On the Characterisation of the Deterministic/Stochastic and Linear/Nonlinear Nature of Time Series

Temujin Gautama[*], Marc M. Van Hulle[*] and Danilo P. Mandic[†]

[*]Laboratorium voor Neuro- en Psychofysiologie
K.U.Leuven, Campus Gasthuisberg, Herestraat 49
B-3000 Leuven, BELGIUM

[†]Department of Electrical and Electronic Engineering
Imperial College of Science, Technology and Medicine
Exhibition Road, SW7 2BT, London, UK

April 29, 2004

Abstract

Most statistical signal nonlinearity analyses adopt the Monte-Carlo approach proposed by Theiler and co-workers, namely the 'surrogate data' method. A surrogate time series, or 'surrogate' for short, is generated as a realisation of the null hypothesis of linearity. A measure ('test statistic') is computed for the original time series and it is compared to those computed for an ensemble of surrogates. If the test statistic computed for the original is significantly different from that computed for the surrogates, the null hypothesis is rejected, and the original time series is judged nonlinear. One of the key issues in signal nonlinearity analysis is the definition of a linear signal. The standard definition is that such a signal is generated by a Gaussian linear stochastic process. This definition, however, is very stringent. Indeed, if a linear signal were to be measured via a zero-memory, nonlinear observation function, or if the driving noise were not Gaussian, the test for linearity would fail, and the signal would be interpreted as nonlinear. Therefore, we extend the definition of linearity to incorporate these *uninteresting* (Theiler *et al.*, 1992) deviations in the null hypothesis and, consequently, the method for generating the surrogate data. We propose a novel method for characterising a time series, the 'Delay Vector Variance' (DVV) method, from which a novel test statistic can be derived. It is shown that, in the context of surrogate data testing, it outperforms a number of established nonlinearity measures. It is based upon the local unpredictability of a time series, which is analysed in a standardised manner, and allows both for a straightforward visualisation, and for a quantitative measure of the nonlinearities present in a time series. A comprehensive comparative study is performed with other nonlinearity analysis methods in a systematic manner and for a wide variety of benchmark time series.

1

Contents

1 Introduction

In real-world applications of statistical signal processing, such as in adaptive filtering or signal/system modelling, it is desirable to verify the presence of an underlying linear or nonlinear signal generation system, *before* the actual filters or models are constructed. Indeed, in the absence of nonlinear behaviour, it is not favourable to use nonlinear models, since these are more difficult to train than their linear counterparts, due to issues such as overfitting and computational complexity. It is important to note that in these cases, the objective is the detection of possible nonlinearity of the *system*, or equivalently of the transfer function. In other applications, for instance the analysis of biomedical signals, such as heart rate variability (HRV), electrocardiogram (ECG), hand tremor and electro-encephalogram (EEG), there is a need to assess the presence or absence of nonlinear behaviour within the *signal*, as opposed to that of the system, because the linear/nonlinear nature of the signal conveys information concerning the health condition of a subject (for an overview, see Schreiber, 1999).

There exist several established methods for performing signal nonlinearity analysis, the differences among them stemming from the underlying assumptions, which strongly impact their outcomes and efficacy. Consequently, the outcome of a test, *e.g.*, the rejection of a null hypothesis of linearity, needs to be interpreted with due caution (see, *e.g.*, Schreiber and Schmitz, 2000; Timmer, 2000). The variety of approaches and the difficulty of interpretation of the results, clearly indicate the need for a unifying approach with straighforwardly interpretable results.

The purpose of this technical report is to introduce such unfying methodology for a comprehensive characterisation of the nature of time series. Before introducing the novel method, we first provide a short tutorial-like overview of the basic concepts, as well as the current state-of-the-art in the domain of signal nonlinearity analysis. The analysis is supported by rigorous simulations on both benchmark and real-world signals.

2 Definitions

Before focusing on patricular methods, we deem it useful to introduc some of the basic concepts and notions from time series analysis.

2.1 System and Signal Nonlinearity

A linear shift-invariant **system**, $f(\cdot)$, is defined as one that obeys the superposition and scaling property, namely for $a, b \in \mathcal{R} : f(ax + by) = af(x) + bf(y)$, together with producing identical outputs for a given input at different instants of time. A system which is shift-invariant, but which violates thesuperposition property is considered nonlinear. In its own right, this allows for a very powerful tool for assessing system nonlinearity, referred to as 'temporal summation', as has been applied, *e.g.*, in the field of neuroimaging (Boynton *et al.*, 1996; Miller *et al.*, 2001; Mechelli *et al.*, 2001), whereby a short and a long pulse are presented to the system, and a system is judged linear if the response to the long pulse can be predicted from a summation of temporally shifted versions of the response to the short stimulus. However, the principle of temporal summation for analysing the nonlinearity of a system implies that the system input is systematically varied, while in typical real-world settings, this is not favourable, or is even physically impossible.

A linear signal, x, is generally defined as the output of a linear shift-invariant system that is driven by Gaussian, white noise. However, in most cases, this definition is somewhat relaxed by allowing the probability distribution of the signal values (the *signal distribution*) to deviate from the Gaussian one. This can be interpreted as a linear signal (following the strict definition) measured by a static (possibly nonlinear) observation function, $h(\cdot)$. Any signal which cannot be generated in such a way is generally referred to as a nonlinear signal. The analysis of the nonlinearity of a signal can often provide insights into the nature of the underlying signal production system. However, care should be taken in the interpretation of the results, since the assessment of nonlinearity within

a signal does not necessarily imply that the underlying signal generation system is nonlinear: the input signal and system (transfer function) nonlinearities are confounded.

2.2 Properties of a Linear Signal

Recall that a linear (shift-invariant) system driven by white Gaussian noise gives rise to a linear signal. Notice, however, that this only yields information about the generation mechanism and the underlying system, and not on the generated signal itself. Therefore, when presented with a time series, the generating mechanism of which is unknown, this definition cannot be used for assessing signal (non-) linearity. For this purpose, before proceeding on to the detection of nonlinear behaviour in a signal, the properties of a linear signal are briefly summarised, since these provide a means for testing whether or not the time series is linear.

A linear signal can be generated by an autoregressive (AR) model[1] driven by normally distributed, white (*i.e.*, uncorrelated over time) noise. An AR-model (linearly) processes a signal by 'shaping' the amplitude spectrum of the input signal. Therefore, a linear signal is obtained by filtering a Gaussian white noise source, and can be characterised by means of its amplitude spectrum. Indeed, since an AR-model can be described by its 'amplitude spectrum shaping' property, and since white noise has a flat amplitude spectrum, the amplitude spectrum of a linear signal conveys all necessary information for determining the parameters of the underlying AR-model. In this context, the amplitude spectrum can be interpreted as a grossly overfit parameterisation of the AR-process, using $n/2$ parameters, where n is the number of samples in the time series (Theiler and Prichard, 1996). The importance of this observation will become clear in Section 3.

Another consequence of this observation is that the phase spectrum is irrelevant for the characterisation of a linear signal. Indeed, the phase relations in a white noise source are random, and do remain random after having been filtered with an AR-model. This intuitively suggests that signals containing sharp transitions, such as the discontinuities in a square wave (see Fig. 1A), are unlikely to be linear: in the frequency domain, such transitions correspond to a *rich* harmonic content with delicately aligned phases. Disrupting this phase alignment will have a strong impact on the signal 'shape'. This is illustrated in Fig. 1C, where a phase randomised version of the square wave is shown, *i.e.*, the amplitude spectrum is retained and the phase relations are randomised. It is evident that the signal's shape has completely changed. Intuitively, if the signal were linear (as is the case for a sine wave, see Fig. 1B), the visual *appearance* of the time series would remain the same (Fig. 1D).

In the remainder of this report, we adopt the terminology of Theiler *et al.* (1994) and use the term 'linear properties' to refer to the mean, variance and autocorrelation function of a time series. The latter is related to the amplitude spectrum of a time series, due to the well-understood Wiener-Khintchin theorem, which states that the autocorrelation function is equal to the inverse Fourier transform of the Power Spectral Density (PSD).

2.3 Deterministic and Stochastic Nature

Apart from the signal linearity and nonlinearity, there are several other properties which can be used for characterisation purposes, two of which we shall briefly describe in particular.

The Wold decomposition theorem (Wold, 1938) states that any discrete, stationary signal can be decomposed into its *deterministic*[2] and *stochastic* (random) components, which are statistically uncorrelated. This theorem forms the basis of many prediction models, since the presence of a deterministic component imposes an upper bound on the performance of these models. Consider, *e.g.*, a sine wave (deterministic) contaminated with white noise (stochastic). In a prediction

---

[1]An autoregressive model of order m is a linear stochastic model of the form: $x_k = \sum_{i=1}^{m} a_i x_{k-i} + \nu$: the current value of the process, x_k, is expressed as a finite, linear combination of the previous values of the process and a random shock, ν.

[2]A deterministic signal is one for which the generation process can be described precisely by a set of equations.
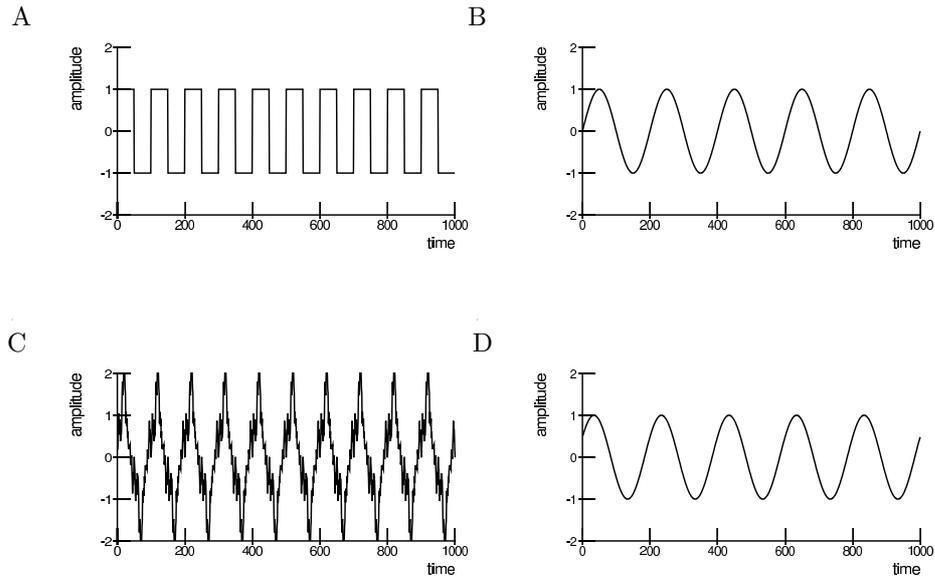
Figure 1: Illustration of the absence of the signal linearity property: A) Square wave; B) Sine wave; C) Phase-randomised version of the square wave; D) Phase-randomised version of the sine wave.

setting, the sine wave portion of the signal can be perfectly predicted using only two preceding samples. However, the prediction performance is degraded due to the presence of the stochastic component, and the portion of the variance that can be accurately predicted equals the variance of the deterministic component (sine wave). In this report, the deterministic component refers to the component of the signal that can be predicted from a number of previous time samples, whereas the stochastic component refers to the component for which such prediction is impossible. Note that this definition makes no statement regarding the number of previous samples required, nor of the model that should be used for the prediction.

Schreiber (1999) has illustrated the deterministic/stochastic and linear/nonlinear properties by interpreting them as conceptual axes, spanning a space in which both time series and modelling approaches can be positioned. His main objective was not to use this space for characterisation purposes, but rather to emphasise that these axes are independent: since determinism is one of the necessary requirements for the presence of deterministic chaos[3], it has often (mistakenly) been used as an indication of nonlinearity. Furthermore, many of the early nonlinearity analysis methods are based upon the methods for analysing chaos. As a consequence, properties such as determinism and the presence of a strange attractor have often been confounded with nonlinear behaviour, *e.g.*, the presence of a strange attractor would lead to the conclusion that the time series is nonlinear, while this is not necessarily so. Therefore, as pointed out by Schreiber and Schmitz (2000), it is important to know what the assumptions of a nonlinearity analysis are, especially regarding deterministic chaos, so as not to confuse cause and effect (chaos implies nonlinearity, but not *vice versa*). A similar interpretation of the assessment of nonlinearity is given by Theiler *et al.* (1992), where it is seen as a first step towards the detection of chaos.

