

Correntropy as a Novel Measure for Nonlinearity Tests

Aysegul Gunduz, Anant Hegde, and Jose C. Principe
 Computational NeuroEngineering Laboratory, Department of Electrical Engineering
 University of Florida, Gainesville, FL 32611

{aysegul,ahegde,principe}@cnel.ufl.edu
 http://www.cnel.ufl.edu

Abstract—Statistical tests have become an essential step in nonlinear system modeling due to the complexities involved in their analysis. Correntropy is a kernel-based similarity measure which includes the information of both distribution and time structure of a stochastic process. The correntropy function’s capability of preserving nonlinear characteristics and high order moments makes it a suitable candidate as a statistic for determining whether a nonlinear structure exists within the system that created the observed time series. Experiments based on surrogate data methods have confirmed that correntropy can be employed as a discriminant measure for detecting nonlinear characteristics in time series.

I. INTRODUCTION

RECENT years have witnessed growing interest in kernel based methods for signal processing applications. Numerous kernel methods have emerged and have successfully been applied to problems of classification, pattern analysis and decomposition [1-5]. Recently introduced, *correntropy* [6] is a similarity measure which takes advantage of the power of kernels to nonlinearly map the input space to a higher dimensional feature space in which inner products can be computed efficiently. The motivation behind correntropy is to combine the statistical distribution and time structure of signals in a single measure. Correntropy and the conventional autocorrelation exhibit common properties, but unlike autocorrelation, correntropy can preserve high order moments and therefore the existing nonlinear characteristics of a signal.

Prior knowledge of the underlying dynamical properties of a natural system can guide and increase accuracy in system modeling. In practical applications nonlinear filters should be avoided if the underlying signals are in fact linear in nature, due to increased complexity of training nonlinear models [7]. Conventional measures for testing nonlinearity include Lyapunov exponents [8] and correlation dimensions [9] which require embedding of signals into higher dimensional and extensive computations. Motivated by its ability to preserve nonlinear characteristics and computational simplicity in high dimensional kernel spaces, we propose to use correntropy as a discriminant measure for determining whether a nonlinear structure exists within an analyzed system.

This paper is organized as follows. In the following section a brief background on kernel functions is given along with the definition and some important properties of correntropy. In Section III, we present surrogate based

methods for nonlinearity tests and explain how correntropy is used as a discriminating statistic. Finally, the main conclusions are summarized in Section IV.

II. CORRENTROPY: A GENERALIZED CORRELATION FUNCTION

Kernel methods are based on the idea of mapping the input space to a higher dimensional feature space in which inner products can be computed efficiently without explicit knowledge of the mapping. The kernel functions employed in these techniques are positive-definite, satisfy Mercer’s conditions [1] and are denoted by

$$\kappa(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad (1)$$

where ϕ is the mapping from the input to the feature space. A widely used Mercer kernel is the Gaussian kernel given by:

$$\kappa(x_i, x_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2)$$

Correntropy is a *similarity measure in feature space* [6], with the motivation to combine the time structure and statistical distribution of signals in a single measure. Thus, although based on kernels, it is different from the conventional kernel methods that work independently with pairs of samples. If $\{x_t, t \in T\}$ is a stochastic process with T being an index set, then the correntropy function $V(s, t)$ is defined as the expected value of the kernel:

$$V(s, t) = E\{\kappa(x_s, x_t)\} \quad (3)$$

a mapping from $T \times T$ into \mathbb{R}^+ . Using a series expansion for the Gaussian kernel, it can be shown that the information provided by the autocorrelation is included within correntropy for $n=1$ [6]:

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

In essence, correntropy generalizes the autocorrelation function to nonlinear spaces. The two functions share many properties: both are symmetric with respect to the origin and take on their maximum value at zero lag. Moreover, from

