19 July 1999

PHYSICS LETTERS A

ELSEVIER

Physics Letters A 258 (1999) 103–114

www.elsevier.nl/locate/physleta

Detecting dynamical change in nonlinear time series

L.M. Hively, P.C. Gailey, V.A. Protopopescu

Oak Ridge National Laboratory, P.O. Box 2009, Oak Ridge, TN 37831-8066, USA

Received 23 February 1999; received in revised form 19 May 1999; accepted 19 May 1999
Communicated by C.R. Doering

Abstract

We present a robust, model-independent technique for measuring changes in the dynamics underlying nonlinear time-serial data. After constructing discrete density distributions of phase-space points on the attractor for time-windowed data sets, we measure the dissimilarity between density distributions via L_1-distance and χ^2 statistics. The discriminating power of the new measures is first tested on the Lorenz model and then applied to EEG data to detect the transition between non-seizure and epileptic activity. We find a clear superiority of the new measures in comparison to traditional nonlinear measures as discriminators of changing dynamics. © 1999 Published by Elsevier Science B.V. All rights reserved.

1. Introduction

A problem frequently encountered in nonlinear time-series analysis is appropriate characterization of changes in the system dynamics. One approach to this problem is to statistically test for stationarity. However, such tests produce a binary result, namely they indicate whether or not a change occurred but provide no information about the extent of departure from one state to another. Stationarity tests also have limited value for inherently nonstationary processes that undergo frequent or continual changes in dynamics (e.g., physiological data, like EEG). For such nonstationary processes, a measure of dissimilarity that quantifies the 'distance' between attractors turns out to be more useful [1,2]. We describe four sensitive measures of dissimilarity, applying them first to model data and then to physiological data. Our method is useful for 1) discriminating between different and possibly close chaotic regimes, and 2) monitoring the extent of departure of a system from an initial dynamical state. Straightforward methods exist [11] for discriminating between regular and chaotic motion, or for detecting the transition between these regimes. However, distinguishing different chaotic regimes can be very difficult, especially when data are limited and noisy.

The paper is organized as follows. Section 2 discusses some traditional nonlinear measures for analysis of time series. Section 3 presents our indicators of dynamical change by comparison of the phase-space distribution functions via distance measures. Section 4 describes the discriminating power of these measures on the Lorenz model, in which changes are easily monitored and controlled. Finally, Section 5 explains the use of these measures on experimental EEG data to detect the transition between non-seizure and epileptic brain activity. While the dissimilarity measures reflect the inherent noise in the EEG data, detection of pre-seizure states is

0375-9601/99/$ - see front matter © 1999 Published by Elsevier Science B.V. All rights reserved.
PII: S0375-9601(99)00342-4

highly superior to that realized using traditional nonlinear measures.

2. Nonlinear measures from time series

Our analysis begins with a scalar signal, x, arriving from a dynamical system whose dimensionality and structure are usually unknown. This variable is sampled at equal time intervals, τ, starting at time, t_0, yielding a sequence of N points, $x_i = x(t_0 + i\tau)$. Dynamical process reconstruction [3] uses d-dimensional time-delay vectors, $y(i) = [x_i, x_{i+\lambda}, \ldots, x_{i+(d-1)\lambda}]$, for a system with d active variables and time lag, λ. The choice of lag and embedding dimension, d, determine how well the phase space reconstruction unfolds the dynamics for a finite amount of noisy data. Takens found that, for a d-dimensional system, $2d+1$ dimensions generally results in a smooth, nonintersecting reconstruction [5]. Sauer et al. [6] showed that, under ideal conditions, the first integer greater than the correlation dimension is often sufficient to reconstruct the system dynamics. This last statement has been confirmed by computing the embedding dimension via the false nearest-neighbors method [7–9]. However, real data have finite length and are affected by noise, implying that too high an embedding dimension may result in over-fitting. We further note that different observables of a system contain disparate levels of dynamical information [10], implying that phase space reconstruction could be easier from one variable, but more difficult (or even impossible) from another. Our subsequent analysis is mindful of the balance between these caveats and the constraints imposed by the limited amount of noisy data.

Based on the phase space reconstruction, various nonlinear measures have been defined to characterize process dynamics. We choose three of these nonlinear measures, against which we compare the new metrics. In particular, we use: (i) the first minimum in the mutual information function as a measure of decorrelation time, (ii) the correlation dimension as a measure of dynamic complexity, and (iii) the Kolmogorov entropy as a measure of predictability. We describe these three measures in the following paragraphs.

(i) The mutual information function (MIF) is a nonlinear version of the (linear) auto-correlation and cross-correlation functions, and was originally developed by Shannon and Weaver [12] with subsequent application to time series analysis by Fraser and Swinney [13]. The MIF measures the average information (in bits) that can be inferred from one measurement about a second measurement, and is a function of the time delay between the measurements. Univariate MIF measures predictability within the same data stream at different times. Bivariate MIF measures predictability of one data channel, based on measurements in a second signal at different times. For the present analysis, we use the first minimum in the univariate MIF, M_1, to indicate the average time lag that makes x_i independent of x_j. The MIF, $I(Q,R)$, and system entropy, H, for two measurements, Q and R, are defined by:

$$I(Q,R) = I(R,Q) = H(Q) + H(R) - H(Q,R), \tag{1}$$

$$H(Q) = -\sum_i P_Q(q_i)\log\left[P_Q(q_i)\right], \tag{2}$$

$$H(Q,R) = -\sum_{i,j} P_{QR}(q_i,r_j)\log\left[P_{QR}(q_i,r_j)\right]. \tag{3}$$

Q is one set of data measurements, $q_1, q_2, \ldots, q_n$, with associated probabilities $P_Q(q_1), P_Q(q_2), \ldots, P_Q(q_n)$. R denotes a second set of data measurements, $r_1, r_2, \ldots, r_n$, with a time delay relative to the q_i values, having associated probabilities $P_R(r_1)$, $P_R(r_2), \ldots, P_R(r_n)$. The function $P_{QR}(q_i,r_j)$ denotes the joint probability of both states occurring simultaneously. H and I are expressed in units of bits if the logarithm is taken in base two.

