On the characterization of the deterministic/stochastic and linear/nonlinear nature of time series

By D. P. Mandic[1], M. Chen[1,*], T. Gautama[2], M. M. Van Hulle[3]
and A. Constantinides[1]

[1]*Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2BT, UK*
[2]*Philips Leuven, Interleuvenlaan 80, 3001 Leuven, Belgium*
[3]*Laboratorium voor Neuro-en, Psychofysiologie, K.U.Leuven, Campus Gasthuisberg, Herestraat 49, 3000 Leuven, Belgium*

The need for the characterization of real-world signals in terms of their linear, nonlinear, deterministic and stochastic nature is highlighted and a novel framework for signal modality characterization is presented. A comprehensive analysis of signal nonlinearity characterization methods is provided, and based upon local predictability in phase space, a new criterion for qualitative performance assessment in machine learning is introduced. This is achieved based on a simultaneous assessment of nonlinearity and uncertainty within a real-world signal. Next, for a given embedding dimension, based on the target variance of delay vectors, a novel framework for heterogeneous data fusion is introduced. The proposed signal modality characterization framework is verified by comprehensive simulations and comparison against other established methods. Case studies covering a range of machine learning applications support the analysis.

Keywords: signal nonlinearity; embedding dimension; delay vector variance; machine learning; data fusion

1. Introduction

Real-world processes comprise both linear and nonlinear components, together with deterministic and stochastic ones, yet it is a common practice to model such processes using suboptimal, but mathematically tractable models. To illustrate the need to asses the *nature of a real-world signal* prior to choosing the actual computational model, figure 1 (modified from Schreiber 1999) shows the enormous range spanned by the fundamental signal properties of 'nonlinear' and 'stochastic'. Despite the fact that real-world processes, due to nonlinearity, uncertainty and noise, are located in areas such as those denoted by (a), (b), (c) and '?', in terms of computational models, only the very specialized cases such as the linear stochastic autoregressive moving average (ARMA) and chaotic (nonlinear deterministic) models are well understood. It is, however, necessary

* Author for correspondence (mo.chen@imperial.ac.uk).
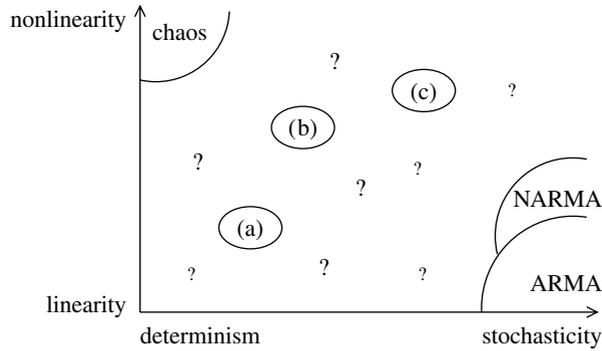
Figure 1. Classes of real-world signals spanned by the properties 'nonlinearity' and 'stochasticity'. Areas where methodologies for the analysis are readily available are highlighted, such as 'chaos' and 'ARMA'.

to verify the presence of an underlying linear or nonlinear signal generation system, for instance, in biomedical applications such as the analysis of the electrocardiogram and the electroencephalogram (EEG), the linear/nonlinear nature of the signal conveys important information concerning the health condition of a subject[1] (Schreiber 1999).

Although extensions of both ARMA models and chaos are also established, such as nonlinear ARMA models from figure 1, the characterization of signal modality for a wide range of real-world signals remains an open issue. Methods for signal nonlinearity analysis exist; these are usually set within a hypothesis-testing framework and are typically based on rigid assumptions, such as the existence of a strange attractor. Consequently, for real-world signals that are subject to uncertainty and noise, the rejection of the null hypothesis of linearity needs to be interpreted with due caution (see Schreiber & Schmitz 2000; Timmer 2000).

Our aim is therefore to introduce a unifying methodology for the simultaneous characterization of real-world signals in terms of nonlinearity and uncertainty, and to highlight the potential benefits of signal modality characterization. The analysis is supported by case studies ranging from *qualitative* assessment of the performance of machine learning algorithms, through to the fusion of heterogeneous data sources for renewable energy applications.

2. Time series model and background theory

Signal nonlinearity characterization emerged in physics in the mid-1990s and is gradually being adopted in practical engineering applications (Ho *et al.* 1997; Mandic & Chambers 2001). Some of the terminologies are given below.

Time series. In signal processing, a time series is a sequence of data points, measured typically at successive times, spaced at (often uniform) time intervals, e.g. $\{x_1, x_2, ..., x_n\}$ where n denotes the time instant.

System versus signal nonlinearity. A linear shift invariant *system*, $f(\cdot)$ obeys superposition, that is $\forall a, b \in \mathbb{R} : f(ax + by) = af(x) + bf(y)$. A system that is shift invariant, but that violates the superposition or scaling, is considered nonlinear.

[1] For instance, the changes in the nature of heart rates from stochastic to deterministic may indicate health hazards (Poon & Merrill 1997).

By contrast, as defined by Theiler *et al.* (1992), a *linear signal* is generated by an autoregressive (AR) model[2] driven by normally distributed, white (uncorrelated over time) noise, thus 'shaping' the flat amplitude spectrum of the white input. Thus, for a linear signal, the amplitude spectrum conveys all the necessary information (Theiler & Prichard 1996).

A consequence of this observation is that the phase spectrum is irrelevant for the characterization of a linear signal. With this in mind, we shall 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.

Deterministic versus stochastic (DVS) signal component. The Wold decomposition theorem (Wold 1938) states that any discrete, stationary signal can be decomposed into its *deterministic* and *stochastic* (random) components that are uncorrelated. This theorem forms the basis for many prediction models, e.g. the presence of a deterministic component imposes an upper bound on the performance of these models. Consider a sine wave (deterministic) contaminated with white noise (stochastic). In a prediction setting, the sine wave portion of the signal can be perfectly predicted using only two preceding samples (sine wave is a marginally stable AR(2) process). 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).

The definition of the 'Nature' of a signal. By the signal nature, we refer to the above two sets of signal properties: linear/nonlinear and deterministic/stochastic. The strict definition of a linear signal, x, can be relaxed by allowing its probability distribution to deviate from Gaussian; this is achieved by viewing it as a linear signal (following the strict definition) measured by a static (possibly nonlinear) observation function, $h(\cdot)$. Any signal that cannot be generated in such a way is referred to as a nonlinear signal.[3] A signal is considered *deterministic* if it can be precisely described by a set of equations, otherwise it is considered *stochastic*.

3. Surrogate data methods

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 linearized 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

---

[2] An AR 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.

[3] 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.

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 realization of the *null hypothesis* of linearity. There are three major aspects of the surrogate data method that need to be considered (Theiler & Prichard 1996; Schreiber & Schmitz 2000): (i) the exact definition of the null hypothesis, (ii) the realization of the null hypothesis, i.e. the generation method for the surrogate data, and (iii) the test statistic. It is crucial to realize that *the rejection of a null hypothesis conveys no information regarding what aspect of the null hypothesis is violated* (see Timmer 2000).

There are two main types of null hypotheses: *simple* and *composite*. A simple null hypothesis asserts that the data are 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 (Schreiber & Schmitz 2000).

