ELSEVIER

Physica D 112 (1998) 361–380

PHYSICA D

Reliable detection of nonlinearity in experimental time series with strong periodic components

C.J. Stam [a,*], J.P.M. Pijn [b,1], W.S. Pritchard [c]

[a] *Department of Neurology and Clinical Neurophysiology, Leyenburg Hospital, PO Box 40551, 2504 LN The Hague, Netherlands*
[b] *Instituut voor Epilepsiebestrijding "Meer en Bosch", Achterweg 5, 2103 SW Heemstede, Netherlands*
[c] *Psychophysiology Laboratory, Bowman Gray Technical Center, 611-12, R.J. Reynolds Tobacco Company, Winston-Salem, NC 27102, USA*

Received 16 March 1997; received in revised form 17 July 1997; accepted 22 July 1997
Communicated by A.M. Albano

Abstract

Testing with phase-randomised surrogate signals has been used extensively to search for interesting nonlinear dynamical structure in experimental time series. In this paper we argue that, in the case of experimental time series with strong periodic components, the method of phase-randomised surrogate data may not be particularly suitable to test for nonlinearity, since construction of such surrogates by FFT requires a time series whose length is a power of 2. We demonstrate that, in the case of (nearly) periodic signals, this approach will almost always produce spurious detection of nonlinearity. This error can be fixed by adjusting the length of the time series such that it becomes an integer multiple of the dominant periodicity. The resulting time series will not be a power of 2, and requires the use of a DFT to generate surrogate data. DFT-based surrogates no longer detect spurious nonlinearity, but cannot be used to detect periodic nonlinearity.

We propose a new test, nonlinear cross-prediction (NLCP), which avoids some of the problems associated with phase-randomised surrogate data, and which allows reliable detection of both periodic and aperiodic nonlinearity. In the test the original data are used to construct a nonlinear model to predict the original data set as well as amplitude-inverted and time-reversed versions of the original data. Lower predictability of the amplitude-inverted or time-reversed copies reflect, respectively, an asymmetric amplitude distribution and time irreversibility. Both of these indicate nonlinearity in the data set.

Keywords: Time series; Periodic; Nonlinearity; Surrogate data; Prediction; Chaos

1. Introduction

1.1. Detection of nonlinear dynamics in periodic time series

Many phenomena of interest in biology and medicine show fluctuating behavior over time. For instance, recordings of the electrical activity of the heart (electrocardiogram, ECG) and the brain (electroencephalogram, EEG) represent a great challenge for time-series analysis and modelling. Reconstructing the dynamics underlying such time series may help to gain a better understanding of the physiology and pathophysiology of the system studied. Improved methods for time-series analysis may yield greater diagnostic information from recorded bioelectric signals. For these reasons, advances in the theory of

* Corresponding author. Tel.: (+31-70) 3592000; fax: (+31-70) 3295046; e-mail: cjstam@compuserve.com.
[1] E-mail: pijn@dds.nl.

PII S0167-2789(97)00183-8

nonlinear dynamical systems, in particular the discovery of deterministic chaos, and the introduction of new methods for nonlinear time-series analysis are of great interest to many investigators in medicine and biology. Application of theoretical concepts and analytical tools from nonlinear dynamics now characterises the rapidly growing field of "Dynamical Disease", a term first coined by Mackey and Glass [9,15]. An excellent overview of current research in this field can be found in [5].

A particularly important breakthrough in this field was achieved when methods became available to reconstruct, from a univariate time series of measurements, a trajectory in the state space of the system [17]. The validity of this reconstruction was proved by a theorem of Takens [31]. Several algorithms were introduced to estimate nonlinear invariant measures, in particular dimensions, Lyapunov exponents and entropies, from reconstructed attractors (for a recent overview see [1]). In physics and related sciences such methods have been used with success to detect and characterise nonlinear dynamics in relatively simple, low-dimensional systems. However, early attempts to apply these methods directly to physiological systems have met with considerable difficulties. Compared to time series recorded in physics experiments, physiological time series typically have a number of problematic characteristics which have to be taken into account when applying nonlinear time-series analysis. First, a mathematical model of the dynamics is seldom available. Second, the underlying dynamics of a biological organism is seldom stable over long periods of time. This introduces the problem of non-stationarity in long time series. Finally, recordings in biology and medicine can contain large amounts of observational noise and probably always contain some amount of dynamical noise. When nonlinear time-series analysis is applied directly to physiological data without taking these problems into account, spurious results are the rule rather than the exception [23,24].

In the analysis of physiological data, rather than attempting to demonstrate the exact nature of the dynamics from the outset, a more realistic approach is first to establish whether the underlying dynamics is in fact nonlinear. Nonlinearity of the dynamics is a necessary, although not sufficient condition for interesting dynamics (nonlinear limit cycles; chaos). In the early 1990s, the method of surrogate data was introduced to test for nonlinear dynamics [34,35]. The method is explained in more detail in Section 2.1. Briefly, the idea is to construct, from an experimental time series, one or more control data sets (surrogate data), which share with the original data all the linear properties (in particular, the power spectrum and (circular)[2] autocorrelation function), but not the nonlinear properties. It was proposed that this could most readily be realised by applying a Fourier transform to the original data, randomising the phases, and then applying an inverse Fourier transform. Both original and surrogate data could be characterised with some nonlinear statistic, which could be chosen freely. If the outcome for the original data were clearly different from the outcome for the surrogate data, then the "null hypothesis" that the data are a linear stochastic process (e.g. linearly filtered Gaussian white noise) could be rejected. Application of the surrogate data method to many time series which were previously claimed to be chaotic has shown that in several cases such claims may have been premature [2,3,8,18,37]. However, in a number of other cases, the method of surrogate data did reveal evidence for nonlinear dynamics in physiological time-series [16,19–22,26,29].

Further complications have to be dealt with before evidence of nonlinear dynamics in biological data can be accepted. For instance, a non-Gaussian amplitude distribution of the time series can give spurious results with surrogate data testing [25]. An algorithm has been proposed to deal with this problem(algorithm II in [34,35]), but this algorithm may itself be a source of new artefacts [37]. A more serious problem, which will be focus of the present investigation, is related to the use of the discrete Fourier transform to generate

[2] According to Theiler et al. [36] at least three different ways to define a sample autocorrelation exist: (1) unbiased estimator $(1/(N-T))\sum_{t=1}^{N-T} x_t x_t + T$; (2) biased estimator $(1/N)\sum_{t=1}^{N-T} x_t x_t + T$; (3) Circular autocorrelation $(1/N)(\sum_{t=1}^{N-T} x_t x_t + T + \sum_{t=N-T+1}^{N} x_t x_t + T - N)$. For time series of finite length these are only approximately equal. FFT-based surrogate data preserve only the circular autocorrelation exactly, and not the other two.

phase-randomised surrogate data. A discrete Fourier transform of a time series of N samples represents $N/2$ discrete frequencies, all of which are integer multiples of the fundamental frequency. Frequencies that do not exactly match one of the $N/2$ frequencies in the discrete spectrum have their power and phase distributed over nearby frequencies in a coupled manner. This simple fact has unfortunate consequences when surrogate data are generated from time series with strong periodicities.

Ideally, the phase-randomised surrogate of a periodic time series should be another periodic time series. If the shape of the individual cycles is complex (showing higher harmonics in the spectrum), this shape will be changed by the phase-randomisation procedure, but the basic periodicity should not be affected. Stated differently, if a time series corresponds to a limit-cycle before phase randomisation, it should still correspond with a limit cycle after phase randomisation. Consequently, in the case of a periodic time series, nonlinear invariant measures could not be expected to change after phase randomisation.

In practice, the fundamental frequency of a periodic or nearly periodic time series will almost never correspond with one of the discrete frequencies in the spectrum. This is especially true when an FFT algorithm is used, and the length of the time series is required to be a power of 2. The consequence is that after phase randomisation the phases of the nearby frequencies will be uncoupled, and a "beating" phenomenon results (Fig. 1). This introduces a waxing and waning component in the surrogate time series that is absent from the original time series, and which can be expected to affect estimates of nonlinear invariant measures such as dimensions.

These problems associated with phase-randomised surrogate data of time series with strong periodicities were already mentioned by Pijn[3] [20] and were also discussed at some length in a paper by Theiler et al. [36] in relation of time series E of the Santa Fe Time Series Prediction and Analysis Competition. Theiler et al. [36] investigated sinusoidal (sine wave plus noise) and non-sinusoidal periodic time series, and concluded that spurious detection of "highly significant" nonlinearity occurs in both cases. The authors did not investigate specific solutions to the problem, and in fact suggested that it might be difficult to find an automatic procedure which could be applied to *all* signals. For instance, quasi-periodic signals present an even greater problem, because it is impossible even in principle to match two or more incommensurate frequencies exactly with a discrete spectrum, because in a discrete spectrum all frequencies are *integer* multiples of the fundamental frequency. Surprisingly, the problems indicated in [36] seem to have been largely ignored in many later experimental studies. This is of particular importance because many time series studied in biology and medicine, for instance the EEG alpha rhythm, do show strong periodicity. Claims for nonlinear dynamics underlying such time series could very well be corrupted by the artefact indicated above.

In this paper we re-investigate the problem of detecting nonlinear dynamics in time series with strong periodicities. First we introduce briefly some basic concepts and definitions. Then, we demonstrate how and when spurious detection of nonlinearity can occur when using phase-randomised surrogate data. The artefact is shown to depend critically upon a matching of the fundamental frequency with the discrete spectrum; even a very small mismatch already produces highly significant spurious results. Next we investigate one of the "ad hoc" solutions to the problem suggested in [36]. The length of the time series is adjusted (shortened) such that it becomes an exact multiple of the period length of the fundamental frequency. An unfortunate consequence of this length adjustment is that the time series will no longer be a power of 2, so a more time consuming DFT is required and some of the data is lost. It is shown that the use of length-adjusted, DFT-based (LA-DFT) surrogate data solves the problem of spurious nonlinearity in the case of sine waves, and strongly periodic linearly filtered noise. However, the LA-DFT surrogate data still fail in the case of quasi-periodic signals.