(ii) The maximum-likelihood correlation dimension, D, is defined by [14,15]:

$$D = \left\{ -\frac{1}{M}\sum_{i,j} \ln\left[\frac{\delta_{ij}/\delta_0 - \delta_n/\delta_0}{1-\delta_n/\delta_0}\right]\right\}^{-1}, \tag{4}$$

where M is the number of randomly sampled point pairs; δ_{ij} is the maximum-norm distance between the (randomly chosen) i–j point pairs, as defined in

Eq. (6) (below). The distance (scale length) δ_n is associated with noise as measured from the time serial data. Note that the distances are normalized with respect to a nominal scale length δ_0, which is chosen as a balance between sensitivity to local dynamics (typically at $\delta_0 \leq 5a$) and avoidance of excessive noise (typically at $\delta_0 \geq a$). Here, the symbol a denotes the absolute average deviation as a robust indicator of variability [15] in the time serial data:

$$a = \frac{1}{w} \sum_{i=1}^{w} |x_i - \underline{x}|, \tag{5}$$

where $\underline{x}$ is the mean of x_i over a window of w points. The distances (δ_{ij}) are defined by:

$$\delta_{ij} = \max_{0 \leq k \leq m-1} |x_{i+k} - x_{j+k}|, \tag{6}$$

where m is the average number of points per cycle.

(iii) The Kolmogorov entropy, K, measures the rate of information loss per unit time, or (alternatively) the degree of predictability. A positive, finite entropy generally is considered to be a clear demonstration that the time series and its underlying dynamics are chaotic. A large entropy indicates a stochastic, non-deterministic (totally unpredictable) phenomenon. One estimates the entropy from the average divergence time for pairs of initially-close orbits. More precisely, the entropy is obtained from the average time for two points on an attractor to go from an initial separation ($\delta < \delta_0$), to a separation of more than a specific distance ($\delta > \delta_0$). The maximum-likelihood entropy is calculated from the method by Schouten, Takens, and van den Bleek [16]:

$$K = -f_s \log\left(1 - \frac{1}{\underline{b}}\right), \tag{7}$$

$$\underline{b} = \frac{1}{M} \sum_{i=j}^{M} b_i, \tag{8}$$

with b_i as the number of timesteps for two points, initially within $\delta < \delta_0$, to diverge to $\delta > \delta_0$. The symbol f_s denotes the data sampling rate.

Entropy and correlation dimension usually are defined in the limit of zero scale length. However, all real data have noise and even noiseless model data is limited by the finite precision of computer arithmetic. Thus, we choose a finite scale length that is slightly larger than the noise ($\delta_0 = 2a$), at which to report the values of K and D, corresponding to finite-scale dynamic structure. Consequently, the values of K and D that we report do not capture the full dynamical complexity and have smaller values than expected for the zero-scale-length limit ($\delta_0 \to 0$).

3. Definition and use of the new measures

The traditional nonlinear measures of the previous section characterize global features of the nonlinear dynamics, but do not reveal slight dissimilarities between dynamical states. The same is true for other global indicators, such as fractal dimension, Lyapunov exponents, etc. This lack of discrimination occurs, because the traditional measures are based on averaged or integrated system properties of the attractor, which provides a global picture of long-term dynamical behavior. Traditional nonlinear measures ultimately provide only one or a few scalar measures as summary descriptors of large data segments.

Greater discrimination is possible by more detailed analysis of the reconstructed dynamics. The natural (or invariant) measure of the attractor is a more refined representation of the reconstruction, describing the visitation frequency of the system dynamics over the phase space. We first represent each signal value, x_i, as a symbolized form, s_i, that is one of S different integers, $0,1,\ldots,S-1$:

$$0 \leq s_i = \mathrm{INT}\left[\frac{S(x_i - x_{\min})}{x_{\max} - x_{\min}}\right] \leq S - 1. \tag{9}$$

Here, $x_{\min}$ and $x_{\max}$ denote the minimum and maximum values of x_i, respectively, over both the base case (reference data) and over the test cases (subsequent data to be tested for departure from the base case). The inequality $0 \leq s_i \leq S-1$ holds trivially for $x_{\min} \leq x_i < x_{\max}$. We require that $s_i(x_i = x_{\max}) = S-1$ in order to maintain exactly S distinct symbols. Thus, the phase space is partitioned into S^d

hypercubes or bins. We then obtain the distribution function as a discretized density on the attractor, by counting the number of phase-space points occurring in each bin. We denote the population of the i-th bin of the distribution function, Q_i, for the base case, and R_i for a test case, respectively. For infinitely precise data, this representation has been used in [17]. The choice of parameters (S, N, and d) depends on the specific data under consideration. For this initial work, we iteratively varied each parameter with the others fixed, to obtain optimum sensitivity of the measures to changes in system dynamics for each class of data evaluated. (All analyses of EEG data described in Section 5 were performed with the same values of S, N, and d). A systematic method to determine optimal values for these parameters is the subject of future work.

We use an embedding window, $M_1 = (d-1)\lambda$, based on the first minimum in the mutual information function, M_1 [13]. Then, we set $\lambda = \mathrm{INT}[0.5 + M_1/(d-1)]$ to obtain an integer value for the reconstruction lag when M_1 is not evenly divisible by $d-1$. The function (INT) converts a decimal number to the next lower integer, and M_1 is measured in timesteps. The reconstruction requires that $\lambda \geq 1$, thus constraining the largest value of dimensionality to $d \leq 2M_1 + 1$ from the above formula.