(a) *Simple null hypothesis*

Owing to the underlying assumption that linear data are generated by a linear stochastic process driven by white Gaussian noise, a simple way to generate surrogate data would be to determine the appropriate AR model that, when fed with different realizations of white Gaussian noise, can be used to generate a number of surrogates. The order of the AR model can be found using a model order selection criterion, such as the minimum description length (MDL, Rissanen 1978)

$$\mathrm{MDL}(p) = N \log(E(p)) + p \log(N), \tag{3.1}$$

where N is the number of samples and $E(p)$ is the squared estimation error for model order p. Although the AR-based approach has the advantage of a well-understood implementation (also surrogate time series of any length can be generated), an important drawback is that the signal distribution of the surrogates becomes approximately Gaussian.[4] Therefore, if the original time series is not Gaussian, rejection of the null hypothesis can be due to a discrepancy in the distributions, rather than due to signal nonlinearity.

(b) *Composite null hypothesis*

To generalize the simple null hypothesis, one possible composite null hypothesis would be 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*. Owing to the Wiener–Khintchin theorem, this constraint corresponds to the original and surrogate time series having identical amplitude spectra.[5] Thus, we can also generate surrogates by randomizing the phase spectrum of the fast Fourier transform (FFT) of the original time series and subsequently retransforming

---

[4] Indeed, the AR model computes a linear combination over a number of data points drawn from a single distribution. For higher model orders, the distribution of the resulting signals becomes approximately Gaussian, due to the central limit theorem.

[5] Note that this is in line with the observation that the phase spectrum is irrelevant for the characterization of a linear signal, as described in §2.
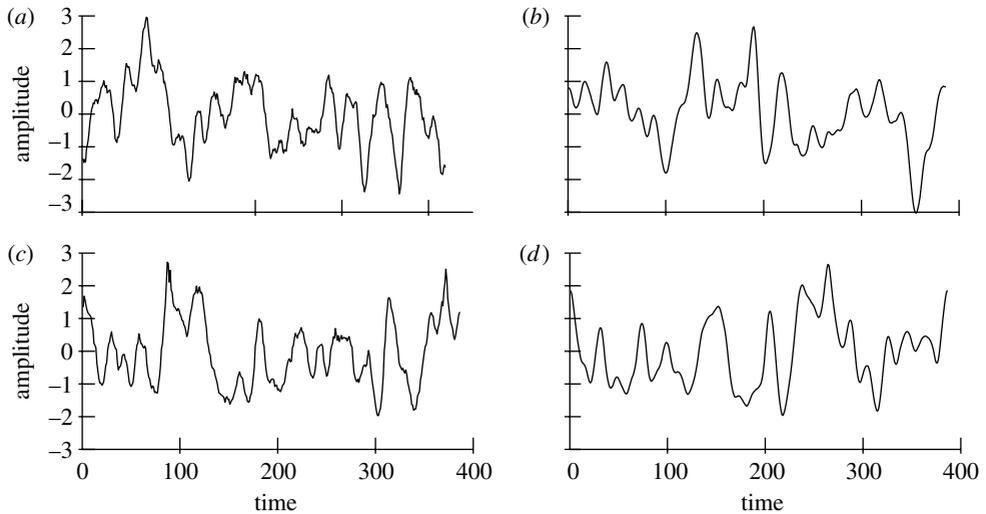
Figure 2. Surrogate realization for the Lorenz series. (*a*) FT-based without endpoint matching, (*b*) FT-based with endpoint matching, (*c*) amplitude-adjusted Fourier transform-based with endpoint matching, and (*d*) iAAFT with endpoint matching.

back into the time domain. This way, the so-called 'FT-based' surrogates are designed to have the same amplitude spectrum and hence the linear properties (mean, variance and autocorrelation function) identical to those of the original time series, but are otherwise random.

Since the FFT assumes the time series to be periodic over a time segment, a mismatch between the start- and endpoint results in a periodic discontinuity that introduces high-frequency artefacts (Theiler *et al.* 1994; windowing can be applied to compensate for this spectral leakage). Examples of FT-based surrogates for the chaotic Lorenz series, without and with endpoint matching, are given in figure 2*a*,*b*. As with the AR-based method (§3*a*), signal distributions of the surrogates do not necessarily resemble those of the original time series, which can lead to false rejections of the null hypothesis. In order to exclude such 'false' rejections, Theiler *et al.* (1992) proposed an amplitude transform of the original time series that makes the distribution Gaussian, *prior* to the application of the FT method, which is transformed back to the original distribution afterwards (amplitude-adjusted Fourier transform method, AAFT). 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 kth sample in a sorted Gaussian series of the same length as the original time series. An example of the AAFT surrogate for the chaotic Lorenz series, including the endpoint matching procedure, is shown in figure 2*c*.

The iterative amplitude-adjusted Fourier transform (*iAAFT*) *method*. The drawback of the AAFT method is that it forces the amplitude spectrum of surrogates to be flatter than that of the original time series, which, again, can lead to false rejections of the null hypothesis. To that end, Schreiber & Schmitz (1996)

---

[6] Sorting a time series refers to sorting the signal amplitudes in an increasing order.

proposed the iAAFT method that produces surrogates with identical ('correct') signal distributions and approximately identical amplitude spectra as the original time series. For $\{|S_k|\}$, 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, two series are generated: (i) $r^{(j)}$, which has the same signal distribution as the original and (ii) $s^{(j)}$, with the same amplitude spectrum as the original. Starting with $r^{(0)}$, a random permutation of the original time series, the iAAFT method is given as follows:

(i) compute the phase spectrum of $r^{(j-1)} \rightarrow \{\phi_k\}$,
(ii) $s^{(j)}$ is the inverse transform of $\{|S_k|\exp(i\phi_k)\}$, and
(iii) $r^{(j)}$ is obtained by rank ordering $s^{(j)}$ so as to match $\{c_k\}$.

The iteration stops when the difference between $\{|S_k|\}$ and the amplitude spectrum of $r^{(j)}$ stops decreasing (Schreiber & Schmitz 2000). An iAAFT surrogate for the Lorenz series, using the endpoint matching is shown in figure 2d. The iAAFT method for generating surrogate time series has become a standard, giving more stable results than the other available methods (Kugiumtzis 1999; Schreiber & Schmitz 2000).

