Available online at www.sciencedirect.com
ScienceDirect
Neural Networks 21 (2008) 222–231
Neural Networks
www.elsevier.com/locate/neunet

2008 Special Issue

A new nonlinear similarity measure for multichannel signals☆

Jian-Wu Xu[a,*], Hovagim Bakardjian[b], Andrzej Cichocki[b], Jose C. Principe[a]

[a] *Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA*
[b] *Advanced Brain Signal Processing, RIKEN Brain Science Institute, Wako-shi, Saitama, 351-0198, Japan*

Received 4 August 2007; received in revised form 29 November 2007; accepted 11 December 2007

Abstract

We propose a novel similarity measure, called the *correntropy coefficient*, sensitive to higher order moments of the signal statistics based on a similarity function called the cross-correntopy. Cross-correntropy nonlinearly maps the original time series into a high-dimensional reproducing kernel Hilbert space (RKHS). The correntropy coefficient computes the cosine of the angle between the transformed vectors. Preliminary experiments with simulated data and multichannel electroencephalogram (EEG) signals during behaviour studies elucidate the performance of the new measure versus the well-established correlation coefficient.

Keywords: Similarly measure; Kernel method; EEG analysis; Biomedical application; Synchronization; Nonlinear dependence

1. Introduction

Quantification of dynamic interdependence in multidimensional complex systems with spatial extent provides a very useful insight into their spatio-temporal organization. In practice, the underlying system dynamics are not accessible directly. Only the observed time series can help decide whether two time series collected from the system are statistically independent or not and further elucidate any hidden relationship between them. Extracting such information becomes more difficult if the underlying dynamic system is nonlinear or the couplings among the subsystems are nonlinear and nonstationary.

There has been extensive research aimed at detecting the underlying relationships in multidimensional dynamic systems. The classical methodology employs a linear approach, in particular, the cross-correlation and coherence analysis (Shaw, 1981). Cross-correlation measures the linear correlation between two signals in the time domain, while the coherence function specifies the linear associations in the frequency domain by the ratio of squares of cross-spectral densities divided by the products of two autospectra. There have been several extensions of correlation to more than two pairs of time series such as directed coherence, directed transfer functions and partial directed coherence (Pereda, Quian Quiroga, & Bhattacharya, 2005). Unfortunately, linear methods only capture linear relationships between the time series, and might fail to detect nonlinear interdependencies between the underlying dynamic subsystems.

Nonlinear measures include mutual information and state-space methods. One technique is the generalized mutual information function (Pompe, 1993). However, a large quantity of noise-free stationary data is required to estimate these measures based on information theory, which restricts their applications in practice. Another method is the phase synchronization where the instantaneous phase using Hilbert transforms is computed and interdependence is specified in terms of time-dependent phase locking (Rosenblum et al., 1996). The state-space methodologies include similarity-index and synchronization likelihood. The similarity-index technique and its modifications compute the ratio of average distances between index points, their nearest neighbours and their mutual nearest ones (Arnhold, Grassberger, Lehnertz, & Elger, 1999; Quian Quiroga, Arnhold, & Grassberger, 2000). Stam et al. proposed the synchronization likelihood to offer a straightforward normalized estimate of the dynamic coupling between interacting systems (Stam & van Dijk, 2002).

☆ An abbreviated version of some portions of this article appeared in Xu, Bakardjian, Cichocki, and Principe (2007) as part of the IJCNN 2007 Conference Proceedings, published under IEE copyright.

* Corresponding author.
E-mail addresses: jianwu@cnel.ufl.edu (J.-W. Xu), hova@brain.riken.jp (H. Bakardjian), cia@brain.riken.jp (A. Cichocki), principe@cnel.ufl.edu (J.C. Principe).

