Asian Journal of Information Technology 3(12), 1165-1171, 2004

Quantification of Nonstationary Structure in High-dimensional Time Series

Andreas Galka[1,2,*], Heiko Hansen[3], Tohru Ozaki[2] and Gerd Pfister[1]

[1] *Institute of Experimental and Applied Physics, University of Kiel, 24098 Kiel, Germany*

[2] *The Institute of Statistical Mathematics, 4–6–7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan*

[3] *Max Planck Institute of Meteorology, Bundesstr. 55, 20146 Hamburg, Germany*

Abstract: We consider the problem of detecting and quantifying nonstationary structure in time series from high-dimensional dynamical systems. This problem is relevant in particular for EEG monitoring, e.g. for the prediction of epileptic seizures, but also for practical data analysis in many other fields. Three groups of measures of nonstationarity are discussed: Correlation dimension, measures based on autoregressive modelling and cross-prediction, and measures based on entropies defined in the spectral or wavelet domains. Results both for simulated and clinical time series are shown.

1. Introduction

As a part of the current interest in characterisation of complex dynamical behaviour the detection and description of nonstationary dynamics has become an important issue (Havstad & Ehlers (1989), Schreiber (1997), West *et al.* (1999), Hegger *et al.* (2000), Mormann *et al.* (2003)). Although in some laboratory experiments it may be possible to observe stationary dynamical states, in observations taken from nature it is usually impossible to separate the object of investigation from unpredictable external influences. Even in well-controlled laboratory experiments unwanted external disturbances may occur which have to be detected by appropriate tools.

There are also systems which might actually evolve according to a stationary dynamics, but for which the necessary state space dimension is too high for reliable identification of such dynamics from time series of reasonable size (and precision); this effect renders them operationally nonstationary.

As a consequence of this, during the analysis of many time series e.g. from medicine, meteorology or economics one has to consider the possibility of nonstationarity, be it genuine or operational.

In various cases the nonstationarity may even be the main object of interest. As an example we mention the work of Lehnertz & Elger (1998, see also Mormann *et al.*, 2003) on the prediction of epileptic seizures from intracranial recordings of the human electroencephalogram (EEG). From estimations of the correlation dimension d_2 based on a moving window they were in several cases able to detect a significant change of brain dynamics several minutes before a seizure occurred.

Characterisation of nonstationarity may be accomplished by explicitly using nonstationary time series models; however, within the framework of nonlinear time series analysis, as developed since the early 1980's in the nonlinear dynamics community (Kantz & Schreiber, 1997) it is frequently preferred to describe nonstationarity by monitoring significant changes of characteristic quantities describing the dynamics, such as fractal dimensions or Lyapunov exponents (Galka, 2000).

In this paper we take a heuristic approach to comparing three approaches for detecting and quantifying nonstationary structure in three given univariate time series. No prior knowledge about the underlying dynamics will be assumed, thereby imitating a common situation in practical data analysis.

2. Description of time series

We will employ three time series, two of which are artificially generated in a way such that their (non-)stationarity properties are precisely known, whereas the third time series represents a clinical recording of brain activity.

The first two time series are obtained by integrating the well-known Mackey-Glass differential-delay equation (Mackey & Glass, 1977)

$$\dot{x}(t) = \frac{ax(t-\tau_d)}{1 + x(t-\tau_d)^c} - bx(t) \quad , \tag{1}$$

where the parameters are chosen as $a = 0.2$, $b = 0.1$ and $c = 10$. In this system the change of the state $\dot{x}(t)$ depends not only on the current state $x(t)$, but also on the state a certain *delay time* τ_d ago. The larger τ_d, the more complex the resulting time series will be, but still the time series remains completely deterministic. This equation is solved numerically by using the same discretisation technique as used by Ding *et al.* (1993). As initial conditions on the interval $[0, \tau_d]$ we choose a constant value of 0.5; but we allow transients in the numerical solution to die out before we begin to sample actual time series. The integration time is chosen as $t_{int} = 0.1$, the sampling time is 20 integration times.