---

[3]Several requirements need to be satisfied for a signal to be chaotic, among which determinism, nonlinearity, and the presence of a strange attractor are the most important ones.

3 Surrogate Data Method

Theiler *et al.* (1992) have introduced the concept of 'surrogate data', which has been extensively used in the context of statistical nonlinearity testing. The surrogate data method tests for a statistical difference between a test statistic computed for the original time series and for an ensemble of test statistics computed on linearised versions of the data, the so-called 'surrogate data', or 'surrogates' for short. In other words, a time series is nonlinear if the test statistic for the original data is not drawn from the same distribution as the test statistics for the surrogates. This is basically an application of the 'bootstrap' method in statistics (Theiler *et al.*, 1992). In the context of nonlinearity testing, the surrogates are a realisation of the *null hypothesis* of linearity. Although this is conceptually appealing, care should be taken when it comes to a correct interpretation of the results. There are three major aspects of the surrogate data method that need to be considered: 1) the exact definition of the null hypothesis; 2) the realisation of the null hypothesis, *i.e.*, the generation method for the surrogate data; and 3) the test statistic. The first two aspects will be explained briefly, mainly following Theiler and Prichard (1996) and Schreiber and Schmitz (2000). For the third aspect, the basic assumptions (chaotic behaviour, determinism, *etc.*) should be examined thoroughly for every test statistic separately, since these determine the type of deviations from the null hypothesis that can be detected. It is crucial to realise that the rejection of a null hypothesis conveys no information regarding to what aspect of the null hypothesis is violated (see, *e.g.* Timmer, 2000).

There are two main types of null hypotheses: *simple* and *composite*. A simple null hypothesis asserts that the data is generated by a specific and known (linear) process. A composite null hypothesis asserts that the unknown underlying process is a member of a certain *family* of processes. An example of a simple hypothesis is that data are drawn from a Gaussian distribution with zero mean and unit variance, whereas an example of a composite null hypothesis is that they are drawn from *a* Gaussian distribution.

3.1 Simple Null Hypothesis

The simple null hypothesis of linearity is that it is generated by a specific linear stochastic process, driven by white Gaussian noise. Thus, for the generation of surrogate data, the appropriate AR-model can be determined, after which it can be used for generating several surrogate time series. The appropriate AR-model is found by subsequently fitting AR-models of increasing order p, observing the squared estimation error $E(p)$, and optimising for an information theory-based model order selection criterion, such as the minimum description length (MDL, Rissanen, 1978):

$$\text{MDL}(p) = N \ \log_{10}(E(p)) + p \ \log_{10}(N), \tag{1}$$

where N is the number of samples. This way, the optimal model order criterion is penalised for computational complexity (roughly speaking, the increase in performance should exceed that expected by the complexity increase of the AR-model).

To illustrate this strategy, consider the first coordinate of the chaotic Lorenz series, shown in Fig. 2A. The optimal model order is determined by gradually fitting increasing-order AR-models to the data (for details, see *e.g.*, Chapter 3 in Mandic and Chambers, 2001). The model order cost function is corrected for complexity using the MDL method, yielding an optimal model of order 3 (Fig. 2B). Figure 2C shows a realisation of the optimal AR-model. Although the original time series is very smooth, the surrogate is rather 'edgy'. This is due to a mismatch between the first and last sample of the time series[4]. As suggested by Theiler *et al.* (1992), a simple *ad hoc* approach

---

[4]The estimation of the AR-model parameters uses the Yule-Walker equations, in which the autocorrelation function is computed as the biased sample autocorrelation. A mismatch in begin- and endpoint introduces false peaks in the sample autocorrelation function, which in turn distorts the power spectral density (PSD) due to the Wiener-Khintchin theorem (see Section 2.2). The so introduced (high) frequency components result in an 'edgy' time series.
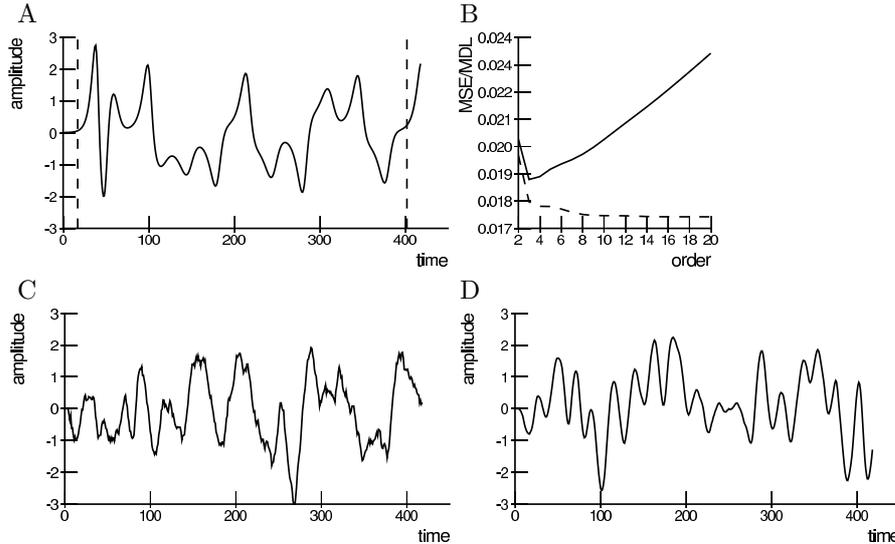
Figure 2: A) First coordinate of a realisation of the Lorenz series. The dashed lines demarcate the region selected by the endpoint matching; B) AR-model order estimation: prediction error (dashed line) and MDL-corrected error (solid line); C) AR-based surrogate time series; D) AR-based surrogate time series with endpoint matching.

can be used to circumvent this problem, namely by choosing a time slice in which the begin- and endpoint closely match. This approach has been applied to the Lorenz series and the resulting surrogate is shown in Fig. 2D. The original time series is first cropped to the time slice between the dashed lines in Fig. 2A. It is evident that the surrogate time series resembles the original time series closer than does that in Fig. 2C.

Although the AR-based approach has the advantage of easy implementation, together with the fact that surrogate time series of any length can be generated, one important side-effect is that the signal distribution of the surrogates becomes approximately Gaussian[5]. Therefore, if the amplitudes of the original time series do not follow a normal distribution, a rejection of the null hypothesis can be due to a discrepancy in signal distribution, rather than to actual nonlinear behaviour.

3.2   Composite Null Hypothesis

One possible composite null hypothesis is that the time series is generated by a linear stochastic process driven by Gaussian white noise, constrained to produce a time series with an autocorrelation function identical to that of the original time series. Due to the Wiener-Khintchin theorem (see Section 2.2), this constraint can be approximated by forcing the original and surrogate time series to have identical amplitude spectra. Note that this is in line with the observation that the phase spectrum is irrelevant for the characterisation of a linear signal, as described in Section 2.2. This allows for a straightforward method for generating surrogate data, namely by phase randomising the Fast Fourier Transform (FFT) of the original time series and retransforming it to the time domain. Indeed, in this way, the 'FT-based' surrogate is constrained to have the same amplitude spectrum thus, to have linear properties (mean, variance and autocorrelation function, see Section 2.2) identical to those of the original time series, but is otherwise random. This is referred to as

---

[5]Indeed, if we assume that the output of the AR-model is stationary (otherwise it would not be stable), and, thus, that the signal distribution remains constant, the AR-model computes the sum of a number of data points drawn from a single distribution. For higher model orders, say exceeding five, the signal distribution of the resulting signals would be approximately Gaussian due to the central limit theorem.
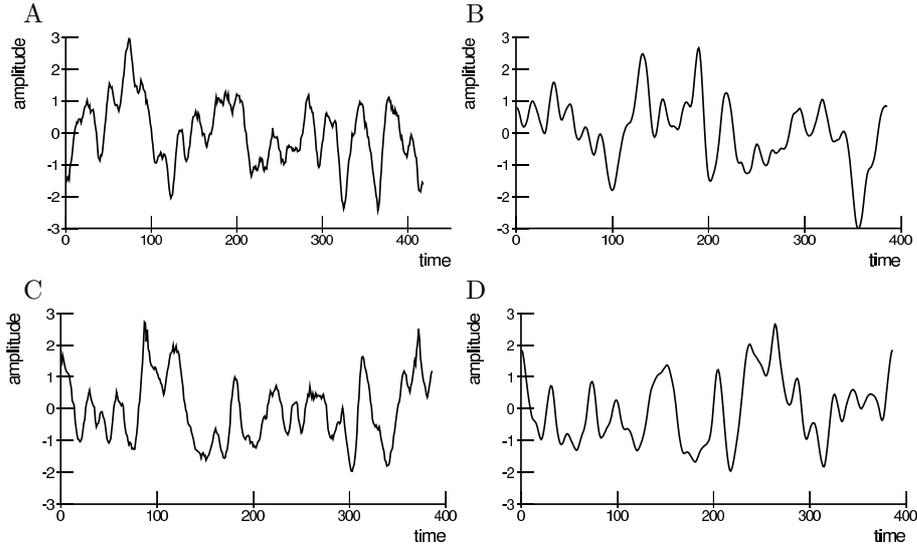
Figure 3: Example surrogates of the Lorenz series: A) FT-based without endpoint matching; B) FT-based with endpoint matching; C) AAFT-based with endpoint matching; D) iAAFT with endpoint matching.

the Fourier Transform (FT) surrogate data method.

In the context of FFT, the endpoint matching procedure is straightforward to interpret: since the FFT assumes the time series to be periodic over the time window under consideration, a mismatch between begin- and endpoint results in a periodic discontinuity, which introduces high frequency components (in classical signal processing, windowing is applied to compensate for this effect of so-called spectral leakage). Examples of FT-based surrogates for the Lorenz series without and with endpoint matching are shown in Fig. 3A and 3B.

As was the case for the AR-based method (Section 3.1), the signal distributions of the surrogates do not necessarily resemble that of the original time series, which can lead to rejections of the null hypothesis. In order to exclude such 'false' rejections, several methods have been proposed to extend the composite null hypothesis to that of an underlying linear stochastic process, driven by white Gaussian noise, *and* followed by a memoryless ('static'), monotonic and possibly nonlinear observation function $h(\cdot)$. The latter element can be used for rendering the signal distributions of the surrogate and the original time series identical. Theiler *et al.* (1992) proposed an amplitude transform of the original time series such that the distribution becomes Gaussian, *prior* to the FT method, and retransforming it to the original distribution afterwards (Amplitude Adjusted Fourier Transform, or AAFT method). Rather than fitting the observation function, $h(\cdot)$, with a parametric model, they employed a rank-ordering procedure, *i.e.*, the time series is sorted[6] and the sample with rank k is set to the same value as the k-th sample in a sorted Gaussian series of the same length as the original time series. An example for the Lorenz series, including the endpoint matching procedure, is shown in Fig. 3C.

The iterative Amplitude Adjusted Fourier Transform Method

However, Schreiber and Schmitz (1996) have shown that the AAFT method biases the amplitude spectrum of the surrogate towards a slightly flatter one than that of the original time series, which, again, can lead to false rejections of the null hypothesis. Therefore, Schreiber and Schmitz (1996) proposed a fixed point iteration scheme, referred to as the iterative Amplitude Adjusted Fourier Transform (iAAFT) method, which produces surrogates with identical ('correct') signal distributions and approximately identical amplitude spectra as the original time series, or *vice*

---

[6]By sorting a time series, we refer to sorting the time samples in increasing order.

versa. The iterative procedure is summarised as follows. Let $\{|S_k|\}$ be the Fourier amplitude spectrum of the original time series, s, and $\{c_k\}$ the sorted version of the original time series. At every iteration j, there are two time series, namely $r^{(j)}$, which has the correct signal distribution, and $s^{(j)}$, which has the correct amplitude spectrum. Starting with $r^{(0)}$ a random permutation of the time samples of the original time series, the procedure is the following:

1. compute the phase spectrum of $r^{(j-1)} \rightarrow \{\phi_k\}$

2. $s^{(j)}$ is the inverse transform of $\{|S_k| \exp(i\phi_k)\}$

3. $r^{(j)}$ is obtained by rank-ordering $s^{(j)}$ so as to match $\{c_k\}$

These steps are iterated until convergence of, *e.g.*, the discrepancy between $\{|S_k|\}$ and the amplitude spectrum of $r^{(j)}$. In our implementation, convergence is assessed as the point at which the MSE between $\{|S_k|\}$ and the amplitude spectrum of $r^{(j)}$ stops decreasing. The algorithm has been shown to converge after a finite number of steps (Schreiber and Schmitz, 2000), which in our simulations is typically 50 iterations for a time series of 1000 samples. An example surrogate for the Lorenz series, using the endpoint matching and retaining $s^{(j)}$ at convergence (thus, original and surrogate time series have identical amplitude spectra), is shown in Fig. 3D. In this case, the method converged after 25 iterations. Unless otherwise stated, we have used the iAAFT method for generating surrogate time series, since it has been observed to yield superior results compared to the other methods (see *e.g.*, Kugiumtzis, 1999; Schreiber and Schmitz, 2000).