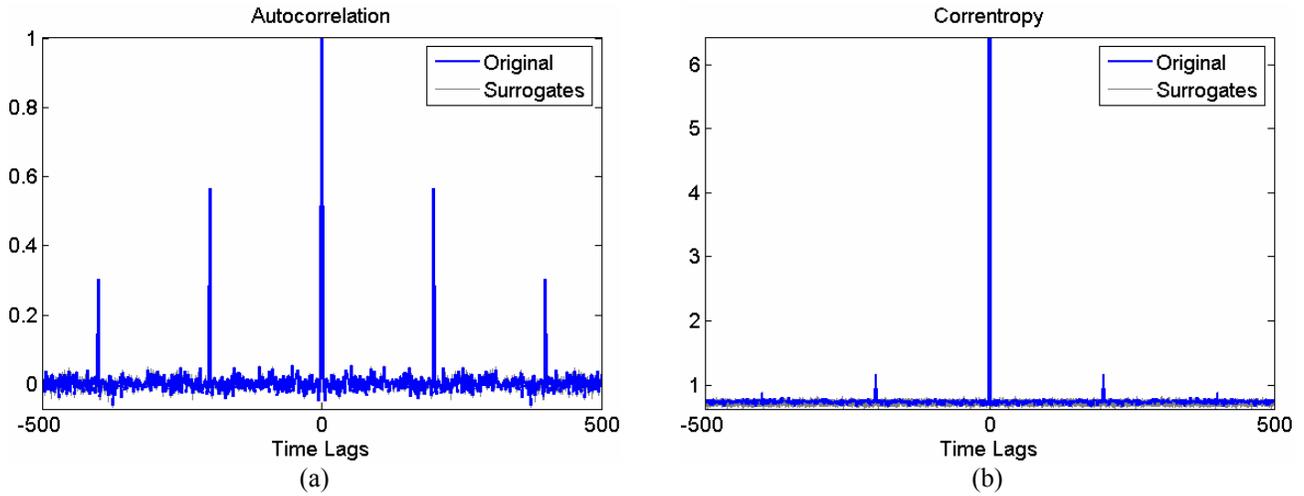


Figure 1. (a) Autocorrelation functions of colored Gaussian noise and of its surrogates, (b) the corresponding correlogram functions.

(4) we infer that correlogram involves higher order moments of $\|x_s - x_t\|$ and exhibits other important properties autocorrelation function does not possess.

For a discrete-time strictly stationary stochastic process correlogram can be estimated through the sample mean:

$$\hat{V}[m] = \frac{1}{N - m + 1} \sum_{n=m}^N \kappa(x_n, x_{n-m}) \quad (5)$$

Finally, the correlogram spectral density is defined in [6] as:

$$P_V[w] = \sum_{m=-\infty}^{\infty} \hat{V}[m] e^{-jwm} \quad (6)$$

which retains all properties of the conventional power spectral density.

As with any kernel method, the choice of kernel size affects the performance of the technique. The kernel size generally is determined empirically. In this study, the bandwidth of the kernel is selected according to Silverman's rule of thumb [10]:

$$\sigma = 0.9AN^{-1/5} \quad (7)$$

where $A = \min\{\text{standard deviation of data samples, data interquartile range}/1.34\}$ and N is the data length.

III. SURROGATE METHODS FOR TESTS OF NONLINEARITY

Methods of surrogate data provide a statistically rigorous framework for tests of nonlinearity [11]. The surrogate data is generated to represent the *null hypothesis* that the examined time series is generated by a Gaussian linear stochastic process. Properly designed surrogate data should only retain up to second order statistics and histogram of the original signal, and be otherwise random [11]. The generated surrogate data are compared to the original data under a discriminating statistic with a confidence level in order to reject the null hypothesis.

One of the widely used surrogate generation methods is Fourier based surrogates [12], which require that the surrogates have the same Fourier amplitudes as the data but with random phases. The key point in these methods is that the squared amplitude of the Fourier transform is a periodogram estimator of the (conventional) power spectral density [13]. Hence, the original time series and its surrogates attained by this method share the same power spectrum and thus the same autocorrelation function regardless the null hypothesis is true or false. However, phase randomization alters the underlying nonlinear dynamic structure within the original data, and thus using a statistic capable of capturing the nonlinear characteristics, such as correlogram, would enable us to reject the null hypothesis within a level of confidence. There are surrogate methods that provide better approximations to the original linear correlations such as amplitude adjusted Fourier transform (AAFT) [12], corrected AAFT (CAFFT) [9, 10] and iterative AAFT (IAAFT) [13]. In this study we generate surrogates via IAAFT.