(c) Hypothesis testing

To describe the fundamental property of signal nonlinearity, a null hypothesis is asserted that the time series is linear; it is rejected if the associate test statistic does not conform with that of a linear signal. Since the analytical form of the probability distribution functions of the metrics ('test statistics') is not known, a non-parametric rank-based test may be used (Theiler & Prichard 1996). For every original time series, $N_s = 99$ surrogates are generated; the test statistics for the original, t_o, and for the surrogates, $t_{s,i}$ ($i = 1, \ldots, N_s$), are computed, the series $\{t_o, t_{s,i}\}$ is sorted and the position index (rank) r of t_o is determined. A right-tailed (left-tailed) test rejects a null hypothesis if rank r of the original exceeds 90 ($r \leq 10$), and a two-tailed test is judged 'rejected' if rank $r > 95$ (or $r \leq 5$). It is convenient to define the symmetric rank r_{symm} as

$$r_{\text{symm}}[\%] = \begin{cases} \dfrac{r}{N_s + 1}, & \text{for right}-\text{tailed tests,} \\[2ex] \dfrac{N_s + 2 - r}{N_s + 1}, & \text{for left}-\text{tailed tests,} \\[2ex] \dfrac{\left| \dfrac{(N_s + 1)}{2} - r \right|}{\dfrac{N_s + 1}{2}}, & \text{for two}-\text{tailed tests,} \end{cases} \quad (3.2)$$

whereby one- or two-tailed tests are rejected if $r_{\text{symm}} > 90\%$.
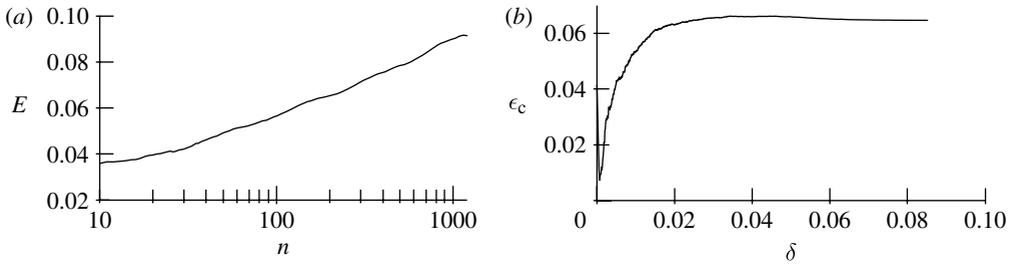
Figure 3. (*a*) DVS and (*b*) cumulative δ–ϵ plots of Mackey–Glass chaotic time series.

4. Established signal characterization methods

Time delay embedding is defined in 'phase space', by a set of delay vectors (DVs) $\boldsymbol{x}(k)$ of a given embedding dimension m and time lag[7] τ, i.e. $\boldsymbol{x}(k) = [x_{k-m\tau}, ..., x_{k-\tau}]^{\mathrm{T}}$. Every DV $\boldsymbol{x}(k)$ has a corresponding *target*, namely the next sample, $x(k)$. It is important to choose the embedding dimension sufficiently large, such that the m-dimensional phase space enables for a 'proper' representation of the dynamical system. For an overview, see Hegger *et al.* (1999).

(*a*) *DVS plots*

The idea underpinning the method introduced by Casdagli (1991), the DVS plots, is to construct piecewise linear approximations of the unknown prediction function that maps the DVs onto their corresponding targets, using a variable number n of neighbouring DVs for generating these approximations. The DVS method examines the (robust) average prediction error $E(n)$ for local linear models as a function of the number of data points, n, within the local linear model. The prediction error as a function of the *locality* of the model conveys information regarding the nonlinearity of the signal. Indeed, a small value of n corresponds to a deterministic model (Farmer & Sidorowich 1987), large values of n correspond to fitting a stochastic linear AR model, whereas intermediate values of n to fitting nonlinear stochastic models. A DVS plot for the Mackey–Glass chaotic (nonlinear deterministic) time series for $m=2$ is shown in figure 3a. The minimum of the delay vector variance (DVV) curve is at the l.h.s. of the plot, indicating correctly a nonlinear and deterministic nature of this time series.

(*b*) *The δ–ϵ method*

This method proposed by Kaplan (1994, 1997) was initially used for a model-free examination of the degree of predictability of a time series. It can be summarized as follows.

— The pairwise (Euclidean) distances between the DVs $\boldsymbol{x}(i)$ and $\boldsymbol{x}(j)$ are computed and denoted by $\delta_{i,j}$. The distance between the 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 $\epsilon_{j,k}$.

[7] For simplicity, τ is set to unity in all simulations.

— 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 bin width Δr:

$$\varepsilon_c(r) = \overline{\epsilon_{j,k}} \quad \text{with } \delta_{j,k} < r, \tag{4.1}$$

where $\overline{\epsilon_{j,k}}$ is, as before, the mean pairwise distance between targets. Figure 3b shows the cumulative plot for the Mackey–Glass time series.

The heuristics for determining parameter E is the Y-intercept of the linear regression of the N_δ (δ, ϵ) pairs with the smallest δ. In the example shown in figure 3b, 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. In our simulations, $N_\delta = 500$.

(c) Traditional nonlinearity metrics

The third-order autocovariance (C3). This is a higher-order extension of the traditional autocovariance and is given by[8]

$$t^{C3}(\tau) = \langle x_k x_{k-\tau} x_{k-2\tau} \rangle. \tag{4.2}$$

In combination with the surrogate data method, this method has been used (Schreiber & Schmitz 1997) as a two-tailed test for nonlinearity.

Reversibility due to time reversal. A time series is said to be reversible if its probability properties are invariant with respect to time reversal, i.e. if the joint probability of $(x_n, x_{n+\tau}, ..., x_{n+k\tau})$ equals the joint probability of $(x_{n+k\tau}, x_{n+(k-1)\tau}, ..., x_n)$, for all k and n (Diks *et al.* 1995). Time reversibility is preserved by a static (possibly nonlinear) transform and Schreiber & Schmitz (1997) proposed the following reversal metric (REV) for measuring the asymmetry due to time reversal,

$$t^{REV}(\tau) = \langle (x_k - x_{k-\tau})^3 \rangle. \tag{4.3}$$

It was shown that, in combination with the surrogate data method, this metric yields a reliable two-tailed test for nonlinearity in figure 4. The time series is judged nonlinear if t_o is significantly different from $t_{s,i}$ (which is not the case in these examples): using the rank-based testing explained in §3c, we obtain $r_{symm} = 28\%$ for C3 and $r_{symm} = 36\%$ for REV, thus not exceeding the significance threshold of 90%.

(d) Correlation exponent

This approach to nonlinearity detection is described by Grassberger & Procaccia (1983) and yields an indication of the local structure of a strange attractor.

The correlation integral is computed as

$$C(l) = \lim_{N \to \infty} \frac{1}{N^2} \{\text{number of pairs } (i, j) \quad \text{for which} \quad \|\boldsymbol{x}(i) - \boldsymbol{x}(j)\| < l\}, \tag{4.4}$$

---

[8] For comparison, time lag τ is set to unity in all simulations.
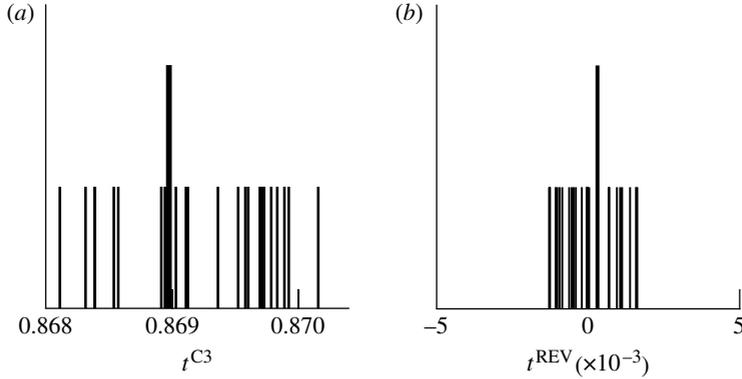
Figure 4. Nonlinearity analysis for the Mackey–Glass time series. (*a*) C3 method and (*b*) REV method. The thick lines represent the test statistics for the original time series (t_o), and the thin lines represent those for 24 surrogates ($t_{s,i}$).

where l is varied and N is the number of DVs available for the analysis. Grassberger & 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 Theiler & Lookman 1993; Hegger *et al.* 1999). 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 the local structure of the attractor yielding two-tailed tests.

