Signal Processing 89 (2009) 14–23

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Correntropy as a novel measure for nonlinearity tests

Aysegul Gunduz *, Jose C. Principe

Computational NeuroEngineering Laboratory, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Article history:
Received 19 September 2007
Received in revised form
26 March 2008
Accepted 1 July 2008
Available online 17 July 2008

Keywords:
Nonlinearity tests
Surrogate methods
Kernel methods
Correntropy

ABSTRACT

Nonlinearity tests have become an essential step in system analysis and modeling due to the computational demands and complexity of analysis involved in nonlinear modeling. Standard nonlinear measures are either too complicated to estimate accurately (such as Lyapunov exponents and correlation dimension), or not able to capture sufficient but not necessary conditions of nonlinearity (such as time asymmetry). Correntropy is a kernel-based similarity measure which contains the information of both statistical and temporal structure of the underlying dataset. The capability of preserving nonlinear characteristics makes correntropy a suitable candidate as a measure for determining nonlinear dynamics. Moreover, since correntropy makes use of kernel methods, its estimation is computationally efficient. Using correntropy as the test statistic, nonlinearity tests based on the null hypothesis that signals of interest are realizations of linear Gaussian stochastic processes are carried out via surrogate data methods. Experiments performed on linear Gaussian, linear non-Gaussian, and nonlinear systems with varying in-band noise levels, data lengths, and kernel sizes confirm that correntropy can be employed as a discriminative measure for detecting nonlinear characteristics in time series. Results of tests performed on data collected from natural systems are in agreement with findings in time series analysis literature.

1. Introduction

Prior knowledge of the underlying dynamical properties of a natural system can guide system modeling, in particular, the selection of linear versus nonlinear models for increased accuracy and better performance. Linear stochastic processes can generate visually complicated signals due to noise, random inputs to the system, or static nonlinear measurement techniques [1]. On the other hand, systems that are anticipated to be nonlinear (such as the brain) might produce signals that do not reflect nonlinear dynamics [2] and our prejudice of the system might be misleading [3]. In practical applications, nonlinear models should be avoided if the underlying signals are in fact linear in nature, due to increased complexity in nonlinear system training and sensitivity analysis in modeling [4]. Hence, it is a good practice to check a priori whether the data suggest the model to be adopted via nonlinearity tests.

The method of surrogate data [5] provides a rigorous framework for nonlinearity tests whose main ingredients are the null hypothesis and a nonlinearity measure. The most commonly used null hypothesis states that the examined time series is generated by a linear Gaussian stochastic process collected through a static nonlinear measurement function. Thus, properly designed surrogate data should only retain the same linear properties (autocorrelation and amplitude distribution) as the original signal, and be otherwise random [6]. The generated surrogate data are compared to the original data under a

* Corresponding author. P.O. Box 116130, Building #33, Center Drive, Room NEB 486, Gainesville, FL 32611, USA. Tel.: +1 352 392 2682; fax: +1 352 392 0044.
E-mail addresses: aysegul@cnel.ufl.edu (A. Gunduz), principe@cnel.ufl.edu (J.C. Principe).

0165-1684/$ - see front matter
doi:10.1016/j.sigpro.2008.07.005

discriminating nonlinear measure. We test if the value of the measure for the original time series is likely to be drawn from the distribution of values of the surrogates within a confidence level. If the measure elicits comparatively different values for the original series, the null hypothesis is rejected [1].

Conventional nonlinearity measures inspired from chaos theory [7] for testing nonlinearity include maximal Lyapunov exponents [8] (a measure of local trajectory divergence), correlation dimension [12] (a measure of the system's effective number of degrees of freedom), correlation entropies [8], time-delayed mutual information [10], information-theoretic redundancies [10], and transfer entropy [11]. The calculation of these statistics depends on several parameters such as embedding dimension, time delays, and radial neighborhoods and therefore requires extensive computations for accurate results in finite sample data. Reliability of the results depends on whether scaling or plateaus regions are present in the plots of these measures versus their parameters and whether the examiner can intelligently interpret the results[1]. Not only do varying parameters complicate interpretations, but also the fact that the procedures need to be repeated for each surrogate data adds to the computational complexity [13].

Predictability of time series has been proposed to distinguish between randomness and chaos, and to detect nonlinear dynamics without constructing actual models [14]. Chaotic systems follow deterministic patterns and thus accurate short-term predictions are possible, although sensitivity to initial conditions exponentially decreases predictability in long-term [17]. Approximate entropy [15,16] measures the likelihood that patterns that are close to each other will remain close in the next incremental comparisons. Transportation distance function [17] measures the difference in long-term behavior of two time series. Delay-vector variance [4,14] yields an inverse measure of predictability through estimating variances of embedded delay-vectors in a neighborhood of a target delay-vector. Although more robust than estimating dimensions, these methods also depend on embedding dimension, time delays, and radial neighborhoods.

Statistical measures proposed for nonlinearity tests are higher order moments which elicit time reversibility [18], higher dimensional autocorrelation functions [19] and higher dimensional spectra [20]. The autocorrelation function defines the properties of linear Gaussian processes [21] and since the autocorrelation function is symmetric around zero time lag, linear Gaussian processes are time reversible [8]. Hence time asymmetry is a sufficient evidence of nonlinearity, but not a requisite [22]. Subba Rao [20] and Hinich [23] suggested linearity tests based on bispectral analysis. However, the computational complexity of estimating the bispectrum is quite extensive.

Correntropy, proposed by Principe et al. [24], is a similarity measure which combines the signal time structure and the statistical distribution of signal amplitudes in a single function. Correntropy and the conventional autocorrelation function exhibit common properties, but unlike autocorrelation, correntropy is sensitive to higher order moments of the amplitude distribution and can identify the nonlinear characteristics of a signal generation. Moreover, it takes advantage of kernel methods [25] to compute inner products efficiently. Motivated by its computational simplicity and its ability to reflect nonlinear characteristics, we propose to use correntropy as a discriminate measure for nonlinearity tests. The discrimination power and false alarm rates of correntropy as a nonlinearity measure are tested on synthetic linear Gaussian, linear non-Gaussian and nonlinear time series with varying data lengths, signal-to-noise ratios and under static nonlinear distortions. Tests performed on the Santa Fe Time Series Competition Data [9] yielded results that are in concord with those performed employing other nonlinear measures which are much more computationally demanding.