0893-6080/$ - see front matter
doi:10.1016/j.neunet.2007.12.039

There are several drawbacks associated with these techniques based on state space embedding. Estimating the embedding dimension of times series corrupted by measurement noise for a valid reconstruction, searching a suitable neighbourhood size and finding a constant number of nearest neighbours are a few of many constraints that severely affect the estimation accuracy.

In this paper, we introduce a novel functional measure, called the *correntropy coefficient*, to characterize dynamic interdependencies between interacting systems. Correntropy is a new concept to quantify similarity based on a reproducing kernel Hilbert space method (Santamaria, Pokharel, & Principe, 2006). Correntropy is sensitive to both the higher order statistical distribution information and temporal structure of the random process. Correntropy can be applied both to one-time series, called the *autocorrentropy*, or a pair of scalar random processes, called the *crosscorrentropy*. In this paper, we work with the *centered crosscorrentropy*, which implicitly subtracts the mean of the nonlinearly transformed signal. The *correntropy coefficient* is defined as the normalized centered cross-correntropy. If two random variables or two time series are independent, then the correntropy coefficient becomes zero; if the two are the same, then it attains maximum value 1; the correntropy coefficient achieves -1 when the two random variables are in the opposite directions. Hence, the correntropy coefficient is a suitable interdependence measure for interacting dynamic systems.

The paper is organized as follows. In Section 2, we briefly introduce the newly proposed correntropy concept and present the method of the correntropy coefficient in details. We also explores the correntropy coefficient from geometrical perspective and other relevant issues in Section 3. Experiments of the correntropy coefficient on simulated data and real EEG signals are presented in Section 4. We conclude the work in Section 5.

2. Method

In function analysis, the symmetrical positive definite kernel is a special type of bivariate function. The most widely used kernel in machine learning and in nonlinear data representation is the Gaussian kernel which is given by

$$\kappa(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-y)^2}{2\sigma^2}\right\}, \tag{1}$$

where σ is the kernel width. According to the Mercer's theorem (Mercer, 1909) of Hilbert space analysis, the symmetrical positive definite kernel function possesses an eigendecomposition as

$$\kappa(x, y) = \sum_{n=0}^{\infty} \lambda_n \varphi_n(x)\varphi_n(y) = \langle \Phi(x), \Phi(y)\rangle$$
$$\Phi : x \mapsto \sqrt{\lambda_n}\varphi_n(x), \quad n = 1, 2, \ldots,$$

where $\{\varphi_n(x), n = 1, 2, \ldots\}$ and $\{\lambda_n, n = 1, 2, \ldots\}$ are sequences of eigenfunctions and corresponding eigenvalues of $\kappa(x, y)$ respectively, and $\langle , \rangle$ denotes the inner product between two infinite dimensional vectors $\Phi(x)$ and $\Phi(y)$. By the Moore–Aronszajn Theorem (Aronszajn, 1950), $\kappa(x, y)$ uniquely determines a high-dimensional reproducing kernel Hilbert space, denoted as $\mathcal{H}_\kappa$, where the nonlinear transformation Φ maps the original signals onto the surface of a sphere in $\mathcal{H}_\kappa$.

Based on the symmetrical positive definite kernel function $\kappa(x, y)$, the "generalized" cross-correlation function, called *cross-correntropy* (Xu et al., 2007; Xu, Pokharel, Paiva, & Principe, 2006), for two given random variables x and y is defined as

$$V(x, y) = E[\kappa(x, y)] = E[\langle \Phi(x), \Phi(y)\rangle],$$

where E denotes the statistical expectation operator. The "generalized" cross-covariance function, called *centred cross-correntropy*, is defined as

$$\begin{aligned} U(x, y) &= E[\kappa(x, y)] - E_x E_y[\kappa(x, y)] \\ &= E[\langle \Phi(x) - E[\Phi(x)], \Phi(y) - E[\Phi(y)]\rangle]. \end{aligned} \tag{2}$$

Therefore the cross-correntropy function might be interpreted as a "conventional" cross-correlation function for the transformed random variables in the high-dimensional RKHS $\mathcal{H}_\kappa$, while the centred cross-correntropy is nothing but the cross-correntropy for the zero mean (centred) random variables $(\Phi(x) - E[\Phi(x)])$. However, if we apply the Taylor series expansion for the Gaussian kernel in the definition of the cross-correntropy, it can be easily noticed that it compactly contains all even moments of the random variables $(x - y)$ (Santamaria et al., 2006). Hence, cross-correntropy includes higher order statistical information about the random variables.