We next compare the distribution function of an unknown process state to that of a basecase. Previous work [18] measured differences between delay vector distributions by the square of the distance between two distribution functions. Here, we measure the difference between Q_i with R_i by the χ^2 statistics and L_1 distance:

$$\chi^2 = \sum_i \frac{(Q_i - R_i)^2}{Q_i + R_i}, \tag{10}$$

$$L = \sum_i |Q_i - R_i|. \tag{11}$$

where the summations in both equations run over all of the populated cells in the phase space. The choice of these measures is dictated by the following considerations. The χ^2 statistics is one of the most powerful, robust, and widely-used statistical tests to measure discrepancies between observed and expected frequencies. The L_1 distance is the natural metric for distribution functions since it is directly related to the total invariant measure on the attractor. To apply these measures properly we have to scale the total population of the unknown distribution function (sum over all the domain populations in R_i) to be the same as the total population of the basecase. The sum in the denominator of Eq. (10) is based on a test for equality of two multinomial distributions [19].

By connecting successive phase space points as indicated by the dynamics, $y(i) \to y(i+1)$, one obtains a discrete representation of the process flow [4]. This approach motivates the extension of the method to capture even more dynamical information using pair-wise connectivity between successive d-dimensional states. We thus form a $2d$-dimensional vector, $Y(i) = [y(i), y(i+1)]$, in the connected phase space. As before, Q and R denote the distribution functions for the basecase and testcase, respectively, in the connected phase space.

We use base S arithmetic to assign an identifier ($I = I_i$) to the ith phase-space state, via $I_i = \sum d^{k-1} s_i(k)$. The sum runs from $k=1$ to $k=d$, corresponding to the successive components of the d-dimensional phase-space vector. The symbol for the kth component of the ith phase-space vector is $s_i(k)$. The numeric identifier for the sequel phase-space point is then $J = I_{i+1}$. Then, we can define the measure of the dissimilarity between these two connected phase space states, as before, via the L_1-distance and χ^2 statistic:

$$\chi_c^2 = \sum_{ij} \frac{(Q_{ij} - R_{ij})^2}{Q_{ij} + R_{ij}}, \tag{12}$$

$$L_c = \sum_{ij} |Q_{ij} - R_{ij}|. \tag{13}$$

The subscript c indicates the connected distribution function measure. The connected phase space approach is consistent with Takens' reconstruction theorem [5] as discussed above, when d is larger than the dimensionality of process for unlimited noise-free data. We note that the value $\lambda = 1$ results in $d-1$ components of $y(i+1)$ being redundant with those of $y(i)$, but we allow this redundancy to accomodate

other data such as discrete points from the Henon map. The measures defined in Eqs. (10)–(13) satisfy the following inequalities:

$$\chi^2 \le L, \tag{14}$$

$$\chi_c^2 \le L_c, \tag{15}$$

$$L \le L_c, \tag{16}$$

$$\chi^2 \le \chi_c^2. \tag{17}$$

To prove Eq. (14), we note that for any non-negative numbers, Q_i and R_i, we have $(Q_i + R_i) \ge |Q_i - R_i|$. Dividing the equality $(Q_i - R_i)^2 = |Q_i - R_i|^2$ by the preceding inequality yields $(Q_i - R_i)^2/(Q_i + R_i) \le |Q_i - R_i|$, which completes the proof when summed over all values of i. One can prove Eq. (15) by the same argument. The proof for Eq. (16) follows directly from the definition of the L_1-norm:

$$L = \sum_i |Q_i - R_i| = \sum_i |\sum_j (Q_{ij} - R_{ij})|$$
$$\le \sum_i \sum_j |Q_{ij} - R_{ij}| = L_c. \tag{18}$$

The proof of Eq. (17) is more cumbersome, but it follows without difficulty from a complete induction argument [20]. These inequalities show that, as expected, the measures of condition change for the connected phase space are stronger than those for the phase space representation.

In the subsequent application of the new phase space measures to discriminate condition change, we note that distribution function values depend on one another due to phase space construction from time delay vectors with dynamical structure [18]. The resulting statistical bias is avoidable by averaging contributions to Eqs. (10) and (12) over values of $y(j)$ or $Y(j)$ which satisfy $|i - j| < \Lambda$ [18], where Λ is some largest typical correlation scale length in the time series. We tested the bias in typical data by sampling every Λ-th connected phase space point for $4 \le \Lambda \le 23$, resulting in Λ different samples for the base case (Q_i) and for each cutset (R_i). We then averaged the sampled χ^2 values over the Λ^2 different combinations of distribution functions for the basecase and testcase cutsets. As expected, a decrease proportional to $1/\Lambda$ occurs in the sampled χ^2 values, because the number of data points contributing to χ^2 decreases in the same proportion. The trend over time in sampled χ^2 values remains the same as in χ^2 values without sampling, showing that no unexpected bias is present. Thus, we use unsampled χ^2 values for the remainder of this work as a *relative* measure, rather than as an unbiased statistic for accepting or rejecting a null statistical hypothesis.

4. Application to the Lorenz model

We assess the discriminating power of the new measures by testing them on the well-known Lorenz model [21]:

$$\frac{dx}{dt} = a(y - x), \quad \frac{dy}{dt} = rx - y - xz,$$
$$\frac{dz}{dt} = xy - bz. \tag{19}$$

Some traditional nonlinear measures are good indicators of a bifurcation or transition to chaos. However transitions between two chaotic regimes are not readily detected by these same measures, especially for relatively small changes in the parameter that underlies the transition. Therefore, the present work concentrates on detecting nonstationarity within a region where the Lorenz system is known to behave chaotically [22]: $a = 10$, $b = 8/3$, and $25 \le r \le 90$. We integrate the model using a multistep, multi-order method by Sampine and Gordon [23].

