band power γ_r, performance values appeared to be significant.

In Fig. 3 results are displayed for evaluation Scheme #2 (*all seizures*, *each channel separately*). This figure shows the distribution of performance values over different channels including maximum, minimum, and median performance. In addition, mean values were calculated from the maximum performance of each patient excluding once again data set A. For a better legibility, distributions over channels are displayed only for measures yielding mean values that were statistically significant. When comparing these values, we found mean performance values for the univariate measures to range from 0.63 to 0.72, while those for the bivariate measures ranged from 0.72 to 0.75. Among the univariate measures, the best performers were the Hjorth parameters HM and HC and the surrogate-corrected algorithmic complexity S–AC. The best bivariate measures were the maximum cross-correlation C_{max} and the measures for phase synchronization based on the Hilbert transform R^H, λ_{cp}^H, ρ_{se}^H. Furthermore results show that while all bivariate measures reached optimum performance for a preictal period of 240 min, the best preictal period for the univariate measures varied among different values. Optimal smoothing parameters were 5 min for all measures. Apart from all bivariate measures, only values for the correlation dimension D^*, the Hjorth complexity HC, and the surrogate-corrected algorithmic complexity S–AC appeared to be statistically significant.

Fig. 4 shows results for evaluation Scheme #3 (*each seizure separately*, *each channel separately*, *constant*

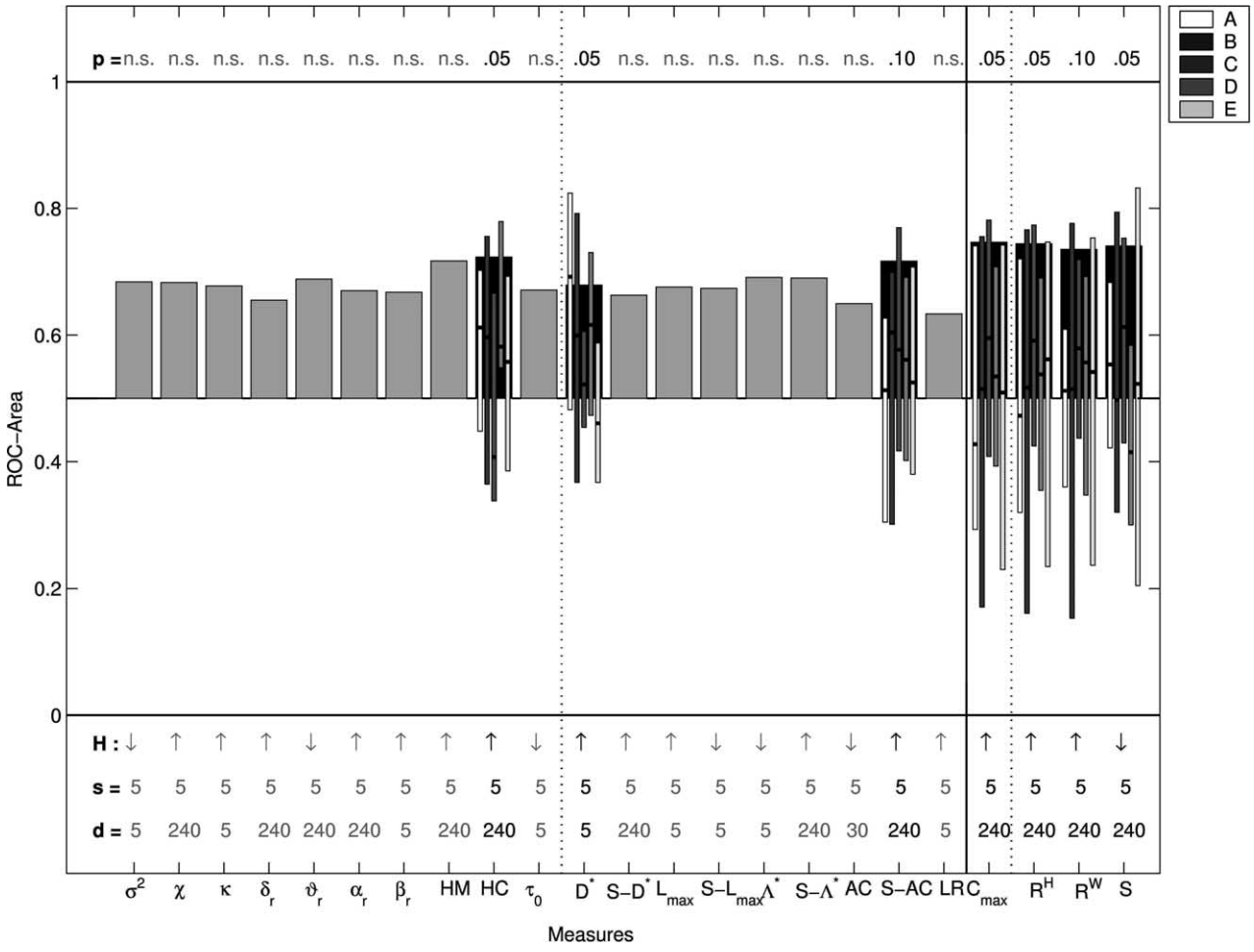

Fig. 3. Same as Fig. 2, but this time for evaluation Scheme #2 (*all seizures*, *each channel separately*). Mean values over patients (represented by black and gray bars in the background) for each measure were obtained by averaging over the maximum performance of different channels for each patient (omitting patient A). For measures with an overall performance that was statistically significant, the distribution of performance values over different channels is depicted by vertical bars with the median of the distribution denoted by a little black square on each bar.

baseline). Since for this evaluation scheme the distributions of performance values for a given patient and measure is two-dimensional (for different seizures and different channels), we determined the median performance of the seizures in each channel and depicted the distribution of these median values for different channels just as for the previous evaluation scheme. Selecting the median seemed convenient as this means that half of the seizures achieved a better performance, while for the other half performance values were lower. Results for this evaluation scheme were similar to those for the previous scheme except that the median performance value for the separate seizures of a channel was generally higher than the performance value for all seizures of a channel. Performance values for univariate measures ranged from 0.74 to 0.80 while values for bivariate measures ranged from 0.81 to 0.86. Among the univariate measures, the best performers were again HM, HC, and *S*–AC. Among the bivariate measures, the measures for non-linear interdependence showed a lower performance than the rest. Optimum parameters for the different measures were similar as for the previous scheme. Statistical significance was found for all bivariate measures except the nonlinear interdependence *S* as well as for the Hjorth parameters, the relative beta band power β_r, and the surrogate-corrected algorithmic complexity *S*–AC.

To investigate the variability in performance for different seizures in the time profile of a given channel, the distributions of performance values among seizures were calculated for each patient and measure for the best channel based on the median performance among seizures, i.e. for the channels with maximum performance in the distributions depicted in Fig. 4. While for almost all measures and patients the median of the distribution among seizures was above 0.5, the minimum was below 0.5. This means that while for the majority of seizures a certain discrimination of the interictal and preictal period could be achieved based on either a preictal decrease or increase in values, there were also seizures for which the opposite effect was observed.

Results for evaluation Scheme #4 (*each seizure separately*, *each channel separately*, *adaptive baseline*) are displayed in Fig. 5 in the same manner as for the previous scheme. It turned out that performance values for