5. The delay vector variance method

The DVV method is somewhat related to the δ–ϵ method and the DVS plots (Casdagli 1991).

For a given embedding dimension m, the DVV approach can be summarized as follows.

— The mean, μ_d, and s.d., σ_d, are computed over all pairwise Euclidean distances between the DVs, $\|\boldsymbol{x}(i) - \boldsymbol{x}(j)\|$ $(i \neq j)$.
— The sets $\Omega_k(r_d)$ are generated such that $\Omega_k(r_d) = \{\boldsymbol{x}(i) \| \boldsymbol{x}(k) - \boldsymbol{x}(i)\| \leq r_d\}$, i.e. sets that consist of all the DVs that lie closer to $\boldsymbol{x}(k)$ than a certain distance r_d, taken from the interval $[\max\{0, \mu_d - n_d\sigma_d\}; \mu_d + n_d\sigma_d]$, 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 $\Omega_k(r_d)$, the variance of the corresponding targets, $\sigma_k^2(r_d)$, is computed. The average over all sets $\Omega_k(r_d)$, normalized by the variance of the time series, σ_x^2, yields the 'target variance', $\sigma^{*2}(r_d)$

$$\sigma^{*2}(r_d) = \frac{\frac{1}{N} \sum_{k=1}^{N} \sigma_k^2(r_d)}{\sigma_x^2}. \tag{5.1}$$

We only consider a variance measurement *valid*, if the set $\Omega_k(r_d)$ contains at least $N_o = 30$ 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.

As a result of the standardization of the distance between the DVs, the resulting 'DVV plots' (target variance, $\sigma^{*2}(r_{\mathrm{d}})$ as a function of the standardized[9] distance, $(r_{\mathrm{d}}-\mu_{\mathrm{d}})/\sigma_{\mathrm{d}}$ are straightforward to interpret. The presence of a strong deterministic component will lead to small target variances for small spans r_{d}. The minimal target variance, $\sigma^{*2}_{\min} = \min_{r_{\mathrm{d}}}[\sigma^{*2}(r_{\mathrm{d}})]$, is a measure for the amount of noise that 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* the DVs belong to the same 'universal' set, and the variance of the targets becomes equal to that of the original time series.

To illustrate the notion of 'signal nature' by the DVV method, consider linear (AR(4)) signal ($n(k) \sim \mathcal{N}(0,1)$) (Mandic & Chambers 2001), given by

$$x(k) = 1.79x(k-1) - 1.85x(k-2) + 1.27x(k-3) - 0.41x(k-4) + n(k) \quad (5.2)$$

and a benchmark nonlinear signal, the Narendra model 3 (Narendra & Parthasarathy 1990), given by

$$z(k) = \frac{z(k-1)}{1 + z^2(k-1)} + x^3(k). \quad (5.3)$$

The average DVV plots, computed over 25 iAAFT-based surrogates for these two benchmark signals are shown in figure 5*a*,*b*. In the following step, the linear or nonlinear nature of the time series is examined by performing the DVV analyses on both the original and a number of surrogate time series. Owing to the standardization of the distance axis, these plots can be conveniently combined in a *scatter diagram*, where the horizontal 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 the 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, as shown in figure 5*c*. Conversely, as is the case in figure 5*d*, where the DVV scatter diagram for the Narendra model 3 time series is shown, the deviation from the bisector line is an indication of signal nonlinearity, which can be quantified by the root mean square error (r.m.s.e.) 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[10]

$$t^{\mathrm{DVV}} = \sqrt{\left\langle \left(\sigma^{*2}(r_{\mathrm{d}}) - \frac{\sum\limits_{i=1}^{N_{\mathrm{s}}} \sigma^{*2}_{\mathrm{s},i}(r_{\mathrm{d}})}{N_{\mathrm{s}}}\right)^2 \right\rangle_{\mathrm{valid}\ r_{\mathrm{d}}}}, \quad (5.4)$$

where $\sigma^{*2}_{\mathrm{s},i}(r_{\mathrm{d}})$ is the target variance at span r_{d} for the ith surrogate, and the average is taken over all spans r_{d} that are valid for both the original and the

[9] Note that we use the term 'standardized' in the statistical sense, namely as having zero mean and unit variance.
[10] Note that while computing this average, as well as with computing the r.m.s.e., only the valid measurements are taken into account.
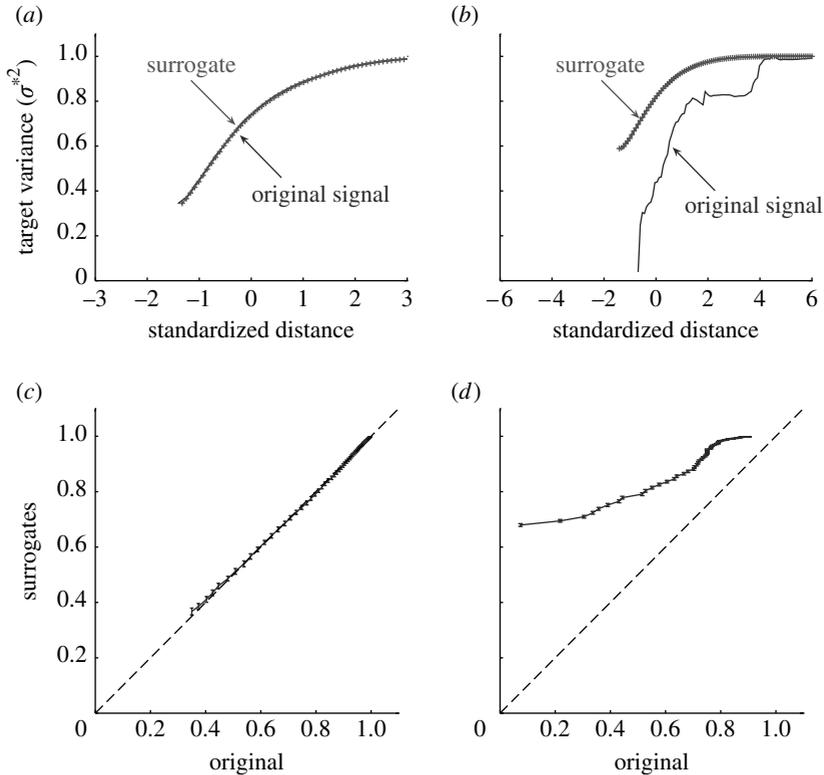
Figure 5. Nonlinear and deterministic nature of signals. DVV plots for (*a*) AR(4) signal and (*b*) Narendra model 3. The plots are obtained by plotting the target variance as a function of standardized distance. DVV scatter diagrams for (*c*) AR(4) model and (*d*) Narendra model 3. The diagrams are obtained by plotting the target variance of the original signal against the mean of the target variances of the surrogate data.

surrogates. The DVV plots represent a single test statistic, allowing for traditional (right-tailed) surrogate testing (the deviation from the average is computed for the original and surrogate time series).