* Corresponding author. Email: andreas@ism.ac.jp

We create two time series from the Mackey-Glass system, each of a total length of 131072 points:

- **Time series A**: The delay time is chosen as $\tau_d = 65$, so that the time series is perfectly stationary. This value corresponds already to a fairly complex dynamics, which can be described by a correlation dimension of $d_2 \approx 5.3$ (Galka, 2000). A typical part of this time series is shown in figure 1.
- **Time series B**: Now we let τ_d depend explicitly on (integration) time during the numerical integration process by a triangle-shaped oscillating function: We start at $\tau_d = 10$ and increase it slowly at constant rate, until $\tau_d = 120$ is reached, then we switch to decreasing the delay time at the same constant rate, until $\tau_d = 10$ is reached, where we switch to increasing again, etc. We choose the rate of changing τ_d such that precisely four full periods of this triangle-shaped oscillation fit into the 131072 points of the entire time series. By this method we create a nonstationary time series for which the quantitative evolution of the nonstationarity is well known in advance. The complexity of the dynamics will be lowest for $\tau_d = 10$ (where the time series in fact becomes periodic) and highest for $\tau_d = 120$. A typical part of this time series is shown in figure 2, from which it can be seen how the periodic dynamics becomes more complex due to the gradual increase of τ_d.

As an example of a time series from actual reality we furthermore employ a clinical time series:

- **Time series C** was recorded by an implanted depth electrode directly from a specific brain region, the gyrus parahippocampus, of a patient suffering from severe temporo-mesial epilepsy. The data set was sampled with 183.8 Hz and covers 25 minutes directly prior to an epileptic seizure. According to results obtained by Lehnertz & Elger (1998) this time series displays a marked change of the dynamics 8 minutes prior to the seizure. A typical part of this time series, chosen from within the last 8 minutes, is shown in figure 3. Earlier parts of the series look quite similar, except for a slightly smaller number of spikes.

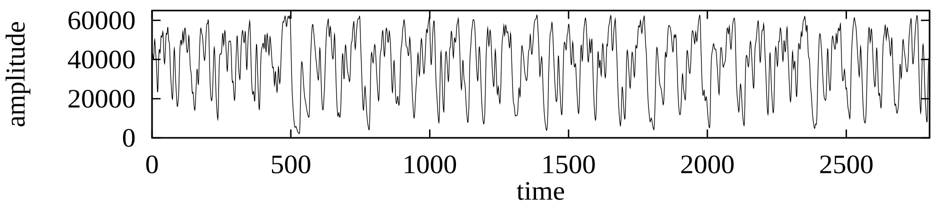

Figure 1: *Part of time series A (2800 points; total length 131072 points), stationary Mackey-Glass system. Amplitudes correspond to a 16bit data format. For more details see text.*

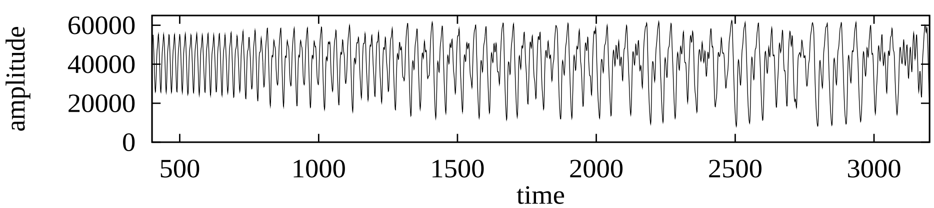

Figure 2: *Part of time series B (2800 points; total length 131072 points), nonstationary Mackey-Glass system, chosen briefly after the point of lowest complexity. Amplitudes correspond to a 16bit data format. For more details see text.*

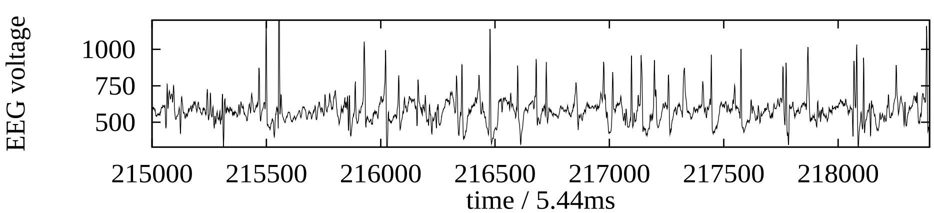

Figure 3: *Part of time series C (3400 points; total length 270000 points), intracranial EEG recording. Amplitudes correspond directly to the output of 12bit A/D conversion. For more details see text.*