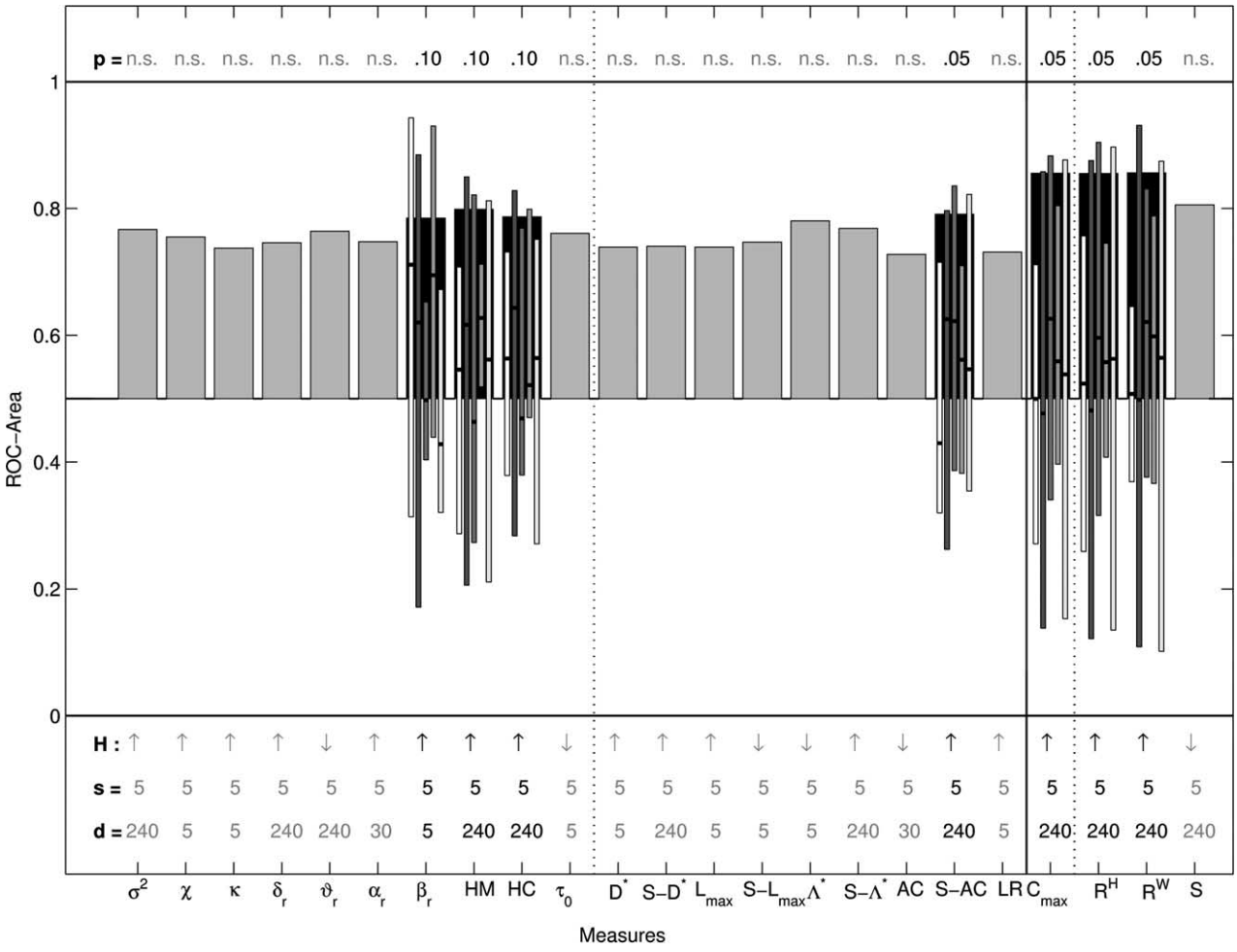

Fig. 4. Same as Fig. 3, but this time for evaluation Scheme #3 (*each seizure separately*, *each channel separately*, *constant baseline*). Mean values over patients (represented by black and gray bars in the background) for each measure were obtained by averaging over the maximum value of the median seizure performance of different channels of each patient (omitting patient A). Instead of one performance value per channel now the median performance of the different seizures per channel is depicted in the distributions for each patient.

the bivariate measures ranged from 0.76 to 0.84 and were thus lower than for the previous scheme and for the most part did not reach statistical significance. The univariate measures on the other hand showed a clear increase in performance with performance values ranging up to 0.90. Among the best univariate performers were the power bands δ_r through γ_r, the local flow Λ^*, the correlation dimension D^*, and the algorithmic complexity AC. Values for these measures and a number of others reached statistical significance.

Concerning the distributions of different seizures for this scheme, it again turned out that in most cases the hypothesis of either a preictal decrease or an increase in values was not valid for *all* seizures of a patient.

4. Discussion

When comparing different measures in terms of their suitability for seizure prediction by testing the ability of these measures to discriminate a presumed preictal state from the interictal period, it is useful to distinguish between univariate and bivariate aproaches.

While univariate measures quantify certain properties of the EEG signal thus possibly reflecting the *state* of a certain region of the brain, the bivariate measures analyzed in this study quantify the degree of synchronization between two EEG signals thereby possibly reflecting the amount of *interaction* between different areas of the brain.

4.1. Univariate measures

As for the *univariate* measures evaluated in this study, the first evaluation scheme (*all seizures*, *all channels*) showed no global effect in the sense of a uniform and similar change found in all channels and for all seizures of a patient. If there was at all a difference in any of the channels, it either occurred in too few channels to have an effect on the overall distributions or it was opposite for different channels (in the sense of a preictal increase in values in some channels

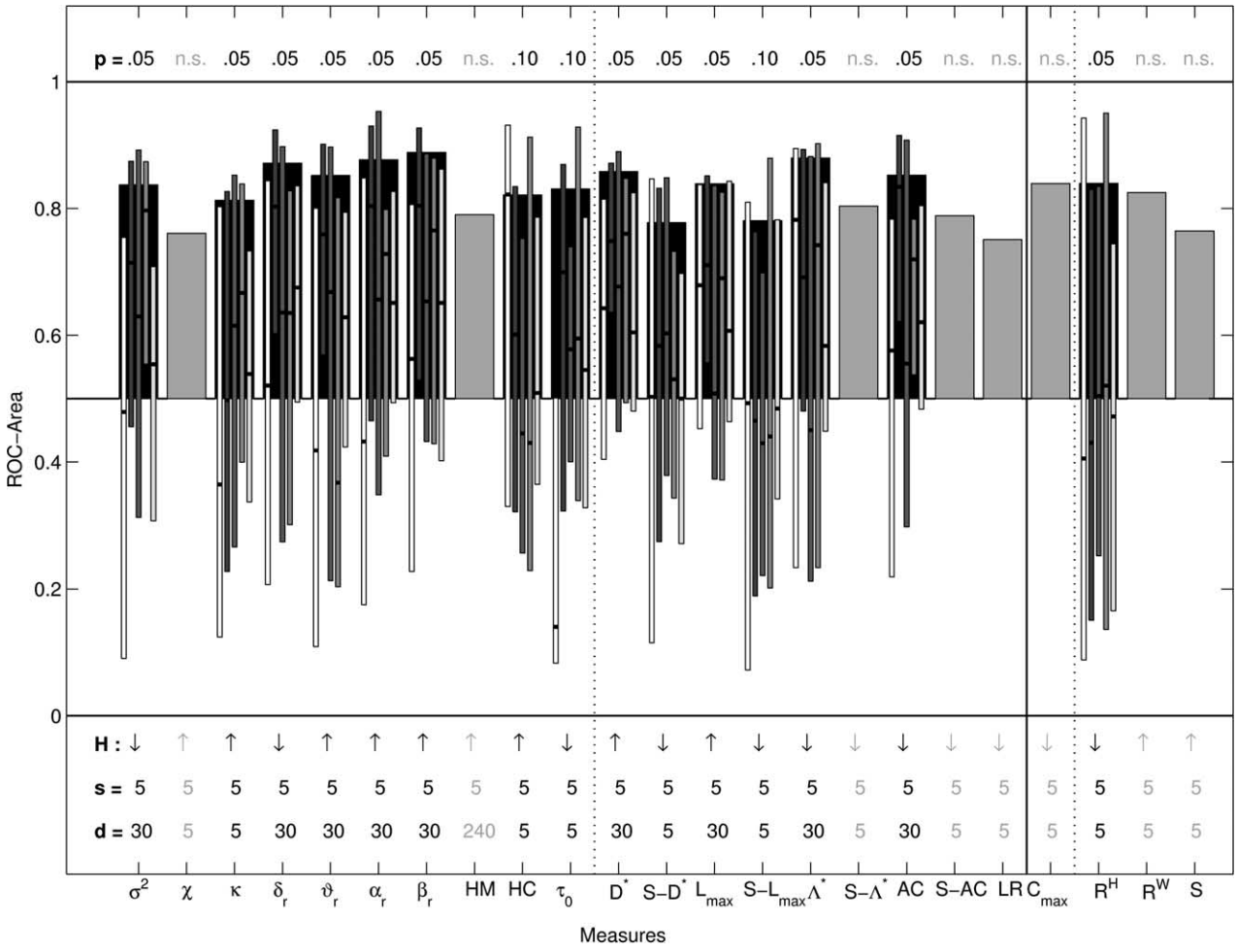

Fig. 5. Same as Fig. 4, but this time for evaluation Scheme #4 (*each seizure separately*, *each channel separately*, *adaptive baseline*). The preictal period of each seizure is tested only against the interictal interval preceding this preictal period.

and a decrease in others) so that effects in different channels may have canceled each other out. The fact that for one of the univariate measures a significance level of $P = 0.05$ was found is not surprising since with the given error probability of 0.05 for falsely classifying a measure's performance as significant, one would expect a false classification among a total number of 21 measures. In this context it is important to point out that the empirical size of the surrogate test for statistical significance used in this study might deviate from its nominal size. An estimation of the empirical size cannot be derived from the data available and will be subject to further research.