3.3 Hypothesis Testing

Since it is nearly impossible to describe the fundamental propery of nonlinearity in a single, unambiguous definition, it is often assessed as the '*absence of linearity*'. In other words, in a statistical context, a null hypothesis is asserted that the time series is linear, and it is rejected if the time series does not conform to the properties associated with a linear signal.

Many nonlinearity analysis methods compare metrics obtained for the original signal to those obtained for an ensemble of surrogates. If the metric of the original time series is significantly different from those of the surrogates, the null hypothesis is rejected and the original time series is hypothesised to be nonlinear. Since the analytical form of the probability distributions of the metrics ('test statistics') is not known, a non-parametric rank-based test is used, as suggested by Theiler and Prichard (1996). For every original time series, we generate $N_s = 99$ surrogates for the nonlinearity tests. The test statistics for the original, t_o, and for the surrogates, $t_{s,i}$ ($i = 1, \ldots, N_s$), are computed and the series $\{t_o, t_{s,i}\}$ is sorted in increasing order, after which the position index (rank) r of t_o is determined. A right-tailed (left-tailed) test is rejected if rank r of the original time series exceeds 90 (is smaller or equal to 10), and a two-tailed test is rejected if rank r is greater than 95, or smaller or equal to 5. For the subsequent analyses, it is convenient to define the symmetrical rank r_{symm} as follows:

$$
r_{\text{symm}}[\%] = \begin{cases} \frac{r}{N_s+1} & \text{for right} - \text{tailed tests} \\[2mm] \frac{N_s+2-r}{N_s+1} & \text{for left} - \text{tailed tests} \\[2mm] \frac{|\frac{N_s+1}{2}-r|}{\frac{N_s+1}{2}} & \text{for two} - \text{tailed tests.} \end{cases}
\tag{2}
$$

In this way, one- or two-tailed tests are rejected if $r_{\text{symm}} > 90\%$.

Many test statistics have been proposed in the open literature (for an overview, see *e.g.* Hegger *et al.*, 1999). For every test statistic, it is important to verify the assumptions on which they are based, or the properties they are examining, since these are important issues in the interpretation of the analysis results. We have selected a number of test statistics, which have been shown to perform well in nonlinearity detection applications. Furthermore, we propose a novel method for
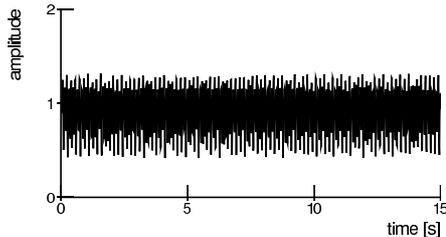
Figure 4: Realisation of the Mackey-Glass chaotic time series.

the characterisation of a time series, the 'Delay Vector Variance' (DVV) method, and show how it can be applied to the testing for nonlinearity. These methods will be explained in the following section. A collection of Matlab functions, which contains implementations of the surrogate data generation methods described in this report, is mad public and is available from `http://134.58.34.50/temu`.

4 Established Nonlinearity Analyses

Throughout the remainder of this report, the chaotic Mackey-Glass time series is used for illustrating the different methods. It is defined by the following equation:

$$\frac{dx}{dk} = \frac{0.2\ x_{k-\tau}}{1 + x_{k-\tau}^{10}} - 0.1\ x_k. \tag{3}$$

The example time series is a realisation of 1500 samples with $x_0 = 0.2$ and $\tau = 17$, and is shown in Fig. 4. Where applicable and unless otherwise stated, the embedding dimension (see further) has been set to $m = 2$.

4.1 Time Delay Embedding

Many of the techniques described in this section rest upon the method of time delay embedding for representing a time series in the so-called 'phase space', *i.e.*, by a set of delay vectors (DVs) of a given embedding dimension m, $\mathbf{x}(k) = [x_{k-m\tau}, \ldots, x_{k-\tau}]^T$, where τ is a time lag, which for simplicity is set to unity in all simulations. In other words, $\mathbf{x}(k)$ is a vector containing m consecutive time samples. Every DV $\mathbf{x}(k)$ has a corresponding *target*, namely the next sample, x_k. For an embedding dimension of two or three, the phase space of a time series can be represented visually in a scatter plot of time samples separated by τ. This is shown in Fig. 5 for the Mackey-Glass time series with an embedding dimension of $m = 2$. Clearly, there is some sort of 'structure' in the scatter plot, indicating the presence of a strange attractor (see Section 4.5). It is important to choose the embedding dimension, m, sufficiently large, such that the m-dimensional phase space enables for a 'proper' representation of the dynamical system. For a general overview, we refer to (Hegger *et al.*, 1999).

4.2 Deterministic *Versus* Stochastic Plots

The idea underpinning the method introduced by Casdagli (1991), the Deterministic *Versus* Stochastic (DVS) plots, is to construct piecewise-linear approximations of the unknown prediction function, which maps the DVs onto their corresponding targets, using a variable number n of neighbouring DVs for generating the approximations. In practice, the DVS method examines the (robust) average prediction error $E(n)$ for local linear models of a given embedding dimension m, as a function of the varying number of data points, n, used for constructing the local linear model. The prediction error as a function of the *locality* of the model conveys information regarding the

10
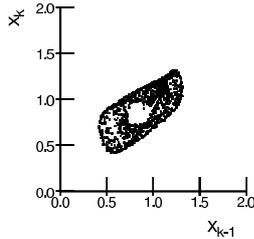
Figure 5: State space representation of the Mackey-Glass chaotic time series.

nonlinearity of the signal, since there is an intimate connection between nonlinear models and local linear ones. Indeed, in terms of a relation between the degree of locality and signal nature, a small value of n corresponds to a deterministic model (Farmer and Sidorowich, 1987), large values of n correspond to fitting a stochastic linear AR model, whereas intermediate values of n to fitting nonlinear stochastic models. Therefore, the position of the minimum in the DVS plot yields information regarding the deterministic or stochastic nature of the time series.

The degree of locality is controlled by the number of nearest DVs (in terms of the the Euclidean distance in the m-dimensional observation space), n, that are used for estimating the model parameters. The complete set of DVs is divided into a test set, $\mathbf{V}_{\text{test}}$, and a training set $\mathbf{V}_{\text{train}}$. For every DV $\mathbf{x}(k)$ in the test set, a subset Ω_k is generated by grouping the n DVs in the training set that are nearest to $\mathbf{x}(k)$. The prediction error, $E(n)$, is computed as the mean (robust) prediction error over $\mathbf{V}_{\text{test}}$, namely

$$E(n) = \langle |x_k - \widehat{x_k}| \rangle_{\mathbf{x}(k) \in \mathbf{V}_{\text{test}}}, \tag{4}$$

where $\widehat{x_k}$ is the output generated by a linear model, when presented with $\mathbf{x}(k)$ as its input, and x_k is the target for $\mathbf{x}(k)$. The model parameters, $A = [a_1, \ldots, a_m]^T$, are determined by solving the linear regression $x_k = A^T \mathbf{x}(k)$, using the set $\mathbf{x}(k) \in \Omega_k$ ($\Omega_k \subset \mathbf{V}_{\text{train}}$, thus $\mathbf{x}(k) \notin \Omega_k$).

The mean prediction error $E(n)$ is then computed as a function of the degree of locality (number n of DVs in the sets Ω_k). The resulting plots, representing E as a function of the number of nearest neighbours n, are referred to as 'DVS plots'. The number of DVs in the sets Ω_k yielding the lowest mean prediction error, n_{opt}, i.e., the position of the minimum in the DVS plot, is used as an indicator of the nature of the time series under examination. A minimum close to the origin of the DVS plot (near the 'local linear extremum') indicates a deterministic nature, and a minimum on the right hand side (near the 'global linear extremum') indicates a linear and stochastic nature. Minima occurring in between the two extrema, for increasing values of n_{opt}, correspond to gradually fitting 'more linear' and 'more stochastic' models, and are an indication of nonlinearity (Casdagli, 1991). The DVS method does not allow for a quantitative analysis.

In our implementation, we divide the time series under study in two contiguous time segments (a $4/5 - 1/5$ split), respectively for estimating the model parameters and for computing the mean prediction error, yielding the two sets of DVs, $\mathbf{V}_{\text{train}}$ and $\mathbf{V}_{\text{test}}$. The ratio $4/5 - 1/5$ is chosen arbitrarily, but is a common choice in cross-validation strategies. Note that this approach assumes that the signal's statistical characteristics in the 'train' and 'test' time segments remain identical. Another possibility, one that would not make this assumption, would be to generate the DVs *before* dividing them (randomly) into a training and test set. In that case, both sets would contain DVs derived from the complete time series. However, such an approach would introduce undesired dependencies between train and test sets, since consecutive DVs (possibly one in $\mathbf{V}_{\text{train}}$ and the other in $\mathbf{V}_{\text{test}}$) share $(m-1)$ signal samples.

An example DVS plot is shown in Fig. 6 for the Mackey-Glass chaotic (nonlinear deterministic) time series, for an embedding dimension of $m = 2$. The minimum of the curve is clearly at the left hand side of the plot, indicating a nonlinear and deterministic nature of the time series.

It should be noted at this point, however, that a *static* nonlinearity, such as that included in the composite null hypothesis explained in Section 3, is interpreted as a genuine nonlinearity
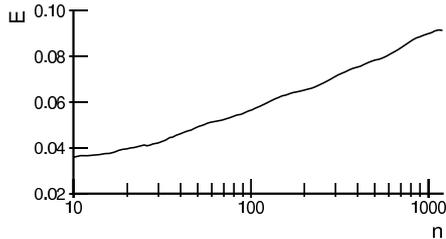
Figure 6: DVS plot of the Mackey-Glass time series.
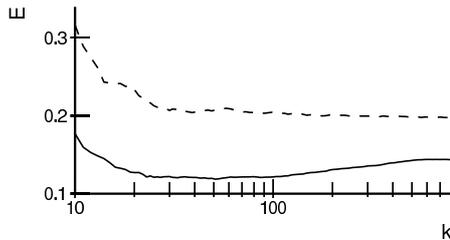
Figure 7: DVS plot for a signal generated by an AR(4) model (dashed curve), and for the same signal passed through a static nonlinearity.

by the DVS method. Thus, a linear stochastic signal, followed by a static nonlinear observation function, will have a local minimum in the DVS-plot in between the linear stochastic and nonlinear deterministic extremum. This is illustrated in Fig. 7, where the DVS plot for a signal with unit variance, generated by a stable AR(4)-model[7], is shown (dashed curve). The DVS plot clearly shows a minimum on the righthand side, at the linear stochastic extremum, indicating a linear stochastic time series. The linear AR-signal is then passed through a static sigmoid nonlinearity (atan), and the resulting DVS plot is also shown in Fig. 7 (solid curve). It is clear that the static nonlinearity results in a local minimum in between the deterministic and linear stochastic extrema, indicating a nonlinear nature of the time series.

4.3 Traditional Nonlinearity Metrics

We now shortly describe two traditional measures of nonlinearity, which have also been used by Schreiber and Schmitz (1997), namely the third-order autocovariance and the asymmetry due to time reversal.

The third-order autocovariance (C3) is a higher-order extension of the traditional autocovariance and is given by:

$$t^{C3}(\tau) = \langle x_k x_{k-\tau} x_{k-2\tau} \rangle, \qquad (5)$$

where τ is a time lag which, for simplicity and convenient comparison, is set to unity in all simulations. In combination with the surrogate data method, it has been used in (Schreiber and Schmitz, 1997) as a two-tailed test for nonlinearity.