The most important part of this paper discusses a new test (nonlinear cross-prediction) which is intended

[3] In fact, the problems that could arise when generating surrogate data from periodic time series were already pointed out to this author by his teacher Van der Tweel in 1988.

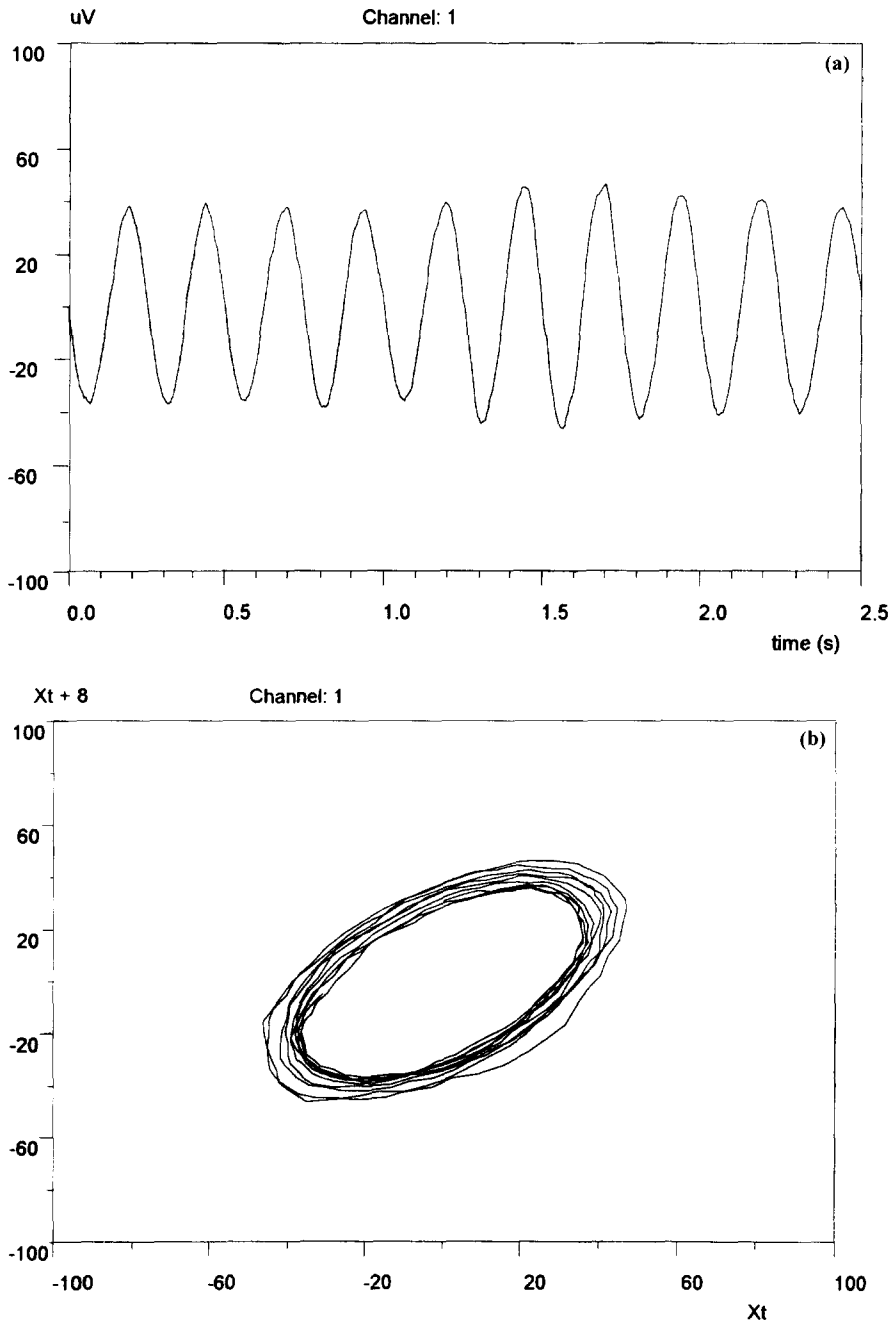

Fig. 1. (A) A sine wave with a period of 51.4844 (which does not quite match with a time-series length of 512) after phase-randomisation (FFT-based without length adjustment). The time series no longer corresponds with a perfect sine, and shows a beating phenomenon. (B) Phase portrait of the sine wave surrogate of (A). The error introduced by the phase-randomisation procedure can be seen clearly.

to avoid the problem of spurious detection of nonlinearity in linear periodic signals. The test consists of applying a nonlinear prediction algorithm (always using the original time series as a "model") to (1) the original time series; (2) a version of the original time series with inverted amplitudes; (3) a time-reversed transform of the original data. The test does not produce spurious results in the case of linear periodic or quasi-periodic signals, and reliably detects nonlinearity in the case of (asymmetric) limit cycle or chaotic dynamics, even in the presence of considerable amounts of observational noise.

1.2. Definitions

In this section we introduce some basic concepts and definitions. We assume a dynamical system that can be described by a set of m coupled ordinary differential equations

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F(X, t, \mu) + \eta. \tag{1}$$

Here and in the rest of this paper X is an m-dimensional state vector. F is an m-dimensional function of X, time t, and μ (a set of control parameters) and η designates (additive) dynamical noise. The state of the dynamical system is not observed directly, but via an observation function G

$$V_t = G(X_t) + \varepsilon, \tag{2}$$

V_t is a k-dimensional vector, where k corresponds with the number of simultaneous measurements. In the rest of the paper we will only consider cases where $k = 1$, and V_t is a scalar time series. Observational or measurement noise is represented by ε.

A dynamical system such as that described by (1) is autonomous when F does not depend explicitly on t. A dynamical system is linear when F is linear, and nonlinear when F is nonlinear. Thus, use of the term "dynamical nonlinearity" implies nonlinearity of F. A dynamical system is deterministic when $\eta = 0$, and stochastic when $\eta > 0$. A time series V_t is nonlinear when at least one of the functions F or G is nonlinear. This implies that nonlinearity of a time series cannot be equated with nonlinearity of the underlying dynamics, since the nonlinearity may be a function of G rather than F.

The goal of nonlinear time-series analysis is to obtain as much information as possible about the function F, when only a time series of measurements V_t is given. The first question to be addressed is whether F is linear or nonlinear. If and only if F is nonlinear, can we expect dynamics with an interesting deterministic component, for instance in the form of nonlinear limit cycle or chaotic attractors (deterministic chaos can never occur in linear systems). Linear systems can have periodic solutions, but these are unstable in the presence of dynamical noise. In other words, linear systems cannot have stable limit-cycle *attractors*. The rest of this paper deals with methods directed specifically at the detection of dynamical nonlinearity underlying observed time series.

2. Phase-randomised surrogate data testing

2.1. Description of the method

A comprehensive account of phase-randomised surrogate data testing for nonlinearity can be found in [34,35]. The method was originally motivated by the search for deterministic chaos in experimental data. Chaos cannot be proven directly in a time series. However, chaos requires nonlinear dynamics. So, as stated above, a more modest but also more realistic goal is to search for nonlinearity in the data. Without nonlinearity, chaos can be excluded. If there is nonlinearity in the data, chaos has not been proven, but at least one necessary condition has been established.

The phase-randomised, surrogate data technique tries to establish the presence of nonlinearity by excluding a reasonable alternative, which is called the "null hypothesis". For this purpose a (usually nonlinear) discriminating statistic Q (for instance, correlation dimension; Lyapunov exponents, entropy, nonlinear predictability) is needed. The null hypothesis is rejected by demonstrating that the observed value of Q for the data is very unlikely when the null hypothesis is true. This requires that the distribution of Q under the null hypothesis (mean value

and standard deviation) be known. These values are obtained from the surogate data. These surrogate data share with the original data only those properties (for instance, linear structure) which are defined by the null hypothesis. If many realisations of the experimental data are available, the distributions of Q for the experimental and surrogate data can be compared directly using conventional statistical tests such as a t-test or Mann–Whitney U-test. Alternatively, it is possible to generate many surrogate data from a single experimental time series and estimates a z-score (which is also referred to as "number of sigmas")

$$Z = \frac{|Q_{\mathrm{d}} - \langle Q_{\mathrm{s}} \rangle|}{\sigma_{\mathrm{s}}}. \tag{3}$$

In this formula Q_{d} is the value of Q for the experimental data, $\langle Q_{\mathrm{s}} \rangle$ is the mean of Q for the surrogate data set and σ_{s} is the standard deviation of Q for the surrogate data. Thus the z-score expresses how many standard deviations ("sigmas") Q of the experimental data deviates from the average Q for the surrogates. Assuming that Q_{s} has a normal distribution, the null hypothesis can be rejected for two sided testing at a significance level of $p < 0.05$ when $z > 1.96$.

The proposed method is quite general in that it allows one to test different null hypotheses, and use any discriminating statistic deemed appropriate. In this study we will use three different statistics: the correlation integral, the E_r statistic and nonlinear predictability (described in Section 2.3). The null hypothesis tested is that the data can be explained by a stationary linear Gaussian model. This implies that all the relevant information is contained in the power spectrum and in the [circular] autocorrelation function. Theiler et al. [34,35] described an algorithm (algorithm I) to construct surrogate data under this null hypothesis. Denote the original time series by V_t with $t = 1, \ldots, N$. A Discrete Fourier Transform of V_t is obtained. The phases are randomised. To guarantee that the inverse Fourier transform will be real, the phases are symmetrised so that $\phi(f) = -\phi(-f)$. An inverse Fourier transform produces the surrogate time series V'_t. The end result of this procedure is a surrogate that has exactly the same power spectrum as the original data, but with random phases. Theiler et al. also consider a more general null hypothesis where the observed time series is a static nonlinear transformation of a linear Gaussian process (algorithm II). For the purposes of the present study we restricted ourselves to algorithm I, but the issue of static nonlinearity will be addressed in Sections 3 and 4.