Results for the second (*all seizures*, *each channel separately*) and third (*each seizure separately*, *each channel separately*, *constant baseline*) evaluation scheme were not very different from those for the first scheme. The only measures with significant performance values in both of these schemes were the Hjorth complexity HC and the surrogate-corrected algorithmic complexity *S*–AC exhibiting a preictal increase in values hours before the seizure onset. These measures possibly reflected changes in dynamics that were also tracked by the bivariate synchronization measures but showed a lower performance than any of these.

In the fourth evaluation scheme (*each seizure separately*, *each channel separately*, *adaptive baseline*) on the other hand the univariate measures showed a clear increase in performance. In addition, for a number of these measure these performance values were statistically significant. The duration of the preictal changes found in this evaluation scheme did not exceed 30 min for any of the measures.

For a possible explanation of the behavior of the different measures with significant performance values in terms of the applied ROC hypothesis, it is useful to take a look at the preictal changes in the spectral bands: Fig. 5 shows that preictally there is a relative decrease of power in the delta band that is accompanied by a relative increase in the remaining bands. The decrease of the variance σ^2 furthermore represents a decrease of the total energy and power of the signal. This relative decrease of power in the lower frequencies of the EEG causes an increase of the Hjorth parameters HM and HC and a decrease of the decorrelation time.

The reduction of linear correlations due to the decrease in delta power could furthermore account for the observed preictal increase of both the correlation dimension D^* and the largest Lyapunov exponent L_{max} since these measures have been previously found to be inversely correlated (Pereda et al., 1999) as well as for the decrease of the local flow Λ^* and the algorithmic complexity AC. Consistently, the decrease in Λ^* and AC, which reflects a high-dimensional stochastic dynamics, is accompanied by a decrease in S–Λ^* and S–AC as well, while for the same reason the increase in D^* and L_{max} is accompanied by a decrease of the surrogate-corrected versions S–D^* and S–L_{max} of these measures.

Locating the channels exhibiting the maximum values for the median performance revealed no uniform picture as optimum channels were found ipsilateral to the focal side for some patients and contralateral for others even for measures with statistically significant results. Furthermore these channels changed from measure to measure for each patient although they were often located within the same region.

Comparing our results with the literature on seizure prediction we find that in contrast to an earlier report about a preictal elevation of the energy of a signal occurring hours before the onset of a seizure (Litt et al., 2001), it turns out that in our study no significant effect on a time scale of hours is observed which is in agreement with another recent study (Maiwald et al., 2004) in which earlier results could not be replicated either. On a shorter time scale we even find the opposite effect, i.e. a decrease in the total energy of a signal within the 30 min prior to seizure onset.

In comparison with earlier reports describing drops in the values of the largest Lyapunov exponent for several minutes prior to the onset of a seizure (Iasemidis and Sackellares, 1991), we here again observe the opposite effect, namely, an increase of L_{max} within the 30 min before seizure onset.

Concerning earlier reports describing a decrease in correlation dimension 5–25 min before seizure onset (Elger and Lehnertz, 1998; Lehnertz and Elger, 1998), we find a discrepancy in the sense that we here observe a preictal *increase* in dimension. Analysis of predictive performance for the ROC hypothesis of a preictal *decrease* (results not shown) indeed yielded values close to 0.5, which means that a distinct preictal drop in dimension was not observed in the time profiles of any of the channels analyzed. These findings are in line with other recent studies (Aschenbrenner–Scheibe et al., 2003; Maiwald et al., 2004) where the correlation dimension was found to perform insufficient for clinical application. This discrepancy to earlier studies may be due to the fact that in these studies the parameters used in the algorithm for the calculation of the correlation dimension were selected in-sample and that the obtained results were not checked for over-training via cross-validation. In addition, the study was performed using comparably short interictal recordings selected by EEG technicians with regard to epileptiform activity, and a posteriori knowledge in the sense of a channel selection was used. Perhaps most importantly, the authors of these studies only looked for a decrease in dimension, so the hypothesis of a preictal increase was not at all evaluated.

With respect to the observed relative decrease in delta power before seizures we would like to note that this phenomenon could also be argued to be (at least in part) an effect of the postictal period. Suppose the well-known phenomenon of a postictal slowing in the EEG lasted longer than the 30 min after a seizure discarded as postictal in this study and then gradually faded away. Then for a cluster of seizures one would observe a continuous decrease of the low frequency content (delta power) while approaching the next seizure. Therefore, the presumably preictal effects found in this evaluation scheme could actually be caused by postictal effects instead. In this case, however, we would expect a low performance of the leading seizures of each patient as compared to the following seizures, which was not found for the patients analyzed in this study. Hence, it appears that the observed effects for the most part indeed reflect features of the preictal state.

The changes in spectral energy before seizures and the accompanying changes of other univariate measures found only in this evaluation scheme thus indicate that univariate measures bear a certain sensitivity to the process of seizure generation on a time scale substantially shorter than those for the bivariate measures and with the need for adjusting the reference level, i.e. the threshold to discriminate between interictal and preictal distributions for each seizure. The poor performance of the univariate measures in the second and third evaluation scheme indicates a high variability of this threshold from seizure to seizure, which makes a prospective implementation of an algorithm for seizure anticipation a non-trivial problem (cf. Mormann et al., 2003b). Furthermore, evaluation Scheme #4 does not allow to estimate the specificity of the observed preictal changes. Suppose, for instance, the level of a measure decreased continuously between two seizures, then the variation of the retrospectively determined threshold for separating the preictal state from the interictal state from seizure to seizure would make it impossible to determine the beginning of a preictal state.

Regarding the performance of the univariate measures in evaluation Scheme #4 for the different seizures of a patient, it should be noted that for every measure there are seizures that exhibit a preictal behavior opposite to the optimum ROC hypothesis for this measure, which means that for all the measures and most of the patients there were certain seizures that would not be predictable based on the proposed methods.

As for the smoothing parameter d, the finding that in all evaluation schemes for the vast majority of measures (univariate and bivariate) a smoothing filter of length $d=5$ min is favored over no smoothing is due to the fact that smoothing of a time series generally decreases the width of its amplitude distribution while leaving the center unchanged. If two distributions have a different center, their separability is thus increased by making them each more

narrow. This effect becomes the more pronounced the less linear correlations exist in the underlying time series.

4.2. Bivariate measures

Regarding the *bivariate* measures examined in this study, the first evaluation scheme (*all seizures*, *all channels*) just like for the univariate measures revealed no general difference between the preictal and interictal period that could be found for all channels and seizures.

Results from the second evaluation scheme (*all seizures*, *each channel separately*) showed that based on a 'best channel performance' using all seizures for the testing, i.e. looking for an effect in a channel (combination) that occurred prior to all seizures and in a similar manner, the bivariate measures showed a substantially higher performance than the univariate measures with performance values that were for the most part statistically significant.

These results show that for the bivariate measures the discrimination of a presumed preictal state from the interictal period is in principle possible using a fixed threshold that remains constant over days so a prospective detection algorithm for a preictal state could be easily implemented (Mormann et al., 2003b) provided that the optimum channel combination is known.

A superior performance of the bivariate measures was also observed in the third evaluation scheme (*each seizure separately*, *each channel separately*, *constant baseline*) where the median performance values of all the seizures of a patient were found to clearly surpass the values from the previous evaluation scheme achieved by testing all seizures together. Some preictal periods, however, could not be discriminated at all due to an opposite ROC hypothesis. Examination of the distribution of performance values for different seizures revealed no particular role of the leading seizure as compared to the other seizures of a patient.

As for the optimum duration of the preictal period, there was a clear tendency for all bivariate measures towards a preictal period of 240 min which is consistent with our earlier findings (Mormann et al., 2003a,b) and accounts for the fact that the changes in dynamics tracked by these measures occur on a rather large time scale.