2. Nonlinearity tests with surrogate methods

The method of surrogate data is a popular tool for testing a null hypothesis on a time series against its temporally random realizations through a discriminating measure. In nonlinearity literature, one widely used null hypothesis is "the examined time series is generated by a Gaussian linear stochastic process" [3]. Linear correlations and time evolution of Gaussian linear processes can be preserved through their autocorrelation functions. For this null hypothesis, thus, constrained realizations of the data are created, which require that the surrogates have the same autocorrelations as the original data, and be otherwise random [3,27]. With this goal in mind, surrogates are created to have same Fourier amplitudes as the data but with random phases. The key point of this methodology is that the squared amplitude of the Fourier transform is a periodogram estimator of the conventional power spectral density [26]. Hence, the original time series and its surrogates attained by this method share the same power spectrum, and therefore the same autocorrelation function. However, any underlying nonlinear dynamic structure within the original data is altered by phase randomization. Thus, in the presence of such dynamic nonlinearities, using a measure capable of yielding distinct values for the original data compared to its surrogate counterparts would enable us to reject the null hypothesis (at a statistical confidence level).

Practically, however, most measured linear processes are likely to be non-Gaussian. This deviation from Gaussianity, leads tests (based on phase randomized surrogates) to routinely reject the linear null hypothesis even though the time series is purely linear [28]. This rejection therefore cannot immediately imply nonlinear dynamics. For a linear process, the non-Gaussian distribution can simply be the nature of the time series, or could be due to a distortive nonlinear measurement function. For the latter probability, Schreiber [19] suggests generalizing the null hypothesis to state that "the examined time series is generated by a linear Gaussian stochastic process, and measured through a monotonic, static nonlinear function" which distorts the normal amplitude distribution [29]. In order to generate surrogates that preserve the amplitude distribution and power spectrum of the original series, the surrogate data are iteratively rank-ordered to the original data magnitudes and spectrally filtered to attain original spectral magnitudes until the discrepancies in distribution and spectra converge to a given accuracy [27]. This is known as the *iterative amplitude adjusted Fourier transform* (IAAFT) method for generating surrogates [27,30] and throughout this paper, the surrogates are generated in this fashion and tested against this generalized null hypothesis.

In the test design, initially a residual probability of a false rejection, α is selected, which corresponds to a confidence level of $(1-\alpha)\times 100\%$. The value of the nonlinearity measure evaluated on each realization and the original series are ranked. The null hypothesis is rejected if the nonlinearity measure evaluated on the original series deviates from the surrogates in a specified direction. For a one-sided test $((1/\alpha)-1)$ surrogate sequences are generated [3]. For example, a confidence level of 95% would require generation of at least 19 surrogates. Engaging more surrogates increases the confidence level of a test.

Discrimination power (DP) of a nonlinearity measure can be estimated through tests performed on deterministic signals whose dynamic properties are known a priori. The rate of rejecting the null hypothesis when the system is in fact nonlinear elicits the discrimination power of the measure $(0\leqslant DP\leqslant 1)$ at the significance level of the test. A deterministic nonlinear system commonly used in nonlinearity tests to determine discrimination powers of measures is the chaotic Lorenz attractor [31]. In addition, the rate of false alarm (type I error) can be measured on linear Gaussian processes and linear non-Gaussian processes.

3. Correntropy

3.1. Definition and properties

Correntropy[1] is a similarity measure of signals mapped nonlinearly into a feature space [24]. In essence, correntropy generalizes the autocorrelation function to nonlinear spaces: If $\{x_t, t\in T\}$ is a strict stationary stochastic process within an index set T, then the autocorrelation and correntropy functions are defined respectively as

$$\begin{aligned} R(s,t) &= E\{\langle x_s, x_t\rangle\} \\ V(s,t) &= E\{\langle \phi(x_s), \phi(x_t)\rangle\} \end{aligned} \tag{1}$$

where ϕ is a nonlinear mapping from the input space to the feature space [24]. Instead of explicitly defining a mapping, computing the mapping, and then taking an inner product of the mapping, correntropy makes use of the “kernel trick” which defines the inner product of the nonlinear mappings as a positive-definite Mercer kernel [32]:

$$\kappa(x_s,x_t) = \langle \phi(x_s), \phi(x_t)\rangle \tag{2}$$

[1] The nomenclature of correntropy was inspired from the facts that it is a similarity measure and that the negative logarithm of the mean of correntropy yields the Parzen estimate of Renyi’s quadratic entropy [48].

The kernel trick provides efficient computations without explicit knowledge of the mapping. Kernel methods have been extensively used in recent years in applications such as prediction [33], classification [34], decomposition [35] and matched filters [36]. With the kernel trick, the definition of correntropy reduces to the expected value of the kernel of choice.

A widely used Mercer kernel is the Gaussian kernel given by

$$\kappa(x_s,x_t) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-\|x_s-x_t\|^2}{2\sigma^2}\right) \tag{3}$$

where σ is the size of the kernel. Using a Taylor series expansion for the Gaussian kernel, it can be shown that the information provided by the autocorrelation is included within correntropy by substituting $n=1$ in the expansion [24]

$$V(s,t) = \sum_{n=0}^{\infty}\frac{(-1)^n}{2^n\sigma^{2n}n!}E\|x_s-x_t\|^{2n} \tag{4}$$

The two functions have many properties in common: both are symmetric with respect to the origin and take on their maximum value at zero lag. Moreover, from (4) it can be observed that for $n>1$, correntropy involves higher order even moments of the term $\|x_s-x_t\|$ and exhibits other important properties autocorrelation function does not possess. (Clearly, the choice of another kernel would lead to a different expansion of correntropy. Note that a kernel that would yield the odd higher order moments[2] would include a term for time reversibility and still require the same amount of computations.)

Fourier transforms of statistical measures yield major functions employed in spectral analysis. Examples of such prominent Fourier transform pairs are the autocorrelation and power spectral density functions (i.e., spectrum), bicorrelation and bispectrum functions. In Principe et al. [24] the Fourier transform of correntropy was introduced as the generalized power spectral density and named *correntropy spectral density* (CSD). Its expression is given as follows:

$$P_V[\omega] = \sum_{m=-\infty}^{\infty} V[m]\,\mathrm{e}^{-\mathrm{j}\omega m} \tag{5}$$

which retains many properties of the conventional power spectral density.