A time series is said to be reversible if its probabilistic properties are invariant with respect to time reversal, i.e., if the joint probability of $(x_n, x_{n+\tau}, \ldots, x_{n+k\tau})$ equals the joint probability of $(x_{n+k\tau}, x_{n+(k-1)\tau}, \ldots, x_n)$, for all k and n (Diks et al., 1995). Furthermore, it has been shown that time reversibility is preserved by a static (possibly nonlinear) transform. The time reversibility has been demonstrated by Weiss (1975) for a linear Gaussian time series, and, thus, for all static transforms thereof. Schreiber and Schmitz (1997) proposed the following metric (REV) for

---

[7]The following stable AR(4)-model is used: $x_k = 1.79\, x_{k-1} - 1.85\, x_{k-2} + 1.27\, x_{k-3} - 0.41\, x_{k-4} + \nu_k$, where ν_k is a white noise source with a standard normal distribution.
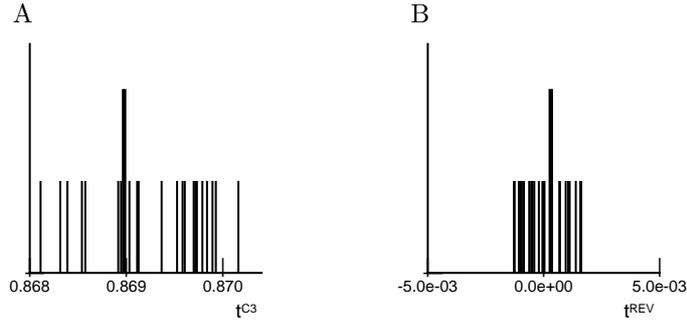
Figure 8: Example C3 (A) and REV (B) analysis for the Mackey-Glass time series. The thick lines represent the test statistics for the original time series, and the thin lines those for the 24 surrogates.

measuring the asymmetry due to time reversal:

$$t^{\text{REV}}(\tau) = \left\langle (x_k - x_{k-\tau})^3 \right\rangle. \tag{6}$$

They have shown that, in combination with the surrogate data method, it yields a reliable two-tailed test for nonlinearity.

The application of these test statistics in a surrogate data setting, is illustrated in Fig. 8. The thick lines represent the test statistics for the original time series, t_{o}. Those for the surrogates, $t_{\text{s},i}$ are drawn as thin lines. The time series is judged nonlinear if t_{o} is significantly different from $t_{\text{s},i}$, which is not the case in these examples: using the rank-based testing explained in Section 3.3, we derive $r_{\text{symm}} = 28\%$ for C3 and $r_{\text{symm}} = 36\%$ for REV, thus, not exceeding the significance threshold of 90%.

4.4 The $\delta - \epsilon$ Method

The method proposed by Kaplan (1994, 1997) was initially used for examining the degree of predictability of a time series without constructing its model, assuming a continuous function that maps the DVs onto their corresponding targets. The analysis can be summarised as follows:

- The pairwise (Euclidean) distances between DVs $\mathbf{x}(i)$ and $\mathbf{x}(j)$ are computed and denoted by $\delta_{i,j}$. The distance between corresponding targets (using the L_2-norm) is denoted by $\epsilon_{i,j}$.

- The ϵ-values are averaged, conditional to δ, i.e., $\epsilon(r) = \overline{\epsilon_{j,k}}$, for $r \leq \delta_{j,k} < r + \Delta r$, where Δr denotes the width of the 'bins' used for averaging $\epsilon_{j,k}$.

- The smallest value for $\epsilon(r)$ is denoted by $E = \lim_{r \to 0} \epsilon(r)$, and is a measure for the predictability of the time series.

The 'cumulative' version of $\epsilon(r)$ avoids the need for setting a binwidth Δr:

$$\varepsilon_c(r) = \overline{\epsilon_{j,k}} \text{ with } \delta_{j,k} < r, \tag{7}$$

where $\overline{\epsilon_{j,k}}$ is, as before, the mean pairwise distance between targets. Figure 9 shows the cumulative plot for the Mackey-Glass time series.

The heuristic for determining E is the Y-intercept of the linear regression of the N_δ (δ, ϵ)-pairs with smallest δ. In the example shown in Fig. 9, this yields $E = 0.0138$, and indicates a deterministic nature. This value can be used as a test statistic for a left-tailed nonlinearity test using surrogate data[8]. In our simulations, we have set $N_\delta = 500$.

---

[8]A Matlab implementation of the $\delta - \epsilon$ method and the test statistic is publicly available from `http://www.macalester.edu/~kaplan/Software/`.
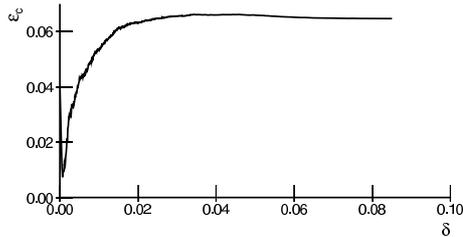
Figure 9: Cumulative $\delta - \epsilon$ plot for the Mackey-Glass time series.

4.5 Correlation Exponent

This approach to nonlinearity detection is described by Grassberger and Procaccia (1983) and computes the correlation exponent, which yields an indication of the local structure of a strange attractor. For this purpose, the correlation integral is computed as

$$C(l) = \lim_{N \to \infty} \frac{1}{N^2} \{\text{number of pairs } (i, j) \text{ for which } \|\mathbf{x}(i) - \mathbf{x}(j)\| < l\}, \tag{8}$$

where l is a length measure which is varied, and N is the number of DVs available for the analysis. Grassberger and Procaccia (1983) established that the correlation exponent, $i.e.$, the slope of the $(\ln (C(l)), \ln (l))$-curve, can be taken as a measure for the local structure of a strange attractor. Several methods exist for determining the range over which the slope is to be computed ('scaling region', see, $e.g.$, Theiler and Lookman, 1993; Hegger et $al.$, 1999). We adopt an ad hoc approach, since the other methods, albeit more appropriate for the estimation of the correlation exponent, require user intervention and we are interested in an autonomous nonlinearity detection method. The slope is computed over the l-interval $[\mu_{\mathrm{d}} \pm \sigma_{\mathrm{d}}]$, where μ_{d} and σ_{d} are the mean and standard deviation of all possible pairwise distances between different DVs. The resulting slope (referred to as COR) should not be interpreted as the actual correlation exponent, but it is proven sufficient in the context of surrogate data testing, since it examines the correlation integral in a standardised scaling region. Indeed, since the surrogate time series have signal distributions identical to that of the original, the distributions of pairwise distances, and thus, the mean and standard deviation, will be similar. Note that this distribution is approximately Gaussian for high embedding dimensions. Therefore, the correlation integral curve is examined in similar regions for both original and surrogate data, and a difference in the slope indicates a difference in local structure. Figure 10B illustrates the analysis for the Mackey-Glass time series, for an embedding dimension of $m = 5$ (this parameter value is determined using Cao's method, as explained in the next paragraph), in the described scaling region and the corresponding slope. The correlation exponent yields two-tailed tests.

The embedding dimension for which the COR analysis is performed, can be determined using Cao's method (Cao, 1997), which is related to the false nearest neighbour method (Kennel et $al.$, 1992). It yields a measure $E_1(d)$ which stops varying when d exceeds the minimum embedding dimension required for reconstructing a possible strange attractor, which is computed in the following manner:

$$E_1(d) = \frac{E(d + 1)}{E(d)}$$

$$E(d) = \frac{1}{N - d} \sum_{k=1}^{N-d} a(k, d)$$

$$a(k, d) = \frac{\|\mathbf{x}^{(d+1)}(k) - \mathbf{x}^{(d+1)}(n(k, d))\|_\infty}{\|\mathbf{x}^{(d)}(k) - \mathbf{x}^{(d)}(n(k, d))\|_\infty},$$

where $\| \cdot \|_\infty$ is the L_∞-norm, $\mathbf{x}^{(d)}(k)$ is the d-dimensional delay vector, starting at time index

14
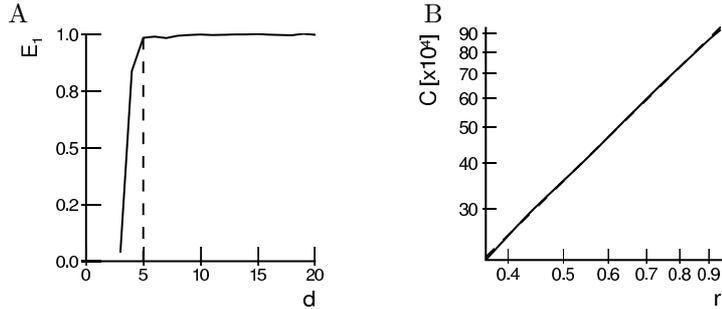
Figure 10: Example analysis for the Mackey-Glass time series. A) Cao's method for determining the optimal embedding dimension ($m = 5$ in this example); B) Grassberger-Procaccia curve in the standardised scaling region, for $m = 5$. The dashed curve represents the output of the regression from which the slope is computed (in this case 1.5169).

$(k - d)$, and $n(k, d)$ is an integer such that $\mathbf{x}^{(d)}(n(k, d))$ is the nearest neighbour of $\mathbf{x}^{(d)}(k)$ in the d-dimensional space in the $\| \cdot \|_\infty$ sense. The measure $E_1(d)$ settles when d exceeds the optimal embedding dimension, and quantifies the degree in which neighbouring DVs have similar targets. In our simulations, it is evaluated for embedding dimensions between 2 and 25. We adopt the following criterion for determining the point of convergence: the value of the differences between consecutive measurements should be lower than 0.01 and the actual measurement should exceed 0.95 times the average of the five preceeding measurements (to compensate for false plateaus). This method is exemplified for the Mackey-Glass time series in Fig. 10A.

The COR analysis is more specific than the other nonlinearity analyses, since it examines the local structure of a strange attractor in a certain scaling region. On the other hand, there exist time series that do not exhibit a strange attractor, but *do* lead to a rejection of a null hypothesis of linearity. Still, if one time series exhibits a strange attractor, and another does not, there should be a difference in COR results. Notice however, that a difference in COR results as such does not provide conclusive evidence for the presence of a strange attractor. Also, as a consequence, linear signals that show geometric structure in phase space can be erroneously judged nonlinear.

5 Proposed Method

We describe a novel characterisation method for a time series which examines the predictability of a time series by virtue of the observation of the variability of the targets (Gautama *et al.*, 2003, 2004). The proposed approach, the 'Delay Vector Variance' (DVV) method, is somewhat related to the $\delta - \epsilon$ method and the Deterministic *Versus* Stochastic plots (Casdagli, 1991), both of which are local prediction techniques described earlier. A Matlab implementation of the method is publicly available from `http://134.58.34.50/temu`. Next, it is described how the proposed method can be used for nonlinearity testing, and, finally, for rigour, the sensitivity of the method to the parameter settings is examined.

5.1 Time Series Characterisation

For a given embedding dimension m, a measure of unpredictability, σ^{*2}, is computed over all sets Ω_k. A set Ω_k is generated by grouping those DVs that are within a certain Euclidean distance to $\mathbf{x}(k)$, which is varied in a manner standardised with respect to the distribution of pairwise distances between DVs. This way, the threshold scales automatically with the embedding dimension m, as well as with the dynamical range of the time series at hand, and thus, the complete range of

15

pairwise distances is examined[9]. For a given embedding dimension m, the proposed '*Delay Vector Variance*' method (DVV) can be summarised as follows:

- The mean, μ_{d}, and standard deviation, σ_{d}, are computed over all pairwise Euclidean distances between DVs, $\|\mathbf{x}(i) - \mathbf{x}(j)\|$ $(i \neq j)$.

- The sets $\Omega_k(r_{\mathrm{d}})$ are generated such that $\Omega_k(r_{\mathrm{d}}) = \{\mathbf{x}(i)|\ \|\mathbf{x}(k) - \mathbf{x}(i)\| \leq r_{\mathrm{d}}\}$, *i.e.*, sets which consist of all DVs that lie closer to $\mathbf{x}(k)$ than a certain distance r_{d}, taken from the interval $[\max\{0, \mu_{\mathrm{d}} - n_{\mathrm{d}}\sigma_{\mathrm{d}}\}; \mu_{\mathrm{d}} + n_{\mathrm{d}}\sigma_{\mathrm{d}}]$, *e.g.*, N_{tv} uniformly spaced distances, where n_{d} is a parameter controlling the span over which to perform the DVV analysis.