In the fourth evaluation scheme (*each seizure separately*, *each channel separately*, *adaptive baseline*) a decrease in performance values as compared to the second and third scheme was observed for the bivariate measures, and performance values for the most part were no longer statistically significant. This indicates that for most seizures the corresponding preictal distribution did not differ significantly from the preceding interictal distribution which could be due to an underestimation of the duration of the presumed preictal state for the bivariate measures. If the changes before seizures started even sooner than 240 min before a seizure, the interictal distributions would contain a substantial amount of preictal values, particularly in case of a clustering of seizures. In the other evaluation schemes, this effect would be much less predominant due to the larger size of the interictal distributions.

The rather large time scale on which preictal changes in synchronization are observed indicates that the bivariate synchronization measures might not be very useful for an actual prediction of seizures but instead rather reflect a disposition, i.e. an increased probability, for the occurrence of a seizure. In this context it would be interesting to examine whether the slow changes reflected by these measures correspond in any way to antiepileptic drug levels of the patients. This would, however, go beyond the scope of this study and thus will be subjected to further research.

The fact that both in the second and in the third evaluation scheme the predominant optimum hypothesis for the bivariate measures is a preictal increase in synchronization, is at first surprising as in previous studies we predominantly found a preictal decrease in synchronization. There is, however, no contradiction to earlier results as it can be seen from Figs. 3 and 4 that the predictive performance based on the opposite hypothesis is of the same magnitude (for data sets B (cf. Fig. 1) and E it is even better). While in some channel combinations we found a preictal increase in synchronization, in other channel combinations a pronounced preictal decrease was observed. Although there was no apparent relation of channel combinations exhibiting maximum performance values based on either ROC hypothesis to the focal region of each patient and maximum values were observed both ipsilateral and contralateral, we frequently found an increase in synchronization to be associated with a decrease in an adjacent channel combination. This phenomenon can be regarded as further evidence for the hypothesis that changes in synchronization are observed when a recording site becomes 'torn out' of its physiological state of synchronization with a neighboring recording site in one direction and is forced into synchronization with some slowly expanding region of pathologically synchronized neuronal tissue emanating from another direction (Mormann et al., 2000, 2003a).

4.3. Summary and outlook

The results obtained in this study differed substantially for univariate and bivariate measures. While univariate measures appeared to be sensitive to changes before a seizure only in relation to the period immediately preceding these changes, bivariate measures were found to reflect changes in dynamics on a longer time scale starting hours before a seizure. A remarkable aspect of our results is that both among the univariate and the bivariate approaches linear measures performed equally good or even better than non-linear measures.

Concerning the future perspective of a prospective algorithm for seizure anticipation, we have shown that

a statistically significant discrimination between a presumed preictal state and the interictal period is in principle possible using bivariate measures for synchronization. Whether results for the univariate measures can be confirmed in a prospective setting remains to be investigated by further studies.

With respect to the performance values obtained for the different measures, we regard the future perspective of an implantable prospective seizure prediction system working with both a sensitivity and a specificity of 100% as unrealistic. Instead we would conjecture the most promising approach to be a combination of bivariate and univariate measures rendering a probability estimate for the occurrence of a seizure within a certain time frame. The statistical validation performed in this study suggests that a prospective seizure anticipation system is in principle possible and would perform better than random. Whether this is sufficient for a clinical application would need to be decided on an individual basis.

It should be noted in this context that although optimum performance of a measure was found to crucially depend on the location of the recording sites, there was no consistency in the optimum locations for different measures or patients that could be used as a priori knowledge for future potential implantations.

A further problem in this context is that any intracranial EEG recordings from epilepsy patients are obtained in a clinical environment during the pre-surgical work-up with patients undergoing medical tapering and other diagnostic procedures so there is no guarantee that results obtained from such an environment are indeed transferable to the everyday life of a patient.

Regarding the multi-center character of this study it is remarkable that in spite of the differences in the pre-surgical work-up among the various centers, the obtained results were quite consistent for the different patients.

For further improvements of the predictive performance of measures it would be promising to investigate circadian changes of these measures with particular regard to their dependence on different vigilance states such as sleep and other influencing factors.

To conclude, our evaluation of a number of univariate and bivariate measures in terms of their ability to discriminate a presumed preictal state from the interictal period has provided statistically significant evidence for the existence of a preictal state that can be discriminated from the interictal state by certain dynamical measures.

Acknowledgements

We are grateful to Peter Grassberger and Martin Kurthen for useful discussion and valuable comments. This work was supported by the Deutsche Forschungsgemeinschaft.

Appendix A. Characterizing measures of the EEG

A.1. Univariate linear measures

The information contained in consecutive amplitude values of a signal that is sampled in the form of a discrete time series $x(t)=x(t_i)=x(t_0+i\Delta t)=x_i$ (with $i=1,\ldots,N$ and Δt denoting the sampling interval) can also be encoded by amplitudes and phases of harmonic oscillations with a range of different frequencies. The map that translates between these representations in the time domain $\{x_i\}$ and the frequency domain $\{s_k\}$ is called Fourier Transform. The power spectrum of a real signal is given by the square of the amplitudes of the Fourier Transform: $\{p_k\}=\{|s_k|^2\}$ with $k=1,\ldots,N/2$ for any frequency where $f_s=1/\Delta t$ is the sampling rate.

The total power of the time series is given by

$$P=\sum_{k=1}^{N/2} p_k=\sum_{f=0}^{f_s/2} p_f.$$

In the following we assume that the time series' mean values were set to zero prior to analysis.

A.1.1. Statistical moments

Statistical moments characterize the amplitude distribution of a signal $\{x_i\}$. The second moment is the variance $\sigma^2=(1/(N-1))\sum_{i=1}^{N} x_i^2$ (of a signal with zero mean), which according to Parseval's theorem is also a measure of the total energy or power of a signal (cf. Litt et al., 2001). The third moment is the skewness $\chi=(1/N)\sum_{i=1}^{N}(x_i/\sigma)^3$ and the fourth moment the kurtosis $\kappa=[(1/N)\sum_{i=1}^{N}(x_i/\sigma)^4]-3$. The skewness is zero for symmetric amplitude distributions and non-zero for asymmetric distributions. Neglecting the direction of the asymmetry, which can be read from the sign of the skewness, we here use only the absolute value $|\chi|$ of the skewness. The kurtosis measures the relative peakedness or flatness of an amplitude distribution.

A.1.2. Spectral band power

Different physiological and pathological processes are reflected by activity in different frequency ranges of the power spectrum $\{p_f\}$ of the EEG. According to these ranges a set of power spectral bands (δ, ϑ, α, β, γ) were defined in classical EEG analysis. The relative power contained in these bands is defined as

$$\delta_r=\frac{1}{P}\sum_{f=0.5\ \mathrm{Hz}}^{4\ \mathrm{Hz}} p_f;\quad \vartheta_r=\frac{1}{P}\sum_{f=4\ \mathrm{Hz}}^{8\ \mathrm{Hz}} p_f;\quad \alpha_r=\frac{1}{P}\sum_{f=8\ \mathrm{Hz}}^{13\ \mathrm{Hz}} p_f;$$

$$\beta_r=\frac{1}{P}\sum_{f=13\ \mathrm{Hz}}^{30\ \mathrm{Hz}} p_f;\quad \gamma_r=\frac{1}{P}\sum_{f=30\ \mathrm{Hz}}^{48\ \mathrm{Hz}} p_f$$

where P is the total power of the signal.

A.1.3. Spectral edge frequency

In a typical EEG signal, most of the power is contained within the frequency band from 0 up to 40 Hz: $P_{40\ \mathrm{Hz}} \approx P$. We here use as a characterizing measure for the power distribution the so-called spectral edge frequency (Stanski et al., 1984) which is defined as the minimum frequency up to which 50% of the spectral power up to 40 Hz is contained in the signal:

$$f_{50} = \min\left\{ f \left| \sum_{\nu=0.5\ \mathrm{Hz}}^{f} p_\nu > P_{40\ \mathrm{Hz}} \cdot 0.50 \right. \right\}$$

A.1.4. Characteristics of the autocorrelation function

The autocorrelation function of a time series with zero mean is defined as

$$A(\tau) = \frac{1}{(N-1)\sigma^2} \sum_{i=1}^{N-\tau} x_i x_{i-\tau}$$

for $\tau = 0,\ldots,N-1$ with σ^2 denoting the variance of the signal. By definition A ranges between -1 and 1 with $A(0) = 1$. Provided that the time series is non-periodic, the autocorrelation function decays from $A(0)$ with increasing values of τ, and fluctuates around zero for larger τ-values. The slower $A(\tau)$ decays initially, the stronger are the linear correlations of the time series. Hence, an estimate of the range of linear correlations can be defined using the first zero crossing

$$\tau_0 = \min\{\tau | A(\tau) = 0\}$$

of the autocorrelation function.