3.2. Surrogate nonlinearity tests with correntropy

Recall that the surrogates are generated in a fashion so that they possess the same spectra as the original time series. Just like the conventional power spectral density, correntropy spectral density represents the distribution of generalized power amongst frequencies. Thus, if normalized by the total generalized power, CSD becomes a

[2] For example, Taylor expansion on the Laplacian kernel yields all of the moments:

$$V_{\mathrm{Lap}}(s,t) = E\left\{\exp\left(\frac{-\|x_s-x_t\|}{\sigma}\right)\right\} = \sum_{n=0}^{\infty}\frac{(-1)^n}{\sigma^n n!}E\|x_s-x_t\|^n$$

probability density function of generalized power over frequencies. If an examined time series does not possess nonlinear dynamics, the underlying distribution of its CSD and that of its surrogates should also be the same. On the other hand, if the two underlying distributions are different, we deduce that the time series contains nonlinear structures not contained in its surrogates. This decomposition of correntropy into probability density functions allows for the use of the two-sample Kolmogorov–Smirnoff goodness-of-fit test [37]. We therefore propose to use normalized CSD as the nonlinear discriminating measure and the Kolmogorov–Smirnoff test for the rejection of the null hypothesis.

The Kolmogorov–Smirnoff goodness-of-fit test is a powerful tool which examines whether two signals are random samples from the same distribution. It is based on the empirical cumulative distribution functions (ecdf) obtained directly from the data samples. For the problem at hand, we want to compare the ecdfs calculated from the CSDs of the original and surrogate series. The maximum difference between the ecdfs of the two series, denoted by F, across all frequencies in range, denoted by f, is the test statistic

$$D = \max_f |F_{\text{orig}}(f) - F_{\text{surr}}(f)| \tag{6}$$

This D-value is compared to a critical value given by

$$D_\alpha = c(\alpha)\sqrt{2/N} \tag{7}$$

where N is the common sample size of the two series and the coefficient $c(\alpha)$ depends on the significance level. For $\alpha = 0.05$, this value equals 1.36.

The Kolmogorov–Smirnoff test states that if the empirical D-value is greater than the critical value, then the hypothesis that the two series were generated from the same distribution should be rejected.

For any kernel method, the choice of kernel size affects the performance of the method. The kernel size generally is determined empirically. In this study, the bandwidth of the kernel is selected according to Silverman's rule of thumb [38]:

$$\sigma = 0.9AN^{-1/5} \tag{8}$$

where N is the data length and A stands for the minimum of the empirical standard deviation of data and the data interquartile range scaled by 1.34 as defined in Silverman's rule. Apart from the stationarity assumption required in the definition, this is the only parameter involved in estimating correntropy apart from the size of the dataset. Reliable estimation of nonlinear measures inspired from chaos theory (such as Lyapunov exponents, correlation dimensions etc.), on the other hand, require calculations over many parameters after which regions of linear or constant trends are sought. If no such regions are evident or if those regions are dependent on the choice of parameters, the tests cannot be conclusive.

For the same number of lags, the computational complexity of correntropy is on the same order as autocorrelation and other statistical nonlinearity measures such as time reversibility and bicorrelation. The Ramsey–Rothman [39] definition of time reversibility uses the following third order autocorrelation:

$$\phi_{\text{TR}}(\tau) = E\{x_t x_{t+\tau}^2 - x_t^2 x_{t+\tau}\} \tag{9}$$

A time series that can be fully characterized by its autocorrelation (such as a linear Gaussian process) is invariant to any time reversals due to the time symmetry property of the autocorrelation function. Thus, any sign of time asymmetry in a time series would lead to the rejection of the null hypothesis.

Bicorrelation is the correlation of a signal with two different time lags [20]:

$$R_{\text{xxx}}(\tau_1, \tau_2) = E\{x_t x_{t+\tau_1} x_{t+\tau_2}\} \tag{10}$$

In order to reduce the number of parameters Schreiber et al. [22] employ $\tau_2 = 2\tau$. Schreiber [22] points out that these two measures are very easy to compute and although work well generally (i.e., elicit good discrimination powers), they can also fail completely. Since we want to emphasize the low computational demands of correntropy, it is only fair that we compare correntropy based tests with these two measures.

4. Simulation results

4.1. Discrimination power analysis on deterministic data

In this section, we present simulations to demonstrate the utility of the proposed correntropy measure as a tool to detect nonlinear structures on three synthetic time series: (i) linear Gaussian, (ii) linear non-Gaussian, (iii) nonlinear, at several noise levels and dataset lengths by running 100 Monte Carlo simulations at each varying parameter. The additive noise in systems (ii) and (iii) is compromised of scaled phase-randomized realizations of the signal in order not to distort the autocorrelation function [22]. The three signal-to-noise ratios (SNR) examined herein are 6, 10 and 20 dB and the data lengths are varied from $N = 500$ to 5000 with increments of 500 samples. The confidence level of the tests is set as 95% (corresponding to significance level of 0.05) and 19 surrogates are generated for each simulation via IAAFT.

4.1.1. Tests on linear Gaussian processes

We create a synthetic linear Gaussian process through a linear feedback system

$$x(n) = 1.5x(n-1) - 0.8x(n-2) + \varepsilon(n) \tag{11}$$

where $\varepsilon(n)$ follows a white Gaussian distribution with zero mean and unit variance. The autocorrelation and correntropy of the original series and its surrogates are shown in Fig. 1. The Kolmogorov–Smirnov test comparing the correntropy spectral densities did not reject that the surrogates were generated from the same distribution. In other words, the null hypothesis of a Gaussian linear source was not rejected. Adding white Gaussian noise and changing the number of samples did not result in the rejection of the null hypothesis with the proposed measure in 100 Monte Carlo simulations. Other examples of linear Gaussian time series are included in [42].

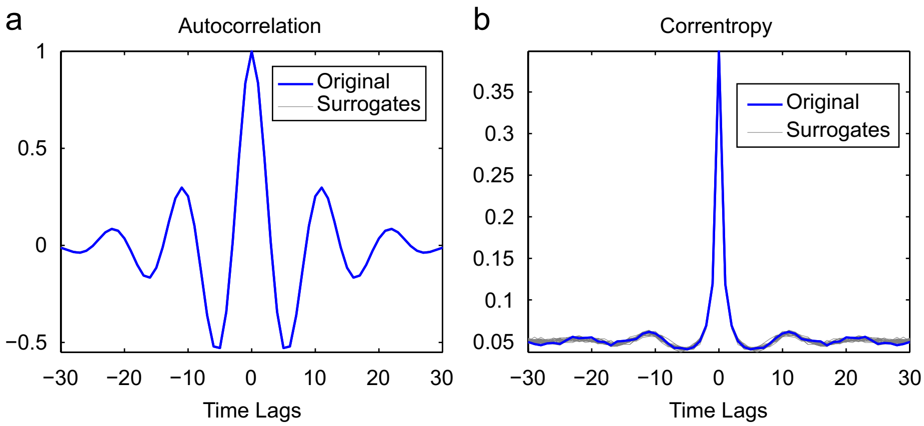

Fig. 1. (a) Autocorrelation and (b) correntropy functions of synthetic linear Gaussian data and surrogate series.