2.2. Length adjusted-DFT surrogates

Usually algorithms I and II [34,35] for generating phase-randomised surrogate data are applied to data sets whose length N is a power of 2. This allows the use of the FFT to estimate the discrete spectrum. However, in the case of strongly periodic time series the length of the data set will almost never be an exact integer multiple of the dominant period. We describe a simple but effective algorithm to adjust the length of the data set such that it becomes an exact integer multiple of the length of the fundamental period (and its higher harmonics). We assume a time series V_t, $t = 1, 2, \ldots, N$. Now we define a frequency mismatch error E_{fmm} as folows:

$$E_{\mathrm{fmm}} = \sum_{i=1}^{m} (V_i - V_{t+i})^2. \tag{4}$$

We start with $t = N - m$, then decrease t in steps of 1, and calculate E_{fmm} for each value of t. The new endpoint of the time series N' is the value of t for which E_{fmm} reaches its first minimum. Intuitively, the length of the time series has been adjusted so that the beginning and the end will fit very smoothly. The goodness of fit can be adjusted by setting m. In this study we choose $m = 10$ which proved sufficient even in the case of periodic time series with very complex waveforms. As an extra bonus, the algorithm also solves the problem of the "jump phenomenon" between the beginning and the end of the time series [34,35]. Because N' will typically not be a power of 2, the DFT, instead of the faster FFT, must be used for generating phase-randomised surrogate data.

2.3. Test statistics

In the following we use three different test statistics to test for nonlinearity with phase-randomised

surrogate data (FFT- or DFT-based): the correlation integral; Kaplan's E_r statistic and nonlinear predictability. Originally, these measures were developed with specific goals in mind. In particular, the correlation integral was used to estimate dimensions and degrees of freedom of the dynamics; the E_r statistic was introduced to detect deterministic dynamics, and nonlinear forecasting was intended as a method to distinguish between different types of dynamics on the basis of their predictability. We should point out that in the present study we use these measures in a much more restricted way, namely only as statistics for nonlinearity. One advantage of this approach is that the assumptions behind the different methods become unimportant [12].

2.3.1. Time-delay embedding

Here we describe the time-delay embedding procedure which was used in combination with the three nonlinear statistics described below. We assume an observed discrete time series $V_t, t = 1, 2, 3, \ldots, N$. From this discrete time series, reconstructed state vectors X_t in an m-dimensional embedding space were obtained with the time-delay procedure [17,28,31]

$$X_t = (V_t, V_{t+L}, V_{t+2\times L}, V_{t+3\times L}, \ldots, V_{t+(m-1)\times L}). \tag{5}$$

Here L is the time delay (lag) and m is the embedding dimension. The sequence of vectors, or points, forms a reconstructed trajectory in state space. We used a combination of the procedure of Rosenstein et al. [27] to choose L, and the procedure of Kennel et al. [13] to choose m. An embedding dimension of 2 and a lag of 1 were taken as a starting point. Then the expansion ("unfolding") of the reconstructed state-space trajectory from the main diagonal in state space was calculated [27]. Subsequently, the lag was increased in steps of 1 until the rate of expansion of the attractor dropped below 40% of its initial value. Next, the percentage of false nearest neighbours was calculated. False nearest neighbours are vectors which lie close together in the reconstructed state space due to insufficient unfolding of the reconstructed trajectory, and not due to dynamic "correlations". Following Kennel et al. [13], two vectors X_i and X_j were considered false nearest neighbours when

$$\frac{|V_{[i+(m\times L)]} - V_{[j+(m\times L)]}|}{R_m(i, j)} > 15. \tag{6}$$

Here, the vertical bars denote absolute value and $R_m(i, j)$ is the Euclidean distance between the two vectors X_i and X_j in dimension m. (Please note that the numerator in (6) refers to $m + 1$.)

Next, the embedding dimension m was increased by 1, and the whole procedure (determining the optimum lag and then the percentage false nearest neighbours) was repeated. This process was continued until either: (1) the percentage false nearest neighbours dropped under 5%, or (2) the percentage false nearest neighbours no longer decreased with further increments of the embedding dimension. In the latter case, we used the value of m which gave the lowest percentage of false nearest neighbours. An impression of the actual choices of L and m made with this algorithm can be obtained from Figs. 2–4, where the values of L and m are indicated on top of the figure. The product of $L \times m$ ("embedding window") usually was on the order of three times the autocorrelation time. (Here and in the rest of the manuscript the autocorrelation time is defined as the time where autocorrelation function drops to $1/e$ of its initial value.)

2.3.2. Correlation integral

For a given r, the correlation integral C_r is defined as the fraction (between 0 and 1) of inter-vector distances less than or equal to r. The correlation integral C_r was determined for a range of 100 values of r (evenly scaled from the smallest to the largest inter-vector distance encountered in the data set) according to the following formula:

$$C_r = 1/N_p \sum_{i=1}^{N-W} \sum_{j=W}^{N} \theta(r - |x_i - x_j|). \tag{7}$$

Here N_p is the number of inter-vector distances considered, θ is the Heaviside unit function, and in this case, $|\ |$ indicates Euclidean distance. The Heaviside function is 0 if the distance between the vectors $> r$, and 1 if the distance between the vectors is $\leq r$. W is a correction factor for spurious influences of autocorrelation [33]. In this study W was chosen equal to three

times the autocorrelation time. In this study we used the correlation integral directly to test for nonlinearity. This was done by comparing C_r of the original time series and the surrogate data, and calculating z-scores for individual values of r (this type of approach, which avoids problems associated with actually calculating dimensions, was used by Takens [32]). Comparisons were made only for the 20 smallest values of r.

2.3.3. E_r statistic

The E_r statistic is a measure of the residual noise in a time series, described by Kaplan [11]. He assumed that if there is a deterministic mechanism underlying the time series, it will have at least the property of (piecewise) continuity. This property is tested by examining whether the images of two very nearby points in the state-space reconstructed from the time series will also by very close together.

We consider distances between pairs of vectors X_j, X_k, obtained from a time series of measurements by delay-time embedding as described above (5)

$$\delta_{j,k} = |X_j - X_k|. \tag{8}$$

The vertical bars denote the Euclidean distance. In [11] it is advised to exclude pairs where $|j-k|$ is small. We have solved this by requiring $|j-k| > 3\times$ autocorrelation time. The image of each vector pair is given by

$$\varepsilon_{j,k} = |X_{j+T} - X_{k+T}|. \tag{9}$$

The parameter T was taken as three times the autocorrelation time of the time series. The test statistic E_r is now defined as

$$E_r \equiv \overline{\varepsilon_{j,k}} \quad \text{for } j,k \quad \text{such that } \delta_{j,k} < r. \tag{10}$$

If a piecewise continuous map underlies the time series, one expects $\lim_{r\to 0} E_r = 0$. Alternatively, if the time series conforms to a linear-stochastic model one expects $\lim_{r\to 0} E_r = E_{\text{noise}}$. Here E_{noise} is the smallest value E_r in the stochastic time series, which equals the average noise level.

To test for small scale deterministic structure one plots E_r as a function of r both for the original time series, and for an ensemble of phase-randomised surrogate data. To reject the null hypothesis that the data can be explained by a random process, we have to prove that, for small r, E_r of the time series is significantly smaller than E_r of the surrogate data. This can be tested with z-scores, as described in Section 2.1. We should note that a significant z-score indicates nonlinearity, even when $E_r > 0$ for very small r. Thus E_r can be used as a statistic to test for nonlinearity, even when the data prove not to be (completely) deterministic.

2.3.4. Nonlinear forecasting

We used the algorithm described by Sugihara and May [30] for nonlinear forecasting with small modifications. Given a starting point in the time series V_t, we would like to predict V_{t+1}, V_{t+2}, V_{t+3}, etc. a number of steps ahead, and compare the predictions, which we will designate P_{t+n}, with the actual time series. First we used the time-delay procedure described above (5) to obtain m-dimensional vectors X_i from the time series V_t. However, for ease of reference, the time index of the vector X_i will now correspond with the last coordinate ($i = t + (m-1) \times L$). Now for each vector X_i we located the $m+1$ nearest neighbours in the m-dimensional state space. We will designate the k^{th} nearest neighbour vectors of X_i as $NN_{k,j}$. The k index indicates the number (from 1 to $m+1$) of the nearest neighbour; the j index its time index in the original time series. We excluded nearest neighbours with time indices j when $|i-j| < 3\times$ autocorrelation time. This procedure is called "within-sample" prediction. Sugihara and May used "out-of-sample" prediction. For a time series of length N, out-of-sample prediction requires $i > 0.5 \times N$ and $j < 0.5 \times N+$ constant. There are no fundamental differences between the two procedures, only within sample prediction may be more suitable for short data sets. Now the predicted value for n steps ahead prediction was given by

$$P_{(i+n)} = \sum_{k=1}^{k=m+1} V_{(k_j+n)} \times W_k. \tag{11}$$

Here w_k is the weight assigned to the kth nearest neighbour according to

$$w_k = \frac{|X_i - NN_{k,j}|^{-2}}{\sum_{k=1}^{k=m+1} |X_i - NN_{k,j}|^{-2}}. \tag{12}$$

Predictions for n ranging from 1 to 20 were made for a number of different V_t evenly distributed along the time series. Next, the Pearson correlation coefficients r between the actual time series V_{t+n} and the predicted values P_{t+n} were calculated and plotted as a function of n. The same procedure was repeated for the surrogate data, and z-scores were obtained by comparing the correlation coefficients for each prediction step. The final z-score used was the average z-score for the 20 prediction steps.

2.4. Test data

All test data time series had an initial length of 512 samples. A discrete sine wave was generated according to the following formula:

$$V_t = A \sin(2\pi(t/T_1)). \tag{13}$$

Here V_t is the amplitude at time t, and T_1 the period of the sine wave, expressed in sample steps. When $512/T_1$ is an integer, the sine wave matches exactly with the length of the time series; otherwise it does not (note that T_1 does not have to be an integer for the sine to match the time series length; in fact, when T_1 is an integer, only a very small number of different vectors can be reconstructed with time-delay embedding!). We explored the range $512/10 <= T_1 <= 512/9$. In the terminology of Section 1.2, the sine wave is an example of deterministic linear dynamics (it can be modelled with a linear second-order, ordinary differential equation).