We calculated 200 000 data values of (x, y, z) at fixed time intervals of $t = 0.01$ for each value of r. We obtained the (connected) phase space measures by partitioning each 200 000-point Lorenz dataset into four non-overlapping subsets of 50 000 points each, for comparison to each of the 50 000-point subsets of basecase at $r = 25$. We compared each of the four testcase subsets to each of the four basecase subsets, yielding sixteen values for each of the four measures of dissimilarity, from which we obtain a mean and the standard deviation of the mean. We found that all three coordinates of the Lorenz system provide some degree of dissimilarity detection. For

example, Fig. 1 shows various nonlinear measures versus r, by analyzing only the y component of the Lorenz system. (We obtain similar results for the x component of the Lorenz system. The measures of dissimilarity increase monotonically for the z component for $25 \leq r \leq 47$, then become constant.) The correlation dimension (Fig. 1a) varies erratically between 1.95 and 2.65, over the whole range. The Kolmogorov entropy (Fig. 1b) also varies somewhat while gradually rising from 0.037 to 0.085. Fig. 1c

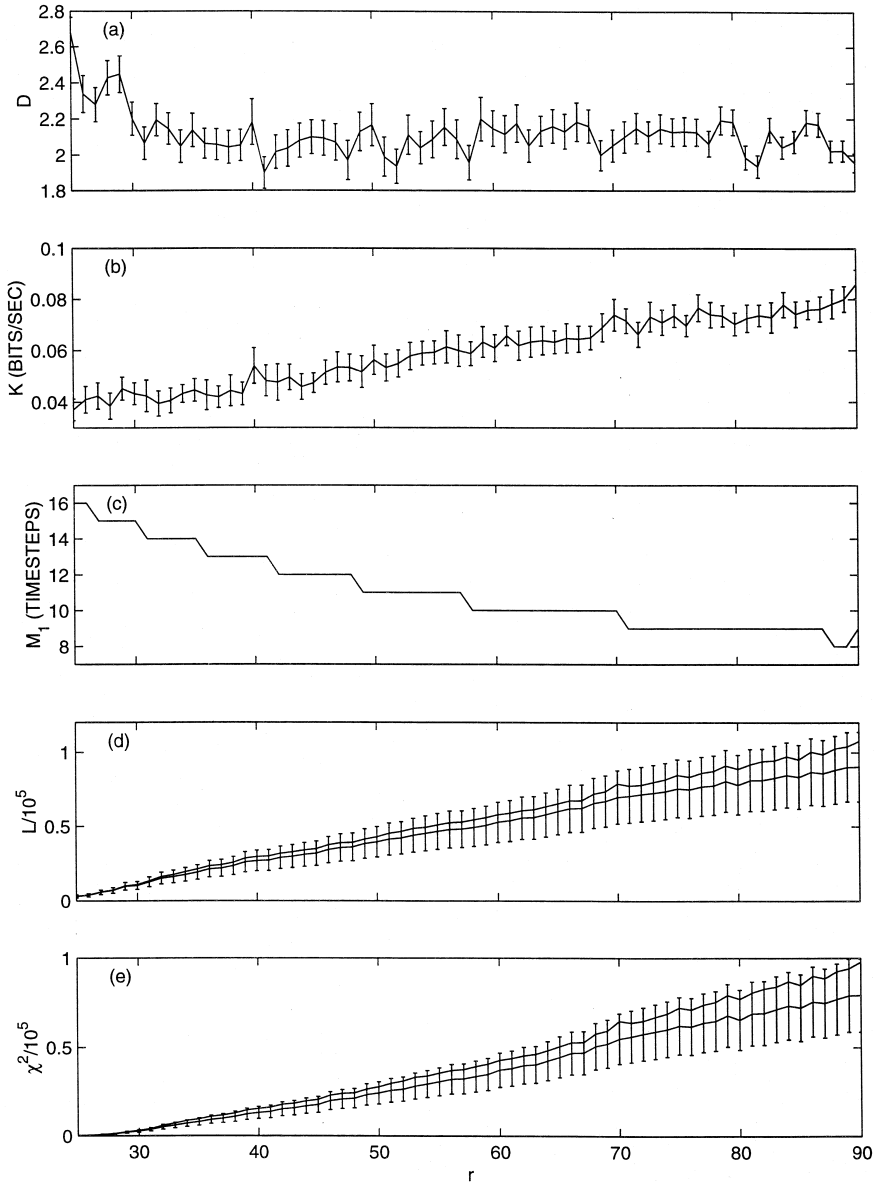

Fig. 1. Nonlinear measures versus r for the Lorenz system, calculated from the y component: (a) correlation dimension, D, (b) Kolmogorov entropy, K, (c) location of the first minimum in the MIF, M_1, (d) $\chi^2/10^5$, and (e) $L/10^5$. The error bars on D and K correspond to 95% confidence intervals. The phase space reconstruction parameters are $S = 12$, $d = 3$, $N = 50\,000$, and $\lambda = 4$. The (connected) phase space measures are the top (bottom) curves, respectively, in Figs. 1d and 1e.

show the location of the first minimum in the mutual information function, M_1 with a monotonic but stepwise decrease as r rises, so that relatively large intervals of r are poorly indicated ($71 \leq r \leq 87$, for instance). A reduction in the sampling interval would reduce the size of these stepwise regions, but this exercise serves as a realistic test of the measures on real data with a limited sampling rate when such a reduction is not possible. In sharp contrast, the (connected) phase space measures (Figs. 1d and 1e) increase almost monotonically from zero to more than 10^5 as r rises from 25 to 90. The values of L and χ^2 essentially coincide over the whole range because the measures are dominated by phase space bins that are populated only for the basecase ($Q_i > 0$ for $R_i = 0$) and only the testcase ($R_i > 0$ for $Q_i = 0$), for which the two measures become analytically equivalent. The curves in Figs. 1d and 1e correspond to the average measure of dissimilarity, while the error bars indicate the standard deviation of the means. We show error bars for the phase-space metrics only, because error bars for the connected phase measures are comparable. As expected from Eqs. (14)–(17), the connected phase-space measures are stronger than their non-connected counterparts.