- For every set $\Omega_k(r_{\mathrm{d}})$, the variance of the corresponding targets, $\sigma_k^2(r_{\mathrm{d}})$, is computed. The average over all sets $\Omega_k(r_{\mathrm{d}})$, normalised by the variance of the time series, σ_x^2, yields the 'target variance', $\sigma^{*2}(r_{\mathrm{d}})$:

$$\sigma^{*2}(r_{\mathrm{d}}) = \frac{\frac{1}{N}\sum_{k=1}^{N} \sigma_k^2(r_{\mathrm{d}})}{\sigma_x^2}. \tag{9}$$

  We only consider a variance measurement *valid*, if the set $\Omega_k(r_{\mathrm{d}})$ contains at least $N_{\mathrm{o}} = 30$ DVs, since having too few points for computing a sample variance yields unreliable estimates of the true (population) variance. A sample of 30 data points for estimating a mean or variance is a general rule-of-thumb. The effect of this parameter choice is rigorously examined in Section 5.3.

As a result of the standardisation of the distance axis, the resulting 'DVV plots' (target variance, $\sigma^{*2}(r_{\mathrm{d}})$ as a function of the standardised[10] distance, $\frac{r_{\mathrm{d}} - \mu_{\mathrm{d}}}{\sigma_{\mathrm{d}}}$) are straighforward to interpret, as illustrated in Fig. 11A for the chaotic Mackey-Glass time series (solid curve) using $m = 2$ and $n_{\mathrm{d}} = 4$. The presence of a strong deterministic component will lead to small target variances for small spans. The minimal target variance, $\sigma_{\min}^{*2} = \min_{r_{\mathrm{d}}}[\sigma^{*2}(r_{\mathrm{d}})]$, is a measure for the amount of noise which is present in the time series (the prevalence of the stochastic component). At the extreme right, the DVV plots smoothly converge to unity, since for maximum spans, *all* DVs belong to the same universal set, and the variance of the targets is equal to the variance of the time series. If this is not the case, the span parameter, n_{d}, should be increased. The parameter sensitivity of the DVV method with respect to nonlinearity testing is discussed in Section 5.3. The average DVV plot, computed over 25 iAAFT-based surrogates for the Mackey-Glass time series is also shown in Fig. 11A (dashed curve). It is evident that the surrogates also have a strong deterministic component, illustrating the dissociation between linear/nonlinear and stochastic/deterministic nature of a time series.

The optimal embedding dimension can be determined by running a number of DVV analyses for different values of m, and choosing that for which the minimal target variance, $\sigma_{\min}^{*2}$, is lowest, *i.e.*, that which yields the best predictability. We have performed this analysis for embedding dimensions ranging from 2 to 25. An example is shown in Fig. 11B for the Mackey-Glass time series, indicating an optimal embedding dimension of $m = 11$.

To summarise, the proposed DVV method can be outlined as follows:

---

[9] For computational reasons, we restrict the analysis by computing the pairwise distances between a 'reference' subset of $N_{\mathrm{sub}} = 500$ DVs and the complete set of DVs, selected by subsampling the DVs. Note that this is not equivalent to subsampling the time series.

[10] Note that we use the term 'standardised' in the statistical sense, namely as having zero mean and unit variance.
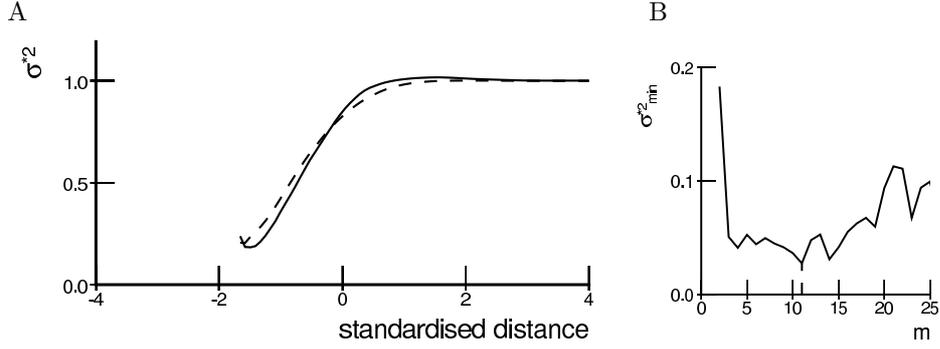
Figure 11: A) DVV plot of the Mackey-Glass time series (solid line) and the average DVV plot, computed over 25 iAAFT-based surrogates (dashed line); B) Minimal target variance, $\sigma^{*2}_{\min}$ as a function of embedding dimension, m.

Delay Vector Variance Method

Input: time series x_k
        embedding dimension, m
        maximal span σ_{d}
        'reference' subset $\mathcal{R}_{\mathrm{e}}$, with N_{sub} elements
        number of target variances, N_{tv}
        minimal set size, N_{o}

Output: DVV plot = target variance σ^{*2} *versus* distance r_{d}

1. $\forall k$: generate delay vector $\mathbf{x}(k) = [x_{k-m}, \ldots, x_{k-1}]^T$
    and corresponding target x_k

2. $\forall i \in \mathcal{R}_{\mathrm{e}}, \forall j$: compute pairwise distances $d(i,j) = \|\mathbf{x}(i) - \mathbf{x}(j)\|$

3. compute standardised axis:

    • $\mu_{\mathrm{d}} = \mathrm{mean}\,(d(i,j))_{ij}$

    • $\sigma_{\mathrm{d}} = \mathrm{std}\,(d(i,j))_{ij}$

    • $r_{\mathrm{d}}(n) = \mu_{\mathrm{d}} - n_{\mathrm{d}}\,\sigma_{\mathrm{d}} + \frac{2\,n_{\mathrm{d}}\sigma_{\mathrm{d}}}{N_{\mathrm{tv}}-1}, n = 1, \ldots, N_{\mathrm{tv}}$

4. $\forall k \in \mathcal{R}_{\mathrm{e}}, \forall n$:
    generate sets $\Omega_k(r_{\mathrm{d}}(n)) = \{\mathbf{x}(i)|\ \|\mathbf{x}(k) - \mathbf{x}(i)\| \leq r_{\mathrm{d}}(n)\}$

5. $\forall n$: if $(r_{\mathrm{d}}(n) > 0)$
        $\sigma^{*2}(r_{\mathrm{d}}(n)) = \mathrm{var}\,(\{x_i|\ \mathbf{x}(i) \in \Omega_k(r_{\mathrm{d}}(n))\ \text{and}\ \#\Omega_k(r_{\mathrm{d}}(n)) \geq N_{\mathrm{o}}\})_k$
    else
        $\sigma^{*2}(r_{\mathrm{d}}(n))$ is marked as invalid

---

5.2 Nonlinearity Analysis

In the following step, the linear or nonlinear nature of the time series is examined by performing DVV analyses on both the original and a number of surrogate time series, using the optimal embedding dimension of the original time series, which is either set manually, or determined using Cao's method, or using the minimal target variance, $\sigma^{*2}_{\min}$. Due to the standardisation of the distance axis, these plots can be conveniently combined in a *scatter diagram*, where the horizontal
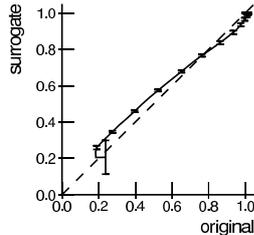
Figure 12: DVV scatter diagram of the Mackey-Glass time series (solid line). The error bars are shown (only one in three, so as not to overload the diagram).

axis corresponds to the DVV plot of the original time series, and the vertical to that of the surrogate time series. If the surrogate time series yield DVV plots similar to that of the original time series, the 'DVV scatter diagram' coincides with the bisector line, and the original time series is judged to be linear. Conversely, as is the case in Fig. 12 where the DVV scatter diagram for the Mackey-Glass time series is shown for $m = 2$ and $n_{\mathrm{d}} = 4$, if the original time series is nonlinear, the curve deviates from the bisector line. Thus, the deviation from the bisector line is an indication of nonlinearity, and can be quantified by the root mean square error (RMSE) between the σ^{*2}s of the original time series and the σ^{*2}s averaged over the DVV plots of the surrogate time series (note that while computing this average, as well as with computing the RMSE, only the valid measurements are taken into account):

$$t^{\mathrm{DVV}} = \sqrt{\left\langle \left(\sigma^{*2}(r_{\mathrm{d}}) - \frac{\sum_{i=1}^{N_{\mathrm{s}}} \sigma_{\mathrm{s},i}^{*2}(r_{\mathrm{d}})}{N_{\mathrm{s}}} \right)^2 \right\rangle_{\mathrm{valid}\ r_{\mathrm{d}}}}, \tag{10}$$

where $\sigma_{\mathrm{s},i}^{*2}(r_{\mathrm{d}})$ is the target variance at span r_{d} for the i–th surrogate, and the average is taken over all spans r_{d} that are valid in all surrogate and DVV plots. In this way, a single test statistic is obtained, and traditional (right-tailed) surrogate testing can be performed (the deviation from the average is computed for the original, and surrogate time series).

5.3 Sensitivity Analysis

To examine the sensitivity of the proposed method to parameter settings with respect to nonlinearity detection, we consider three nonlinear time series (the Mackey-Glass time series, the Laser and the Model2 series) and one linear (Model5). For details regarding the time series, we refer to Section 6.1. For each of the time series, we perform a set of DVV-based nonlinearity analyses for a range of parameter values, using a set of $N_{\mathrm{s}} = 99$ surrogates, which is identical across analyses. Unless otherwise stated, the default parameter settings are the following: $m = 3$, $n_{\mathrm{d}} = 4$, $N_{\mathrm{tv}} = 25\ n_{\mathrm{d}}$, $N_{\mathrm{sub}} = 500$ and $N_{\mathrm{o}} = 30$. To evaluate the performance of the DVV nonlinearity detection, we use a heuristic measure for the probability with which the null hypothesis is rejected, namely:

$$d = \frac{|t_{\mathrm{o}} - \mu_{\mathrm{s}}|}{\sigma_{\mathrm{s}}}, \tag{11}$$

where t_{o} is the RMSE measure for the original time series, and μ_{s} and σ_{s} are the mean and standard deviations of the RMSE measures for the surrogates. Note that a corresponding p-value would be given by $\mathrm{erf}(d/\sqrt{2})$ if the distribution of d were Gaussian, and $d \geq 3$ could, $e.g.$ be used as a rejection threshold. This measure has also been used in the nonlinearity detection context, $e.g.$, in (Theiler $et\ al.$, 1992). Observing this measure as a function of the parameters yields more detailed information than the rejection rate, and is related to the power[11] and size[12] of the test.

---

[11]The power of a statistical test, $(1 - \beta)$, is the probability of correctly rejecting the null hypothesis.
[12]The size of a statistical test, $(1 - \alpha)$, is the probability of erroneously rejecting the null hypothesis.
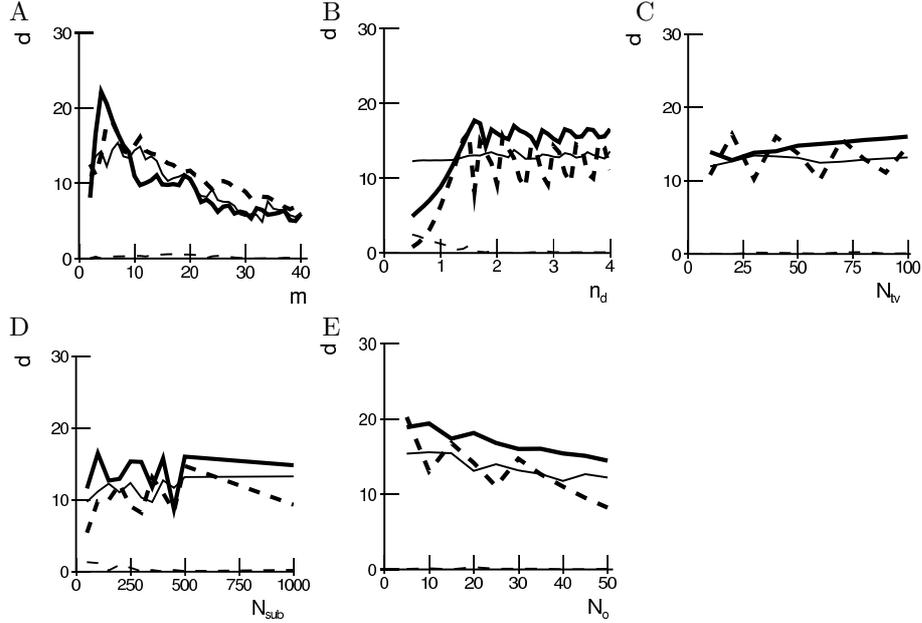
Figure 13: Sensitivity analysis results of the DVV method. Significance measure d as a function of the embedding dimension m (A), the maximal span, n_d (B), the number of evaluation points, N_{tv} (C), the size of the subset, N_{sub} (D), and the minimal set size, N_o (E). The plot conventions are the following: the Mackey-Glass time series (thick, solid curve), the Laser series (thick, dashed curve), the Model2 (thin, solid curve) and the Model5 series (thin, dashed curve).