In the test design, initially a residual probability of a false rejection, α , is selected, which corresponds to a confidence level of $(1 - \alpha) \times 100\%$. For a one-sided test $M = 1/\alpha - 1$ surrogate sequences are generated [12]. For example, a confidence level of 95% would require generation of at least 19 surrogates. Discrimination power can be further increased by engaging more surrogates.

A. Surrogate Data Tests with Correlogram

We herein propose to use the correlogram spectral density, i.e. the Fourier transform of correlogram, as the nonlinear discriminating measure for the rejection of the null hypothesis. The Fourier Transform amplitudes of a random process are known to possess a chi squared distribution at each frequency with the number of degrees of freedom given by the length of the window [14]. The correlogram spectral density represents the distribution of total power amongst frequencies. Thus, if normalized by the total power, it is essentially the probability density function of correlogram power over frequencies. If an examined time

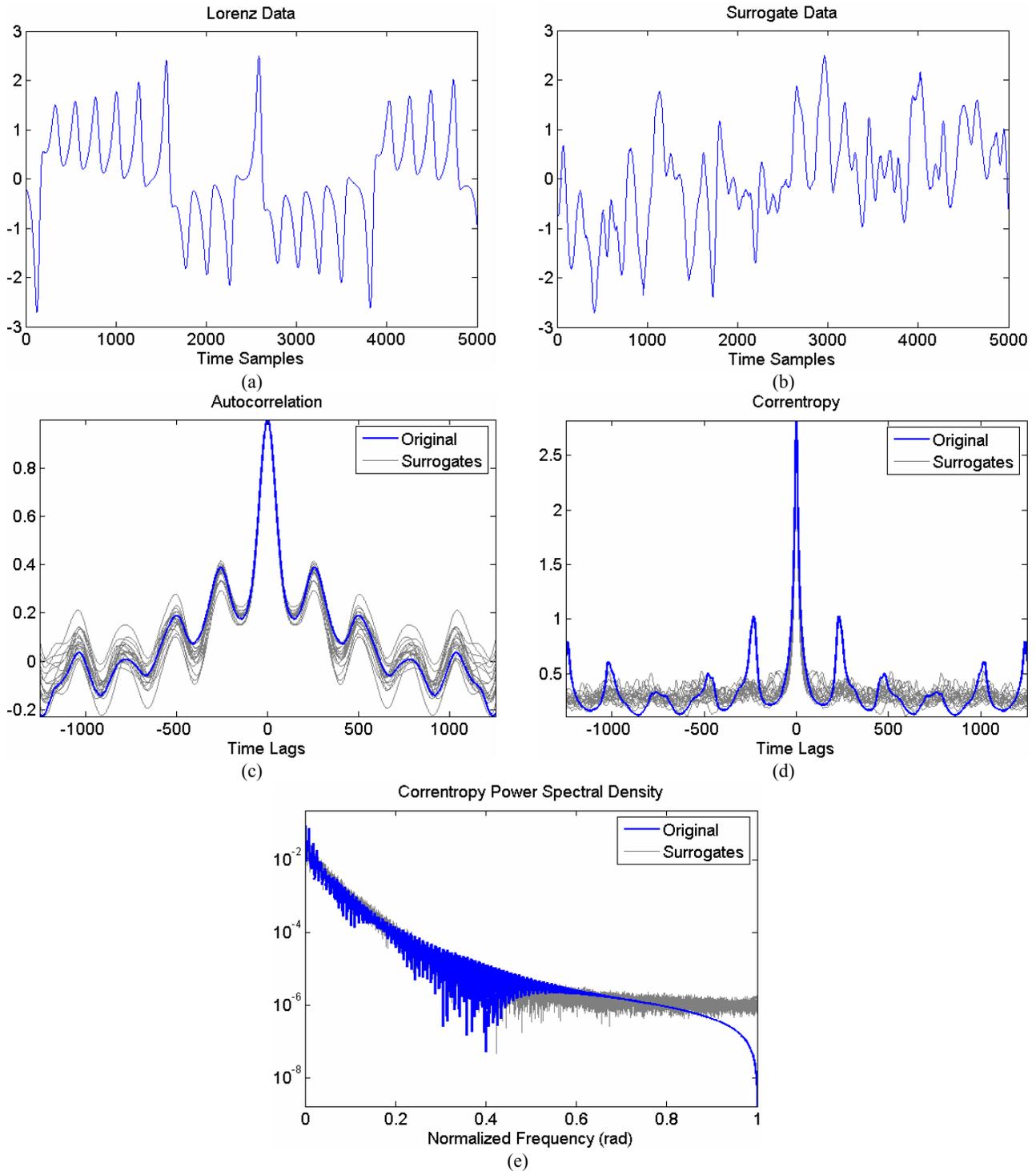


Figure 2. (a) Lorenz series, (b) one of its surrogates. (c) Autocorrelation functions of Lorenz data and of its surrogates, (d) the corresponding correntropy functions, (e) the correntropy power spectral density plots in logarithmic scale.

series does not possess nonlinear dynamics, the underlying distribution of its correntropy power and that of its surrogates should 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. Therefore, the null hypothesis that the time series is generated by a Gaussian

linear stochastic process is rejected. (It should be noted that if the underlying distributions for the series and its surrogates are found to be the same, this does not imply that we can accept the null hypothesis. We rather conclude that the null hypothesis cannot be rejected for this measure.)