3. Correlation dimension estimation

As our first approach to quantifying nonstationarity we present correlation dimension estimation on a "moving window", as employed by Lehnertz & Elger (1998): A series of segments is extracted from the time series, where consecutive segments may overlap to some degree, then the correlation dimension is estimated for each segment, yielding a single value for each segment.

Such moving-window approaches have been criticised, since they may be statistically inefficient and anticipate quasi-stationarity within each segment, and modelling approaches have been proposed that do not employ segments (Krystal *et al.* (1998), Hegger *et al.* (2000)); however, so far these approaches lack generality, since they rely on specific assumptions about the dynamics or the structure of the nonstationarity, such as assuming that it can be described by a random walk, therefore we prefer to employ the standard moving-window approach. We shall give only a brief description of correlation dimension estimation here, since this technique is widely used and well documented elsewhere (see e.g. Kantz & Schreiber (1997), Galka (2000)).

Given a (univariate) time series $x_i,\ i = 1,\dots N$ (which in this case is limited to the data within a given segment), the corresponding state space is reconstructed by time-delay embedding, which yields a set of vectors $\mathbf{x}_i = (x_i, x_{i-\tau}, \dots, x_{i-(m-1)\tau})$, where m and τ denote the embedding dimension and the time delay, respectively. In the case of our Mackey-Glass time series an appropriate value for τ is 3, as determined by the ILD criterion of Buzug & Pfister (1992); for the EEG time series we obtain $\tau = 5$ by the same criterion. The embedding dimension m is chosen to range from 2 to 20. For each value of m the correlation sum is calculated:

$$C(r,m) = \frac{2}{(N_v + 1 - W)(N_v - W)} \sum_{i=W+1}^{N_v} \sum_{j=1}^{N_v - i} H(r - \|\mathbf{x}_j - \mathbf{x}_{i+j}\|) \ , \tag{2}$$

where $H(x) = 1$ for positive x and $H(x) = 0$ otherwise. $\|.\|$ denotes maximum norm. The bias correction parameter W (also known as Theiler correction parameter) is chosen as $W = 20$; its choice is not critical, provided it is sufficiently large. Then the *radius-dependent* correlation dimension estimate follows from $C(r,m)$ by

$$d_2(r,m) = \frac{\partial \log C(r,m)}{\partial \log r} \ . \tag{3}$$

A radius-independent estimate (which still depends on m) is obtained by evaluating the slope of $\log C(\log r)$ on a suitably chosen radius interval (see Galka (2000) for technical details of the implementation).

Finally one has to study the behaviour of $d_2(m)$; if it converges to a constant value for increasing m, this value is accepted as an estimate of the correlation dimension. But in this application we simply plot the values of $d_2(m)$ as a function of segment number (i.e. time) and omit this final step, thereby obtaining a generalised measure of complexity even in cases where a proper convergence of $d_2(m)$ for increasing m is not found. This procedure has been found to be useful in EEG analysis, although from the viewpoint of dimension theory so far no theoretical justification has been established.

4. Segment distances from AR modelling

We shall now present an approach to quantifying nonstationarity, which is based on linear autoregressive (AR) modelling. Again we divide the time series into segments, then we choose a constant model order p and fit an AR(p) model to each segment by standard least squares estimation.. Let x_{ik} , $i = 1, \dots, N_s$, denote the subseries forming segment k, then the corresponding AR(p) model is given by

$$x_{ik} = a_{0k} + \sum_{j=1}^{p} a_{jk} x_{(i-j),k} + \epsilon_i \ , \tag{4}$$

where a_{jk} , $j = 0, \dots, p$, denote the coefficients of the AR(p) model of this segment, and ϵ_i denotes a stochastic noise term. Given a set of segments there is an equal number of models, and in the spirit of previous work by Hernández *et al.* (1995) and Schreiber (1997) each of the models can be employed to predict each of the segments. Using the model from segment k to predict segment l yields a cross-prediction residual variance ν_{kl} defined by

$$\nu_{kl} = \frac{1}{N_s - p} \sum_{i=p+1}^{N_s} \left(x_{il} - a_{0k} - \sum_{j=1}^{p} a_{jk} x_{(i-j),l} \right)^2 \ . \tag{5}$$

Even if the entire time series has been normalised to have zero mean, it is important to include a constant term a_{0k} for each segment.