A.1.5. Hjorth parameters

Hjorth defined activity, mobility and complexity as 'a set of parameters intended as a clinically useful tool for the quantitative description of an EEG' (Hjorth, 1970). Since the activity is proportional to the variance of a signal, we here use the mobility, defined as the variance of the slopes of the EEG normalized by the variance of the amplitude distribution of the time series, and the complexity, defined as the variance of the rate of slope changes with reference to an ideal sine curve. In the frequency domain, the mobility and complexity can be estimated from the second and fourth statistical moment of the power spectrum, respectively:

$$\mathrm{HM} = \frac{1}{P} \sum_{k=1}^{N/2} p_k k^2; \quad \mathrm{HC} = \frac{1}{P} \sum_{k=1}^{N/2} p_k k^4$$

A.2. Univariate non-linear measures

While linear measures are calculated directly from the time series or its power spectrum, a number of non-linear measures have been derived from the theory of dynamical systems (Kantz and Schreiber, 1997; Ott, 1993; Schuster, 1989) that are designed to quantify different properties of so-called state space trajectories. Calculation of these measures therefore requires reconstruction of the state space trajectory from the scalar time series $\{x_i\}$ where $i = 1,\ldots,N$. This reconstruction can be achieved by means of delay coordinates $\vec{x}_i = (x_i, x_{i-\tau}, \ldots, x_{i-(m-1)\tau})$ (Takens, 1981) with $i = 1,\ldots,M = N - m\tau$ where $\{\vec{x}_i\}$ defines the reconstructed state space trajectory. Here, τ is a time delay and m is the embedding dimension which, according to Whitney's theorem, must be chosen as $m \geq 2d+1$ (where d is the dimension of the geometrical object formed by the genuine trajectory in state space) if any exact determinism present in the original (multi-variate) system is to be preserved (Whitney, 1936).

*A.2.1. Estimate of an effective correlation dimension D**

For deterministic dynamics the correlation dimension (Grassberger and Procaccia, 1983) allows to estimate the number of active degrees of freedom. To this purpose, the correlation sum that counts the number of pairs of vectors in state space that are closer than a given hypersphere radius ε is calculated as a function of ε

$$C(\varepsilon) = \frac{2}{(M-W)(M-W-1)} \sum_{i=1}^{M} \sum_{j=i+W}^{M} \Theta(\varepsilon - |\vec{x}_i - \vec{x}_j|)$$

where $|\cdot|$ indicates the maximum norm in m dimensions and Θ is the Heaviside step function ($\Theta(a) = 0$ for $a \leq 0$ and $\Theta(a) = 1$ for $a > 0$). The exclusion of pairs closer in time than the length of the so-called Theiler window W is essential to reduce the unwanted influence of temporal correlations on $C(\varepsilon)$ (Theiler, 1986). From the local slope of the correlation sum $\mathrm{d}(\varepsilon) = \mathrm{d} \ln C(\varepsilon)/\mathrm{d} \ln \varepsilon$, the correlation dimension is defined as $D_2 = \lim_{N\to\infty} \lim_{\varepsilon\to 0} \mathrm{d}(\varepsilon)$. From the limites it follows that the calculation of the correlation dimension would require an infinite length N and an unlimited accuracy of the time series. However, an estimate of an *effective* correlation dimension (Grassberger et al., 1991) can be obtained if an almost constant value of $\mathrm{d}(\varepsilon)$ is found at least for a range of ε values, the so-called quasi-scaling region.

Here, we calculate $\mathrm{d}(\varepsilon)$ for embedding dimensions of $m = 1$ and $m = 25$ using a fixed time delay ($\tau = 1 \cdot \Delta t$) and Theiler window ($W = 5 \times \Delta t$). The range of ε is chosen to match the resolution of the analog-to-digital converters and is divided into 128 intervals. A quasi-scaling region $[\varepsilon_l, \varepsilon_u]$ is defined by $\varepsilon_u = \max\{\varepsilon | \mathrm{d}(\varepsilon)_{|m=1} > 0.975\}$ and $\varepsilon_l = \min\{\varepsilon |\ |\mathrm{d}(\varepsilon_u)_{|m=25} - \mathrm{d}(\varepsilon)_{|m=25}| \leq 0.05\ \mathrm{d}(\varepsilon)_{|m=25} \wedge \varepsilon < \varepsilon_u\}$. If ε_u and ε_l exist and the number n_r of ε values in $[\varepsilon_l, \varepsilon_u]$ is greater than 4, the estimate $D^* = (1/n_r) \cdot \sum_{\varepsilon=\varepsilon_l}^{\varepsilon_u} \mathrm{d}(\varepsilon)_{|m=25}$ is computed (cf. Lehnertz and Elger, 1998). If no quasi-scaling behavior is found for $\mathrm{d}(\varepsilon)$ or if $D^* \geq 9.5$, an arbitrary but fixed value of $D^* = 10$ is set (cf. Lehnertz and Elger, 1995).

A.2.2. Largest Lyapunov exponent

The exponential divergence or convergence of nearby trajectories in state space is conceptually the most basic

indicator of deterministic chaos and can be sufficiently estimated using the largest Lyapunov exponent $L_{\max}$. The first proposed algorithm to compute $L_{\max}$ from a time series (Wolf et al., 1985) suffers from severe drawbacks that occur particularly with short and noisy time series, strongly depends on parameters used for the state space reconstruction, and is computationally highly expensive (Rosenstein et al., 1993). In order to avoid these shortcomings we here use a combination of improved algorithms (Kantz, 1994; Rosenstein et al., 1993) according to which the $L_{\max}$ can be estimated from $d_j(i) \approx C_j \, e^{L_{\max} \cdot i \cdot \Delta t}$ where $d_j(i)$ denotes the average divergence between two trajectory segments at time t_i. C_j with $j=1,\ldots,M$ is a constant that is given by the initial separation of a reference vector $\vec{z}_j$ in state space and its nearest neighbor. In order to improve statistics we follow Kantz (1994) and search for *all* neighbors starting within a hypersphere of radius ε around $\vec{z}_j$ using a box-assisted algorithm (Schreiber, 1995). Based on the relation

$$\ln d_j(i) \approx \ln C_j + L_{\max} \cdot i \cdot \Delta t$$

the largest Lyapunov exponent is then calculated using a least-squares fit to an average line defined by $y(i) = (1/\Delta t)\langle \ln d_j(i) \rangle$, where $\langle \cdots \rangle$ denotes the average over all values of j.

We here use a linear regression to the longest possible linear region with a goodness-of-fit of 0.001. For our analyses we estimate $L_{\max}$ using an embedding dimension of $m=7$ and a fixed time delay of $\tau = 5 \cdot \Delta t$. In order to reduce the unwanted influence of temporal correlations we follow Rosenstein et al. (1993) and choose a Theiler window of a length given by the reciprocal of the mean frequency of the power spectrum.

A.2.3. Local flow Λ^*

The local flow, a measure derived from the coarse-grained flow average (Kaplan and Glass, 1992), aims at discriminating deterministic from stochastic dynamics. For this technique, the reconstructed m-dimensional state space is divided into b^m non-overlapping hyper-cubes. If the hyper-cube with index j is passed n_j times by the trajectory, a normalized vector $v_{j,k}$ will be generated for each pass ($k=1,\ldots,n_j$) whose direction is determined by connecting the points where the trajectory enters and leaves the hyper-cube. Summing up all vectors of passes through hyper-cube j, the resultant vector V_j, normalized by the number of passes n_j, is $V_j = (1/n_j) \sum_{k=1}^{n_j} v_{j,k}$. The coarse-grained flow average Λ is then defined as

$$\Lambda = \sum_j \frac{V_j^2 - R^2}{1 - R^2}$$

with $R \propto (1/\sqrt{n})$ being the expected value for a vector addition of n vectors of unit length yielded by a random walk in m dimensions.