6. Comparative study of the signal characterization methods

For generality, consider a unit variance deterministic signal (sum of three sine waves, scaled to unit variance) contaminated with uniformly distributed white noise with s.d. σ_n. After standardizing to unit variance, the resulting signal, n_k, is passed through a nonlinear system

$$x_k = \arctan(\gamma_\mathrm{nl} \boldsymbol{C}^\mathrm{T} \boldsymbol{x}(k)) + \frac{n_k}{2}, \tag{6.1}$$

where γ_nl controls the degree of nonlinearity, $\boldsymbol{C} = [0.2, -0.5]^\mathrm{T}$, and $\boldsymbol{x}(k)$ are the DVs of embedding dimension $m = 2$. This is a benchmark nonlinear system referred to as model 2 (Narendra & Parthasarathy 1990). The predictability is influenced by σ_n, whereas the degree of nonlinearity is controlled by γ_nl.

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. C3, third-order cumulant; COR, correlation exponent; DVV, delay vector variance; REV, reversal metric.)

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

To illustrate the potential of the DVV method, a 'tilt set' of nine time series is generated, defined by $\sigma_n \in \{0, 0.25, 0.5\}$ and $\gamma_{nl} \in \{0, 0.5, 1.0\}$, which spans the signal space from figure 1.

For the tile set, we used an embedding dimension of $m=2$, and the maximal span, n_d, was determined by visual inspection so that the DVV plots converged to unity at the extreme right, yielding $n_d=3$. The results of the rank tests for the tile set (6.1) are shown in table 1 (significant rejections at the level of 0.1 are shown in boxes). The DVS method is not included, since it does not allow for a quantitative analysis. From table 1, in the absence of noise ($\sigma_n=0$), only the $\delta-\epsilon$, the COR and DVV methods detected nonlinearities for slopes $\gamma_{nl} \geq 2.0$. When noise was added, the time REV was able to detect the nonlinear nature for high slopes γ_{nl}. The third-order cumulant (C3) metric was unable to detect nonlinearities in this type of signal, whereas the correlation exponent (COR) analysis was detecting (wrongly) nonlinearity even for the linear case with $\gamma_{nl}=0$. The $\delta-\epsilon$ method failed to detect nonlinearities within the signals when the noise was present, since 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 (right most column in table 1).

The results for the DVS and DVV analyses are illustrated in figures 6 and 7, 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 figure 6 show that, as γ_{nl} increases, the error discrepancy between the best local linear model and the global linear model becomes larger, indicating a higher degree of nonlinearity. In the DVV scatter diagrams (figure 7), the effect of an increasing degree of nonlinearity 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. For instance, in the first column of the tile figures, the lowest detected level of uncertainty increased from top to bottom
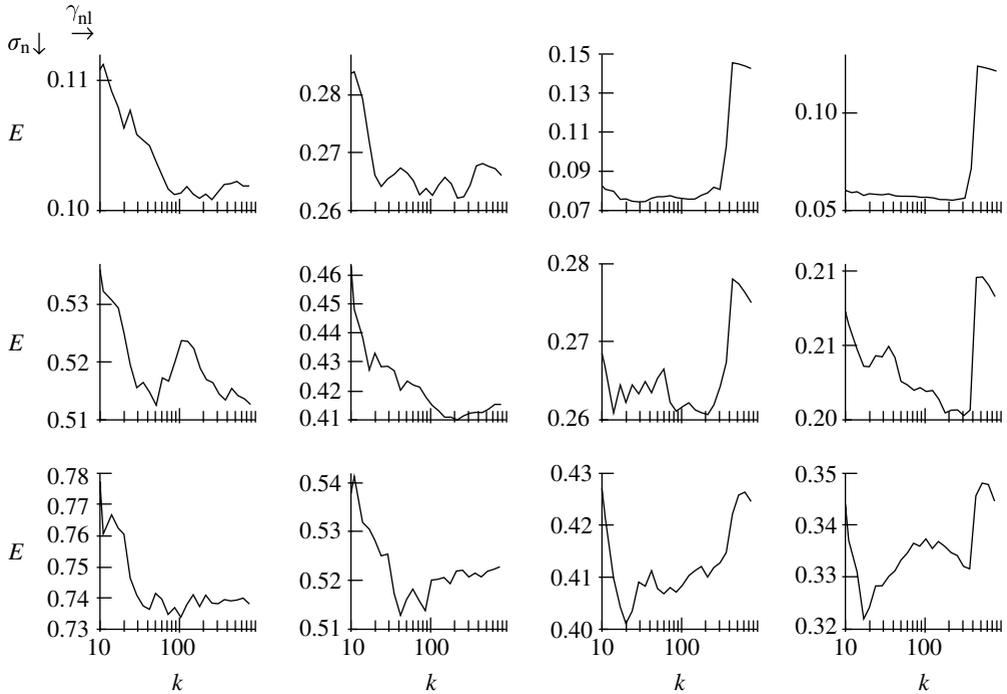
Figure 6. DVS plots for the tile set. The degree of nonlinearity increases from left to right and the noise level from top to bottom.

(figure 6). Conversely, for increasing degrees of nonlinearity (first row in figures), the minimum of the DVV curve becomes more pronounced in figure 6, and the deviation from the bisector line becomes more emphasized in figure 7.

7. Applications of the DVV method

This section illustrates the applications of the DVV method for a range of machine learning problems.

(a) Recursive versus iterative learning

We shall now consider standard adaptive filtering architectures and illustrate the effects of recursive and iterative learning on signal nature preservation. Recursive training refers to traditional *online* processing methods, where the set of adaptive learning parameters (weights) $\boldsymbol{w}(k)$ is updated in a generic form of

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \text{update},$$

where the update $\boldsymbol{\Delta w}(k)$ for gradient descent linear setting is given by