5. Application to EEG data

We turn next to analysis of brain wave data, which have been described in terms of nonlinear dynamics [24]. Nonlinear EEG measures are not stationary [25], displaying marked transitions between normal and epileptic states. EEG data display low-dimensional features [24,26] with at least one positive Lyapunov exponent [24,27] and hence finite Kolmogorov entropy. EEG data also displays clear phase space structure [24,27], which the above analysis shows is useful for measuring condition change in the Lorenz attractor. Thus, use of the new L_1 and χ^2 phase space measures is plausible for detection of nonlinear condition changes in brain wave data.

We obtained sixteen channels of analog EEG data in the bipolar montage from archival VHS tapes [28]. We converted this data to 12-bit digital form at a sampling rate of 512 Hz. We choose $w = 20480$ data points for each unknown cutset, consistent with Manuca et al. [25] who found a one-dimensional 'meta phase-space' for windows of ≤ 40 seconds. We use the first 400 seconds of EEG data to construct ten 40-second basecase cutsets, each with $w = 20480$ points. Then, we compare each of the ten basecase cutsets to every unknown cutset to obtain average values for χ^2 and L (and the corresponding standard deviation of the mean). We overlap adjacent unknown cutsets by 50% for an optimal mix of new and old data [29] for smooth time-history trending.

Brain dynamics in EEG are obscured by muscular activity due to eye blinks, facial twitches, etc. One can avoid these artifacts by obtaining EEG data from depth and subdural electrodes, although such methods are invasive and non-ambulatory. Furthermore, standard high-pass filtering techniques introduce phase distortions that are unacceptable for nonlinear analysis. Instead, in order to retain nonlinear amplitude and phase relationships, we remove most of the low frequency artifacts corresponding to eye blinks with a zero-phase quadratic filter [28,20].

Our previous analysis of EEG data [28,31] found correlation dimension values of 1 to 2.6 for non- and pre-seizure activity, and ≥ 6 during a seizure, consistent with others' work [24,26]. These results suggest a choice of $d \leq 7$ for the connected phase space reconstruction. However, we find that $d = 7$ overfits the EEG data due to noise, modest cutset size, and the finite precision. This finding does not contradict the Takens' reconstruction theorem [5], which applies to an infinite amount of arbitrarily precise, noise-free data. We find that $d = 3$ and $S = 34$ are adequate for the present work, with the value of M_1 for the phase space measures taken from the first 400 seconds of (non-seizure) data.

Fig. 2 shows a typical plot of nonlinear EEG measures versus time. For the first 800 seconds, variability in all of the measures is modest, representing the dynamics of normal brain activity. The clinical seizure occurs at 2620–2726 seconds, as indicated by the vertical bars at these times in this (and the subsequent) figure. A pre-seizure indication in D occurs as a small decrease at 1150 seconds (Fig. 2a). Likewise, a pre-seizure indication in K is a small peak at 1150 seconds (Fig. 2b). Later variations in D and K provide no clear pre-seizure indication in Fig. 2. The value of M_1 rises slowly from 12 to 20 during the non-seizure period (0–1100 seconds), then falls gradually to an average value of