Using an embedding dimension of $m=6$, the number of hyper-cubes per state space axis is determined from the range and the variance of the time series according to $b = 0.875 \cdot (\max\{x_i\} - \min\{x_i\})/\sigma^2$ resulting in values of b from 6 to 20. Rather than using a fixed time delay τ the local flow is determined by summing up the coarse-grained flow average for different values of τ in accordance with (Andrzejak et al., 2001):

$$\Lambda^* = \sum_{\tau=5}^{20} \Lambda(\tau)$$

A.2.4. Algorithmic complexity

Another approach to characterize time series is based on the theory of symbolic dynamics (Hao, 1989). For this approach, the time series is transformed into a symbol sequence by partitioning the range of sampling values and assigning a different symbol S to each interval of this binning. Then each value of the time series is replaced by the symbol of its interval. The thresholds of the partition are chosen separately for each time series to yield a homogenous distribution of symbols. To achieve a good statistics as well as a good representation of the time series, the number of different symbols was set to $A=16$. The resulting symbol sequence $\{S_i\}$ with $i=1,\ldots,N$ was then investigated for its complexity by estimating the size $c(\{S_i\})$ of its vocabulary. This size was defined as the number of different words in a Lempel–Ziv parsing (Lempel and Ziv, 1976) of the symbol sequence. In this algorithm, the symbol sequence is scanned from the beginning to its end, and its complexity $c(\{S_i\})$ is increased by one unit as soon as a new subsequence of consecutive symbols is encountered in the scanning process (Kasper and Schuster, 1987) and the following symbol is regarded as the beginning of the next symbol sequence. This value is normalized by the expected asymptotic value for a random sequence of symbols of length N to yield the algorithmic complexity:

$$\mathrm{AC} = \frac{\log_A N}{N} c(\{S_i\})$$

A.2.5. Surrogate time series and surrogate correction

The method of surrogate time series allows to test a specified null hypothesis about the dynamics underlying a given time series. For this purpose, an ensemble of surrogate time series is constructed from the original time series in such a way that the surrogates have all properties included in the null hypothesis in common with the original, but are otherwise random. Then a certain measure, which has to be sensitive to at least one property that is not included in the null hypothesis, e.g. non-linearity (cf. Theiler et al., 1992), is calculated for the original and the surrogates. If the result for the original time series deviates significantly from the distribution of the surrogates, the null hypothesis can be rejected. The probability of false rejections, i.e. the nominal size of the test, is adjustable by the number of surrogates. For the present study, we applied a technique for generating iterative amplitude-adjusted surrogates (Schreiber and

Schmitz, 1996). These types of surrogates allow testing of the null hypothesis that the time series was measured from a Gaussian linear stochastic and stationary dynamics by means of a static and invertible but possibly non-linear measurement function. Starting from a random permutation of the original amplitudes of the time series, the surrogates are constructed by an iterative algorithm that alternately adjusts the power spectrum and the amplitude distribution to the original values, resulting in a deviation of the respective other quantity. After a sufficient number of iterations (typically 20–50) deviations of both quantities from values of the original time series will be reduced to negligibly small values.

Here we define

$$S\text{–}NM = \begin{cases} NM_{EEG} - \overline{NM_{SUR}} & if \quad NM_{EEG} - \overline{NM_{SUR}} > 0 \\ 0 & else \end{cases}$$

where NM is a placeholder for any non-linear univariate measure and the over-bar denotes mean values of four surrogates. This equation is valid for measures which reflect non-linear determinism by high values (Λ^*, AC). For the other measures (D^*, L_{max}), which reflect non-linear determinism by low values, the signs in the equation should be inverted to ensure that all four surrogate-corrected measures attain positive values for nonlinear deterministic dynamics. Note that this use of surrogates differs from their original purpose (Theiler et al., 1992) since taking these differences can be regarded as an 'offset correction' rather than as a hypothesis test, where the offset is given by linear properties of the dynamics. Surrogate correction was applied to all previous non-linear univariate measures, resulting in the corresponding measures *S*–D^*, *S*–L_{max}, *S*–Λ^*, and *S*–AC.

A.2.6. Loss of recurrence

The loss of recurrence LR is used to quantify the degree of non-stationarity within a time series (Rieke et al., 2002). This measure analyzes the distribution of distances in time between reference vectors and their neighboring vectors in state space. A system is regarded as stationary if the time index of a neighbor is statistically independent from that of the reference. For non-stationary systems we expect the absence of distant time indices in the neighborhood of the reference, i.e. a *loss of recurrence*.

Let $U_\varepsilon(\vec{x}_r) = \{\vec{x}_n | \|\vec{x}_r - \vec{x}_n\| \le \varepsilon\}$ define a set of vectors in the ε-neighborhood of $\vec{x}_r$ in an m-dimensional reconstructed state space. We here use $m=10$ and define ε_r in dependence of the reference $\vec{x}_r$ using a fixed number $k=4$ of nearest neighbors and the maximum-norm as a metric. The lags $l_r^i = |n_r^i - r|$ of the ith nearest neighbor of $\vec{x}_r$ are transformed using the distribution function $\tilde{l} = \Phi_r(l)$, i.e. the a priori probability under stationary conditions that the observed distance in time is less than or equal to l. The distribution $f(\tilde{l})$ of all transformed time distances $\tilde{l}_n$ reflects the non-stationarity of the system. For a stationary system $f(\tilde{l})$ is uniformly distributed in the interval [0,1], and the median μ equals 0.5, whereas in case of non-stationarity, the recurrence of related state space vectors is reduced, i.e. the neighborhood of $\vec{x}_r$ depends on the time index r and furthermore the indices of the neighboring vectors n_r^i are clustered around r. For non-stationary signals, the observed distances in time are therefore on average smaller than expected and thus the amount of lower values $\tilde{l}$ is increased whereas higher values are reduced, and the median μ of this distribution $f(\tilde{l})$ is less than 0.5. There is no need for a surrogate correction of this measure as surrogate time series are stationary by construction and the median μ of the distribution $f(\tilde{l})$ always matches 0.5 up to statistical fluctuations.

A.3. Bivariate linear measures

A.3.1. Maximum linear cross-correlation

In order to quantify the similarity of two signals $\{x_i\}$ and $\{y_i\}$ we use the maximum of a normalized cross-correlation function as a measure for lag synchronization (Rosenblum et al., 1997):

$$C_{max} = \max_\tau \left\{ \left| \frac{C(x,y)(\tau)}{\sqrt{C(x,x)(0) \cdot C(y,y)(0)}} \right| \right\}$$

where

$$C(x,y)(\tau) = \begin{cases} \dfrac{1}{N-\tau} \displaystyle\sum_{i=1}^{N-\tau} x_{i+\tau} y_i & \tau \ge 0 \\ C(y,x)(-\tau) & \tau < 0 \end{cases}$$

is the well-known linear cross-correlation function. C_{max} is confined to the interval [0, 1] with high values indicating that the two signals have a similar course in time (though possibly shifted by a time lag τ) while dissimilar signals will result in values close to zero.

A.4. Bivariate non-linear measures

A.4.1. Non-linear interdependence

The non-linear interdependence (Arnhold et al., 1999) as a measure for generalized synchronization (Rulkov et al., 1995) between two EEG signals $\{x_i\}$ and $\{y_i\}$ is calculated after reconstruction of the state space trajectories $\{\vec{x}_i\}$ and $\{\vec{y}_i\}$ for these signals. Let α_{ij} and β_{ij} with $j=1,\ldots,k$ denote the time indices of the k nearest neighbors in state space of $\vec{x}_i$ and $\vec{y}_i$, respectively. For each $\vec{x}_i$ the squared mean Euclidean distance to its k nearest neighbors is given by

$${}^{x}R_i^{(k)} = \frac{1}{k} \sum_{j=1}^{k} (\vec{x}_i - \vec{x}_{\alpha_{ij}})^2$$

while the *y-conditioned* mean squared Euclidean distance is constructed by replacing the nearest neighbors by

the equal time partners of the closest neighbors of $\vec{y}_i$:

$$ {}^{x|y}R_i^{(k)} = \frac{1}{k}\sum_{j=1}^{k}(\vec{x}_i - \vec{x}_{\beta_{ij}})^2. $$

${}^{y}R_i^{(k)}$ and ${}^{y|x}R_i^{(k)}$ are defined accordingly.