$$\boldsymbol{\Delta w}(k) = \mu e(k)\boldsymbol{x}(k),$$

where μ is the learning rate; $e(k)$ is the instantaneous output error; and $\boldsymbol{x}(k)$ is the regressor vector in the filter memory.
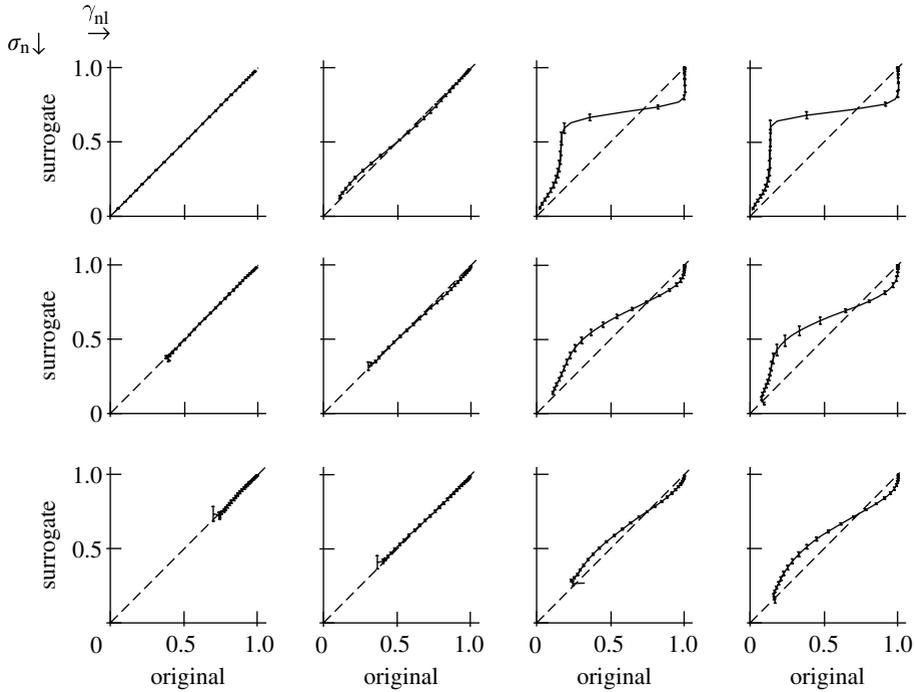
Figure 7. DVV scatter diagrams for the tile set. The degree of nonlinearity increases from left to right and the noise level from top to bottom. The error bars indicate the s.d. from the mean of $\boldsymbol{\sigma}^{*2}$.

On the other hand, iterative *a posteriori* (or data reusing) techniques naturally employ prior knowledge, by reiterating the above update for a fixed tap input vector $\boldsymbol{x}(k)$, whereby the data-reusing update is governed by the refined estimation error $e_i(k)$ for every *a posteriori* update $i = 1, \ldots, L$.

Figure 8 illustrates the quantitative performance measure (the prediction gains R_p) and qualitative performance by means of the DVV scatter diagrams, for three different learning algorithms and various orders of data reusing iterations. The least mean square (LMS) algorithm was used to train a linear finite impulse response (FIR) filter, while the nonlinear gradient descent (NGD) and real-time recurrent learning (RTRL) were used to train a nonlinear neural dynamical perceptron and recurrent perceptron, respectively. This was achieved for the prediction of a nonlinear benchmark signal (5.3). The three columns in figure 8*a–c* illustrate different orders of data reusing (0 times, 3 times and 9 times, from left to right), and the three rows in figure 8(i)–(iii) represent the three different algorithms (LMS, NGD and RTRL from top to the bottom). In terms of the prediction gain R_p and nature preservation, there is a tendency along each row, that with the increase in the order of data reusing, the quantitative performance index R_p increased as well, and the DVV scatter diagrams for the filter output approached those for the original signal (dotted line, figure 8).

(*b*) *Functional magnetic resonance imaging applications*

The general linear model is still widely used in neuroscience, exhibiting suboptimal performance, which also depends on the recording method Vanduffel *et al.* (2001). Our aim was to show that the different recording methods convey
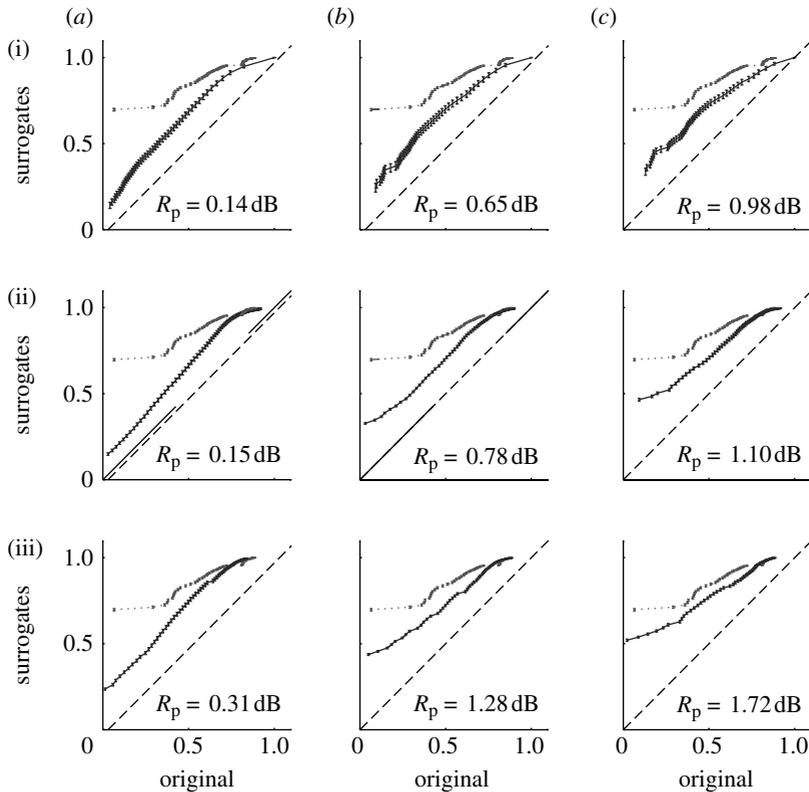
Figure 8. DVV analysis of the qualitative performance of recursive and iterative algorithms for benchmark nonlinear signal (5.3). (*a*) Standard algorithm, (*b*) three iterations of DR, and (*c*) nine iterations of DR. (i) FIR filter (LMS), (ii) feedforward perception, and (iii) recurrent perception. The dotted lines represent the original signal, while the solid lines represent DVV scatter diagrams for the one-step ahead prediction.

different degrees of nonlinearity, and hence the signal nonlinearity analysis should be undertaken prior to the actual signal modelling. To achieve this, 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). The latter is expected to be dependent on fewer physiological variables that possibly interact in a nonlinear fashion, and should, therefore, display less nonlinearity than the BOLD signals (Friston *et al.* 2000).

In figure 9, the DVV scatter diagrams for BOLD signals were nonlinear, whereas those for MION signals were linear and noisy, which conforms with the underlying null hypothesis.

(*c*) Data fusion for sleep psychology applications

The goal of the data fusion is to combine data collected from different sensors in order to make better use of the available information and achieve improved performance that could not be achieved by the use of a single sensor only
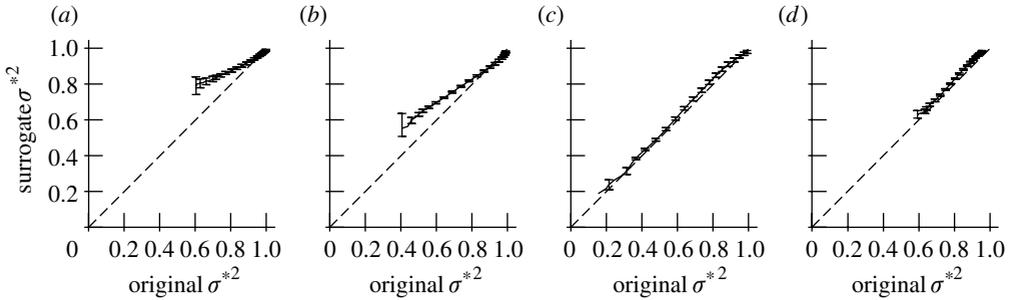
Figure 9. DVV scatter diagrams for the fMRI time series (*a*) B1, (*b*) B2, (*c*) B3, and (*d*) B4 (Gautama *et al.* 2003).