An important observation here is that when two random variables x and y are independent, that is the joint probability density function $P(x, y)$ equals to the product of marginal probability density functions $P(x)P(y)$, then $E[\kappa(x, y)] = E_x E_y[\kappa(x, y)]$. Therefore the centred cross-correntropy reduces to zero only if the two random variables are independent. This is a much stronger condition than uncorrelatedness, as required by the conventional covariance function in order to achieve zero value.

The centred cross-correntropy has similar properties to the covariance function such as $U(x, x) \geq 0$ and $U(x, y) = U(y, x)$. One of the most important properties is that the centred cross-correntropy is symmetrical and nonnegative definite.

Proposition 1. *The centered cross-correntropy $U(x, y)$ is a symmetrical nonnegative definite function defined in $\mathcal{X} \times \mathcal{X} \longrightarrow \mathcal{R}$.*

Proof. The symmetry of $U(x, y)$ is easily seen since the kernel function used in the definition is symmetrical. Given any positive integer n, any set of $x_1, x_2, \ldots x_n \in \mathcal{X}$ and any not all zero real numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$, by definition we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j U(x_i, x_j)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j E\left[\langle \Phi(x_i) - E[\Phi(x_i)], \Phi(x_j) - E[\Phi(x_j)]\rangle\right]$$

$$= E\left[\left\|\sum_{i=1}^{n}\alpha_i(\Phi(x_i) - E[\Phi(x_i)])\right\|^2\right] \geq 0. \quad (3)$$

Therefore, $U(x, y)$ is symmetrical and nonnegative definite. □

By normalizing the centred cross-correntropy, we can define the "generalized" correlation coefficient, called the *correntropy coefficient*, as

$$\eta = \frac{U(x, y)}{\sqrt{U(x, x)U(y, y)}}, \quad (4)$$

where $U(x, x)$ and $U(y, y)$ are the centred autocorrentropy functions for variables x and y respectively. The absolute value of the correntropy coefficient is bounded by 1. This property can be proved by the following proposition:

Proposition 2. *The centred cross-correntropy $U(x, y)$ satisfies*

$$|U(x, y)| \leq \sqrt{U(x, x)U(y, y)}, \quad (5)$$

hence the absolute value of the correntropy coefficient $|\eta| \leq 1$. Property (5) is called the Cauchy–Schwartz inequality in the RKHS $\mathcal{H}_\kappa$.

Proof. Let $n = 2$ in (3), the expression reduces to

$$\alpha_1^2 U(x, x) + \alpha_2^2 U(y, y) \geq 2\alpha_1\alpha_2|U(x, y)|. \quad (6)$$

When both $U(x, x)$ and $U(y, y)$ are nonzero, we can substitute $\alpha_1^2 = \frac{U(y,y)}{2\sqrt{U(x,x)U(y,y)}}$ and $\alpha_2^2 = \frac{U(x,x)}{2\sqrt{U(x,x)U(y,y)}}$ into (6) to obtain (5). On the other hand, if at least one of the two autocorrentropy functions is zero, we claim that the validity of (6) implies $U(x, y) = 0$. Therefore (5) must hold because the left-hand side is zero and the right-hand side is nonnegative. Hence we conclude the proof. □

Unlike the conventional correlation coefficient, the correntropy coefficient will produce a non-zero value (which depends on the kernel width used in the Gaussian kernel) for two uncorrelated but not independent random variables. In the context of generalized synchronization, the correntropy coefficient is able to characterize both higher-order relationship and nonlinearity between interacting systems.