As measures for non-linear interdependence we use

$$ {}^{x|y}S = \frac{1}{M}\sum_{i=1}^{M}\frac{{}^{x}R_i^{(k)}}{{}^{x|y}R_i^{(k)}} $$

and

$$ {}^{x|y}H = \frac{1}{M}\sum_{i=1}^{M}\log\frac{{}^{x}R_i^{(M)}}{{}^{x|y}R_i^{(k)}} $$

with

$$ {}^{x}R_i^{(M)} = \frac{1}{M-1}\sum_{j=1,j\neq i}^{M}(\vec{x}_i - \vec{x}_j)^2. $$

${}^{y|x}S$ and ${}^{y|x}H$ are defined accordingly. Both measures yield high values for high degrees of non-linear interdependence and values near zero for independent time series. While ${}^{x|y}S$ is restricted to the interval [0, 1], ${}^{x|y}H$ is not normalized and might also have slightly negative values. In order to have the same number of channel combinations as for the other bivariate measures, we here use a symmetrized version of the two measures for non-linear interdependence:

$$ S = \frac{{}^{x|y}S + {}^{y|x}S}{2} \quad \text{and} \quad H = \frac{{}^{x|y}H + {}^{y|x}H}{2} $$

A.4.2. Measures for phase synchronization

Phase synchronization (Huygens, 1673) is traditionally defined as *phase locking* ($\phi_x(t)-\phi_y(t)=\text{const}$) or, in the case of noisy and/or chaotic systems (Rosenblum et al., 1996), as *phase entrainment* ($\phi_x(t)-\phi_y(t)<\text{const}$), with $\phi_x(t)$ and $\phi_y(t)$ denoting the phase variables of two oscillating signals $x(t)$ and $y(t)$.

In this study we use three different measures for phase synchronization. The first measure, the *mean phase coherence* (Hoke et al., 1988; Lachaux et al., 1999; Mormann et al., 2000) is defined as

$$ R = \left|\frac{1}{N}\sum_{j=1}^{N} e^{i[\phi_x(t_j)-\phi_y(t_j)]}\right| $$

The second and third measure are termed *index based on conditional probability* and *index based on Shannon entropy* (Tass et al., 1998). For these measures, an equidistant binning of the interval $[0, 2\pi]$ is required where the number of bins is given by $L = e^{0.626+0.4\ln(N-1)}$ as in (Rosenblum et al., 2001).

The index based on conditional probability is then defined as

$$ \lambda_{cp} = \frac{1}{L}\sum_{l=1}^{L}|r_l| $$

where

$$ r_l = \frac{1}{M_l}\sum_{j} e^{i\phi_y(t_j)} $$

$$ \phi_x(t_j) \in [(l/L)2\pi, ((l+1)/L)2\pi] $$

with

$$ M_l = \left|\left\{\phi_x(t_j)|\phi_x(t_j) \in \left[\frac{1}{L}2\pi, \frac{l+1}{L}2\pi\right]\right\}\right| $$

denoting the number of phase values $\phi_x(t)$ contained in the bin l. ($|\{\cdots\}|$ denotes the number of elements contained in the set $\{\cdots\}$.)

The index based on Shannon entropy is given by

$$ \rho_{se} = 1 + \frac{1}{\ln L}\sum_{l=1}^{L} p_l \ln p_l $$

with

$$ p_l = \frac{\left|\left\{\phi_x(t_j)-\phi_y(t_j)|\phi_x(t_j)-\phi_y(t_j) \in \left[\frac{l}{L}2\pi, \frac{l+1}{L}2\pi\right]\right\}\right|}{|\{\phi_x(t_j)-\phi_y(t_j)\}|} $$

denoting the relative frequency of finding a phase difference in a certain bin l.

All three phase synchronization measures are confined to the interval [0, 1] where high values indicate a high degree of phase synchronization and low values correspond to unsynchronized signals.

In order to measure changes in phase synchronization of two signals $x(t)$ and $y(t)$ over time, it is necessary to determine their phases $\phi_x(t)$ and $\phi_y(t)$. To this aim, two different approaches have been proposed.

One is the *analytic signal* approach (Gabor, 1946; Panter, 1965) which defines an *instantaneous phase*

$$ \phi(t) = \arctan\frac{\tilde{s}(t)}{s(t)} $$

for an arbitrary signal $s(t)$ using the *Hilbert Transform*

$$ \tilde{s}(t) = \frac{1}{\pi}\mathrm{pv}\int_{-\infty}^{+\infty}\frac{s(t')}{t-t'}\mathrm{d}t' $$

(pv denoting the Cauchy principal value).

The second approach (Lachaux et al., 1999) uses a definition based on the Wavelet Transform. Here the phase variable is defined as

$$ \phi(t) = \arctan\frac{\mathrm{Im}\ W(t)}{\mathrm{Re}\ W(t)} $$

using the Wavelet coefficients

$$ W(t) = \int_{-\infty}^{+\infty}\Psi(t-t')s(t')\mathrm{d}t' $$

of a complex Morlet Wavelet

$$ \Psi(t) = (e^{i\omega_0 t} - e^{\omega_0^2\sigma^2/2})e^{-t^2/2\sigma^2} $$

where $\omega_0 = 3$ Hz is the center frequency and σ the decay rate of the wavelet determined by setting the number of cycles $nc = 6\omega_0\sigma$ to a value of 3.

Note that in a recent study it has been shown that both definitions of a phase are mathematically related to each other (Quian Quiroga et al., 2002).

References

Andrzejak RG, Widman G, Lehnertz K, Rieke C, David P, Elger CE. The epileptic process as nonlinear deterministic dynamics in a stochastic environment—an evaluation of mesial temporal lobe epilepsy. Epilepsy Res 2001;44:129–40.

Andrzejak RG, Mormann F, Kreuz T, Rieke C, Kraskov A, Elger CE, Lehnertz K. Testing the null hypothesis of the non-existence of the pre-seizure state. Phys Rev E 2003;67:010901.

Arnhold J, Grassberger P, Lehnertz K, Elger CE. A robust method for detecting interdependencies: application to intracranially recorded EEG. Physica D 1999;134:419–30.

Aschenbrenner-Scheibe R, Maiwald T, Winterhalder M, Voss HU, Timmer J, Schulze-Bonhage A. How well can epileptic seizures be predicted? An evaluation of a nonlinear method. Brain 2003;126: 2616–26.

Blanco S, Garcia H, Quian Quiroga R, Romanelli L, Rosso OA. Stationarity of the EEG series. IEEE Eng Med Biol 1995;4:395–9.

Duckrow RB, Spencer SS. Regional coherence and the transfer of ictal activity during seizure onset in the medial temporal lobe. Electroencephalogr Clin Neurophysiol 1992;82:415–22.

Elger CE. Future trends in epileptology. Curr Opin Neurol 2001;14:185–6.

Elger CE, Lehnertz K. Seizure prediction by non-linear time series analysis of brain electrical activity. Eur J Neurosci 1998;10:786–9.

Gabor D. Theory of communication. Proc IEEE London 1946;93:429–57.

Grassberger P, Procaccia I. Characterization of strange attractors. Phys Rev Lett 1983;50:346–9.

Grassberger P, Schreiber T, Schaffrath C. Nonlinear time sequence analysis. Int J Bifurc Chaos 1991;1:521–47.

Hao B-L. Elementary symbolic dynamics and chaos in dissipative systems. Singapore: World Scientific; 1989.

Hjorth B. EEG analysis based on time domain properties. Electroencephalogr Clin Neurophysiol 1970;29:306–10.

Hoke M, Lehnertz K, Pantev C, Lütkenhöner B. Spatiotemporal aspects of synergetic processes in the auditory cortex as reavealed by the magnetoencephalogram. In: Basar E, Bullock T, editors. Dynamics of cognitive and sensory processing in the brain. Berlin, Heidelberg, New York: Springer; 1988. p. 84–105.

Huygens C. Horoloquium Oscilatorium. Parisis, France: Apud F. Muguet; 1673 [engl. translation: The Pendulum Clocks, Iowa State University Press, Ames, 1986].

Iasemidis LD, Sackellares JC. The evolution with time of the spatial distribution of the largest Lyapunov exponent on the human epileptic cortex. In: Duke D, Pritchard W, editors. Measuring chaos in the human brain. Singapore: World Scientific; 1991. p. 49–82.