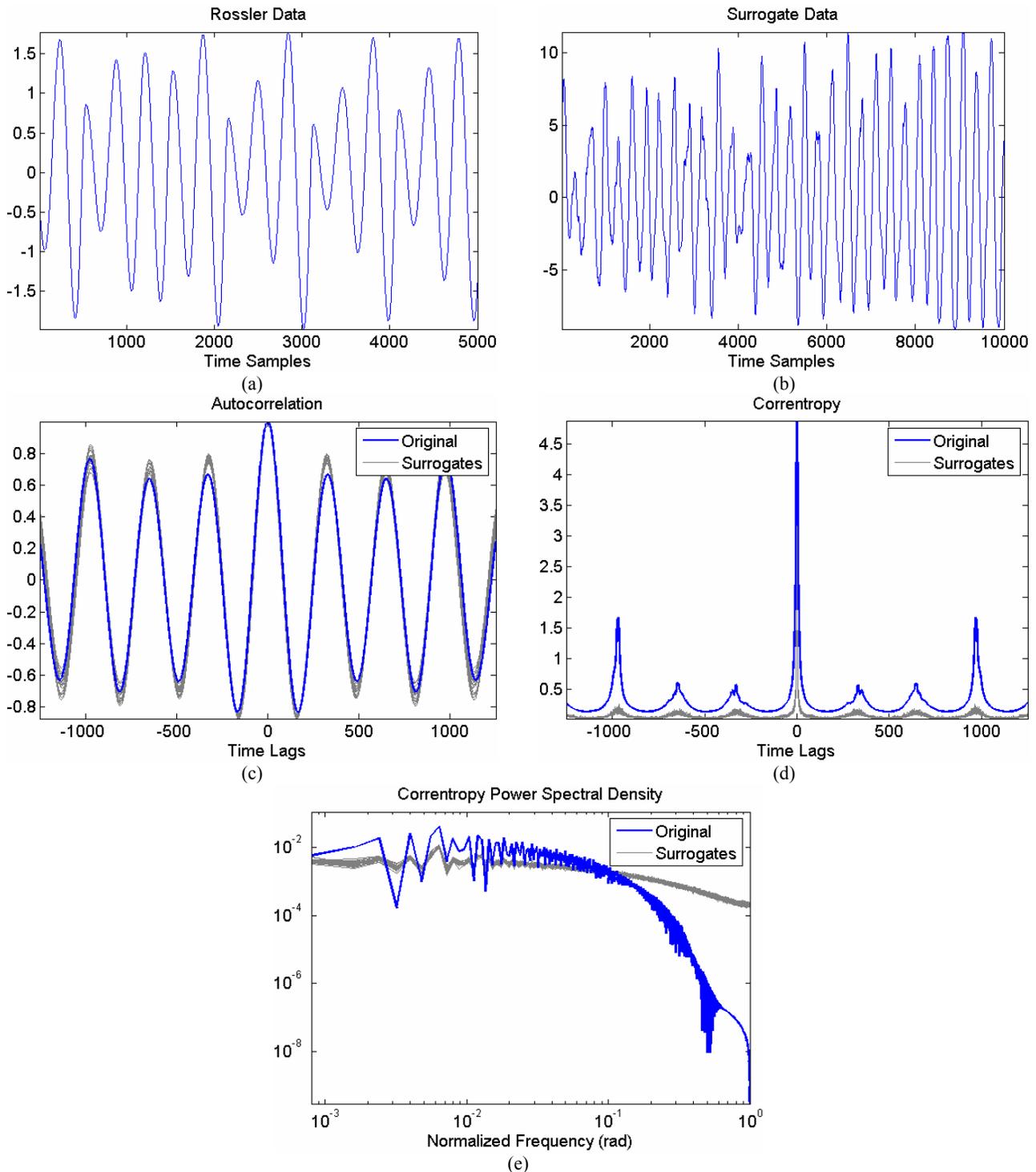


Figure 3. (a) Rössler series, (b) one of its surrogates. (c) Autocorrelation functions of Rössler data and of its surrogates, (d) the corresponding correntropy functions, (e) the correntropy power spectral density plots in logarithmic scale.

B. Kolmogorov-Smirnov Goodness-of-Fit Test

The two-sample Kolmogorov-Smirnov goodness-of-fit is a powerful tool that tests whether two time series are random samples from the same distribution [15]. 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

correntropy power densities of the original and surrogate series.

The maximum difference between the ECDFs of the two series is the test statistic:

$$D = \max_f |F_{orig}(f) - F_{surr}(f)|. \quad (8)$$

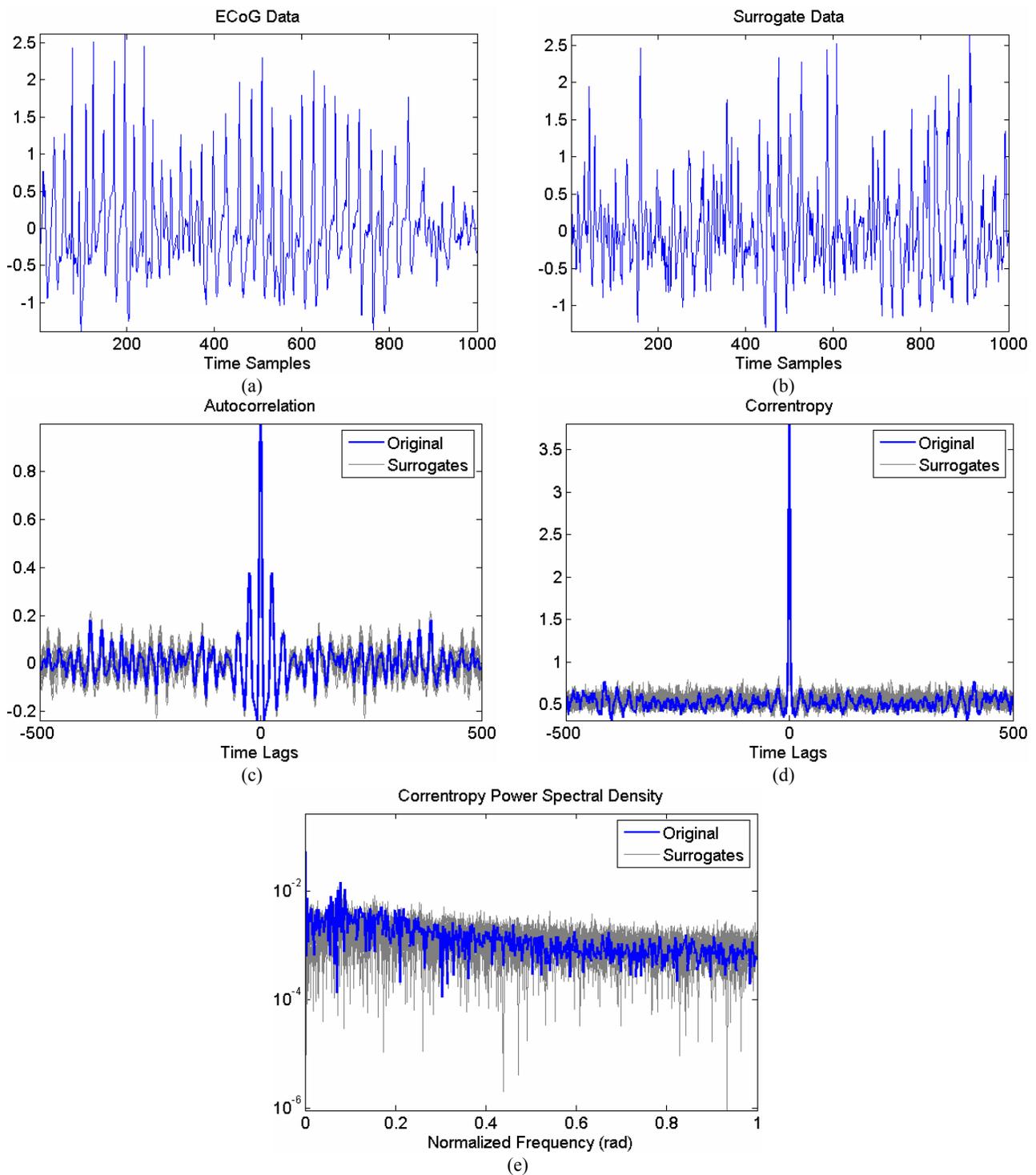


Figure 4 (a) ECoG data, (b) one of its surrogates. (c) Autocorrelation functions of ECoG data and of its surrogates, (d) the corresponding correntropy functions, (e) the correntropy power spectral density plots in logarithmic scale.

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

$$D_{\alpha} = c(\alpha) \sqrt{\frac{2}{N}} \quad (9)$$

where the N is the common sample size of the two series and the coefficient $c(\alpha)$ depends on the significance level. For $\alpha = 0.05$, $c(\alpha) = 1.36$ [15].

The Kolmogorov-Smirnov test states that if the empirical D -value is greater than the critical value, than the hypothesis that the two series were generated from the same distribution