4.1.2. Tests on non-Gaussian linear processes

As discussed in Section 2, when the null hypothesis is rejected on an examined series through a nonlinearity measure, we are not certain whether to attribute the rejection to nonlinear dynamics or the non-Gaussian nature of the process. Nagarajan [16] has extensively studied Fourier based surrogate techniques in the presence of non-Gaussian linear processes and instantaneous nonlinearities. The tests all resulted in the rejection of the null hypothesis due to the non-Gaussian nature of the series rather than nonlinearity. Hinich et al. [28] have also reported high false alarm rates when nonlinearity tests were conducted on linear exponential processes using Fourier based surrogates. A measure that does not reject the null hypothesis in the presence of non-Gaussian linear processes would be considered to have low discrimination power from the perspective of testing for Gaussianity, however, from the perspective of nonlinearity tests, it would increase the liability of rejections suggested by this measure.

For the tests, the linear system described in (11) is injected with $\varepsilon(n)$ sampled from a white exponential distribution with zero mean and unit variance.[3] The original series and its realizations are compared against correntropy based measures, time reversibility and bicorrelation for varying data lengths ($N = 500-5000$ with increments of 500 samples). One hundred Monte Carlo simulations were performed with in-band Gaussian noise levels of SNR = 6, 10, and 20 dB at each data length. The false alarm rates for the three statistics at SNR = 20 dB are plotted in Fig. 2. For correntropy based tests, the mean and standard deviation of the false alarm rates for the three SNR levels across the varying data lengths are found to be $\{10.0 \pm 6.9\%, 8.8 \pm 4.8\%, 4.4 \pm 7.2\%\}$. Hence, the rejection of the null hypothesis in the absence of nonlinear dynamics is low for the proposed measure.

On the other hand, significant false alarm rates are attained for measures of time reversibility and bicorrelation. The rejection of the null hypothesis with these measures is due to the non-Gaussianity of the synthesized linear process.[4]

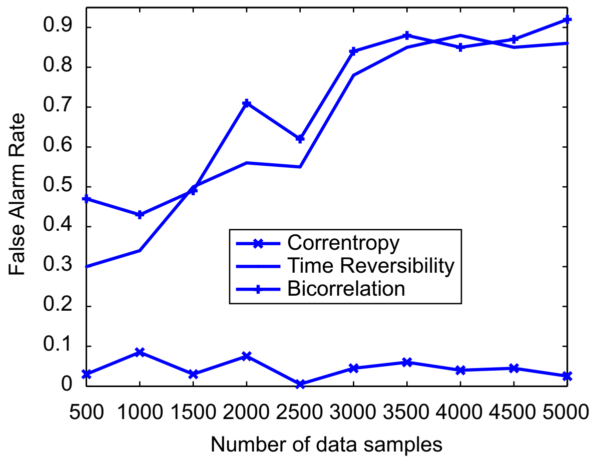

Fig. 2. False alarm rates of the three measures on noisy synthetic linear non-Gaussian data (SNR = 20 dB).

4.1.3. Tests on nonlinear processes

Finally, we generate nonlinear data via the chaotic Lorenz system governed by the equations

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z \end{aligned} \tag{12}$$

with $\rho = 28, \beta = 8/3$ via the fourth-order Runge–Kutta method (with integral step 0.005). The results of the test for the y-component are depicted in Fig. 3. Decreased discrimination powers with increased noise levels and fewer data samples are expected. Schreiber [22] questions the usefulness of a measure with discrimination powers less than 0.7. Therefore, we infer that at high SNR levels

[3] Kolmogorov–Smirnoff tests were used to verify that the amplitude distribution still deviated from Gaussianity after the filtering process.

[4] In fact, Weiss [40] showed that all non-Gaussian linear processes are non-time reversible.

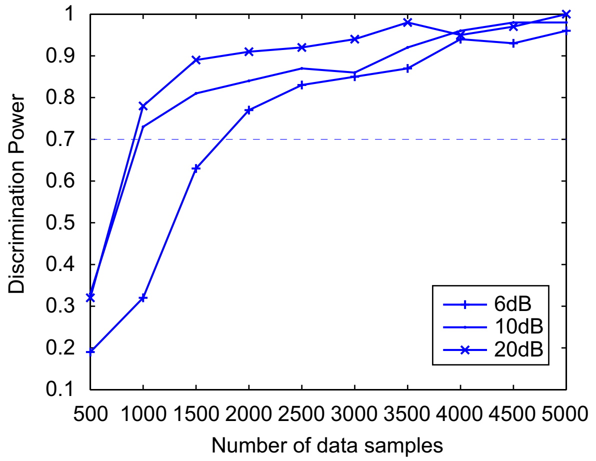

Fig. 3. Discrimination power of the correntropy based measure on noisy Lorenz series (SNR = 20 dB) for three levels of noise and varying data lengths.

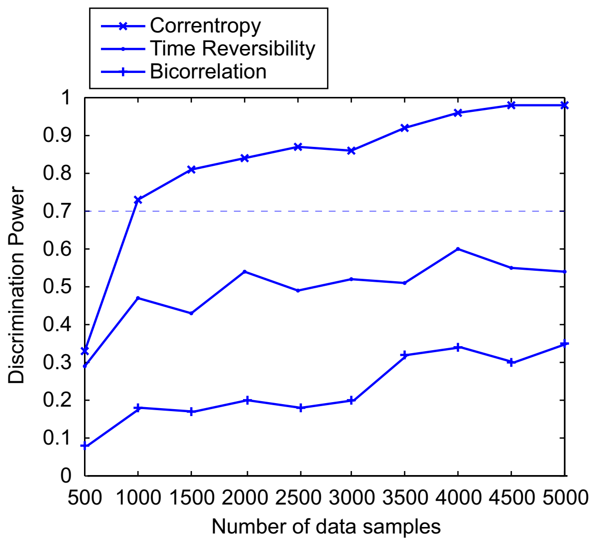

Fig. 4. Discrimination power of the three measures on noisy Lorenz series (SNR = 20 dB) as a function of varying data lengths.

(10–20 dB) data lengths of $N = 1000$ are sufficient for adequate discrimination powers.

Further, the discrimination powers of the three measures at SNR = 20 dB are given in Fig. 4. We observe that the other two measures fail to yield discrimination power levels of 0.7. These results are in accord with Schreiber and Schmitz [22] who reported no significant discrimination powers with time reversibility and bicorrelation measures for the x-component of the Lorenz series with in-band noise of 6 dB (with parameter $\tau = 1$).