A time series corresponding to a static nonlinear transformation of linear dynamics was obtained as follows:

$$V_t = 2 \times |A \sin(2\pi(t/T_1))|. \tag{14}$$

Here the vertical bars denote taking the absolute value. For T_1 a value of 52.3378 was used, to ensure V_t would not match with a time series length of 512.

A quasi-periodic signal with two incommensurate frequencies was generated according to

$$V_t = B \sin(2\pi(t/T_1)) + B \sin(2\pi(t/T_2)). \tag{15}$$

In this case $T_1 = 24.9904$, and $T_2 = T_1/\sqrt{3}$.

The Rossler system is defined by three coupled nonlinear differential equations:

$$\begin{aligned} dX/dt &= -(Y + Z), \\ dY/dt &= X + 0.2Y, \\ dZ/dt &= 0.2 + Z(X - C). \end{aligned} \tag{16}$$

We studied a time series of the X variable for different values of the control parameter $C = 3.5$; 4.5; 5.7 and 3.0 + dynamical noise. In the last case, Gaussian white noise was added to the differential equation for the X variable.

Baker and Gollub [4] give a model for a forced, damped pendulum:

$$\begin{aligned} d\omega/dt &= -\omega/q - \sin\theta + g\cos\varphi, \\ d\theta/dt &= \omega, \\ d\varphi/dt &= \omega_d. \end{aligned} \tag{17}$$

The parameters q and ω_d were fixed at $q = 2$ and $\omega_d = 2/3$. The time series of ω was studied for $g = 1.05$ (limit cycle with a simple shape); $g = 1.35$ (limit cycle with a more complex shape) and $g = 1.2$ (chaotic dynamics).

The logistic map

$$X_{t+1} = CX_t(1 - X_t) \tag{18}$$

was studied for $C = 3.6$, because, according to Theiler et al. [36], the system displays almost periodic chaotic behaviour for this parameter value.

The Henon system in a chaotic regime is given by

$$\begin{aligned} X_{t+1} &= 1.4 + 0.3Y_t - (X_t)2, \\ Y_{t+1} &= X_t. \end{aligned} \tag{19}$$

A time series of the X variable of the Henon was used.

The Lorenz system in a chaotic regime is given by

$$\begin{aligned} dX/dt &= 10(Y - X), \\ dY/dt &= X(28 - Z) - Y, \\ dZ/dt &= XY - 8/3Z. \end{aligned} \tag{20}$$

Finally, a simple linear stochastic model was studied,

$$X_{t+1} = 0.925(X_t + 1.07(X_t - X_{t-1}) + 0.25\varepsilon). \tag{21}$$

Here ε is white noise between -0.5 and 0.5. The parameters of this model were chosen to give an

almost periodic time series with a narrow band power spectrum.

Differential equations were solved with numerical integration using a fourth order Runge–Kutta algorithm [4]. Intial values were chosen randomly. The first 1024 iterations were discarded.

2.5. Experimental results with phase-randomised surrogate data

First, spurious detection of nonlinearity in the case of a simple sine wave was examined. Sine waves were generated according to (13). Twenty different sine waves were generated, with periods ranging from 512/10 to 512/9. Only for periods of 512/10 and 512/9 does the the sine match with the length of the time series (512 samples). For the 18 intermediate values studied, the time-series length is not an integer multiple of the sine period. Surrogate-data testing was applied to each of the 20 sine waves using three different statistic: the correlation integral, the E_r statistic and nonlinear prediction. In each case, 10 phase-randomised surrogate-data time series were generated using the full length of the time series and FFT-based surrogates. The resulting z-scores are shown in Fig. 2. Spurious detection of nonlinearity is absent only when the period of the sine matches the length of the time series exactly ($T_1 = 512/10$ or $T_1 = 512/9$). In almost all other cases, all three statistics spuriously detect significant nonlinearity. The artefact is already present (and significant) for a very small mismatch ($T_1 = 51.4844$; see Fig. 1.) but tends to become greater when the mismatch increases. In the case of a simple sine wave, testing for nonlinearity with FFT-based phase-randomised surrogate data will almost always produce spurious results, independent of the type of statistic used.

Next, we examined whether the use of length-adjusted DFT-based surrogates could prevent the spurious detection of nonlinearity. Test data with linear dynamics described under Section 2.5 were examined for nonlinearity using both FFT-based and LA-DFT-based surrogate data. In each case, 10 surrogate-data time series were generated. For the sine wave a period of 53.7600 was used (for this period length there is a large discrepancy with the time series length), and for the quasi-periodic signal periods of $T_1 = 24.9904$ and $T_2 = T_1/\sqrt{3}$. These periods

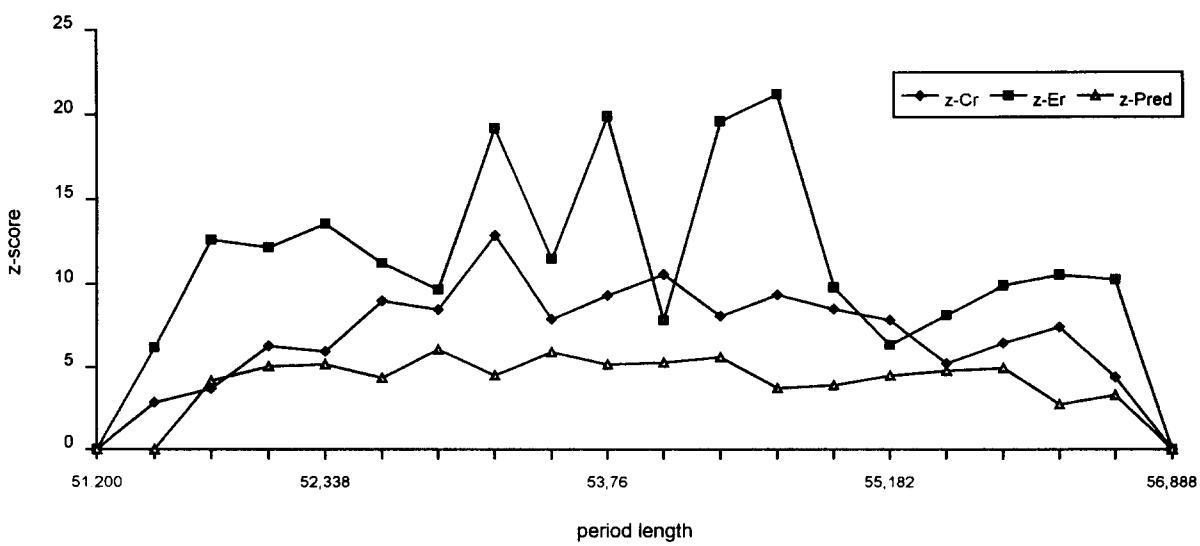

Fig. 2. Results of testing for nonlinearity with phase-randomised FFT-based surrogate data. The formula for the discrete sine wave was $V_t = A\sin(2\pi(t/T_1))$. Length of the time series was 512 samples. Period length T_1 (indicated on the X-axis) matches with time-series length only in the cases of period length of 51.200 (512/10) or 56.888 (512/9). Non-linearity was tested using the correlation integral (z-Cr), the E_r statistic (z-Er), and nonlinear predictability (z-Pred). In each case, z-scores were based upon 10 surrogate data. z-scores for three different statistics are shown on the Y-axis.

Table 1
Testing for nonlinearity with either FFT-based or length-adjusted DFT-based phase-randomised surrogate data

	Cr		Er		Pred	
	FFT	DFT	FFT	DFT	FFT	DFT
Sine	9.30	0.00	19.86	0.01	5.15	0.00
Sine-Abs.	9.56	0.00	12.03	1.88	7.37	0.00
Quasi-per.	4.20	1.46	9.20	2.38	5.59	0.00

Z-scores, based upon 10 surrogate data, using the correlation integral (C_r), the E_r statistic (E_r) or nonlinear predictability (Pred) as a test statistic. All time series had a length of 512 samples. Sine: sine wave with a period length of 53.7600; Sine Abs.: time series of absolute values of a sine with period length 52.3378; quasi-per.: quasi-periodic signal with period lengths of 24.994 and $24.994/\sqrt{3}$ samples.

were chosen such that a large mismatch between period length and time-series length was obtained. The results are shown in Table 1. For the sine wave, FFT surrogates produce spurious detection of nonlinearity with all three statistics, in agreement with Fig. 2. In this case, using the LA-DFT-based surrogate data no longer results in erroneous detection of nonlinearity with any of the three statistics. The sine wave with a static nonlinear transformation resulted in spurious detection of nonlinearity with all three statistics; with the LA-DFT surrogates this no longer occurred. In the case of the quasi-periodic signal, FFT surrogates produce false positive results with all three statistics. Although the z-scores decrease when LA-DFT surrogate data are used, only for the nonlinear prediction statistics is the correction successful ($z = 0.00$).

Finally, 20 different realisations of a simple linear stochastic model (21) were investigated. The parameters of this model were chosen such that the time series displayed strong periodicity and a narrow-band power spectrum. In all cases nonlinear prediction was used as a test statistic, either in combination with FFT-based surrogates, or in combination with LA-DFT surrogates. The results are shown in the last two columns of Table 4. Using a z-score cut-off of 1.96 (which corresponds to a false positive rate of 1 out of 20 for a normally distributed discriminated statistic) spurious detections of nonlinearity occurred in 9 out of 20 realisations of the linear model. Even with a conservative z-score cut-off of 3.00, spurious nonlinearity was still detected in 6 out of 20 realisations. Contrary to this, testing with the LA-DFT-based surrogate data only produced one z-score exeeding the 1.96 limit in accordance with statistical theory.