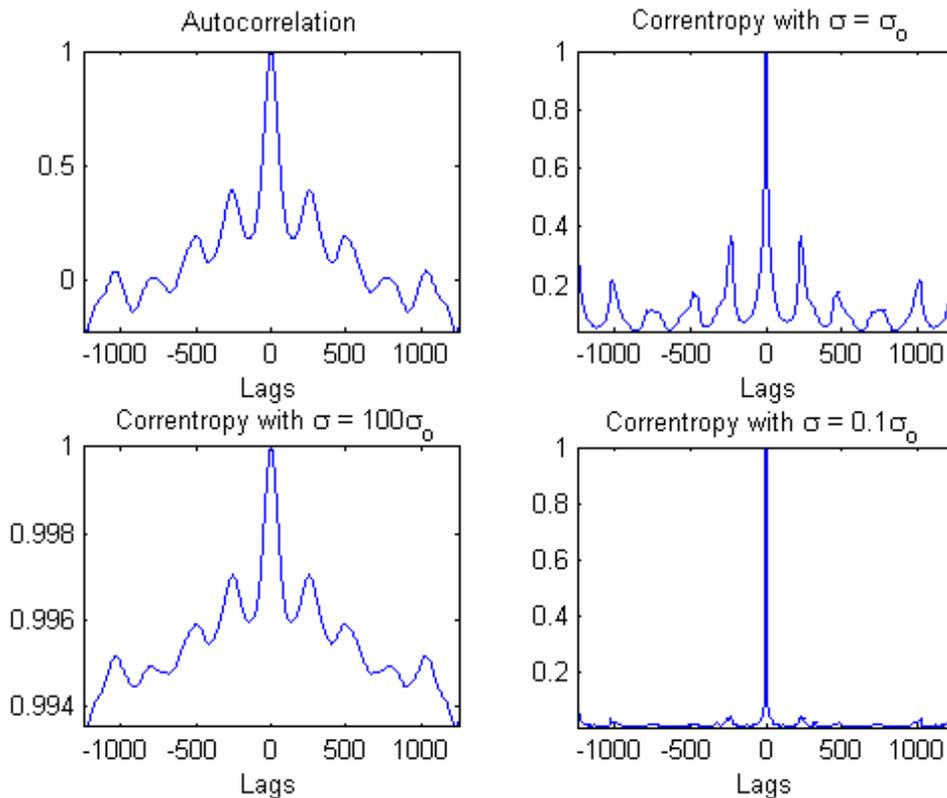


Figure 5. Comparison of autocorrelation, and normalized correntropies of Lorenz data with kernel widths σ_0 (based on Silverman's rule), $100\sigma_0$, $0.1\sigma_0$.

should be rejected. (Once again, if the test fails we cannot conclude that they were generated from the same distribution).

C. Simulation Results

In this section, we present a couple of synthetic simulations to demonstrate the utility of the proposed correntropy measure as a tool to detect non-linear structures in time-series. Firstly, we examine the simple case of a colored Gaussian noise time-series generated from a 1st order linear autoregressive model with a nonzero tap at the 200th lag. The autocorrelation and correntropy of the original series and its surrogates are shown in Figure 1. Observe that the injected high correlation is observed in both functions. Because of the linear construction, it is easy to expect that all the surrogate realizations share the same autocorrelation and correntropy functions as the original data. (Note that the kernel size is obtained for the original data via equation (7) and is also employed in correntropy calculations of the surrogates). As expected, comparisons between the correntropy powers of the original series with each one of the surrogates always failed the Kolmogorov-Smirnov test. In other words, the null hypothesis of a Gaussian linear source was not rejected. Passing the Gaussian data through higher order linear filters did not alter the results.

Lorenz [16] and Rössler [17] attractors are nonlinear chaotic systems which are widely used as reference inputs for nonlinearity tests with surrogate methods [13]. Fig. 2

presents the time series of the first component of a Lorenz segment series, along with one of its generated surrogates. The correlation coefficient between the autocorrelation of the Lorenz series and its surrogates averages at 0.9913 over all surrogates. The correntropies, on the other hand, show deviations between the original and the surrogates (refer to Figure 2). As the system nonlinearities were lost in surrogate