We now define a measure of distance between segments k and l by defining

$$D_{kl} = \log \frac{\nu_{kl}}{\nu_{ll}} \ . \tag{6}$$

However, the numbers defined by this equation do not actually represent distances, since they will generally neither be symmetric, nor define a metric space; nevertheless they are useful for our purpose. We could obtain a symmetric definition by using $(D_{kl} + D_{lk})/2$, but we have observed that such symmetrisation step renders this approach less useful.

The denominator in equation 6 provides a reasonable normalisation; alternatively, D_{kl} (or more precisely: D_{kl}/N_s) can also be regarded as the difference between two values of the Akaike Information Criterion *AIC* (Akaike, 1974), the first corresponding to using the model from segment k for predicting segment l, and the second being defined in the usual way for segment l. Since *AIC* becomes minimal for the optimal model (for given p), we can expect the distances according to equation 6 to be non-negative.

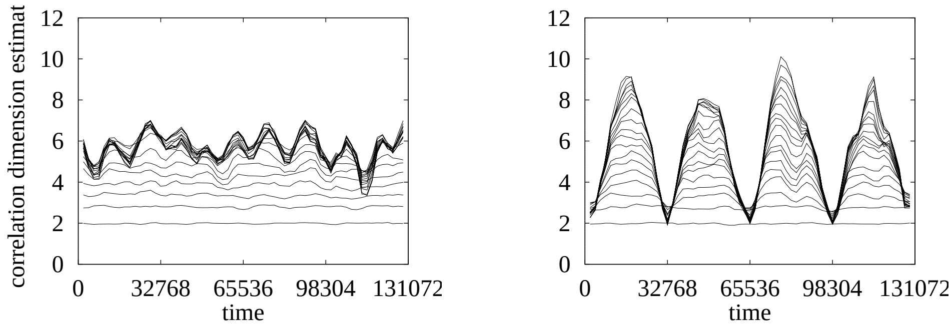

Figure 4: *Correlation dimension estimate versus time using a moving window of 4096 points length for time-delay reconstructions of time series A (left panel) and time series B (right panel). The time delay is* $\tau = 3$, *the embedding dimension ranges from* $m = 2$ *to* $m = 20$ *(curves in ascending order). The Theiler correction parameter is* $W = 20$. *Time is measured directly by counting sample points (for a sampling time of* $t_s = 20\,t_{int}$, *where* $t_{int} = 0.1$*).*

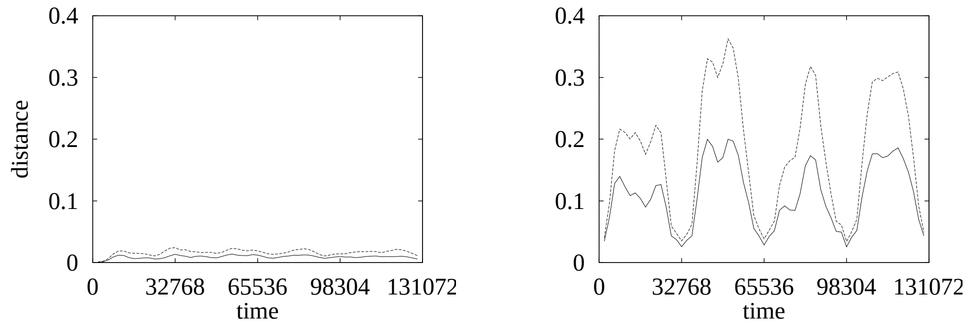

Figure 5: *Segment distances* D_{1j} *versus time using a moving window of 4096 points length for linear AR models of order* $p = 12$ *(solid lines) and* $p = 24$ *(dashed lines) of time series A (left panel) and time series B (right panel). Time is measured directly by counting sample points.*

We remark that a similar measure, as given by equation 6, has been found useful for the comparison between linear and nonlinear autoregressive models, as part of a parametric approach for testing for nonlinearity (Galka & Ozaki, 2001).