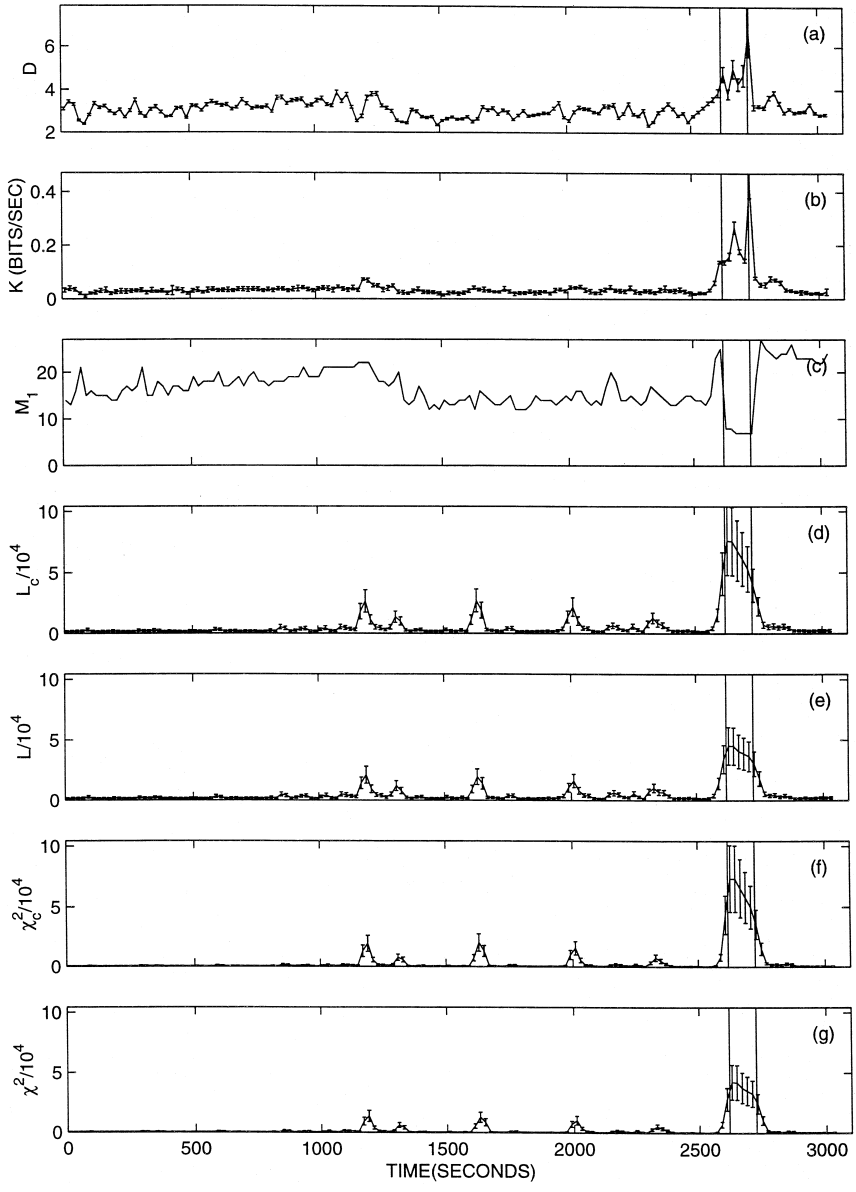

Fig. 2. Nonlinear measures versus time for dataset #62723t: (a) correlation dimension, (b) Kolmogorov entropy, (c) first minimum in the MIF (in timesteps), (d) L_c, (e) L, (f) χ_c^2, and (g) χ^2.

14 (range: 11–20) during the pre-seizure period (Fig. 2c). During the seizure, D and K show clear peaks, while M_1 falls abruptly. In sharp contrast to these weak pre-seizure indications, clear peaks occur in the PS measures (Figs. 2d–g) at 1150, 1300, 1600, and 2350 seconds. We next discuss conversion of these measures to a renormalized metric.

Since these metrics are very disparate in range and interpretation we present the results in a renormalized framework to facilitate comparison. For a given nonlinear measure V, we define V_i as the value of nonlinear measure for the i-th cutset; $\underline{V}$ is the mean value of V_i over the first 400 (non-seizure) seconds of the dataset, with a corresponding sample

standard deviation, σ. The renormalized measure is defined as $U(V) = (V_i - \underline{V})/\sigma$. In our case, V is in turn D, M_1, K, χ^2, etc. Fig. 3 compares these change metrics for the same dataset as Fig. 2. To avoid spurious change indications, we define a positive indication of change as two or more consecutive values of the change metric, $U > U_c$. The horizontal line in each subplot of Fig. 3 shows the limit, $U_c = 4.75$, corresponding to a false positive probability of $< 10^{-6}$ in Gaussian random data. The change

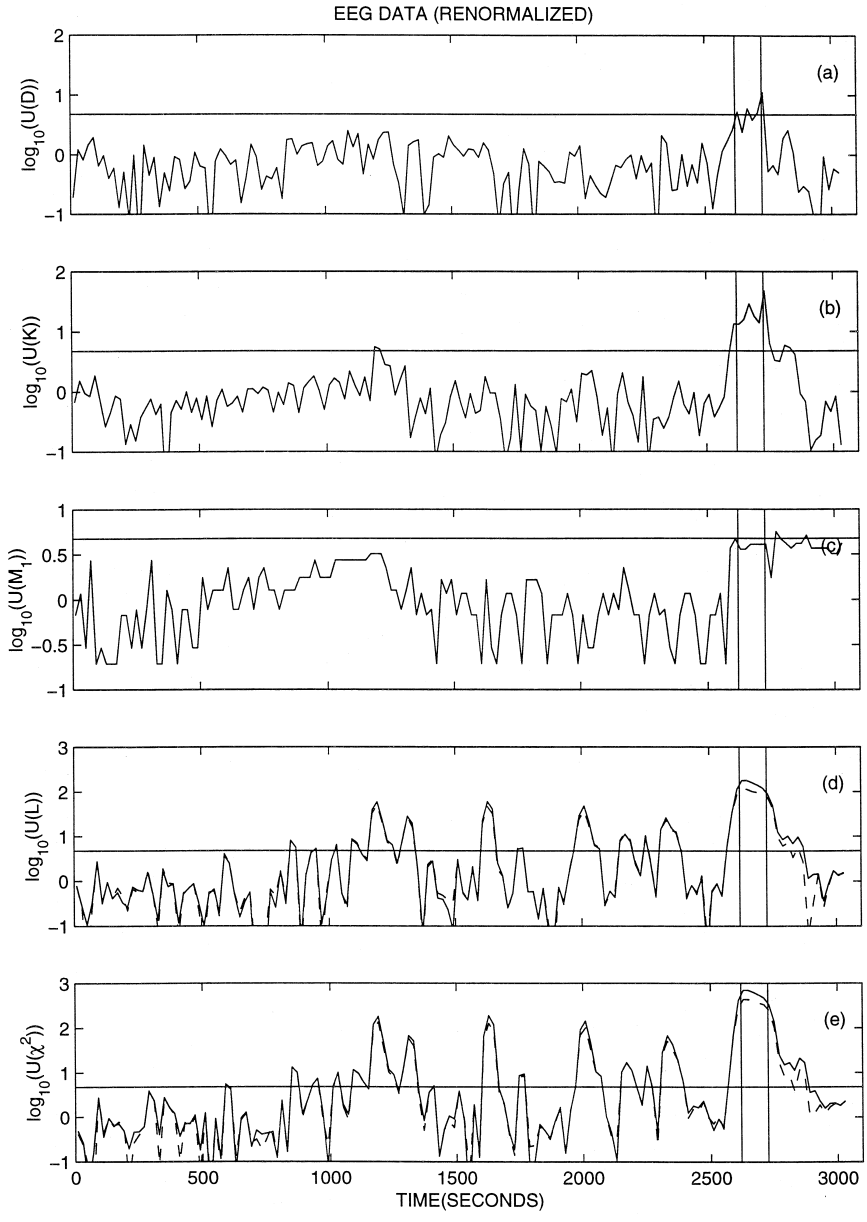

Fig. 3. Change metrics versus time for dataset #62723t: (a) correlation dimension, (b) Kolmogorov entropy, (c) first minimum in the MIF, (d) L_1 measure for CPS (solid) and PS (- -), and (e) χ^2 measure for CPS (solid) and PS (- -). The ordinate values of the change metric (U) are in units of standard deviations from the mean.