(Mandic *et al.* 2005*a*,*b*). Sleep stage scoring (Anderson & Horne 2005; Morrell *et al.* 2007) based on the combination of data obtained from multichannel EEG and other body sensors is one such application. To illustrate the usefulness of the DVV method in the fusion of heterogeneous data, a set of publicly available (http://ida.first.fraunhofer.de/∼jek/publications.html) physiological recordings of five healthy humans during three consecutive afternoon naps were used. Three physiological signals for every patient and each nap were considered: the EEG, electrooculogram (EOG) and respiratory trace (RES).

Manual scoring by medical experts divides these signals into six classes: (i) awake, eye open (W1), (ii) awake, eye closed (W2), (iii) sleep stage I (S1), (iv) sleep stage II (S2), (v) no assessment (NA), and (vi) artefact in EOG signal (AR). In order to score sleep stages, the DVV features from EEG, EOG and RES were concatenated to give a fused feature vector, and then the classification was performed using a simple neural network classifier.

Figure 10*a*,*b* illustrates the labels assigned by the medical expert and a simple perceptron, respectively, using a classifier based on the DVV features, for the first nap of patient 1. From the figure, it can be observed that these two labels are a close match and that a simple concatenation of the DVV features provides a rich information source in heterogeneous data fusion. Fusion of linear (power spectrum) and nonlinear (DVV) features has also found application in the modelling of awareness of car drivers, namely the detection of microsleep events (Golz & Sommer 2007; Sommer *et al.* 2007).

(*d*) *Complex DVV and testing for the 'complexity of a complex process'*

A straightforward extension of the univariate iAAFT method to cater for complex-valued signals is based on the matching of the amplitude spectra of the surrogate and the original complex-valued signal (Gautama *et al.* 2004). This can be achieved by rank ordering the moduli of the complex-valued signal, rather than on the real and imaginary parts separately (bivariate iAAFT method).

Since for the complex-valued time series a delay vector is generated by concatenating time delay embedded versions of the two dimensions (real and imaginary), in the complex-valued versions of the DVV method, the variance of such variables is computed as the sum of the variances of each variate, i.e. $\sigma_s^{*2}(r_d) = \sigma_{s,r}^{*2}(r_d) + \sigma_{s,i}^{*2}(r_d)$, where $\sigma_{s,r}^{*2}(r_d)$ denotes the target variance for the real part of the original signal s, and $\sigma_{s,i}^{*2}(r_d)$ denotes that for the imaginary

(*a*)

(*b*)
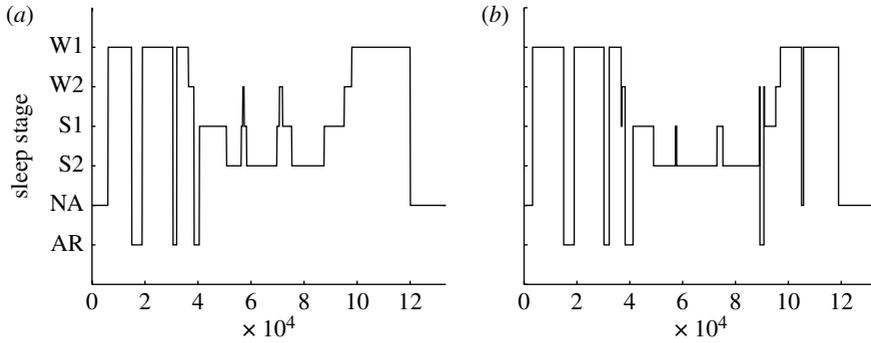
Figure 10. Labels for sleep stage scoring. (*a*) Labels assigned by a medical expert. (*b*) Labels calculated using the DVV methodology.

(*a*)

(*b*)
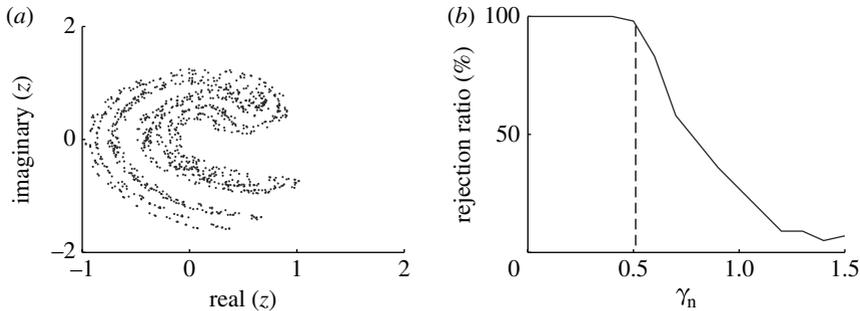
Figure 11. Judging the complex nature of an Ikeda map. (*a*) The original signal. (*b*) Behaviour in the presence of noise.

part. A statistical test[11] for judging whether the underlying signal is complex or bivariate is based upon the complex DVV method (Schreiber & Schmitz 2000). Figure 11*b* shows the result of the statistical test for an Ikeda map contaminated with complex white Gaussian noise (CWGN) over a range of power levels. The complex Ikeda map and CWGN had equal variances, and the noisy signal Ikeda$_{noisy}$ was generated according to Ikeda$_{noisy}$ = Ikeda$_{original}$ + $\gamma_n \times$ CWGN, where γ_n denotes the ratio between the s.d. of the complex Ikeda map and that of the additive white Gaussian noise. For low levels of noise, the time series was correctly judged complex.

Table 2 illustrates the results of a statistical test on the different regions of wind data recorded over 24 hours.[12] It is shown in the table that the more intermittent and unpredictable the wind dynamics ('high'), the greater the advantage obtained by the complex-valued representation of wind (Goh *et al.* 2006), while for a relatively slowly changing case, it is not. Also, there are stronger indications of a complex-valued nature when the wind is averaged over shorter intervals, as represented by the respective percentage values of the ratio of the rejection of the null hypothesis of a real bivariate nature.

[11] We consider the underlying complex-valued time series bivariate if the null hypothesis is rejected in the statistical test.

[12] The wind data with velocity v and direction θ are modelled as a single quantity in a complex representation space, namely $v(t)e^{j\theta(t)}$.

Table 2. The rejection rate (the higher the rejection rate the greater the benefit of complex-valued representation) for the wind signal.

	region containing high dynamics (%)	region containing low dynamics (%)
averaged over 1 s	96	71
averaged over 60 s	83	58

8. Summary

We have highlighted the need to characterize a time series based on different aspects of the signal, providing a unique 'signal fingerprint'. These criteria can be the local predictability, nonlinearity, smoothness, sparsity or linearity. It has been shown that the fundamental problem of choosing an appropriate criterion or test statistic for the nonlinearity analysis needs to be addressed 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 and 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 characterizing a time series, the DVV method, and have evaluated its performance in the context of nonlinearity detection. This has been supported by comprehensive simulations on a large number of time series, both synthetic and real world. The case studies provided illustrate the usefulness of this methodology for the qualitative assessment of machine learning algorithms, analysis of functional magnetic resonance imaging (fMRI) data and the DVV method as a tool for heterogeneous data and feature fusion.

References

Anderson, C. & Horne, J. A. 2005 Electroencephalographic activities during wakefulness and sleep in the frontal cortex of healthy older people: links with "thinking". *Sleep* **26**, 968–972.