Iasemidis LD, Pardalos P, Sackellares JC, Shiau DS. Quadratic binary programming and dynamical system approach to determine the predictability of epileptic seizures. J Comb Optimization 2001;5:9–26.

Kantz H. A robust method to estimate the maximal Lyapunov exponent of a time series. Phys Lett A 1994;185:77–87.

Kantz H, Schreiber T. Nonlinear time series analysis. Cambridge: Cambridge University Press; 1997.

Kaplan DT, Glass L. Direct test for determinism in a time series. Phys Rev Lett 1992;68:427–30.

Kasper F, Schuster HG. Easily calculable measure for the complexity of spatio-temporal patterns. Phys Rev A 1987;36:842–8.

Kreuz T, Andrzejak RG, Mormann F, Kraskov A, Stögbauer H, Elger CE, Lehnertz K, Grassberger P. Measure profile surrogates: a new method to validate the performance of epileptic seizure prediction algorithms. Phys Rev E 2004;69:061915.

Lachaux JP, Rodriguez E, Martinerie J, Varela FJ. Measuring phase synchrony in brain signals. Hum Brain Mapp 1999;8:194–208.

Le Van Quyen M, Martinerie J, Baulac M, Varela F. Anticipating epileptic seizure in real time by a nonlinear analysis of similarity between EEG recordings. NeuroReport 1999;(10):2149–55.

Le Van Quyen M, Adam C, Martinerie J, Baulac M, Clémenceau S, Varela F. Spatio-temporal characterization of non-linear changes in intracranial activities prior to human temporal lobe seizures. Eur J Neurosci 2000;12:2124–34.

Le Van Quyen M, Martinerie J, Navarro V, Boon P, D'Have M, Adam C, Renault B, Varela F, Baulac M. Anticipation of epileptic seizures from standard EEG recordings. Lancet 2001;357:183–8.

Lehnertz K, Elger CE. Spatio-temporal dynamics of the primary epileptogenic area in temporal lobe epilepsy characterized by neuronal complexity loss. Electroencephalogr Clin Neurophysiol 1995;95:108–17.

Lehnertz K, Elger CE. Neuronal complexity loss in temporal lobe epilepsy: effects of carbamazepine on the dynamics of the epileptogenic focus. Electroencephalogr Clin Neurophysiol 1997;103:376–80.

Lehnertz K, Elger CE. Can epileptic seizures be predicted? Evidence from nonlinear time series analysis of brain electrical activity Phys Rev Lett 1998;80:5019–23.

Lempel A, Ziv J. On the complexity of finite sequences. IEEE Trans Inform Theory 1976;22:75–81.

Litt B, Esteller R, Echauz J, D'Alessandro M, Shor R, Henry T, Pennell P, Epstein C, Bakay R, Dichter M, Vachtsevanos G. Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. Neuron 2001;30:51–64.

Lopes da Silva FH. EEG analysis: theory and practice. In: Niedermayer E, Lopes da Silva FH, editors. Electroencephalography, basic principles, clinical applications and related fields. Baltimore, MD: Urban and Schwarzenberg; 1987. p. 871–97.

Maiwald T, Winterhalder M, Aschenbrenner-Scheibe R, Voss HU, Schulze-Bonhage A, Timmer J. Comparison of three nonlinear prediction methods by means of the seizure prediction characteristic. Physica D 2004;194:357–68.

Martinerie J, Adam C, Le Van Quyen M, Baulac M, Clemenceau S, Renault B, Varela FJ. Epileptic seizures can be anticipated by non-linear analysis. Nat Med 1998;4:1173–6.

Mormann F, Lehnertz K, David P, Elger CE. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. Physica D 2000;144:358–69.

Mormann F, Kreuz T, Andrzejak RG, David P, Lehnertz K, Elger CE. Epileptic seizures are preceded by a decrease in synchronization. Epilepsy Res 2003a;53:173–85.

Mormann F, Andrzejak RG, Kreuz T, Rieke C, David P, Elger CE, Lehnertz K. Automated detection of a pre-seizure state based on a decrease in synchronization in intracranial EEG recordings from epilepsy patients. Phys Rev E 2003b;67:021912.

Navarro V, Martinerie J, Le Van Quyen M, Clemenceau S, Adam C, Baulac M, Varela F. Seizure anticipation in human neocortical partial epilepsy. Brain 2002;125:640–55.

Ott E. Chaos in dynamical systems. Cambridge: Cambridge University Press; 1993.

Panter P. Modulation, noise, and spectral analysis. New York: McGraw-Hill; 1965.

Pereda E, Gamundi A, Nicolau MC, Rial R, Gonzalez J. Interhemispheric differences in awake and sleep human EEG: a comparison between non-linear and spectral measures. Neurosci Lett 1999;263:37–40.

Quian Quiroga R, Kraskov A, Kreuz T, Grassberger P. Performance of different synchronization measures in real data: a case study on electroencephalographic signals. Phys Rev E 2002;65:041903.

Rajna P, Clemens B, Csibri E, Dobos E, Geregely A, Gottschal M, Gyorgy I, Horvath A, Horvath F, Mezofi L, Velkey I, Veres J, Wagner E. Hungarian multicentre epidemiologic study of the warning and initial symptoms (prodrome, aura) of epileptic seizures. Seizure 1997;6:361–8.

Rieke C, Sternickel C, Andrzejak RG, Elger CE, David P, Lehnertz K. Measuring nonstationarity by analyzing the loss of recurrence in dynamical systems. Phys Rev Lett 2002;88:244102.

Rogowski Z, Gath I, Bental E. On the prediction of epileptic seizures. Biol Cybern 1981;42:9–15.

Rosenblum MG, Pikovsky AS, Kurths J. Phase synchronization of chaotic oscillators. Phys Rev Lett 1996;76:1804–7.

Rosenblum MG, Pikovsky AS, Kurths J. From phase to lag synchronization in coupled chaotic oscillators. Phys Rev Lett 1997;78:4193–6.

Rosenblum M, Pikovsky A, Kurths J, Schäfer C, Tass PA. Phase synchronization: from theory to data analysis. In: Moss F, Gielen S, editors. Handbook of biological physics. Amsterdam: Elsevier; 2001. p. 279–321.

Rosenstein MT, Collins JC, De Luca CJ. A practical method for calculating the largest Lyapunov exponents from small data sets. Physica D 1993; 65:117–34.

Rulkov NF, Sushchik MM, Tsimring LS, Ababarnel HDI. Generalized synchronization of chaos in directionally coupled chaotic systems. Phys Rev E 1995;51:980–94.

Schreiber T. Efficient neighbor searching in nonlinear time series analysis. Int J Bifurc Chaos 1995;5:349–58.

Schreiber T, Schmitz A. Improved surrogate data for nonlinearity tests. Phys Rev Lett 1996;77:635–8.

Schuster HG. Deterministic chaos: an introduction. 2nd revised version. Weinheim: VCH; 1989.

Stanski DR, Hudson RJ, Homer TD, Saidman LJ, Meathe E. Pharmacodynamic modeling of thiopental anesthesia. J Pharmacokinet Biopharm 1984;12:223–40.

Takens F. Detecting strange attractors in turbulence. In: Rand DA, Young LS, editors. Dynamical systems and turbulence lecture notes on mathematics 898. New York: Springer; 1981. p. 366–81.

Tass P, Rosenblum MG, Weule J, Kurths J, Pikovsky AS, Volkmann J, Schnitzler A, Freund HJ. Detection of n:m phase locking from noisy data: application to magnetoencephalography. Phys Rev Lett 1998;81: 3291–4.

Theiler J. Spurious dimension from correlation algorithms applied to limited time series data. Phys Rev A 1986;34:2427–33.

Theiler J, Eubank St, Longtin A, Galdrikian B, Farmer JD. Testing for nonlinearity in time series: the method of surrogate. Physica D 1992;58: 77–94.

Whitney H. Differentiable manifolds. Ann Math 1936;37:645–80.

Winterhalder M, Maiwald T, Voss HU, Aschenbrenner-Scheibe R, Timmer J, Schulze-Bonhage A. The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. Epilepsy Behav 2003;4:318–25.

Wolf A, Swift JB, Swinney L, Vastano A. Determining Lyapunov exponents from a time series. Physica D 1985;16:285–317.