Finally, we examined discrimination power of the test with varying kernel sizes for the Lorenz series. The data length and signal-to-noise ratio were fixed at $N = 2500$ and SNR = 20 dB, respectively. Once again 100 Monte Carlo simulations were performed at each kernel size. The

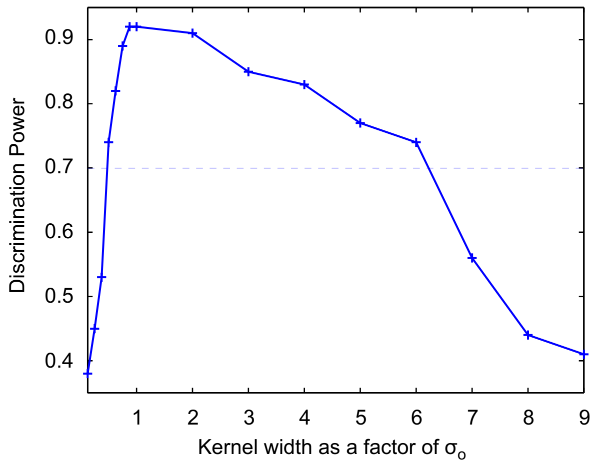

Fig. 5. Discrimination power as function of kernel sizes for the Lorenz series with $N = 2500$ and (SNR = 20 dB).

kernel widths were chosen to be factors of the kernel width $\sigma_o = 0.1956$ attained by Silverman's rule. The results of this set of simulations are given in Fig. 5. The reason we witness decreased discrimination powers when the kernel size is too small can be attributed to the fact that correntropy converges to a delta-Dirac function for both the original time series and the surrogates and therefore the measure loses its ability to discriminate. When the kernel size is increased too much, the scaling factor for the higher order moments becomes negligible and the scaling factor for the autocorrelation term increases. The shape of the correntropy function converges to the autocorrelation function [41]. Fig. 6 demonstrates these converging properties of the correntropy function for the noiseless Lorenz series [42]. The skewness in Fig. 5 is due to the fact that correntropy converges faster to the delta-Dirac function than it converges to the autocorrelation function.

4.2. *Nonlinearity tests on Santa Fe Competition data with correntropy*

We performed nonlinearity tests based on our proposed measure on time series which were used in a competition held at the NATO Advanced Research Workshop on Comparative Time Series Analysis which today is widely known as the Santa Fe Competition data [9]. This is still today a defacto standard to compare different nonlinearity tests. The three datasets we have chosen are dataset A: Lorenz-like chaos in NH3-FIR lasers [43], dataset B: multichannel physiological data [44], and dataset E: whole Earth telescope observations of the White Dwarf star [45]. Some of the attributes of the datasets pointed out in [9] are as follows: Dataset A is a stationary, low-dimensional, clean time series collected in a laboratory experiment in one trial and it shows nonlinear behavior. Dataset B is a natural system, which was recorded in episodes over multiple channels and well documented. It consists of heart rates, respiration rates, and blood oxygen concentrations. The analysis of this

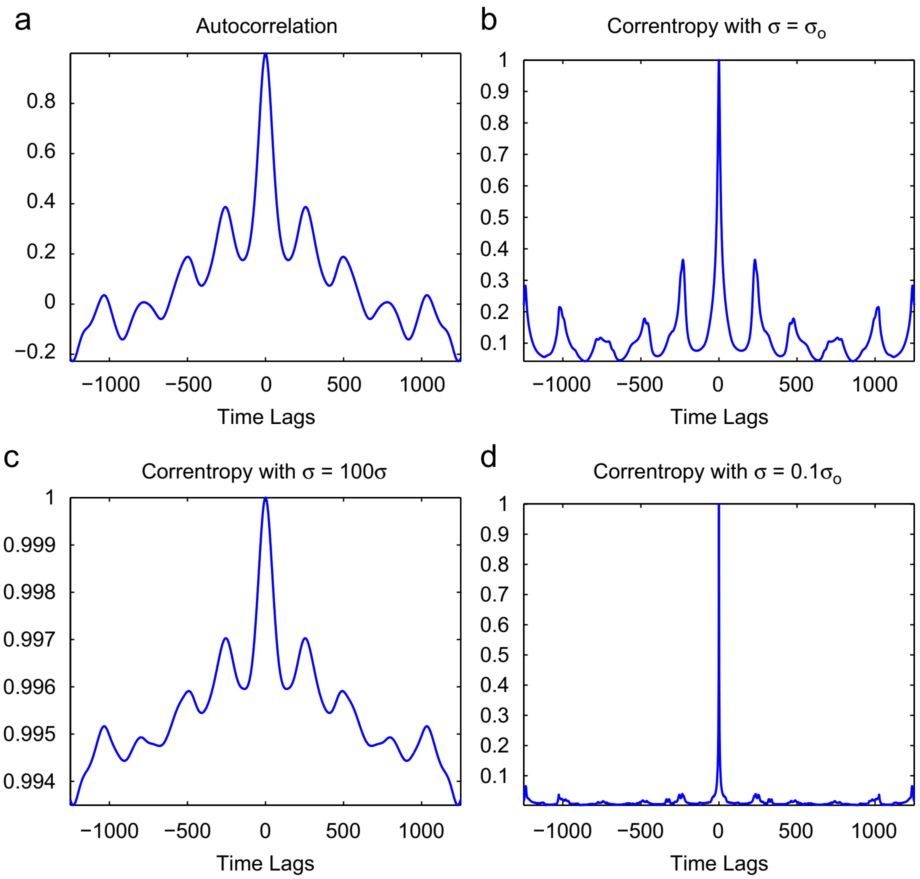

Fig. 6. Comparison of autocorrelation and normalized correntropies of Lorenz data with kernel widths σ_o (based on Silverman's rule), $100\sigma_o$, $0.1\sigma_o$.

dataset can potentially save lives. Finally, dataset *E* is a noisy well-documented natural system, which is most probably linear.

Dataset *A* which consists of 1000 data points is given in Fig. 7 along with one of its surrogates. Autocorrelation and correntropy functions of the dataset and its surrogates are also given. It can be observed that the autocorrelation of the dataset belongs to the distribution of the autocorrelation functions of the surrogates. The correntropy function of the dataset, on the other hand, is distinctly different than its surrogates and this observation is confirmed by the Kolmogorov–Smirnoff tests in which the normalized CSD of the dataset rejected hypothesis of belonging to the same distribution as all of the 19 surrogates. Hence, the proposed measure rejected the null hypothesis with 95% confidence. Because the dataset was not heavily contaminated with noise, 1000 data samples were sufficient for correct discrimination. The same result (nonlinearity of dataset *A*) was also reported by Casdagli et al. [46] using nonlinear prediction errors (which is against our goal of avoiding nonlinear modeling with no a priori information on the system) and by Palus [10] using redundancies and linear redundancies (which require extensive computations).