generation, the surrogate data exhibit different correntropies and thus different correntropy power spectra. Application of the KS test as before revealed the presence of non-linearity in the Lorenz time-series. Specifically, the hypothesis that the correntropy power of the Lorenz series and all of its surrogates are samples from the same distribution were rejected by the Kolmogorov-Smirnov test with a confidence level of 95%.

Similar to the experiment with Lorenz system, we analyzed for non-linear structures in the 1st component of the Rössler system. A segment of the original time-series and a surrogate realization are presented in Figure 3. Correlation coefficients between autocorrelation of the Rössler series and autocorrelation of its surrogates have an average of 0.9946. The correntropies, once again, show variation amongst the original and the surrogates which is also reflected in the distribution of correntropy power spectra (Fig 4). The KS test rejects that the correntropy power of Rössler series and its surrogates were generated by the same power distribution at 95% significance level.

Hence, we conclude that the series contains nonlinear dynamics that is absent in its surrogates.

Our last experiment consists of testing this tool on a real-world data. We apply the methodology to 5 seconds of multielectrode array recordings gathered from Sprague Dawley rats (200 samples per sec). Two seconds of the data and one of its surrogates are presented in Figure 4. The proposed methodology rejected the null hypothesis at a 95% significance level suggesting non-linear structure in the rat data.

The effects of kernel size can be summarized as follows. As the kernel size approaches infinity, the higher order moment terms ($m > 2$) in the correntropy definition converge to zero. Thus, correntropy does not contain more information than autocorrelation [18]. On the other hand, when too small a kernel size is chosen, most of the power is concentrated in the zeroth lag, and information regarding other lags are lost. Figure 5 shows the effect of kernel size on the correntropy of Lorenz data.