A test for stationarity then can be based either on the complete matrix D_{kl} or on just one row of this matrix, by fixing a reference segment, say the first, and observing the fluctuations of D_{1l} as a function of l.

5. Entropies based on spectral or Wavelet representations

Finally we mention a third class of measures which recently has been proposed for the purpose of characterising neurophysiological time series (Inouye *et al.* (1991), Quian Quiroga *et al.* (1999)). The frequency-domain representation of a time series x_i (in our case the subset forming a segment, x_{ik}, but the segment index k is not needed in this section) through the corresponding set of Fourier coefficients c_j, resulting from the Discrete Fourier Transform, is well known; the distribution of energy across different frequencies can be described by a distribution

$$p_j = \frac{|c_j|^2}{\sum_k |c_k|^2} \quad , j = 1, \ldots, N_s/2 \quad . \tag{7}$$

The properties of this distribution can be described by its Shannon entropy

$$H = -\sum_j p_j \log p_j \quad . \tag{8}$$

Recently the same concept was applied to wavelet representations of time series (Quian Quiroga *et al.*, 1999). The discrete wavelet transform of a given function $f(t)$ (sampled by $x_i\,, i = 1, \ldots, N_s$) can be expressed as

$$f(t) = \sum_{j=0}^{K} \sum_{k=1}^{2^{-j} N_s} d_{kj}\, \psi(2^{-j}t - k + 1) + \sum_{k=1}^{2^{-K} N_s} a_{kK}\, \phi(2^{-K}t - k + 1) \quad , \tag{9}$$

where $\psi(t)$ denotes the chosen wavelet (in this paper always the "Daubechies 6" wavelet), $\phi(t)$ the corresponding scaling function and K the depth of the decomposition. The first term constitutes a sum of consecutively

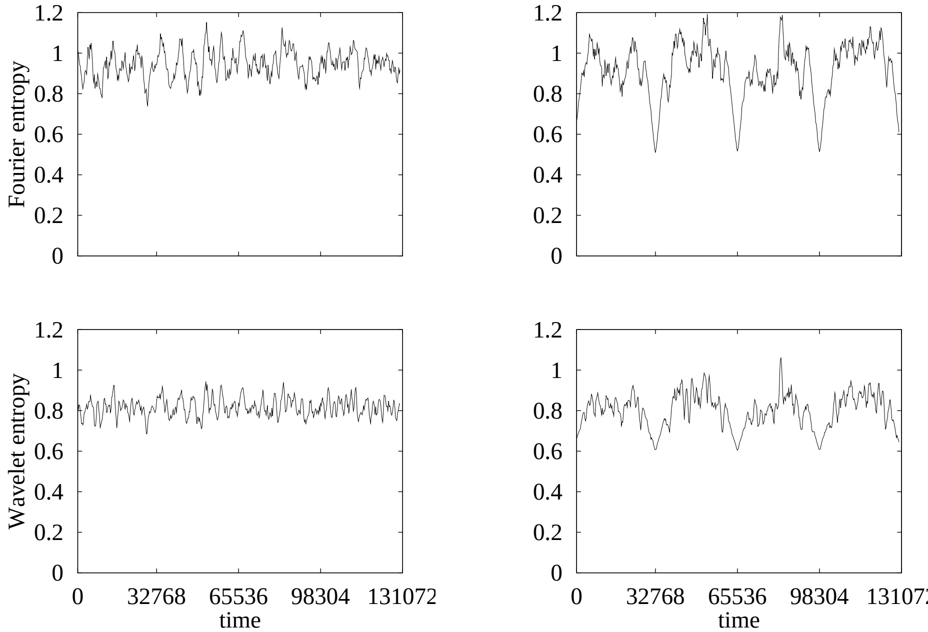

Figure 6: *Fourier entropy (upper panels) and wavelet entropy (lower panels) versus time using a moving window of 2048 points length for time series A (left panels) and time series B (right panels). Time is measured directly by counting sample points.*