Fig. 8 presents the correntropies of one episode of the heart rate recordings in dataset *B* and its generated surrogate series. The system was recorded at 2 Hz sampling rate with data length of $N = 17\,000$. Our correntropy-based tests applied to 4 min of data rejected the null hypothesis with 95% confidence over 19 surrogates. This result is confirmed by Casdagli et al. [46] who found 10–25% more accurate prediction results with nonlinear models compared to global linear ones.

Finally, we examined dataset *E*.14 and its surrogates. The data length of this set is $N = 2602$ and the dataset was contaminated with noise. The corresponding autocorrelation and correntropy function are provided in Fig. 9. The correntropy of the original series is not distinguishable from those of its surrogates. Moreover, the CSD based tests do not reject the null hypothesis and thus the test is inconclusive about whether there are nonlinear dynamics in this time series. Since this is a natural system, we would anticipate that there are nonlinearities present. Nevertheless, other measures are in accord with this finding. For the same dataset, Palus [10] could not find significant differences between linear and nonlinear measures of redundancy and therefore reported that a linear stochastic process is consistent with the data. Theiler et al. [47] state that evidence of nonlinearity in this dataset should be dismissed as an artifact of long coherence time. Thus, we have shown that correntropy is a suitable measure for nonlinearity tests, not only on synthetic data, but also on

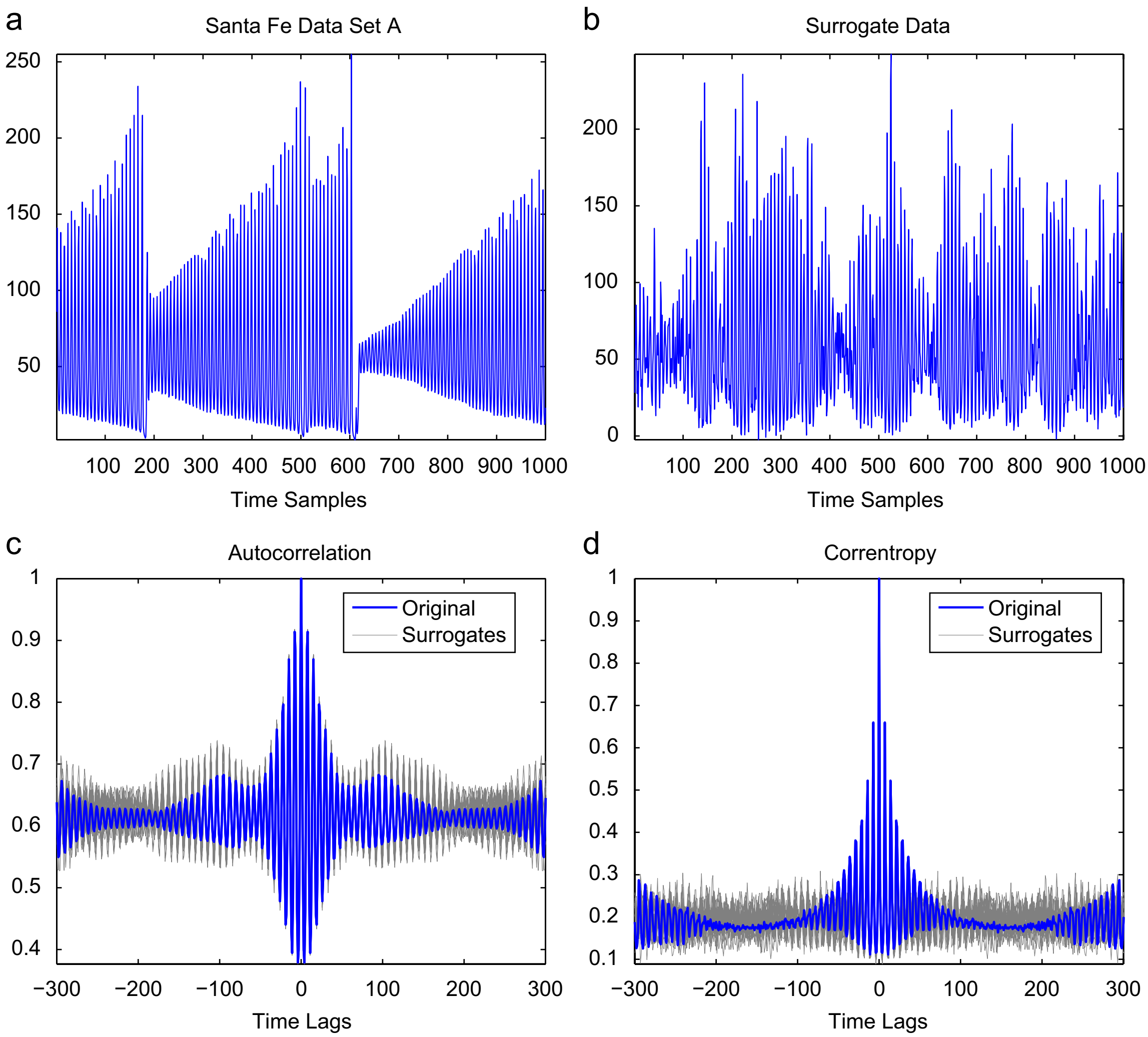

Fig. 7. (a) Original dataset *A*, (b) one of the generated surrogates, (c) autocorrelation, and (d) correntropy functions of original and surrogate series.