IV. CONCLUSION

A generalized power spectral measure for Fourier based surrogate nonlinearity tests has been proposed. The methodology relies on the fact that correntropy captures nonlinearities of time series through nonlinear mappings of kernels, without resorting to any of the more conventional nonlinear measures such as Lyapunov exponents or correlation dimension, which are much more time consuming to compute. The method rejects the null hypothesis that the signal at hand 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 correntropy power spectrum distribution.

The methodology has been applied to synthesized Gaussian signals passed through linear filters, Lorenz and Rössler data generated through nonlinear attractors. The results have verified that, in fact correntropy based measures were able to distinguish between linear and nonlinear time series. Finally, the methodology was tested on local field potentials of behavioral rats and the results suggested the presence of nonlinear sources. The accuracy of the methodology can only be stated within a confidence level.

ACKNOWLEDGMENT

This work is supported by NSF grants ECS-0422718 and ECS-0300340.

REFERENCES

- [1] Vapnik, V., *The Nature of Statistical Learning Theory*, New York, Springer Verlag, 1995.
- [2] Friess, T., Cristianini, N., Campbell, C., "The Kernel Adatron Algorithm, A Fast and Simple Learning Procedure for Support Vector Machine", *Proceedings of the 15th International Conference on Machine Learning*, Morgan Kaufman Publishers, 1998.
- [3] Scholkopf, B., Smola, A., Muller K.R., "Nonlinear Component Analysis as a Kernel Eigenvalue Problem", *Neural Computation*, (10), pp. 1399-1319, 1998.
- [4] Mika, S., Ratsch, G. Wetson, J., Scholkopf, B, Fisher, "Discriminant Analysis with Kernels", *IEEE International Workshop on Neural Networks for Signal Processing*, IX, pp. 1-48, Madison, WI, 1999.
- [5] Bach, F.R., Jordan, M.I., "Kernel Independent Component Analysis", *Journal of Machine Learning Research*, (3), pp. 1-48, 2002.
- [6] Santamaria, I., Pokharel, P., Principe, J.C., "Generalized Correlation Function, Definition, Properties and Application to Blind Equalization", *to be published in IEEE Transactions on Signal Processing*.
- [7] Gautama, T., Mandic, D.P., Van Hulle, M.M., "A Novel Method for Determining the Nature of Time Series", *IEEE Transactions on Biomedical Engineering*, 51, (5), pp. 728-36, 2004.
- [8] Berge, P., Pomeau, Y., Vidal, C., *Order within Chaos*, Wiley, New York, 1984.
- [9] Baker, G. L. and Gollub, J. B., Cambridge, England, Cambridge University Press, 1996.
- [10] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, pp. 48-49, 1986.
- [11] Kugiumtzis, D., "Surrogate Data Test on Time Series", *Nonlinear Deterministic Modeling and Forecasting of Economic and Financial Time Series*, 2000.
- [12] Schreiber, T. Schmitz, A., "Surrogate Time Series", *Physical D*, 142, pp. 346-382, 2000.
- [13] Schreiber, T. Schmitz, A., "Improved Surrogate Data for Nonlinearity Tests", *Physical Review Letters*, 77 (4), pp. 635-638, 1996.
- [14] Bendat, J.S., Piersol, A.G., *Measurement and Analysis of Random Data*, John Wiley & Sons Co., New York, 1966.
- [15] Press, W.H., Teukolsky, S.A, Vetterling, W.T., Flannery, B.P., *Numerical Recipes C: The Art of Scientific Computing*, Cambridge University Press; Second edition, pp. 623-628, 2002.
- [16] Lorenz, E. N., "Deterministic Nonperiodic Flow", *Journal of Atmospheric Sciences*, 20, pp. 130-141, 1963.
- [17] Peitgen, H.-O., Jurgens, H.; and Saupe A., *Chaos and Fractals, New Frontiers of Science.*, New York, Springer-Verlag, pp. 686-696, 1992.
- [18] Erdogmus, D., *Information Theoretic Learning: Renyi's Entropy and Its Applications to Adaptive System Training*, Ph.D. Thesis, Gainesville, Florida, 2002.