extracted details (in terms of a multiscaling decomposition), whereas the second represents the remainder not explained by these details.
Following a suggestion of Quian Quiroga (1999), the two sets of coefficients d_{kj}, a_{kK} can be used for defining a distribution by forming the average energy per step of the decomposition:

$$E_j = \frac{1}{2^{-j}N_s} \sum_{k=1}^{2^{-j}N_s} |d_{kj}|^2 \ , \ E_{K+1} = \frac{1}{2^{-K}N_s} \sum_{k=1}^{2^{-K}N_s} |a_{kK}|^2 \ , \tag{10}$$

which leads to the distribution

$$p_j = \frac{E_j}{\sum_{k=1}^{K+1} E_k} \ . \tag{11}$$

The weights employed in equation 10 are needed since even for white noise the distribution of energy across the steps of the decomposition will not be uniform.
Finally the Shannon entropy can be calculated for the p_j according to equation 8, and the results can be compared for various segments.
As for the case of segment distances, for both definitions of entropy, based on Fourier or wavelet decompositions, it is important to allow for a nonvanishing mean value of each segment, i.e. to include the zero-frequency component of the spectra into the distributions.

6. Results

In this section we present the results of applying the approaches presented in sections 3, 4 and 5 to the three time series described in section 2. For brevity, only a qualitative discussion of results will be given. First we show in figures 4, 5 and 6 the results for the two simulated time series from the Mackey-Glass system (time series A and B).
It can be seen that correlation dimension performs well in resolving the triangle-shaped nonlinearity pattern in time series B, whereas for time series A the estimates converge to a roughly constant level across segments which, however, shows considerable fluctuations. Segment distances from cross-prediction (using model orders $p = 12$ and $p = 24$) are less successful in resolving the triangle-shaped pattern except for segments of low complexity; but the concept of segment distance yields a very good result for the stationary time series, since in this case all distances have values close to zero. The comparison of the results for time series A and B is helpful for assessing this success; in cases where only one time series is available a comparison with surrogate data sets (Theiler *et al.*, 1992) can help to detect significant nonstationarities. The Fourier and wavelet entropies show a similar performance as the segment distances, both succeed in discriminating between low and high complexity, but within the regime of high complexity they provide very little discrimination, as compared to correlation dimension estimation; the results seem rather to be dominated by fluctuations. For the case of