To summarise: use of FFT-based phase-randomised surrogate data typically results in spurios detection of nonlinearity in the case of sine waves, quasi-periodic signals and nearly periodic time series generated by a stochastic linear model. Using length-adjusted DFT-based phase-randomised surrogates prevents the spurious detection of nonlinearity in the case of a sine wave, realisations of a linear model, but not in the case of quasi-periodic signals.

3. Nonlinear cross-prediction of amplitude and time-reversed control signals

3.1. Theoretical background

In this section we describe a new test, which is aimed specifically at detecting weak nonlinearity in periodic or nearly periodic time series. Also, the test aims to avoid the problems associated with phase-randomised surrogate data. We will first discuss some theoretical considerations. A formal description of the test is given in Section 3.2.

The basic idea behind the test is that nonlinearity in a periodic time series can be reflected in asymmetry in the amplitude distribution, time irreversibility [6], or both. Asymmetries in amplitude distribution and time directionality may occur independently, and should be characterised separately. We introduce some formal notions to explain the relationship between time series irreversibility and nonlinearity. According to Tong [38] time reversibility (of a time series) can be formally defined as follows: A stationary time series $\{X_t\}$ is time reversible if for every positive integer n, and every $t_1, t_2, \ldots, t_n \in \mathbb{Z}$, the vectors $(X_{t1}, X_{t2}, \ldots, X_{tn})$ and $(X_{-t1}, X_{-t2} \ldots, X_{tn})$ have the same joint distributions. A stationary time series that is not time reversible is, by definition, time irreversible. Usually, X_t is not observed directly, but undergoes some static transformation, either in the system itself, or in the measuring process. Formally: the time series $\{X_t\}$ can be transformed instantaneously to another time series $\{Y_t\}$ by a 1–1 function

$g: Y_t = g(X_t)$ for each t. Now, Y_t is stationary and time reversible if and only if $\{X_t\}$ is stationary and time reversible. In other words, *the function g cannot induce time irreversibility in Y_t if it is not already present in X_t*. Any time series that is a realisation of a stationary, linear Gaussian model is time reversible, because the covariance functions of Gaussian distributions are symmetrical. Thus, if an observed time series Y_t is time irreversible, the dynamics underlying X_t cannot be explained by a stationary, linear Gaussian model. However, a non-Gaussian amplitude distribution of Y_t could be due to a static nonlinear transformation $g(X_t)$, and by itself is no proof of nonlinear dynamics. Time irreversibility of a time series is thus incompatible with a static nonlinear transformation of a stationary linear Gaussian process. Amplitude asymmetry indicates nonlinearity of the time series, but it could, at least in principle, be explained by a static nonlinear transformation of linear dynamics.

One should be aware that a symmetric signal (time reversible and amplitude symmetric) does not imply that the underlying system must be linear. An example of a nonlinear system which generates a reversible and amplitude symmetric signal is the Poincare oscillator. When the system moves on its attractor it generates a sine wave. Therefore, the amount of asymmetry of a signal does not express something like "the amount of nonlinearity of the underlying system". It is only that strong asymmetry of the signal is a strong indication for nonlinearity in the system.

3.2. Description of the method

The test consists of two elements: (A) a nonlinear model (actually: a local linear approximation) of the dynamics, which is based explicitly upon the original time series. We used the Sugihara and May [30] algorithm, but without any loss of generality other algorithms for nonlinear forecasting can be used, as long as they use the original time series explicitly to construct a nonlinear model. (B) Amplitude- and time-reversed versions of the data. The amplitude-reversed version is obtained as follows: $Y_t = 2(2 \times \langle X \rangle) - X_t$ for all t. Here X_t is the original time series, $\langle X \rangle$ is the mean of X_t, and Y_t is the amplitude-reversed time series. The time-reversed version is obtained as follows: $Y_t = X_{N-t+1}$ (for $t = 1, 2, \ldots, N$), where N is number of samples in the time series.

Now the model based upon the original time series is used to predict: (A) the original time series, (B) the amplitude-inverted version and (C) the time-reversed version. Because in the case of (B) and (C) the time series used for constructing the model and the time series that has to be predicted are *different*, we designate this procedure as *nonlinear cross-prediction (NLCP)*. In the specific case of the Sugihara and May algorithm for prediction, a plot of the correlation between predicted and actual time series as a function of prediction horizon is obtained for (A)–(C). Now, whatever the shape of (A), if the time series is *asymmetric* around its mean value, predictability of (B) will be less than for (A). Also, if the time series is *time irreversible*, predictability of (C) will be less than for (A). Then can be seen by comparing the predictability plots for (A)–(C).

We obtain simplified quantitative measures by averaging the correlation coefficient over the 20 prediction steps. This averaged correlation coefficient is designated "pred" for (A). We define "ama" (amplitude asymmetry) as the *difference* between the average correlation coefficient for (A) and (B), and "tir" (time irreversibility) as the *difference* between the averaged correlation coefficient for (A)–(C). Now if the time series is reversible and has a symmetric amplitude distribution, we expect ama $= 0$ and tir $= 0$. Amplitude asymmetry in the time series will result in ama > 0; time irreversibility in tir > 0. As indicated above, both amplitude asymmetry and time irreversibility are indications of nonlinearity of the time series. However, when amplitude asymmetry is found in the absence of time irreversibility this nonlinearity could, in principle, be due solely to a simple static nonlinear transformation of some underlying linear process. The presence of time irreversibility excludes this possibility, but cannot exclude more complex types of nonlinear transformations of linear dynamics.

3.3. Results with test signals

Nonlinear cross-prediction was applied to all the test data described in Section 2.5. For each time series,

pred, ama, and tir were calculated. The results are shown in Table 2. The sine wave and the quasi-periodic signal showed perfect predictability (pred = 1), and no sign of either static or dynamic nonlinearity (ama and tir <= 0) (Fig. 3). The sine wave with a static nonlinear transformation also showed a perfect predictability, clear amplitude asymmetry, but no time irreversibility. The Rössler system was examined for different values of the control parameter C. For C = 3.5 there is a mildly asymmetric limit cycle. Amplitude asymmetry (ama = 0.07) and time irreversibility (tir = 0.05) were detected. For C = 4.5 (Fig. 4) there is a more complex limit cycle. In this case, strong amplitude asymmetry and irreversibility were detected.

Table 2
Testing for nonlinearity using nonlinear cross-prediction

	Pred	Ama	Tir
Sine	1.00	0.00	0.00
Sine-Abs.	1.00	0.10	0.00
Quasi-per.	0.99	−0.01	0.00
Rossler 3.5	1.00	0.07	0.05
Rossler 4.5	1.00	0.34	0.35
Rossler 5.7	0.99	0.36	0.38
Rossler d.n.	0.97	0.16	0.13
Pend 1.05	1.00	0.13	0.11
Pend 1.35	1.00	0.24	0.03
Pend 1.2	0.80	0.00	0.12
Logist 3.6	0.98	0.05	0.06
Henon	0.44	0.41	0.34
Lorenz	0.89	−0.04	0.55

Pred, Ama and Tir: correlation coefficients between predicted and actual time series, averaged for predicting 1–20 samples ahead. In all cases the original time series is used to predict (1) the original time series (Pred); (2) the amplitude-reversed time series (Ama); (3) the time-reversed time series (Tir). All time series had a length of 512 samples. Sine: sine wave with a period length of 53.7600; Sine-Abs: time series of absolute values of a sine with period length 52.3378; quasi-per.: quasi-periodic signal with period lengths of 24.994 and $24.994/\sqrt{3}$ samples. Rossler: time series of X-variable; the second number indicates the value of the control parameter C. For C = 3.5 or 4.5, the time series is periodic; for C = 5.7 the time series is chaotic. Pend.: time series of the X-variable of the pendulum model. The second number is the control parameter C. For C = 1.05 or C = 1.35 the time series is periodic; for C = 1.2 the time series is chaotic. Logist: time series of logistic equation. Control parameter value is 3.6. Henon: time series of the X-variable of the Henon system. Lorenz: time series of the X-variable of the Lorenz system. Equations for all systems can be found under Section 2.5 in the text.

In the case of a chaotic Rössler (C = 5.7), there was a minimal drop in predictability (from 1.00 to 0.99), and clear amplitude asymmetry and irreversibility. Both of these could still be detected when dynamical noise was added to a Rössler limit cycle (C = 3). For simple and complex limit cycles, the pendulum time series also displayed both types of asymmetry. The chaotic pendulum showed a clear drop in predictability (from 1.00 to 0.80), and irreversibility in the absence of amplitude asymmetry. The logistic function showed weak amplitude asymmetry and irreversibility. The Henon showed very low predictability, and strong amplitude asymmetry and irreversibility. The Lorenz showed a strong irreversibility in the absence of amplitude asymmetry.

Next, the influence of measurement noise on the ability of the nonlinear cross-prediction algorithm to detect nonlinearity was examined. We considered time series of the Rossler system for three values of the control parameter C: 3.5, 4.5 and 5.7, corresponding with weak periodic nonlinearity, strong periodic nonlinearity and chaos. To each of these time series Gaussian white noise was added with a SNR (defined as average signal amplitude/average noise amplitude) varying from 100 to 1. The results are shown in Table 3. In all cases, predictability of the time series decreased with lower SNR, as expected. In the weakly nonlinear time series, amplitude asymmetry and irreversibility could still be detected with a SNR of 3. In the strongly nonlinear and chaotic time series, amplitude asymmetry and irreversibility could still be detected with a SNR of 2.

Finally, the test was applied to 20 different realisations of a simple linear model (21). The parameters of this model were chosen such that the resulting time series showed a strong periodicity and a single narrow peak in the power spectrum. The results are shown in Table 4 and an example is shown in Fig. 5. Pred ranged between 0.77 and 0.94, with an average of 0.88. The high predictability reflects the regular sinusoidal character of the time series. Ama was very low, ranging between −0.03 and 0.03, with an average of 0.00. Tir was even lower, ranging between −0.05 and 0.00, with an average of −0.02.