In practice, we only have a finite number of data points or time series samples available from the dynamic system. So we have to work with the estimate of the correntropy coefficient. Substituting the definition of the centred cross-correntropy (2) into the correntropy coefficient (4) and approximating the ensemble average by the sample mean, we can obtain an estimate of correntropy coefficient directly from data,

$$\hat{\eta} = \frac{\frac{1}{N}\sum_{i=1}^{N}\kappa(x_i, y_i) - \frac{1}{N^2}\sum_{i,j=1}^{N}\kappa(x_i, y_j)}{\sqrt{\kappa(0) - \sum_{i,j=1}^{N}\frac{\kappa(x_i,x_j)}{N^2}}\sqrt{\kappa(0) - \sum_{i,j=1}^{N}\frac{\kappa(y_i,y_j)}{N^2}}}, \quad (7)$$

where N is the total number of samples, $\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\kappa(x_i, y_j)$ is called the *cross-information potential* between x and y, $\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\kappa(x_i, x_j)$ and $\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\kappa(y_i, y_j)$ are the information potential for x and y respectively (Principe, Xu, & Fisher, 2000), and $\kappa(0)$ is the value of Gaussian kernel (1) when the argument $(x - y) = 0$.

3. Discussion

In this section, we explore more details about the correntropy coefficient both in theoretical analysis and practical implementation.

3.1. Geometrical interpretation

Since the centred cross-correntropy is symmetrical and nonnegative definite, it also has a direct eigendecomposition by the Mercer's theorem,

$$U(x, y) = \sum_{n=0}^{\infty}\gamma_n\psi_n(x)\psi_n(y) = \langle\Psi(x), \Psi(y)\rangle$$
$$\Psi : x \mapsto \sqrt{\gamma_n}\psi_n(x), \quad n = 1, 2, \ldots, \quad (8)$$

where γ_n and ψ_n are eigenvalues and eigenfunctions for the centred cross-correntropy respectively. According to the Moore–Aronszajn Theorem, $U(x, y)$ also uniquely induces a high-dimensional reproducing kernel Hilbert space, denoted as $\mathcal{H}_U$. Notice that the nonlinear map Ψ has implicitly embedded the expectation operator so that every vector in $\mathcal{H}_U$ becomes deterministic and contains statistical information of signals, hence it is data dependent. While in the RKHS $\mathcal{H}_\kappa$ induced by the Gaussian kernel, each vector is still stochastic and the RKHS is data independent. Thus, the nonlinear map Ψ provides a natural link between stochastic and deterministic functional analysis.

Substituting Eq. (8) into the definition of the correntropy coefficient Eq. (4), we obtain

$$\eta = \frac{\langle\Psi(x), \Psi(y)\rangle}{\|\Psi(x)\|\|\Psi(y)\|} = \cos\theta,$$

where $\|\Psi(x)\|$ and $\|\Psi(y)\|$ are the length of two vectors $\Psi(x)$ and $\Psi(y)$ in $\mathcal{H}_U$ respectively, and θ is the angle between these two vectors. With this geometrical interpretation, the correntropy coefficient essentially computes the cosine of the angle between two nonlinear transformed vectors in RKHS $\mathcal{H}_U$ induced by the centred cross-correntropy. In particular, if two vectors are orthogonal, then θ is 90° and η equals 0; if two vectors are in the same direction, then θ is 0° and η equals 1, while two vectors are in the opposite direction, θ becomes 180° and η equals −1. Orthogonality between vectors $\Psi(x)$ and $\Psi(y)$ in $\mathcal{H}_U$ corresponds to independence between random variables x and y. When two vectors are in the same or opposite directions, this suggests a strong dependence between two random variables x and y.