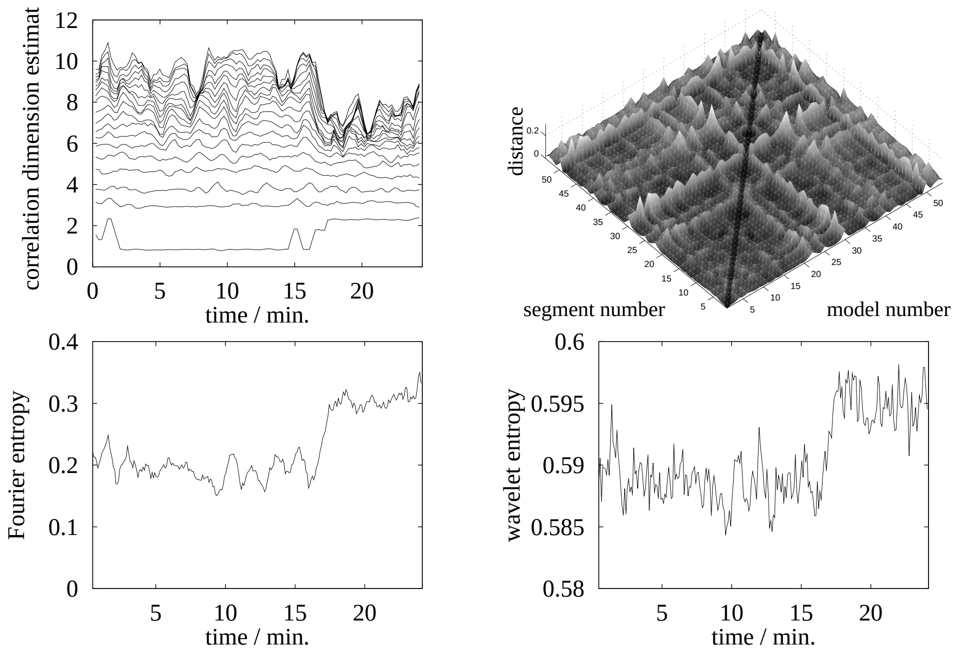

Figure 7: *Upper left panel: Correlation dimension estimate versus time using a moving window of 5000 points length for time-delay reconstructions of time series C, using a time delay of* $\tau = 5$*, embedding dimensions ranging from* $m = 2$ *to* $m = 20$ *(curves in ascending order) and a Theiler correction parameter of* $W = 20$*. Upper right panel: Segment distances* D_{ij} *versus model number* i *and segment number* j *using a moving window of 5000 points length for linear AR models of order* $p = 16$ *of time series C. Time is measured by counting segments; segments have 50 % overlap. Lower panels: Fourier entropy (left panel) and wavelet entropy (right panel) versus time using a moving window of 8192 points length (with 87.5 % overlap) for time series C. Note that in the right panel the vertical axis does not start at zero.*

stationarity wavelet entropy shows a slightly better performance, since it produces less fluctuations; obviously the wavelet decomposition is better suited for adapting to the waveforms present in the individual segments at high complexity.

Next we show in figure 7 the results for time series C. The estimates of correlation dimension reproduce the results of Lehnertz & Elger (1998); a marked transition from a state of higher dimension towards a state of lower dimension (which still resides at fairly high values around $d_2 \approx 7$) can be seen after approximately 17 minutes. This result is commonly regarded as indicating a significant loss of "neuronal complexity" and henceforth increased probability of the onset of an epileptic seizure; the dynamics during seizures can be regarded as possessing low complexity due to the presence of highly synchroneous oscillations.

However, the Fourier and wavelet entropies show that this interpretation may be overly simplified, since here these measures succeed in detecting the same state transition, but now the second state is found to have *increased* spectral complexity. It should also be noted that wavelet entropy rates this time series as essentially stationary, since the difference of entropy between both states is very small in comparison to absolute values of entropy (of course, this may be a consequence of a suboptimal implementation; e.g. by careful adaptation of the basis wavelet to the data a different performance may be obtained).

Finally we note that segment distances essentially fail to detect this state transition at all. In this case we have chosen to display the full matrix D_{ij}, but the only prominent features it contains are certain individual segments that have a comparatively large distance to all other segments.

7. Conclusion

In this paper we have investigated the issue of detecting nonstationary structure in high-dimensional time series. Clearly the immense variety of potential dynamics and nonstationarities which could be present in actual time series precludes a general solution of this problem. It is for the same reason that time series models containing explicit nonstationarity will remain of limited use.

The two examples which we have considered, the Mackey-Glass system and the intracranial EEG, are highly different in various ways, but still correlation dimension estimation seems to provide reasonable results for both cases. For the regime of high complexity of the underlying dynamics it seems to be a very useful tool; approaches relying on linear properties, such as segment distances based on linear AR models and entropies defined on spectral or wavelet distributions, cannot cope well with changes of complexity which do not produce sizable changes of the spectral content of the signal. Both time series A and B were generated in a way such that they are high-dimensional almost always.

On the other hand we have seen that the pronounced state transition which seems to be present in the EEG time series does also produce changes of linear properties, although this series itself is also high-dimensional, such that the sophisticated (and time-consuming) nonlinear technique of correlation dimension estimation is not the only way to detect and describe the nonstationary structure.

Obviously the main drawback both of correlation dimension and entropies is the tendency of these quantities to display considerable fluctuations even for stationary time series, if the dynamics is high-dimensional; this will easily result in spurious detection of nonstationary structure. This problem can be avoided by the concept of segment distances. However, the failure of segment distances to detect the state transition in time series C demonstrates that this technique depends to large extent on the underlying modelling approach. Obviously plain linear AR modelling is an unsuitable choice for spike-dominated EEG time series. More refined modelling would probably yield improvements, but the computational time demands would soon become sizable.