Casdagli, M. 1991 Chaos and deterministic *versus* stochastic non-linear modeling. *J. R. Stat. Soc. B* **54**, 303–328.

Diks, C., Van Houwelingen, J., Takens, F. & DeGoede, J. 1995 Reversibility as a criterion for discriminating time series. *Phys. Lett. A* **201**, 221–228. (doi:10.1016/0375-9601(95)00239-Y)

Farmer, J. & Sidorowich, J. 1987 Predicting chaotic time series. *Phys. Rev. Lett.* **59**, 845–848. (doi:10.1103/PhysRevLett.59.845)

Friston, K., Mechelli, A., Turner, R. & Price, C. 2000 Nonlinear responses in fMRI: the balloon model, volterra kernels, and other hemodynamics. *NeuroImage* **12**, 466–477. (doi:10.1006/nimg.2000.0630)

Gautama, T., Mandic, D. & Van Hulle, M. 2003 Signal nonlinearity in fMRI: a comparison between BOLD and MION. *IEEE Trans. Med. Imaging* **22**, 636–644. (doi:10.1109/TMI.2003.812248)

Gautama, T., Mandic, D. & Van Hulle, M. 2004 A non-parametric test for detecting the complex-valued nature of time series. *Int. J. Know. Based Intell. Eng. Syst.* **8**, 99–106.

Goh, S. L., Chen, M., Popovic, D. H., Aihara, K., Obradovic, D. & Mandic, D. P. 2006 Complex-valued forecasting of wind profile. *Renew. Energy* **31**, 1733–1750. (doi:10.1016/j.renene.2005.07.006)

Golz, M. & Sommer, D. 2007 Automatic knowledge extraction: fusion of human expert ratings and biosignal features for fatigue monitoring applications. In *Signal processing techniques for knowledge extraction and information fusion* (eds D. P. Mandic, M. Golz, A. Kuh, D. Obradovic & T. Tanaka), pp. 299–316. Berlin, Germany: Springer.

Grassberger, P. & Procaccia, I. 1983 Measuring the strangeness of strange attractors. *Physica D* **9**, 189–208. (doi:10.1016/0167-2789(83)90298-1)

Hegger, R., Kantz, H. & Schreiber, T. 1999 Practical implementation of nonlinear time series methods: the TISEAN package. *Chaos* **9**, 413–435. (doi:10.1063/1.166424)

Ho, K., Moody, G., Peng, C., Mietus, J., Larson, M., Levy, D. & Goldberger, A. 1997 Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics. *Circulation* **96**, 842–848.

Kaplan, D. 1994 Exceptional events as evidence for determinism. *Physica D* **73**, 38–48. (doi:10.1016/0167-2789(94)90224-0)

Kaplan, D. 1997 Nonlinearity and nonstationarity: the use of surrogate data in interpreting fluctuations. In *Frontiers of blood pressure and heart rate analysis* (eds M. Di Rienzo, G. Mancia, G. Parati, A. Pedotti & A. Zanchetti), pp. 15–28. Amsterdam, The Netherlands: IOS Press.

Kugiumtzis, D. 1999 Test your surrogate data before you test for nonlinearity. *Phys. Rev. E* **60**, 2808–2816. (doi:10.1103/PhysRevE.60.2808)

Mandic, D. & Chambers, J. 2001 *Recurrent neural networks for prediction: learning algorithms architecture and stability*. Chichester, UK: Wiley.

Mandic, D., Goh, S. L., Aihara, K. 2005*a* Sequential data fusion via vector spaces: complex modular neural network approach. In *Proc. IEEE Workshop on Machine Learning for Signal Processing*, pp. 147–151.

Mandic, D. *et al.* 2005*b* Data fusion for modern engineering applications: an overview. In *Proc. Int. Conf. on Artificial Neural Networks, ICANN '05, Warsaw, Poland*, vol. 2, pp. 715–721.

Morrell, M. J., Meadows, G. E., Hastings, P., Vazir, A., Kostikas, K., Simonds, A. & Corfield, D. R. 2007 The effects of adaptive servo ventilation on cerebral vascular reactivity in patients with congestive heart failure and sleep-disordered breathing. *Sleep* **30**, 648–653.

Narendra, K. S. & Parthasarathy, K. 1990 Identification and control of dynamical systems using neural networks. *IEEE Trans. Neural Netw.* **1**, 4–27. (doi:10.1109/72.80202)

Poon, C. S. & Merrill, C. 1997 Decrease of cardiac choas in congestive heart failure. *Nature* **389**, 492–495. (doi:10.1038/39043)

Rissanen, J. 1978 Modelling by shortest data description. *Automatica* **14**, 465–471. (doi:10.1016/0005-1098(78)90005-5)

Schreiber, T. 1999 Interdisciplinary application of nonlinear time series methods. *Phys. Rep.* **308**, 1–64. (doi:10.1016/S0370-1573(98)00035-0)

Schreiber, T. & Schmitz, A. 1996 Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.* **77**, 635–638. (doi:10.1103/PhysRevLett.77.635)

Schreiber, T. & Schmitz, A. 1997 On the discrimination power of measures for nonlinearity in a time series. *Phys. Rev. E* **55**, 5443–5447. (doi:10.1103/PhysRevE.55.5443)

Schreiber, T. & Schmitz, A. 2000 Surrogate time series. *Physica D* **142**, 346–382. (doi:10.1016/S0167-2789(00)00043-9)

Sommer, D., Chen, M., Golz, M., Trutschel, U. & Mandic, D. 2007 Feature fusion for the detection of microsleep events. *Int. J. VLSI Signal Process. Syst.* **49**, 329–342. (doi:10.1007/s11265-007-0083-4)

Theiler, J. & Lookman, T. 1993 Statistical error in a chord estimator of correlation dimension: the "rule of five". *Int. J. Bifurcat. Chaos* **3**, 765–771. (doi:10.1142/S0218127493000672)

Theiler, J. & Prichard, D. 1996 Constrained-realization Monte-Carlo method for hypothesis testing. *Physica D* **94**, 221–235. (doi:10.1016/0167-2789(96)00050-4)

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. 1992 Testing for nonlinearity in time series: the method of surrogate data. *Physica D* **58**, 77–94. (doi:10.1016/0167-2789(92)90102-S)

Theiler, J., Linsay, P. & Rubin, D. 1994 Detecting nonlinearity in data with long coherence times. In *Time series prediction: forecasting the future and understanding the past* (eds A. Weigend & N. Gershenfeld), pp. 429–455. Reading, MA: Addison-Wesley.

Timmer, J. 2000 What can be inferred from surrogate data testing? *Phys. Rev. Lett.* **85**, 2647. (doi:10.1103/PhysRevLett.85.2647)

Vanduffel, W., Fize, D., Mandeville, J., Nelissen, K., Van Hecke, P., Rosen, B., Tootell, R. & Orban, G. 2001 Visual motion processing invesigated using contrast-agent enhanced fMRI in awake behaving monkeys. *Neuron* **32**, 565–577. (doi:10.1016/S0896-6273(01)00502-5)

Wold, H. 1938 *A study in the analysis of stationary time series.* Uppsala, Sweden: Almqvist & Wiksell.