5.3.1 Embedding Dimension, m

The embedding dimension, m, determines how many previous time samples are used for examining the local predictability. For modelling purposes, this parameter is very important, since it determines the 'tap input memory', and conversely, the number of parameters of the model. However, in the DVV nonlinearity detection context, m is not as critical, as is shown in Fig. 13A. Albeit the significance measure d is influenced by the embedding dimension, m, for the three nonlinear signals, the symmetrical rank remains $r_{symm} = 100\%$ over the whole range of dimensions tested. Furthermore, the linear signal does not lead to a rejection of the null hypothesis over the range of dimensions tested ($r_{symm} \in [31, 73]$).

5.3.2 Maximal Span, n_d

The maximal span parameter, n_d, determines the range of standardised distances to consider. Note that the number of evaluation points $N_{tv} = 25\, n_d$, due to which the spacing between consecutive standardised distances remains constant. The results are shown in Fig. 13B. Except for the Model2 time series, there is a clear effect of the maximal span parameter on the significance measure, d. The corresponding symmetrical ranks remain $r_{symm} = 100$ for the Mackey-Glass and Model2 series. At a significance level of $\alpha = 0.10$, the null hypothesis is erroneously accepted for the Laser series for $n_d \leq 0.6]$, and it is erroneously rejected for the Model5 series, for $n_d \leq 0.9$. For n_d-values exceeding 0.9, the nonlinearity for all time series are correctly assessed.

5.3.3 Number of Evaluation Points, N_{tv}

The number of standardised distances for which the target variances are computed, N_{tv}, has been systematically varied from 10 to 100, using increments of 10. The results are shown in Fig. 13C for

the four time series. Over the complete range of parameter values tested, the symmetrical ranks for the three nonlinear time series remains $r_{\text{symm}} = 100\%$, and $r_{\text{symm}} \in [47, 72]$ for the linear signal.

5.3.4 Size of Subset, N_{sub}

Reducing the size of the subset of DVs to which the pairwise Euclidean distances are computed, N_{sub}, greatly speeds up the DVV analysis. The effect on the significance measure is shown in Fig. 13D. Again, $r_{\text{symm}} = 100\%$ over the range of parameter values tested ($N_{\text{sub}} \in [50, 1000]$) for the nonlinear signals, and $r_{\text{symm}} \in [3, 70]$ for the linear signal. Note that the first two values of the significance measure for the linear signal ($d(N_{\text{sub}} = 50) = 1.33$ and $d(N_{\text{sub}} = 100) = 1.20$) are relatively high, but they correspond to RMSE values in the left half of the distribution, while we use a right-tailed test (note the absolute value in Eq. 11).

5.3.5 Minimal Set Size, N_{o}

The minimal set size parameter, N_{o}, is normally set to a default value of $N_{\text{o}} = 30$, such that the variance estimates are reliable. The effect of this parameter is now examined by analysing the nonlinearity detection performance for values between $N_{\text{o}} = 5$ and $N_{\text{o}} = 50$ in increments of 5. For the nonlinear signals, $r_{\text{symm}} = 100\%$ over the range tested, and for the linear signal, $r_{\text{symm}} \in [46, 64]$. The results are shown in Fig. 13E. There is a clear downward trend of d for increasing N_{o}, which is to be expected: at the limit (N_{o} equals the number of delay vectors), a set is only valid if it contains *all* delay vectors, and, consequently, yielding a target variance will be unity (only at the right extremum of a DVV plot). However, for *reasonable* values, say, between 10 and 30), the performance is reliable.

The sensitivity of the proposed DVV method to parameter settings has been analysed for four different time series, three nonlinear and the fourth linear. It was found that the embedding dimension, m, and the maximal span, n_{d}, were the only parameters with a noticeable effect with respect to nonlinearity detection. Furthermore, the effects were minor for *reasonable* parameter values, say, for $m \in [3, 10]$ and $n_{\text{d}} \geq 1$.

6 Comparative Study

To verify the proposed DVV-analysis, a number of time series with different natures are generated, for which we can control the predictability, *i.e.*, the prevalence of the deterministic component, and the degree of nonlinearity. Furthermore, we consider five benchmark signals, which have been considered in a comparative study (Barnett *et al.*, 1997), and a number of commonly used ('standard') test signals, the properties of which have been studied in the literature.

6.1 Signals

A unit-variance deterministic signal (sum of three sine waves, scaled to unit variance) is contaminated with uniformly distributed white noise with standard deviation σ_{n}. After standardising to unit variance, the resulting signal, n_k, is passed through a second-order nonlinear system, described by:

$$x_k = \arctan(\gamma_{\text{nl}} \mathbf{C}^T \mathbf{x}(k)) + \frac{n_k}{2},$$

where γ_{nl} controls the degree of nonlinearity, $\mathbf{C} = [0.2, -0.5]^T$, and $\mathbf{x}(k)$ are DVs of embedding dimension $m = 2$. This is a benchmark nonlinear system referred to as model II in (Narendra and Parthasarathy, 1990). In this way, the predictability is influenced by σ_{n}, whereas the degree of nonlinearity is controlled by γ_{nl}. In total, we generate 9 time series, defined by $\sigma_{\text{n}} \in \{0, 0.25, 0.5\}$ and $\gamma_{\text{nl}} \in \{0, 0.5, 1.0\}$. We refer to this set of signals as the 'tile' set, since, in a way, it tiles the
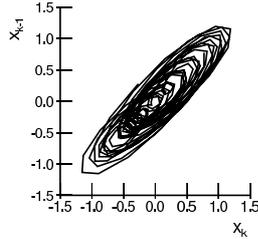
Figure 14: Phase space representation of a linear signal from the tile set, using an embedding dimension of $m = 2$.

space mentioned by Schreiber (1999), formed by a deterministic/stochastic and a linear/nonlinear axes. It should be noted that, since the driving deterministic signal is a sum of sines, all time series display attractor dynamics in phase space, as illustrated in Fig. 14 showing the phase diagram of one of the linear signals of the tile set.

The algorithms are further tested on five **benchmark** time series that have also been used in (Barnett *et al.*, 1997).

Model1 is a fully deterministic, chaotic Feigenbaum recursion of the form:

$$x_k = 3.57\, x_{k-1}(1 - x_{k-1}), \tag{12}$$

where the initial condition is $x_0 = 0.7$.

Model2 is a Generalised Autoregressive Conditional Heteroscedastic (GARCH) process of the following form:

$$
\begin{aligned}
x_k &= \sqrt{h_k}\nu_k, \\
h_k &= 1 + 0.1\, x_{k-1}^2 + 0.8\, x_{k-1},
\end{aligned}
\tag{13}
$$

where $h_0 = 1$ and $x_0 = 0$.

Model3 is a nonlinear moving average (NLMA) process of the following form:

$$x_k = \nu_k + 0.8\nu_{k-1}\nu_{k-2}. \tag{14}$$

Model4 is an Autoregressive Conditional Heteroscedastic (ARCH) process of the following form:

$$x_k = (1 + 0.5\, x_{k-1}^2)^{1/2}\nu_k, \tag{15}$$

with the value of the initial observation set to $x_0 = 0$.

Model5 is an autoregressive moving average (ARMA) model of the form:

$$x_k = 0.8\, x_{k-1} + 0.15\, x_{k-2} + \nu_k + 0.3\, \nu_{k-1}, \tag{16}$$

with $x_0 = 1$ and $x_1 = 0.7$.

Each of the generated signals consists of 2000 samples, and for the four stochastic models (Model2–Model5), the white noise disturbances, ν_k, are sampled independently from a standard normal distribution. Thus, only the Model5 series is linear (for a more detailed description, see Barnett *et al.*, 1997).

Finally, to complete the comparison, we include four **standard** time series which have been analysed frequently in the context of nonlinearity, namely the Sunspots series (280 samples), the Laser data from the Santa Fe Competition (1000 samples), the first coordinate of a realisation of the Lorenz series (1000 samples), and a realisation of a Hénon series (1000 samples).

6.2 Simulations

6.2.1 Tile Set

For the tile set, we used an embedding dimension of $m = 2$, for the benchmark set, all analyses were performed with $m = 3$, and for the standard time series, $m = 2$. In all DVV analyses, the maximal span, n_d, was determined by visual inspection such that the DVV plots converged to unity at the extreme right, yielding $n_d = 3$. This convenience did not influence the generality of our results. The results of the rank tests for the tile set are shown in Table 1 (significant rejections at the level of 0.1 of the null hypothesis, *i.e.*, an underlying Gaussian linear stochastic model, the output of which is amplitude transformed, are shown in boxes). Note that the DVS method is not included in this table, since it does not allow for a quantitative analysis. In the absence of noise ($\sigma_n = 0$), only the $\delta - \epsilon$, COR and DVV methods detected nonlinearities for slopes $\gamma_{nl} \geq 2.0$. When noise was added to the driving signals, the time reversal metric (REV) was able to detect the nonlinear nature for high slopes. The third-order cumulant (C3) was unable to detect nonlinearities in this type of signals. The COR analysis detects nonlinearity even in the cases where $\gamma_{nl} = 0$. This could be due to the presence of an attractor in all signals from this set, as explained in Section 6.1. The $\delta - \epsilon$ method fails to detect nonlinearities in the signals in the presence of noise. This could be attributable to the decreasing deterministic component when noise is added, which, in turn, decreases the sensitivity of the method, as it is based on the deterministic properties of a time series. Only the DVV method consistently detected nonlinear behaviour for $\gamma_{nl} \geq 2$, for all noise levels.

The results for the DVS and the DVV analyses are illustrated in Fig. 15 and 16, respectively. The degree of nonlinearity, γ_{nl}, increases from left to right, and the noise level, σ_n, increases from top to bottom. The DVS plots in Fig. 15 show that, as γ_{nl} increases, the error discrepancy between the best local linear model and the global linear model becomes larger, indicating, indeed, a higher degree of nonlinearity. In the DVV scatter diagrams (Fig.16), the effect of increasing nonlinearity as described above, corresponds to a stronger deviation from the bisector line (dashed line). The effect of increasing σ_n in the DVS plots is a higher error value at the optimal degree of locality. The span on the horizontal axes of the DVV scatter diagrams becomes smaller as σ_n increases. Both methods are in agreement and show a gradual change as a function of the degree of nonlinearity and the noise level. Thus, for instance, in the first columns of the tile figures, the lowest error increases (Fig. 15), and the horizontal range spanned by the DVV scatter diagrams decreases (Fig. 16) from top to bottom, *i.e.*, for increasing noise levels. Conversely, considering the first row in the tile figures, from left to right, *i.e.*, for increasing degrees of nonlinearity, the minimum becomes more pronounced in Fig. 15, and the deviation from the bisector line becomes more emphasised in Fig. 16.

6.2.2 Benchmark and Standard Sets

The results for the remaining time series under study are shown in Table 2. The DVS plots (which do not allow for a quantitative analysis) for the benchmark and standard sets are shown respectively in Fig. 17 and Fig. 19. For comparison, the corresponding DVV scatter diagrams are visualised in Fig. 18 and 20. It is clear from the Figures and Tables that the different methods yielded different results. All the methods detected nonlinearities in the Hénon and Model4 time series. The $\delta - \epsilon$, COR, and DVV methods consistently rejected the null hypothesis, and the DVS method showed indications of nonlinearity, for all other chaotic series (Laser, Lorenz and Model1). Nonlinear behaviour was detected in the Sunspots time series by DVS and REV. The COR and DVV methods were the only ones to reject all nonlinear signals described in (Barnett *et al.*, 1997) (Model1–Model4)[13]. None of the methods detected nonlinearities in the linear time series, Model5.