metric for correlation dimension, $U(D)$, occurs above this limit only briefly during the seizure (Fig. 3a). $U(K)$ briefly rises above this limit (Fig. 3b) at 1200 seconds, and also later during the seizure. $U(M_1)$ is above the limit only for two brief periods following the seizure. In sharp contrast, the change metrics for the PS measures rise well above this limit at 1130 seconds (for χ^2 and for χ_c^2) and at 1210 seconds (for L and L_c). Table 1 summarizes the time values (measured in seconds prior to the onset of the clinical seizure), at which a positive indication of change occurs for each measure over nine EEG datasets. A negative value for this time corresponds to positive indication of change after the clinical seizure onset. Analysis of normal EEG data shows no positive indication of condition change. All of the phase space measures show positive indications of change for all of the seizure datasets. As expected, the connected phase space measures are better, with an average pre-seizure indication time of 967 and 954 seconds for L_c and χ_c^2, respectively, as compared to average indication times of 916 and 927 seconds for L and χ^2, respectively. Although the average times for positive indication are larger for D, K, and M_1 by 17–300 seconds, several instances exist for no positive indication of change in D, K, and M_1. Since later indication of change is better than no indication, we conclude that the phase space measures are superior to the conventional nonlinear measures (D, K, and M_1) as indicators of condition change. A topic of future research is improvement of the method to provide the earliest possible indication of condition change.

Table 1
Times prior to seizure at which condition change is detected (in seconds)

Dataset #	D	K	M_1	L_c	L	χ_c^2	χ^2	best
109310	**2435**	1135	*	1135	815	1155	−65	2435
109314	1080	1920	**2180**	**2180**	**2180**	1940	**2180**	2180
119230	**2031**	471	*	891	891	891	891	2031
119234	2040	2040	**2100**	**2100**	**2100**	**2100**	**2100**	2100
62723t	−160	1360	*	**1700**	**1700**	**1700**	**1700**	1700
69212	696	**976**	676	716	716	716	716	976
73305D	*	*	*	745	**765**	645	**765**	765
c8492D	*	186	146	366	**566**	366	**566**	566
wm12sD	*	*	*	21	21	21	**501**	501
Maximum time	2435	2040	2180	2180	2180	2100	2180	
Minimum time	−160	186	146	21	21	21	−65	
Average time	1354	1155	1276	1095	1084	1059	1039	
# no indications	3	2	5	0	0	0	0	

Entries denoted by an asterisk (*) show no positive indication of condition change. For each dataset, bold entries denote the earliest time of dissimilarity indication.

Our approach differs markedly from earlier work [26,27] in several respects. First, previous investigations used data from subdural and depth electrodes, while we use scalp EEG data that allows non-invasive, ambulatory, long-term monitoring. Second, those studies obtained EEG data from the epileptogenic area, while we used data only from channel 13 in the bipolar montage, showing the robustness of our method, and allowing easy electrode placement by a patient for non-clinical monitoring. Third, prior effort used invasive monitoring to avoid low-frequency artifacts, which we removed from scalp data with a zero-phase quadratic filter, dramatically improving the data quality and the nonlinear measures. Fourth, previous investigations focused only on temporal lobe epilepsy, while we earlier determined no consistent trends in D and K for various seizure types and instead focused on these new measures of condition changes for any seizure. These advantages will facilitate routine use by patients in a non-clinical setting.

Practical use of this method for forewarning of epileptic events will require characterization of several seizures for each patient, and detailed determination of detection criteria. Future work will involve: (i) statistical evaluation of false positives and negatives in epileptic patients, and of false positives in normal patients and (ii) clinical monitoring of each patient to determine optimal reconstruction parameters, which subsequently would be fixed for ambulatory monitoring.

6. Discussion

We presented two new model-independent indicators to detect condition change (dissimilarity) in nonlinear time series. These new measures have a firm theoretical basis in the invariant measure that

the dynamics determines on the attractor [22]. The connected phase space distribution function also contains information about the dynamical flow from one attractor state to the next. The new indicators of condition change measure the difference between these density functions for a base case and a test case, as χ^2 statistics and L_1 distance. Thus, these indicators integrate (and magnify) the *differences* between the process dynamics, and avoid the inner cancellation effects due to averaging over many orbits (as one does, for instance, when calculating the correlation dimension and Kolmogorov entropy). Moreover, this approach represents the dynamics of noisy, finite-length, finite-precision data, in a way that does not contradict the Takens' reconstruction theorem [6], as discussed above. Changes in the Lorenz attractor are clearly detected by these new measures and by the mutual information, as r increases from 25 to 90 (Fig. 1). On the other hand, these changes go almost undetected by the correlation dimension and Kolmogorov entropy. These new measures also indicate significant pre-seizure changes in all nine analyzed EEG datasets (Table 1). However, correlation dimension, Kolmogorov entropy, and mutual information fail to detect any change at all in three, two, and five EEG datasets, respectively. These results show that the new measures are superior to traditional nonlinear measures for detection of condition change.