To summarise: the nonlinear cross-prediction test does not detect spurious nonlinearity in the case of

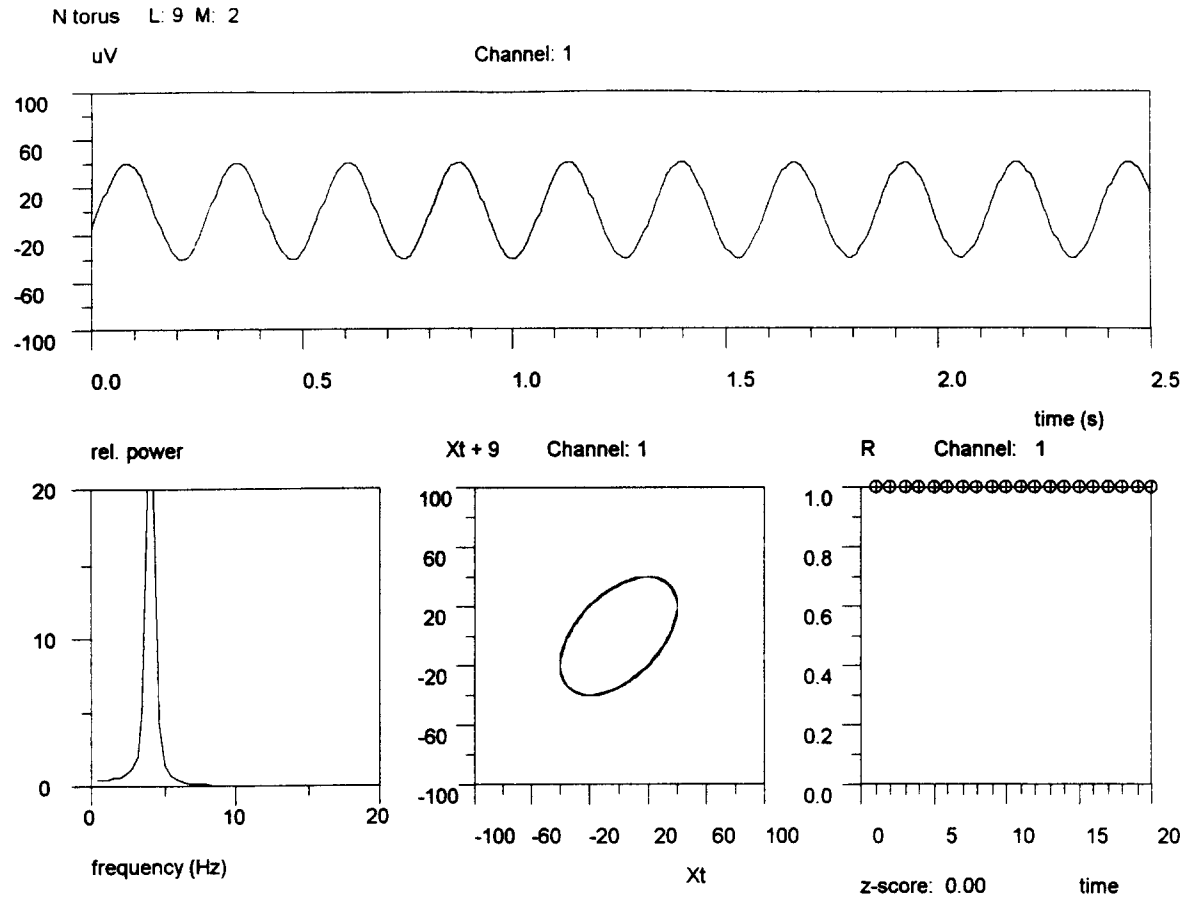

Fig. 3. Non-linear cross-prediction of a sine wave with period length 53.760. The upper panel shows the time series. The lower left panel shows the relative power spectrum. The panel in the middle shows the phase portrait. The panel at the lower right shows the nonlinear cross-prediction plot. The X-axis corresponds with the prediction horizon; the Y-axis is the correlation coefficient between predicted and actual time series. The plot with the open circles corresponds with prediction of the original time series. The continuous line corresponds with the prediction of the amplitude-inverted data. The plot with the vertical markers corresponds with the time-reversed data. In the case of a sine wave all three plots coincide, and predictability is perfect, independent of the prediction horizon.

a sine wave, a quasi-periodic signal, or a time series generated by a stochastic linear model. Nonlinearity which is solely due to a static transformation of a time series with linear dynamics is recognised as such (amplitude asymmetry only). For the time series presented here which correspond with a limit cycle or a chaotic attractor, time irreversibility indicating intrinsic nonlinearity is always detected. Some time series with time irreversibility also show amplitude asymmetry (Rossler, periodic pendulum, logistic map, Henon), but others do not (chaotic pendulum; Lorenz). The test performs well in the presence of considerable amounts of measurement noise.

4. Discussion

4.1. Phase-randomised surrogate data

The first part of the present study confirmed and extended results obtained earlier by Theiler et al. [36]. When phase-randomised surrogate data are

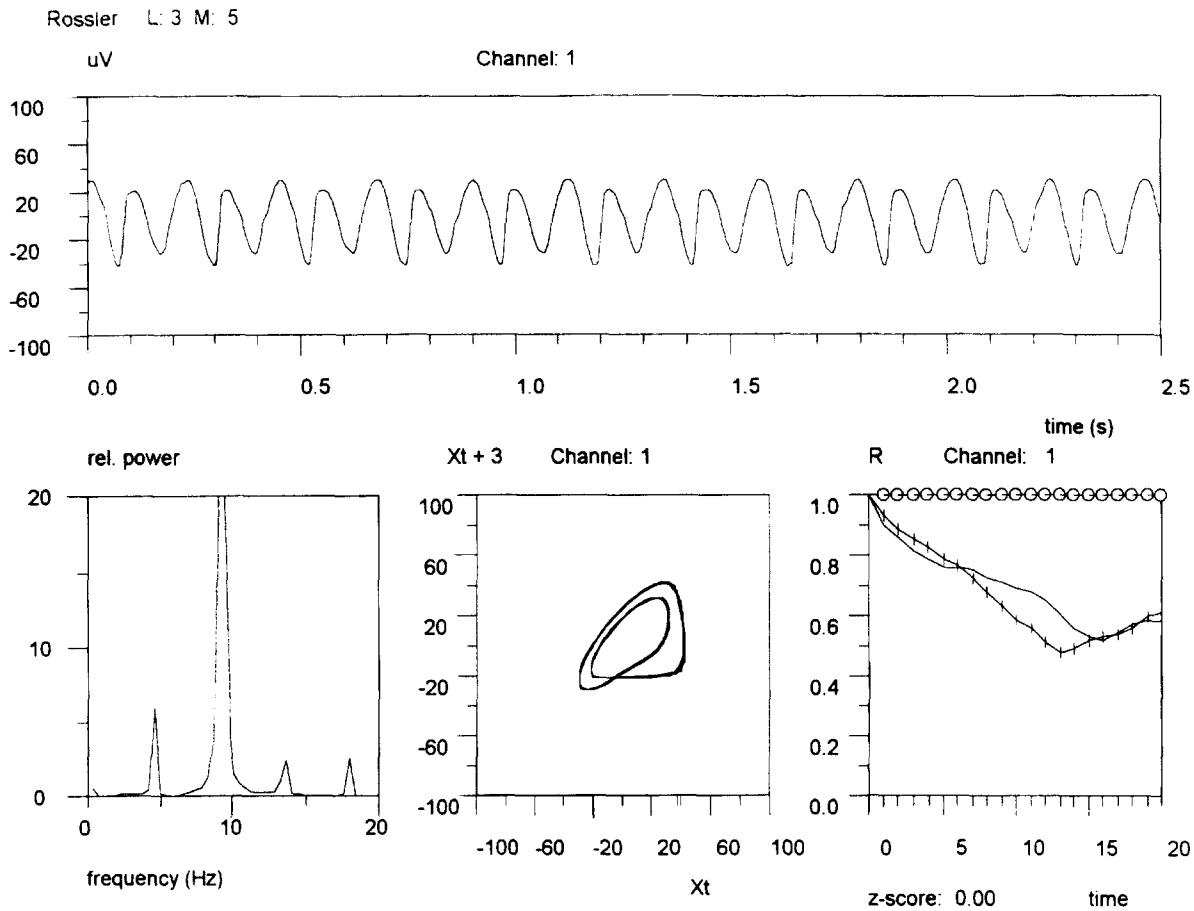

Fig. 4. Non-linear cross-prediction of a time series of the X variable of the Rossler system, for $C = 4.5$. For this value of the control parameter the time series is periodic, with a fairly complex waveform. Panels as in Fig. 3. The power spectrum shows a dominant frequency and a few (sub)harmonic frequencies. The phase portrait shows the complex shape of the limit cycle. The predictability of the original time series is perfect and does not depend on the prediction horizon. Predictability of amplitude-inverted and time-reversed data is substantially worse, indicating nonlinearity.