---

[13]Note that in (Barnett *et al.*, 1997), the $\delta - \epsilon$ method also rejected the linearity hypothesis for Model2, whereas in our simulations, the null hypothesis was accepted, albeit marginally so. This could be due to our choice of the time lag τ, which was optimised for in (Barnett *et al.*, 1997), but was set to unity in all our simulations, or to a

γ_{nl}	σ_{n}	C3	REV	$\delta - \epsilon$	COR	DVV
0.0	0.0	31	38	99	3	22
1.5	0.0	45	59	6	88	100
2.0	0.0	54	73	2	1	100
2.5	0.0	65	52	2	1	100
0.0	0.5	36	81	56	83	52
1.5	0.5	52	82	73	54	98
2.0	0.5	54	100	94	1	100
2.5	0.5	43	87	95	1	100
0.0	1.0	34	82	52	100	28
1.5	1.0	57	89	52	10	82
2.0	1.0	71	24	11	1	100
2.5	1.0	38	41	76	1	100

Table 1: Results of the rank tests for the tile set. Significant rejections of the null hypothesis at the level of 0.1 are indicated by boxes.
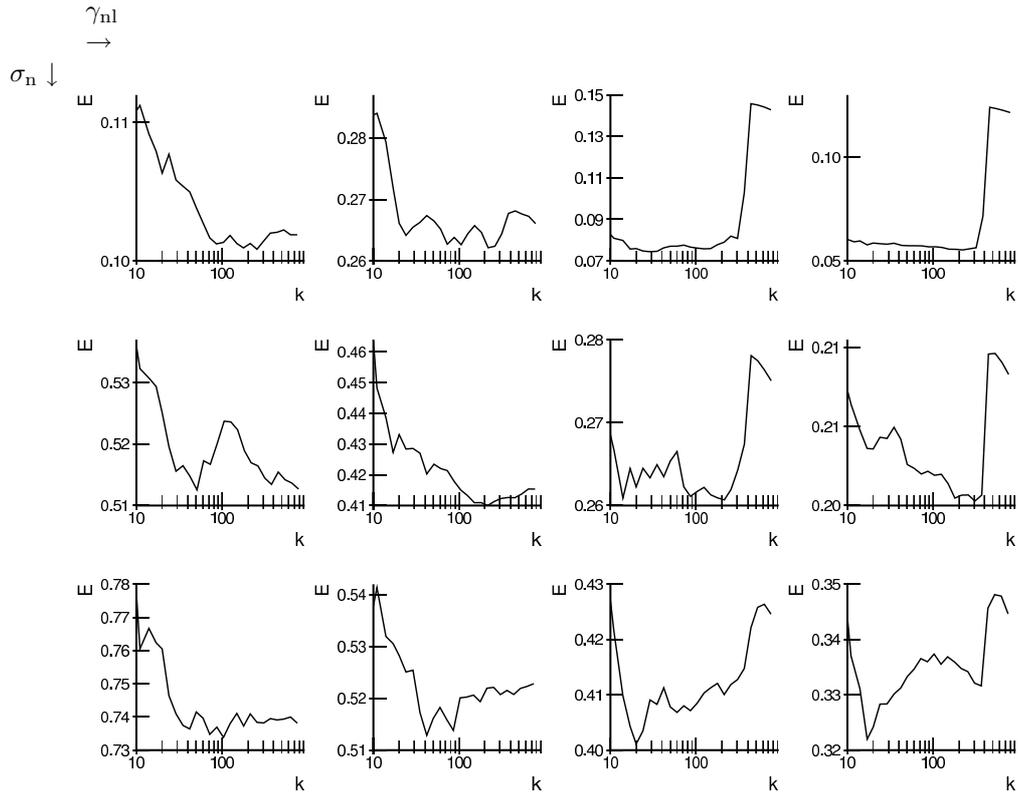
Figure 15: DVS plots for the tile set. The degree of nonlinearity increases from left to right, the noise level from top to bottom.

23
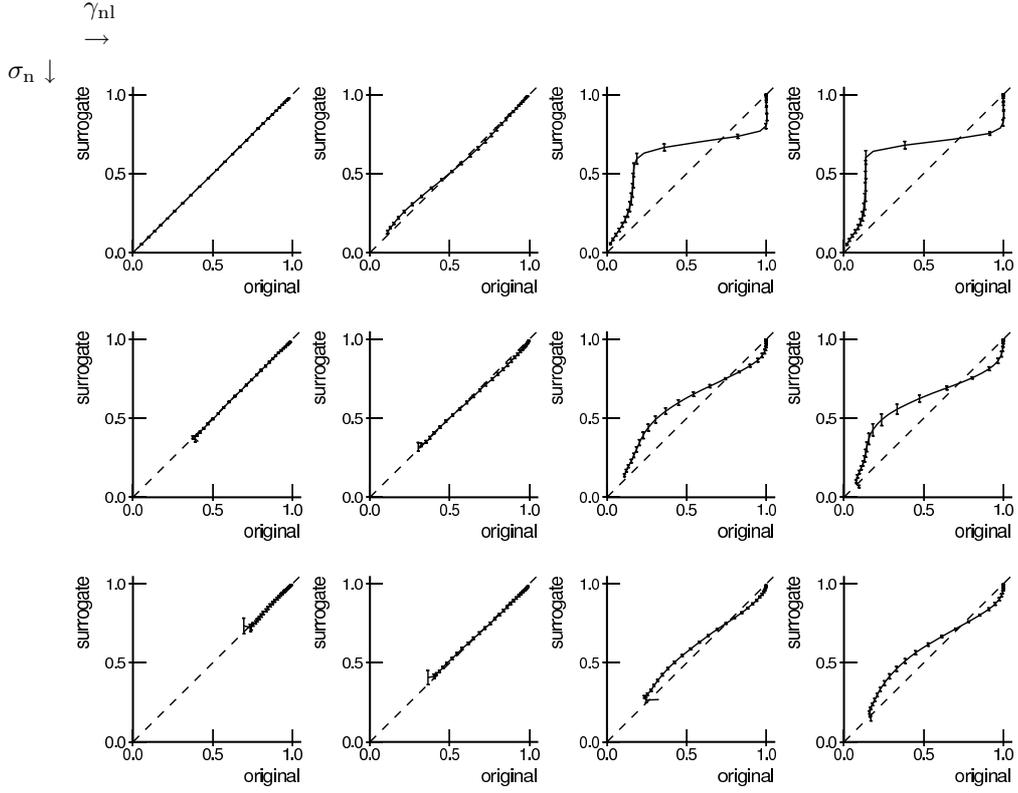
Figure 16: DVV scatter diagrams for the tile set. The degree of nonlinearity increases from left to right, the noise level from top to bottom. The error bars indicate the standard deviation from the mean of σ^{*2}.

signal	C3	REV	$\delta - \epsilon$	COR	DVV
Model1	13	1	3	1	100
Model2	84	97	12	1	100
Model3	100	73	8	1	100
Model4	100	97	1	1	100
Model5	32	41	57	28	16
Sunspots	8	100	30	20	43
Laser	1	45	1	1	100
Lorenz	19	1	1	1	100
Hénon	100	1	1	1	100

Table 2: Results of the rank tests for the benchmark and standard time series. Significant rejections of the null hypothesis at the level of 0.1 are indicated by boxes.

24
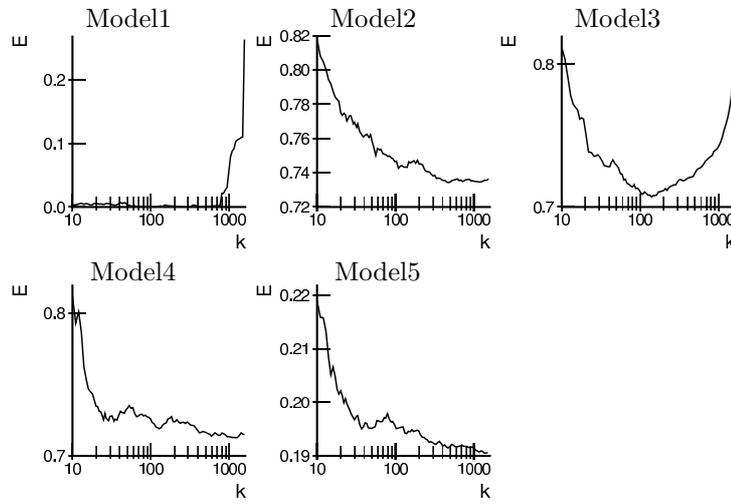
Figure 17: DVS plots of the five benchmarks signals used by Barnett *et al.* (1997).
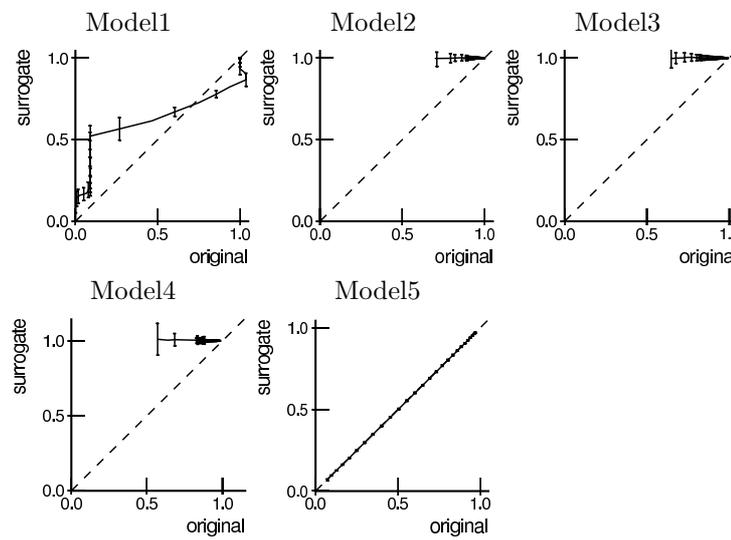
Figure 18: DVV scatter diagrams of the five benchmarks signals used by Barnett *et al.* (1997).
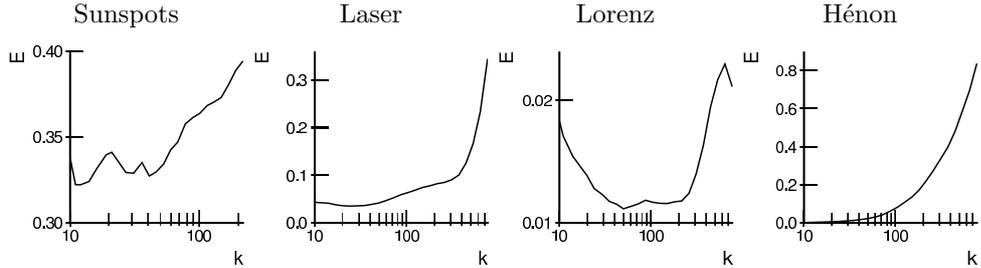
Figure 19: DVS plots of the four standard signals.
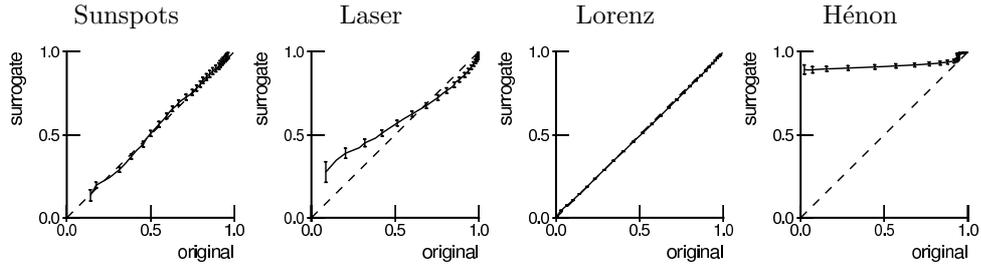
Figure 20: DVV scatter diagrams of four standard signals.

7 Case Studies

To illustrate the proposed DVV method in the context of biomedical signal processing, two case studies are considered, namely heart rate variability (HRV) and functional magnetic resonance imaging (fMRI) time series. In all results shown, the optimal embedding dimension has been obtained by selecting that for which the minimal target variance is smallest.