We found that indications of condition change exist over a range of embedding parameters (d, S, and λ), for the following reason. If the full distribution function in the actual phase space is $\rho(Z)$, then a partially reconstructed distribution function, $\rho(Y)$, is the sum over regions of $\rho(Z)$ that map into Y, multiplied by the Jacobian determinant for the change in variables. Thus, the information in every component of Z will affect the Y space, unless the projection, $Z \to Y$, is chosen carefully to avoid dependency, which we have not done. Kennel [30] used this approach to test for dynamical nonstationarity in experimental data without resolving the dimensionality in the time series. The important question of finding the *optimal* embedding parameters for detecting nonstationarity with our measures will be addressed elsewhere [20].

An earlier version [31] of this approach was successful for various physical processes. Examples include: distinguishing different drilling conditions from spindle motor current of a machining center; detecting balanced and unbalanced centrifugal pump conditions from motor power; predicting failure of a bellows coupling in a rotating drive train from motor power; and discerning the difference in microcantilever vibrations with and without mercury on the sensor. Success for such diverse applications suggests that this technique can be reliably used for detecting condition change in nonlinear dynamical processes.

Acknowledgements

This work was sponsored by the U.S. Department of Energy, Office of Energy Management. V.P. was partially supported by the U.S. Department of Energy, Office of Basic Energy Sciences. Oak Ridge National Laboratory is managed for the United States Department of Energy by Lockheed Martin Energy Research Corporation, under Contract No. DE-AC05-96OR22464. We are grateful to members of the staff at the Knoxville Neurology Clinic at St. Mary's Medical Center (Knoxville, TN), including Dr. M.L. Eisenstadt, Dr. R.E. Leppanen, W.W. Holland, and H.L. Leach, who provided expert assistance with the EEG data acquisition and interpretation. We thank N.E. Clapp, Jr. (ORNL) for converting the data from archival VHS tapes to digital form, and for determining the times of the clinical seizures from the paper records.

References

[1] A.T. Schreiber, Phys. Rep. 308 (1999) 1.
[2] B.R. Moeckel, B. Murray, Physica D 102 (1997) 187.
[3] J.-P. Eckmann, D. Ruelle, Rev. Mod. Phys. 57 (1985) 617.
[4] H.D.I. Abarbanel, Analysis of Observed Chaotic Data, Springer Publ., New York, 1996.
[5] F. Takens, Lecture Notes in Mathematics 898 (1981) 366, Springer, New York.
[6] T. Sauer, J. Yorke, M. Casdagli, J. Stat. Phys. 65 (1991) 579.
[7] H.D.I. Abarbanel, M.B. Kennel, Phys. Rev. E 47 (1993) 3057.
[8] H.D.I. Abarbanel, R. Brown, J.J. Sidorowich, L.Sh. Tsimring, Rev. Mod. Phys. 65 (1993) 1331.
[9] L. Cao, Physica D 110 (1997) 43.

[10] C. Letellier, J. Maquet, L. Le Sceller, G. Gouesbet, L.A. Aguirre, J. Phys. A 31 (1998) 7913.
[11] A. Cover, J. Reneke, S. Lenhart, V. Protopopescu, Math. Models, Meth. in Appl. Sci. 7 (1997) 823.
[12] C.E. Shannon, W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana, 1949.
[13] A.M. Fraser, H.L. Swinney, Phys. Rev. A 33 (1986) 1134.
[14] F. Takens, Lecture Notes in Mathematics 1125 (1984) 99, Springer-Verlag, Berlin.
[15] J.C. Schouten, F. Takens, C.M. van den Bleek, Phys. Rev. E 50 (1994) 1851.
[16] J.C. Schouten, F. Takens, C.M. van den Bleek, Phys. Rev. E 49 (1994) 126.
[17] C. Grebogi, E. Ott, J.A. Yorke, Phys. Rev. A 37 (1988) 1711.
[18] C. Diks, W.R. van Zwet, F. Takens, J. de Goede, Phys. Rev. E 53 (1996) 2169.
[19] A.M. Mood, S.A. Graybill, D.C. Boes, Introduction to the Theory of Statistics, McGraw Hill, 1974.
[20] L.M. Hively, P.C. Gailey, V.A. Protopopescu, to be published.
[21] E.N. Lorenz, J. Atmos. Sci. 30 (1963) 130.
[22] E.A. Jackson, Perspectives of Nonlinear Dynamics, vol. 2, Cambridge University Press, Cambridge, 1990.
[23] L.F. Sampine, M.K. Gordon, Computer Solution of Ordinary Differential Equations: the Initial Value Problem, Freeman, 1975.
[24] T. Elbert, W.J. Ray, Z.J. Kowalik, J.E. Skinner, K.E. Graf, N. Birbaumer, Physiol. Rev. 74 (1994) 1.
[25] R. Manuca, M.C. Casdagli, R.S. Savit, Math. Biosci. 147 (1998) 1.
[26] K. Lehnertz, C.E. Elger, Phys. Rev. Lett. 80 (1998) 5019.
[27] L.D. Iasemidis, J.C. Sackellares, Neuroscientist 2 (1996) 118.
[28] L.M. Hively, N.E. Clapp, C.S. Daw, W.F. Lawkins, ORNL/TM-12961, Oak Ridge National Laboratory, Oak Ridge, TN 1995.
[29] D.F. Elliott, K.R. Rao, Fast Transforms, Analyses, Applications, Academic Press, 1982.
[30] M.B. Kennel, Phys. Rev. E 56 (1997) 316.
[31] L.M. Hively, in: T.E. Shannon et al. (Eds.), Proc. Maintenance and Reliability Conf., Univ. of Tennessee, Knoxville, vol. 1, 1997, 16.01.