Table 3
Influence of additive Gaussian white noise on the ability of the nonlinear cross-prediction algorithm to detect nonlinearity in a time series ($N = 512$) of the X-variable of the Rössler system (C: control parameter)

SNR	Rössler $C = 3.5$			Rössler $C = 4.5$			Rössler $C = 5.7$		
	Pred	Ama	Tir	Pred	Ama	Tir	Pred	Ama	Tir
∞	1.00	0.07	0.04	1.00	0.31	0.34	0.99	0.39	0.43
100	1.00	0.07	0.04	1.00	0.31	0.34	0.99	0.45	0.51
50	1.00	0.07	0.04	1.00	0.33	0.36	0.99	0.44	0.50
25	1.00	0.07	0.04	1.00	0.30	0.33	0.98	0.44	0.48
10	0.99	0.07	0.04	0.99	0.30	0.31	0.98	0.43	0.45
5	0.97	0.07	0.04	0.97	0.30	0.30	0.93	0.34	0.32
4	0.96	0.05	0.02	0.93	0.23	0.25	0.89	0.34	0.34
3	0.93	0.06	0.05	0.89	0.20	0.20	0.86	0.33	0.32
2	0.82	0.04	0.01	0.78	0.21	0.21	0.71	0.19	0.14
1	0.50	0.01	−0.01	0.33	−0.03	0.06	0.41	0.06	0.01

Table 4
Testing for nonlinearity in 20 different realisations of a simple linear stochastic model

Realisation	Pred	Ama	Tir	z-FFT	z-LA-DFT
1	0.85	0.00	−0.04	−0.07	−0.14
2	0.81	−0.03	−0.04	1.94	1.27
3	0.86	0.01	−0.02	0.27	0.39
4	0.87	−0.02	−0.03	1.13	−0.66
5	0.82	0.02	−0.03	2.02	−1.07
6	0.88	0.02	0.00	3.41	−0.64
7	0.92	−0.01	−0.01	4.42	1.24
8	0.81	0.00	−0.05	0.30	−0.79
9	0.94	0.00	0.00	3.36	0.64
10	0.92	0.01	−0.01	1.13	−0.45
11	0.84	0.02	−0.03	2.30	−0.19
12	0.92	0.00	−0.02	3.51	0.00
13	0.92	0.01	−0.02	0.00	−0.27
14	0.85	0.01	−0.02	1.63	1.14
15	0.84	−0.02	−0.03	1.36	−0.16
16	0.77	0.03	−0.02	3.08	−0.24
17	0.93	0.01	−0.02	2.18	0.00
18	0.90	−0.01	−0.02	1.7	−2.13
19	0.94	0.00	−0.01	1.95	0.00
20	0.91	0.02	−0.02	5.68	−0.84

Results for the nonlinear cross-prediction test are shown in the second to the fourth column (Pred, Ama, and Tir). The fifth column shows the z-scores obtained using FFT-based surrogates and using the prediction statistic. The sixth column shows the same for LA-DFT-based surrogates.

constructed from periodic time series, and the periodicity does not match with the length of the time series, errors are introduced into the surrogate data and spurious detection of nonlinearity occurs. The problem affects sine waves, quasi-periodic signals, and narrow-band filtered noise. The cause of this artefact is that at least one of the frequencies in the time series does not match exactly with one of the discrete frequencies in the discrete spectrum. Consequently, the power is distributed over nearby frequencies, and the phases of these nearby frequencies are coupled. After phase-randomisation, the phases become uncoupled, and a beating phenomenon (not present in the original time series) results.

Using length-adjusted, DFT-based surrogate data solved the problem, but only for sine waves and narrow band filtered noise. In the case of a quasi-periodic signal the problem could not be solved. So adjusting the length of the time series to fit with the length of the basic periodicity in the signal does not constitute a universal solution to the problem. In fact, in the case of a quasi-periodic signal one can reason that the problem cannot be solved even in principle because two incommensurate frequencies can never both fit exactly with discrete frequencies in the spectrum (which are all integer multiples of the fundamental frequency).

In agreement with [36], we conclude that there is a fundamental problem when phase-randomised surrogate data are applied to detect nonlinearity in (nearly) periodic time series. One may question whether this result has any consequences for practical research. After all, completely noise-free sine waves, quasi-periodic signals and time series corresponding with limit cycles are unlikely to be encountered in experiments. Even in the exceptional case where one would obtain such a time series in an experiment, it would probably not be too difficult to recognise the nature of the time series with simple methods (visual inspection; phase portraits; powere spectra). In such cases the need for a formal test of nonlinearity may not arise in the first place. We should not forget that testing with phase-randomised surrogate data was originally introduced to investigate whether an aperiodic time series, with a broad-band spectrum, should best be modelled as a linear stochastic or deterministic chaos. In this context

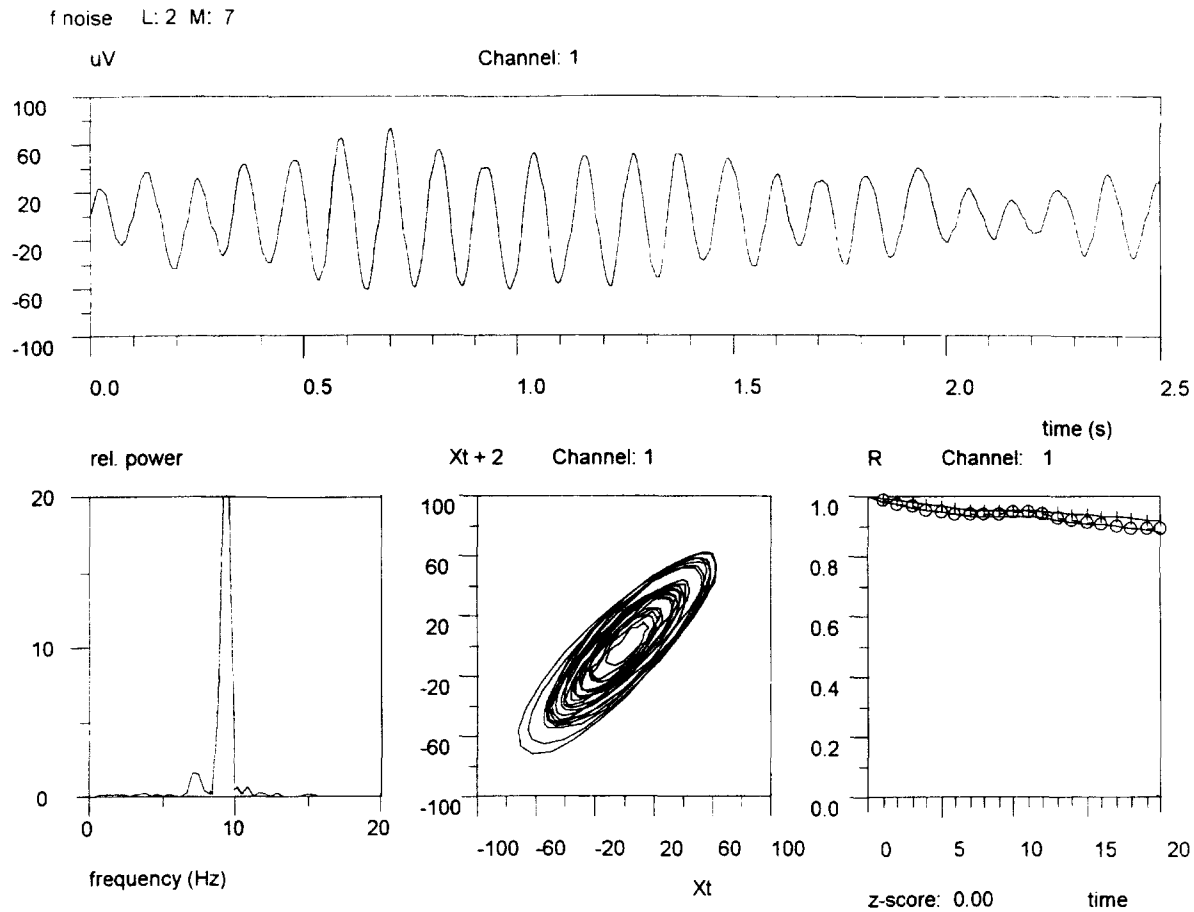

Fig. 5. Nonlinear cross-prediction of a time series generated by a stochastic linear model (21) with a very narrow power spectrum. The nature of the dynamics cannot be easily deduced from the phase portrait, which shows multiple nearby seemingly ordered trajectories. Predictability for the original time series is good for one or two steps ahead, but then slowly drops. The prediction plots for both the amplitude-inverted and time-reversed data coincide almost completely with the plot for the original data, indicating that the time series is linear.

the method works well, as is clear from the results obtained with chaotic data.

However, many time series in biology and medicine show noisy rather than perfect periodicity. Examples of such time series are the electrocardiogram, continuous records of blood pressure or flow velocity in cerebral arteries, and some types of EEG rhythms, notably some very regular types of 8–12 Hz alpha rhythm. In such cases, deciding whether the underlying dynamics is to be explained by a linear-stochastic model with narrow band filtering or nonlinear dynamics with some amount of dynamic and observational noise may not be a trivial problem. The fact that FFT-based surrogate data can and do produce spurious detection of nonlinearity in narrow-band filtered noise (Table 4) implies in our opinion that this method cannot be used to detect reliably nonlinear dynamics in physiological time series with noisy periodicity. So, the frequency mismatch artefact is not one which only arises with model signals, but it can be expected to play a serious role in the analysis of experimental time series. Even if it would be possible to construct perfect phase-randomised surrogate data for all types of data, this only leads to a more fundamental problem. Per definition, the phase-randomised surrogate data of any periodic time series is another periodic time series. Limit

cycles change shape, but still remain limit cycles after phase randomisation. A nice illustration of this can be found in [14, Figs. 25.6 B and C]. Consequently, nonlinear invariant measures will be similar for the original and surrogate time series, and the nonlinearity cannot be detected. These practical and theoretical problems motivated us to search for an alternative method.

4.2. Nonlinear cross-prediction

Nonlinear cross-prediction of amplitude and time-reversed control data did not give rise to spurious detection of nonlinearity in sine waves, a quasi-periodic time series and narrow band filtered noise. In all time series presented here with true nonlinearity in the underlying dynamics (limit cycle or chaotic attractor) time irreversibility was detected. Time irreversibility cannot be explained by simple static nonlinear transformations, and is inconsistent with a linear stationary stochastic model of the dynamics. Thus, time irreversibility may be taken as a measure of intrinsic nonlinearity in the time series. Unfortunately, the reverse is not true: systems with nonlinear dynamics can produce reversible time series. An example of such a system is the Poincare limit cycle. In this case, our test cannot detect the nonlinearity in the dynamics.

Asymmetry of the amplitudes around the mean value was never found in the time series with linear dynamics and was detected in some but not all of the time series with nonlinear dynamics. In and by itself amplitude asymmetry cannot be taken as proof of nonlinearity in the dynamics. Absence of amplitude asymmetry does not exclude nonlinear dynamics. Obviously, amplitude asymmetry is a much weaker finding than dynamic nonlinearity. However, it is perfectly possible that in some cases amplitude asymmetry in an experimental time series will have physiological meaning. Therefore we prefer to include detection of both amplitude asymmetry and irreversibility in the test.