7.1 Heart Rate Variability

Four Heart Rate Variability (HRV) time series have been derived from long-term electrocardiogram recordings (14 to 22 hours each), with manually reviewed beat annotations, taken from the MIT-BIH Long-Term database[14]. The labelling convention is shown in Table 3. The signals consist of recordings of male patients suffering from different heart diseases. For efficiency, the time series have been limited in size to 8192 samples. It is generally accepted that the heart condition influences the nonlinear nature of the HRV signal Christini *et al.* (1995); Guzzetti *et al.* (1996); Ho *et al.* (1997); Poon and Merrill (1997).

The DVV scatter diagrams (Fig. 21 for A1–A4) show that there is a difference between the four HRV signals in the deviation from the bisector line. Clearly, A2 yields the smallest deviation

---

different approach to the estimation of E. However, this does not impact the generality of our analysis

[14]Publicly available from `http://www.physionet.org/physiobank/database/ltdb/`

label	record number
A1	14046
A2	14149
A3	14157
A4	14172

Table 3: Labelling conventions of the HRV time series.
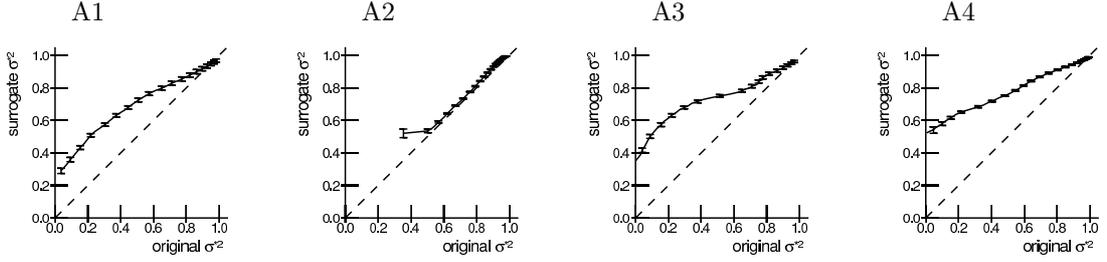
Figure 21: DVV scatter diagrams for the HRV time series (A1–A4). The error bars correspond to the upper and lower quartiles of the target variances for the surrogates.
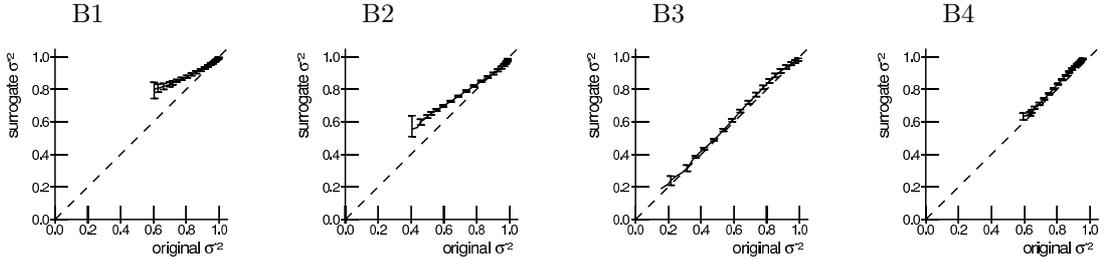
Figure 22: DVV scatter diagrams for the fMRI time series (B1–B4). The error bars correspond to the upper and lower quartiles of the target variances for the surrogates.

from the linearity hypothesis.

7.2   Functional Magnetic Resonance Imaging

The functional Magnetic Resonance Imaging (fMRI) time series have been taken from two experimental macaque motion studies Vanduffel *et al.* (2001). We consider four time series, taken from the left and right middle temporal area (MT/V5), recorded using two different contrast agents: one set (time series labelled B1 and B2, 1920 samples) is recorded using the traditional Blood Level Oxygen Dependent (BOLD) contrast agent, and the other (time series B3 and B4, 1200 samples) using an exogenous contrast agent, namely monocrystalline iron oxide nanoparticle (MION), which has been recently introduced for application in fMRI. The latter is expected to be dependent on fewer physiological variables which possibly interact in a nonlinear fashion, and should, therefore, display less nonlinearity than the BOLD signals Friston *et al.* (2000).

The proposed DVV method detects nonlinearity in B1, B2 and B4 using the surrogate data test, and, additionally, the method reveals that the deviations from the bisector line are smaller for the MION signals (B3 and B4) than for the BOLD ones (B1 and B2), which complies with the recording conditions. This can be observed in the DVV scatter diagrams shown in Fig. 22: the diagrams for B3 and B4 almost coincide with the bisector line, whereas those for B1 and B2 do not.

8 Summary

A time series can be characterised by a variety of different criteria, based on different aspects of the signal. Hence, the fundamental problem of choosing an appropriate criterion or 'test statistic' for the nonlinearity analysis needs to be done with due caution. Indeed, nonlinearity analysis results ought to be interpreted with respect to the definition of linearity that has been adopted (which is reflected in the surrogate data generation method), and the aspects of the time series on which the test statistic is based, such as time reversal asymmetry, phase space geometry, correlation exponent, to mention just a few.

To provide a unifying approach to detecting the nature of real-world signals, we have introduced a novel way for characterising a time series, called the 'Delay Vector Variance' (DVV) method, and have evaluated its performance in the context of nonlinearity detection. We have performed comprehensive simulations on a large number of time series, both synthetic and real-world, and have found that the proposed DVV method outperforms some well-established methods. Finally, two case studies of nonlinearity detection have been described, one on heart rate variability (HRV) and the other on functional magnetic resonance imaging (fMRI) time series.

Acknowledgments

References

Barnett, W., Gallant, A., Hinich, M., Jungeilges, J., Kaplan, D. and Jensen, M. (1997), A Single-Blind Controlled Competition Among Tests for Nonlinearity and Chaos, *J. of Econometrics*, **77**, 297–302.

Boynton, M., Engel, S., Glover, G. and Heeger, D. (1996), Linear Systems Analysis of Functional Magnetic Resonance Imaging in V1, *J. Neurosci.*, **16**(13), 4207–4221.

Cao, L. (1997), Practical Method for Determining the Minimum Embedding Dimension of a Scalar Time Series, *Physica D*, **110**, 43–50.

Casdagli, M. (1991), Chaos and Deterministic *versus* Stochastic Non-linear Modeling, *J. R. Statist. Soc. B*, **54**(2), 303–328.

Christini, D., Bennett, F., Lutchen, K., Ahmed, H., Hausdorff, J. and Oriol, N. (1995), Applications of Linear and Nonlinear Time Series Modeling to Heart Rate Dynamics Analysis, *IEEE Trans. Biomed. Eng.*, **42**, 411–415.

Diks, C., van Houwelingen, J., Takens, F. and DeGoede, J. (1995), Reversibility as a Criterion for Discriminating Time Series, *Phys. Lett. A*, **201**, 221–228.

Farmer, J. and Sidorowich, J. (1987), Predicting Chaotic Time Series, *Phys. Rev. Lett.*, **59**, 845–848.

Friston, K., Mechelli, A., Turner, R. and Price, C. (2000), Nonlinear Responses in fMRI: The Balloon Model, Volterra Kernels, and Other Hemodynamics, *NeuroImage*, **12**, 466–477.

Gautama, T., Mandic, D. and Van Hulle, M. (2003), Indications of Nonlinear Structures in Brain Electrical Activity, *Phys. Rev. E*, **67**(4), 046204.

Gautama, T., Mandic, D. and Van Hulle, M. (2004), The Delay Vector Variance Method for Detecting Determinism and Nonlinearity in Time Series, *Physica D*, **190**(3–4), 167–176.

Grassberger, P. and Procaccia, I. (1983), Measuring the Strangeness of Strange Attractors, *Physica D*, **9**, 189–208.

Guzzetti, S., Signorini, M., Cogliati, C., Mezzetti, S., Porta, A., Cerutti, S. and Malliani, A. (1996), Non-linear Dynamics and Chaotic Indices in Heart Rate Variability of Normal Subjects and Heart-Transplanted Patients, *Cardiovasc. Res.*, **31**, 441–446.

Hegger, R., Kantz, H. and Schreiber, T. (1999), Practical Implementation of Nonlinear Time Series Methods: The TISEAN Package, *Chaos*, **9**, 413–435.

Ho, K., Moody, G., Peng, C., Mietus, J., Larson, M., Levy, D. and Goldberger, A. (1997), Predicting Survival in Heart Failure Case and Control Subjects by Use of Fully Automated Methods for Deriving Nonlinear and Conventional Indicaes of Heart Rate Dynamics, *Circulation*, **96**, 842–848.

Kaplan, D. (1994), Exceptional Events as Evidence for Determinism, *Physica D*, **73**, 38–48.

Kaplan, D. (1997), Nonlinearity and Nonstationarity: The Use of Surrogate Data in Interpreting Fluctuations, *in* M. Di Rienzo, G. Mancia, G. Parati, A. Pedotti and A. Zanchetti, eds, 'Frontiers of Blood Pressure and Heart Rate Analysis', IOS Press, Amsterdam.

Kennel, M., Brown, R. and Abarbanel, H. (1992), Determining Embedding Dimension For Phase-Space Reconstruction Using a Geometrical Construction, *Phys. Rev. A*, **45**, 3403–3411.

Kugiumtzis, D. (1999), Test Your Surrogate Data Before You Test For Nonlinearity, *Phys. Rev. E*, **60**, 2808–2816.

Mandic, D. and Chambers, J. (2001), *Recurrent Neural Networks for Prediction: Learning Algorithms Architecture and Stability*, Wiley: Chichester.

Mechelli, A., Price, C. and Friston, K. (2001), Nonlinear Coupling between Evoked rCBF and BOLD Signals: A Simulation Study of Hemodynamic Responses, *NeuroImage*, **14**, 862–872.

Miller, K., Luh, W., Liu, T., Martinez, A., Obata, T., Wong, E., Frank, L. and Buxton, R. (2001), Nonlinear Temporal Dynamics of the Cerebral Blood Flow Response, *Hum. Brain Mapp.*, **13**, 1–12.

Narendra, K. and Parthasarathy, K. (1990), Identification and Control of Dynamical Systems Using Neural Networks, *IEEE Trans. Neural Networks*, **1**(1), 4–27.

Poon, C.-S. and Merrill, C. (1997), Decrease of Cardiac Choas in Congestive Heart Failure, *Nature*, **389**, 492–495.

Rissanen, J. (1978), Modelling By Shortest Data Description, *Automatica*, **14**, 465–471.

Schreiber, T. (1999), Interdisciplinary Application of Nonlinear Time Series Methods, *Phys. Rep.*, **308**(1), 1–64.

Schreiber, T. and Schmitz, A. (1996), Improved Surrogate Data for Nonlinearity Tests, *Phys. Rev. Lett.*, **77**, 635–638.

Schreiber, T. and Schmitz, A. (1997), On the Discrimination Power of Measures for Nonlinearity in a Time Series, *Phys. Rev. E*, **55**, 5443–5447.

Schreiber, T. and Schmitz, A. (2000), Surrogate Time Series, *Physica D*, **142**(3-4), 346–382.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. and Farmer, J. (1992), Testing for Nonlinearity in Time Series: The Method of Surrogate Data, *Physica D*, **58**, 77–94.

Theiler, J., Linsay, P. and Rubin, D. (1994), Detecting Nonlinearity in Data With Long Coherence Times, *in* A. Weigend and N. Gershenfeld, eds, 'Time Series Prediction: Forecasting the Future and Understanding the Past', Addison-Wesley, pp. 429–455.

Theiler, J. and Lookman, T. (1993), Statistical Error in a Chord Estimator of Correlation Dimension: The "Rule of Five", *Int. J. Bif. Chaos*, **3**, 765–771.

Theiler, J. and Prichard, D. (1996), Constrained-Realization Monte-Carlo Method for Hypothesis Testing, *Physica D*, **94**, 221–235.

Timmer, J. (2000), What Can be Inferred from Surrogate Data Testing?, *Phys. Rev. Lett.*, **85**(12), 2647.

Vanduffel, W., Fize, D., Mandeville, J., Nelissen, K., Van Hecke, P., Rosen, B., Tootell, R. and Orban, G. (2001), Visual Motion Processing Invesigated Using Contrast-agent Enhanced fMRI in Awake Behaving Monkeys, *Neuron*, **32**(4), 565–577.

Weiss, G. (1975), Time-Reversibility of Linear Stochastic Processes, *J. Appl. Prob.*, **12**, 831–836.

Wold, H. (1938), *A Study in the Analysis of Stationary Time Series*, Almqvist and Wiksell: Uppsala.