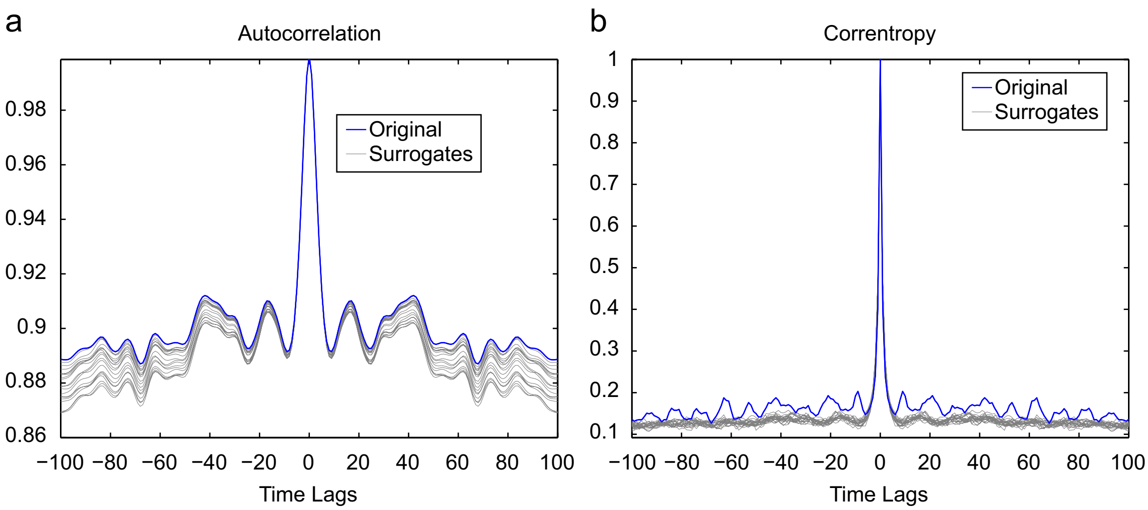

Fig. 8. (a) Autocorrelation and (b) correntropy functions of dataset *B* and surrogate series.

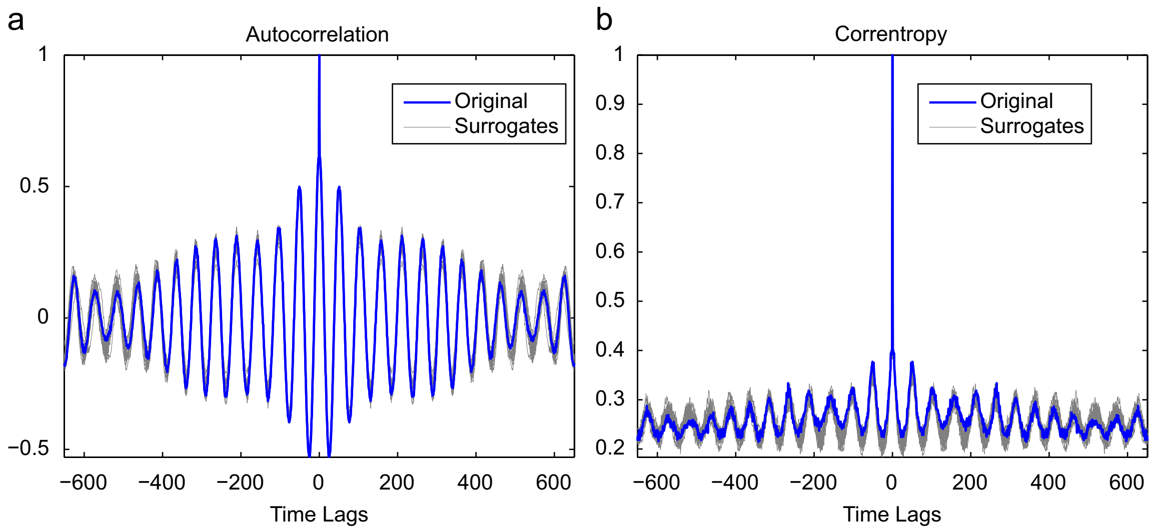

Fig. 9. (a) Autocorrelation and (b) correntropy functions of dataset *E*.14 and surrogate series.

limited samples of noisy measurements from real physical systems.

5. Conclusions

Herein, a generalized power spectral measure based on correntropy for surrogate based nonlinearity tests has been proposed. The methodology relies on the fact that correntropy captures dynamic properties of time series through nonlinear mappings of kernels, without resorting to any conventional nonlinear test statistics such as Lyapunov exponents or correlation dimension, which rely on many parameters and are computationally time consuming. The method rejects the null hypothesis that the observed signal is of Gaussian and linear nature if the correntropy power spectral densities of the signal and all of its surrogates fail the two-sample Kolmogorov–Smirnoff test. This test scheme is based on the loss of dynamic nonlinear properties of the original series through the process of generating surrogate data. If no such properties exist, the series and its surrogates share the same CSD distribution. It should be pointed out that this nonlinearity measure, just like other statistical nonlinearity measures, relies on the method of surrogate data and is not a stand alone measure. Therefore, the measure would fail if the generation of proper surrogates failed.

The methodology has been applied to synthetic linear Gaussian, linear non-Gaussian, and nonlinear data and analyzed with varying in-band noise levels, data lengths and kernel widths. The rejection of the null hypothesis in the two linear cases is insignificant. Not rejecting the null hypothesis in the linear case of non-Gaussian sources demonstrates the liability of the proposed statistic as a nonlinearity measure. Discrimination power of the measure was examined with a nonlinear Lorenz attractor for various SNR levels, dataset lengths and kernel sizes. Significant discrimination powers were attained in the presence of in-band noise at data lengths as low as 1000 samples and a wide range of kernel widths. Finally, we applied our nonlinearity measure to real natural systems with various data lengths and corruptive noises. Our results were in accord with the results published in literature on the same datasets, supporting the liability of the proposed measure. Overall, the proposed correntropy based nonlinearity measure is a computationally efficient and reliable test statistic which captures not only temporal correlations but also dynamical properties of time series.

Acknowledgments

The authors would like to thank Dr. Anant Hegde and Weifeng Liu for their insightful discussions.

References

[1] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press, Cambridge, 1997.
[2] W.J. Freeman, Mass Action in the Nervous System, Academic Press, London, 1975.
[3] T. Schreiber, A. Schmitz, Surrogate time series, Physica D 142 (2000) 346–382.
[4] T. Gautama, D.P. Mandic, M.M. van Hulle, A novel method for determining the nature of a time series, IEEE Trans. Biomed. Eng. 51 (2004) 728–736.
[5] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, J.D. Farmer, Physica D 58 (1992) 77–94.
[6] D. Kugiumtzis, in: A. Soofi, L. Cao (Eds.), Nonlinear Deterministic Modeling and Forecasting of Economic and Financial Time Series, Kluwer Academic Publishers, Dordrecht, 2002.
[7] K.T. Alligood, Chaos: An Introduction to Dynamical Systems, Springer, New York, 1997.
[8] C. Diks, Nonlinear Time Analysis: Methods and Applications, World Scientific, Singapore, 1999.
[9] A.S. Weigend, N.A. Gershenfeld (Eds.), Time Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley, Reading, MA, 1994.
[10] M. Palus, Identifying and quantifying chaos using information—theoretic functionals, in: A.S. Weigend, N.A. Gershenfeld (Eds.), Time Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley, Reading, MA, 1994, p. 387 (Chapter III).