Our test uses only two "surrogate" time series (amplitude-inverted and time-reversed) for each original time series (we should note, however, that these are surrogate data which preserve both the power spectrum and the amplitude distribution *exactly*), and thus does not allow a formal statistical test in the case of a single time series. However, when multiple records are available (and in many biological and medical experiments this is the typical rather than the exceptional case), our method can be used for inferential testing of hypothesis using conventional statistical tests (t-test; Mann–Whitney U-test; ANOVA, etc.). Furthermore, while statistics give information about the *likelihood* of a null hypothesis, it may sometimes also be useful to have information about the *magnitude* of the asymmetry (as provided by the ama and tir measures of the cross-prediction test). The test proposed by Diks et al. [6] and applied in van der Heyden et al. [10] gives information about the likelihood of the null hypothesis, but does not further characterise the data. The method of nonlinear cross-prediction always gives a characterisation of the data in terms of their predictability, irrespective of the absence or presence of nonlinearity.

4.3. Conclusion

Many interesting time series in biology and medicine show strong periodicity with some amount of dynamical and obervational noise. Deciding whether or not such time series reflect nonlinear dynamics is nontrivial, and requires the use of sophisticated time-series techniques. As shown in this paper, and previously in [36], the commonly used method of phase-randomised surrogate data may be unsuited for detecting nonlinearity in this type of data. In some cases the problem can be solved by adjusting the length of the time series, but this approach is cumbersome, and cannot be guaranteed to work in all cases. Moreover, from a purely theoretical point of view, phase-randomisation should not affect the invariant measures of periodic nonlinear time series.

The method of nonlinear cross-prediction, proposed in this paper, avoids spurious detection of nonlinearity, while it allows detection of nonlinearity in the case of limit cycle and chaotic systems that produce asymmetric signals, even in the presence of considerable amounts of observational noise. It may be important to re-investigate physiological time series with

strong periodic components which were previously claimed to show nonlinearity using the method of phase-randomised surrogate data. If such claims can be substantiated with methods like the one proposed in the present paper or in [6], we can exclude the possibility that such claims were based solely upon the frequency mismatch-artefact.

Acknowledgements

This study was inspired by many discussions on nonlinear dynamics and time-series analysis. We would like to acknowledge the historical contribution of Van der Tweel. Daniel Kaplan suggested, in essence, the LA-DFT algorithm. Cees Diks and Marcel van der Heijden gave valuable comments on parts of the manuscript. Finally, we would like to thank the anonymous reviewer for his valuable comments and suggestions.

References

[1] H.D.I. Abarbanel, Analysis of Observed Chaotic Data (Springer, New York, 1996).

[2] P. Achermann, R. Hartmann, A. Gunzinger, W. Guggenbühl and A.A. Borbély, All-night sleep EEG and artificial stochastic control signals have similar correlation dimensions, Electroenceph. clin. Neurophysiol. 90 (1994) 384–387.

[3] A.M. Albano and P.E. Rapp, On the reliability of dynamical measures of EEG signals, in: Proc. 2nd Ann. Conf. on Nonlinear Dynamical Analysis of the EEG, eds. B.H. Jansen and M.E. Brandt (World Scientific, Singapore, 1993) pp. 117–139.

[4] G.L. Baker and J.P. Gollub, Chaotic Dynamics. An Introduction (Cambridge University Press, Cambridge, 1996).

[5] J. Belair, L. Glass, U. van der Heiden and J. Milton, Dynamical Disease, Mathematical Analysis of Human Illness (AIP Press, Woodbury, New York, 1995).

[6] C. Diks, J.C. van Houwelingen, F. Takens and J. DeGoede, Reversibility as a criterion for discriminating time series, Phys. Lett. A 201 (1995) 221–228.

[7] J. Fell, J. Roschke and C. Schaffner, Surrogate data analysis of sleep electroencephalograms reveals evidence for nonlinearity, Biol. Cybernet. 75 (1996) 85–92.

[8] L. Glass, D.T. Kaplan and J.E. Lewis, Tests for deterministic dynamics in real and model neural networks, in: Proc. 2nd Ann. Conf. on Nonlinear Dynamical Analysis of the EEG, eds. B.H. Jansen and M.E. Brandt (World Scientific, Singapore, 1993) pp. 233–249.

[9] L. Glass and M.C. Mackey, Pathological conditions resulting from instabilities in physiological control systems, Ann. N.Y. Acad. Sci. 316 (1979) 214–235.

[10] M.J. van der Heyden, C. Diks, J.P.M. Pijn and D.N. Velis, Time irreversibility of intracranial human EEG recordings in mesial temporal lobe epilepsy, Phys. Lett. A 216 (1996) 283–288.

[11] D.T. Kaplan, Exceptional events as evidence for determinism, Physica D 73 (1994) 38–48.

[12] D.T. Kaplan and L. Glass, Understanding Nonlinear Dynamics (Springer, New York, 1995).

[13] M. Kennel, R. Brown and H. Abarbanel, Determining embedding dimension for phase space reconstruction using a geometrical reconstruction, Phys. Rev. A 45 (1992) 3403–3411.

[14] F.H. Lopes da Silva, The generation of electric and magnetic signals of the brain by local networks, in: Comprehensive Human Physiology, eds. R. Greger and U. Windhorst, Vol. 1 (Springer, Berlin, 1996) pp. 509–531.

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

[16] J. Müller-Gerking, J. Martinerie, S. Neuenschwander, L. Pezard, B. Renault and F.J. Verela, Detecting nonlinearities in neuro-electric signals: A study of synchronous local field potentials, Physica D 94 (1996) 65–91.

[17] N. Packard, J. Crutchfield, D. Farmer and R. Shaw, Geometry from a time series, Phys. Rev. Lett. 45 (1980) 712–715.

[18] M. Palus, Testing for nonlinearity in the EEG, in: Proc. 2nd Ann. Conf. on Nonlinear Dynamical Analysis of the EEG, eds. B.H. Jansen and M.E. Brandt (World Scientific, Singapore, 1993) pp. 100–114.

[19] L. Pezard, J.L. Nandrino, B. Renault, F. El Massioui, F. Allilaire, J. Muller, F.J. Varela and J. Martinerie, Depression as a dynamical disease, Biol. Psychiatry 39 (1996) 991–999.

[20] J.P.M. Pijn, Quantitative evaluation of EEG signals in epilepsy: Nonlinear association time delays and nonlinear dynamics, Ph.D. Thesis, University of Amsterdam (1990).

[21] J.P.M. Pijn, J. van Neerven, A. Noest and F.H. Lopes da Silva, Chaos or noise in EEG signals; dependence on state and brain site, Electroenceph. clin. Neurophysiol. 79 (1991) 371–381.

[22] W.S. Pritchard, D.W. Duke and K.K. Krieble, Dimensional analysis of resting human EEG II: Surrogate data testing indicates nonlinearity but not low-dimensional chaos, Psychophysiology 32 (1995) 486–491.

[23] P.E. Rapp, Chaos in the neurosciences: Cautionary tales from the frontier, Biologist 40 (1993) 89–94.

[24] P.E. Rapp, A.M. Albano, T.I. Schmah and L.A. Farwell, Filtered noise can mimic low-dimensional chaotic attractors, Phys. Rev. E 47 (1993) 2289–2297.

[25] P.E. Rapp, A.M. Albano, I.D. Zimmerman and M.A. Jimenez, Phys. Lett. A 192 (1994) 27–33.

[26] S.A.R.B. Rombouts, R.W.M. Keunen and C.J. Stam, Investigation of nonlinear structure in multichannel EEG, Phys. Lett. A 202 (1995) 352–358.
[27] M.T. Rosenstein, J.J. Collins and C.J. De Luca, Reconstruction expansion as a geometry-based framework for choosing proper delay times, Physica D 73 (1994) 82–98.
[28] T. Sauer, J.A. Yorke and M. Casdagli, Embedology, J. Stat. Phys. 65 (1991) 579–616.
[29] C.J. Stam, B. Jelles, H.A.M. Achtereekte, S.A.R.B. Rombouts, J.P.J. Slaets and R.W.M. Keunen, Investigation of EEG non-linearity in dementia and Parkinson's disease, Electroenceph. clin. Neurophysiol. 95 (1995) 309–317.
[30] G. Sugihara and R.M. May, Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, Nature 344 (1990) 734–740.
[31] F. Takens, Detecting strange attractors in turbulence, Lecture Notes in Math. 898 (1981) 366–381.
[32] F. Takens, Detecting nonlinearities in stationary time series, Int. J. Bifurc. and Chaos 3 (1993) 241–256.
[33] J. Theiler, Spurious dimension from correlation algorithms applied to limited time-series data, Phys. Rev. A 34 (1986) 2427–2432.
[34] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian and J.D. Farmer, Testing for nonlinearity in time series: The method of surrogate data, Physica D 58 (1992) 77–94.
[35] J. Theiler, B. Galdrikian, A. Longtin, S. Eubank and J.D. Farmer, Using surrogate data to detect nonlinearity in time series, in: Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity, Proc. Vol. XII, eds. Casdagli and S. Eubank (Addision-Wesley, Reading, MA, 1992) pp. 163–188.
[36] J. Theiler, P.S. Linsay and D.M. Rubin, Detecting nonlinearity in data with long coherence times, in: Time series prediction, Forecasting the future and understanding the past, SFI Studies in the sciences of Complexity, Proc. Vol. XV, eds. A.S. Weigend and N.A. Gershenfeld (Addison-Wesley, Reading, MA, 1993).
[37] J. Theiler and P.E. Rapp, Re-examination of the evidence for low-dimensional, nonlinear structure in the human electroencephalogram, Electroenceph. clin. Neurophysiol. 98 (1996) 213–222.
[38] H. Tong, Non-linear Time series, A Dynamical Systems Approach (Oxford University Press, New York, 1995).