Currently we therefore suggest that a test for nonstationarity of a given time series from unknown dynamics should be performed by a combination of different methods which are sensitive for various properties of the data; correlation dimension, Fourier entropy and segment distances based on AR modelling may form a first selection for a useful set of such methods.

Acknowledgements

We would like to thank Klaus Lehnertz and Christian Elger from the Klinik für Epileptologie, Bonn, Germany, for kindly providing the clinical time series. The work of AG has been supported by Deutsche Forschungsgemeinschaft (DFG) through project GA 673/1-1 and by the Japanese Society for the Promotion of Science (JSPS) through fellowship ID No. P 03059.

References

Akaike, H. (1974): A new look at the statistical model identification. *IEEE Trans. Autom. Contr.*, **19**, 716–723.

Buzug, T. and Pfister, G. (1992): Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global and local dynamical behavior of strange attractors. *Phys. Rev. A*, **45**, 7073–7084.

Ding, M., Grebogi, C., Ott, E., Sauer, T. and Yorke, J.A. (1993): Plateau onset of correlation dimension: when does it occur? *Phys. Rev. Lett.*, **70**, 3872–3875.

Galka, A. (2000): *Topics in Nonlinear Time Series Analysis - With Implications for EEG Analysis.* World Scientific, Singapore.

Galka, A. and Ozaki, T. (2001): Testing for nonlinearity in high-dimensional time series from continuous dynamics. *Physica D*, **158**, 32–44.

Havstad, J.W. and Ehlers, C.L. (1989): Attractor dimension of nonstationary dynamical systems from small data sets. *Phys. Rev. A*, **39**, 845–853.

Hegger, R., Kantz, H., Matassini, L. and Schreiber, T. (2000): Coping with non-stationarity by over-embedding. *Phys. Rev. Lett.*, **84**, 4092.

Hernández, J.L., Biscay, R., Jimenez, J.C., Valdes, P. and Grave de Peralta, R. (1995): Measuring the dissimilarity between EEG recordings through a non-linear dynamical system approach. *Int. J. Bio-Med. Computing*, **38**, 121–129.

Inouye, T., Shinosaki, K., Sakamoto, H., Toi, S., Ukai, S., Iyama, A., Katsuda, Y. and Hirano, M. (1991): Quantification of EEG irregularity by use of the entropy of the power spectrum. *Electroenc. Clin. Neurophys.*, **79**, 204–210.

Kantz, H. and Schreiber, T. (1997): *Nonlinear time series analysis.* Cambridge University Press, Cambridge.

Krystal, A.D., Prado, R. and West, M. (1998): New methods of time series analysis for non-stationary EEG data: Eigenstructure decompositions of time varying autoregressions. *Clin. Neurophysiol.*, **110**, 1–10.

Lehnertz, K. and Elger, C.E. (1998): Can epileptic seizures be predicted? Evidence from nonlinear time series analysis of brain electrical activity. *Phys. Rev. Lett.*, **80**, 5019–5022.

Mackey, M.C. and Glass, L. (1977): Oscillation and chaos in physiological control systems. *Science*, **197**, 287–289.

Mormann, F., Andrzejak, R.G., Kreuz, T., Rieke, C., David, P., Elger, C.E. and Lehnertz, K. (2003): Automated detection of a preseizure state based on a decrease in synchronization in intracranial electroencephalogram recordings from epilepsy patients. *Phys. Rev. E*, **67**, 021912.

Quian Quiroga, R., Rosso, O.A. and Başar, E. (1999): Wavelet entropy: a measure of order in evoked potentials. *Electroenc. Clin. Neurophys.*, **49**, 298–302.

Schreiber, T. (1997): Detecting and analysing nonstationarity in a time series using nonlinear cross predictions. *Phys. Rev. Lett.*, **78**, 843–847.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. and Farmer, J.D. (1992): Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, **58**, 77–94.

West, M., Prado, R. and Krystal, A.D. (1999): Evaluation and comparison of EEG traces: Latent structure in nonstationary time series. *J. Amer. Stat. Assoc.*, **94**, 1083–1095.