[11] J.M. Nichols, M. Seaver, S.T. Trickey, M.D. Todd, C. Olson, L. Overbey, Detecting nonlinearity in structural systems using the transfer entropy, Phys. Rev. E 72 (2005) 046217.
[12] P. Grassberger, I. Procaccia, Measuring the strangeness of strange attractors, Physica D 9 (1983) 189–208.
[13] H. Kantz, T. Schreiber, Dimension estimates and physiological data, Chaos: An Interdisciplinary J. Nonlinear Sci. 5 (1995) 143–154.
[14] T. Gautama, D.P. Mandic, M.M. Van Hulle, The delay vector variance method for detecting determinism and nonlinearity in time series, Physica D 190 (2004) 167–176.
[15] S.M. Pincus, I.M. Gladstone, R.A. Ehrenkranz, A regularity statistic for medical data analysis, J. Clinical Monitoring 7 (1991) 335–345.
[16] R. Nagarajan, Surrogate testing of linear feedback processes with non-Gaussian innovations, Physica A: Statistical Mechanics Applications 366 (2006) 530–538.
[17] S. Basu, E. Foufoula-Georgiou, Detection of nonlinearity and chaoticity in time series using the transportation distance function, Phys. Lett. A 301 (2002) 413–423.
[18] C. Diks, J.C. van Houwelingen, F. Takens, J. DeGoede, Reversibility as a criterion for discriminating time series, Phys. Lett. A 201 (1995) 221–228.
[19] T. Schreiber, Interdisciplinary application of nonlinear time series methods, Phys. Reports 308 (1999) 1–64.
[20] T. Subba Rao, M.M. Gabr, An Introduction to Bispectral Analysis and Bilinear Time Series Models, Springer, New York, 1984.
[21] D.T. Kaplan, L. Glass, Coarse-grained embeddings of time series: random walks, Gaussian random processes, and deterministic chaos, Physica D 64 (1993) 431–454.
[22] T. Schreiber, A. Schmitz, Discrimination power of measures for nonlinearity in a time series, Phys. Rev. E 55 (1997) 5443–5447.
[23] M.J. Hinich, Testing for Gaussianity and linearity of a stationary time series, J. Time Ser. Anal. 3 (1982) 169–176.
[24] I. Santamaria, P.P. Pokharel, J.C. Principe, Generalized correlation function: definition, properties and application to blind equalization, IEEE Trans. Signal Process. 54 (2006) 2187–2197.
[25] M. Aizerman, E. Braverman, L. Rozonoer, Theoretical foundations of the potential function method in pattern recognition learning, Automat. Control 25 (1964) 821.
[26] P. Stoica, R. Moses, Spectral Analysis of Signals, Prentice-Hall, Upper Saddle River, NJ, 2005.
[27] T. Schreiber, A. Schmitz, Improved surrogate data for nonlinearity tests, Phys. Rev. Lett. 77 (1996) 635–638.
[28] M.J. Hinich, E.M. Mendes, L. Stone, Detecting nonlinearity in time series surrogate and bootstrap approaches, Studies in Nonlinear Dynamics & Econometrics 9 (4) (2005) (art. 3).
[29] A. Galka, T. Ozaki, Testing for nonlinearity in high-dimensional time series from continuous dynamics, Physica D: Nonlinear Phenomena 158 (2001) 32–44.
[30] R. Hegger, H. Kantz, T. Schreiber, Practical implementation of nonlinear time series methods: the TISEAN package, Chaos 9 (1999) 413–435.
[31] E.N. Lorenz, Deterministic nonperiodic flow, J. Atmospheric Sci. 20 (1963) 130–148.
[32] B. Schölkopf, A. Smola, Learning with Kernels, MIT Press, Cambridge, 2002.
[33] T. Friess, N. Cristianini, C. Campbell, The Kernel-Adatron: a fast and simple learning procedure for support vector machines, in: Proceedings of the 15th International Conference on Machine Learning, Madison, WI, 1998, pp. 188–196.
[34] S. Mika, G. Rätsch, J. Wetson, B. Schölkopf, B. Fisher, Fisher discriminant analysis with kernels, in: IEEE Workshop on Neural Networks for Signal Processing, Madison, WI, 1999, pp. 41–48.
[35] F.R. Bach, M.L. Jordan, Kernel independent component analysis, J. Machine Learning Res. 3 (2002) 1–48.
[36] P. Pokharel, R. Agrawal, J.C. Principe, Correntropy based matched filtering, in: IEEE Workshop on Machine Learning for Signal Processing, Mystic, CT, 2005, p. 341.
[37] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, Cambridge, 1992.
[38] S.J. Sheather, Density Estimation, Statistical Sci. 19 (2004) 588–597.
[39] J.B. Ramsey, P. Rothman, Time Irreversibility and Business Cycle Asymmetry, J. Money, Credit Banking 28 (1996) 1–21.
[40] G. Weiss, Time reversibility of linear stochastic processes, J. Appl. Probability 12 (1975) 831–836.
[41] D. Erdogmus, Information theoretic learning: Renyi's entropy and its applications to adaptive system training, Ph.D. Thesis, Gainesville, FL, USA, 2002.
[42] A. Gunduz, A. Hegde, and J.C. Principe, Correntropy as a novel measure for nonlinearity tests, in: Proceedings of the International Joint Conference on Neural Networks, Vancouver, 2006.
[43] Hubner, N.B. Abraham, C.O. Weiss, Dimensions and entropies of chaotic intensity pulsations in a single-mode far–infrared NH_3 laser, Phys. Rev. A 40 (1989) 6354–6365.
[44] D.R. Rigney, A.L. Goldberger, W.C. Ocasio, Y. Ichimaru, G.B. Moody, R.G. Mark, Multi-channel physiological data: description and analysis, in: A.S. Weigend, N.A. Gershenfeld (Eds.), Time Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley, Reading, MA, 1994, p. 105 (Chapter I).
[45] J.C. Clemens, Whole Earth telescope observations of the White Dwarf Star, in: A.S. Weigend, N.A. Gershenfeld (Eds.), Time Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley, Reading, MA, 1994, p. 139 (Chapter I).
[46] M.C. Casdagli, A.S. Weigend, Exploring the continuum between deterministic and stochastic modeling, in: A.S. Weigend, N.A. Gershenfeld (Eds.), Time Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley, Reading, MA, 1994, p. 347 (Chapter III).
[47] J. Theiler, P.S. Lindsay, D.M. Rubin, Detecting nonlinearity in data with long coherence times, in: A.S. Weigend, N.A. Gershenfeld (Eds.), Time Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley, Reading, MA, 1994, p. 429 (Chapter III).
[48] E. Parzen, Statistical methods on time series by Hilbert space methods, Technical Report no. 23, Applied Mathematics and Statistics Laboratory, Stanford University, USA, 1959.