The RKHS approach to analyzing the conventional correlation function was originally proposed by Parzen (1959) because the correlation function is also nonnegative definite, thus it determines a unique reproducing kernel Hilbert space, denoted as $\mathcal{H}_R$. Grenander analyzed the standard correlation coefficient from RKHS perspective in Grenander (1981). Both $\mathcal{H}_R$ and $\mathcal{H}_U$ are data dependent reproducing kernel Hilbert spaces, however $\mathcal{H}_U$ implicitly embeds $\mathcal{H}_\kappa$ which incorporates

higher order statistics intrinsic in the data. Therefore the correntropy coefficient requires independence of two signals to make two corresponding vectors in $\mathcal{H}_U$ orthogonal, while the standard correlation coefficient only needs uncorrelatedness.

3.2. Kernel width

The variance of the Gaussian kernel function is normally called kernel width or kernel size. The kernel size should be considered a scale parameter controlling the metric in the projected space. From the geometrical perspective, the kernel size decides the length of each of the nonlinearly transformed vectors and the angle between them in the RKHS $\mathcal{H}_U$ because $\|\Psi(x)\| = \sqrt{\kappa(0) - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\kappa(x_i, x_j)}$. It can be easily seen that the vector length monotonically decreases as the kernel width increases, while the centred crosscorrentropy exhibits more complex pattern since it also depends on the nature of the relationship between two random variables. The kernel width controls the ability of the centred correntropy to capture the nonlinear couplings intrinsic in the signals. If the kernel width is too large, the correntropy coefficient loses the ability to detect the nonlinearity and approaches the conventional correlation coefficient; when the kernel width is too small, the nonlinear transformations $\Phi(x)$ in (2) and $\Psi(x)$ in (8) cannot interpolate between data points. This can also be verified by applying the Taylor series expansion to the Gaussian kernel where the kernel width appears as the weighting parameter in both second and higher- order moments. The effect of the kernel size on different order moments is scaled by the $2k$ power where k is the moment order. When the kernel size is too large, the contribution of higher-order statistics decays rapidly and the centred cross-correntropy approaches the conventional cross-covariance function; on the other hand, when the kernel size is too small, the effect of higher order moments overweighs the second order one. An appropriate kernel size should maintain the balance of second-order and higher-order statistics of the signal, which is a challenge and depends upon the application.

Therefore a good choice of the kernel parameter is crucial for obtaining the good performance of the proposed method. There are two ways of handling the selection of kernel size. One is to seek an optimal kernel size. The cross-validation has been one of the mostly used methods in machine learning field to choose an appropriate kernel width. Other simpler approaches include the Silverman's rule of thumb which is given by (Silverman, 1986):

$$\sigma = 0.9AN^{-1/5}, \tag{9}$$

where A is the smaller value between standard deviation of data samples and data interquartile range scaled by 1.34, and N is the number of data samples. The Silverman's rule is easy to apply and frequently selects a good kernel size, hence we will set the kernel width according to (9) throughout the paper. Alternatively, the kernel size can be thought as a scale parameter that provides different looks to the dependence among the variables. Just like in wavelets, the kernel size is able to analyze the dependencies at different resolutions. Since many real-world signals are very complex, this multiresolution analysis may elucidate better the relationships.

3.3. Scaling effect

Because the transformations $\Phi(x)$ in (2) and $\Psi(x)$ in (8) are nonlinear, any scaling of the original random variables results in different performance of the correntropy coefficient. Unlike the conventional correlation coefficient which is insensitive to the amplitude scaling of the signals and only measures the similarity of the signals through time, the correntropy coefficient measures both the time and amplitude similarities between two signals. Therefore, in certain applications, it is vital to normalize both signals before applying the correntropy coefficient. For example, the amplitudes of EEG signals are highly dependent on the different electrode impedances. It is critical to normalize all channels of EEG signals to the same dynamic range.

4. Experiments

We test the correntropy coefficient on simulated data set and real world EEG signals in this section.

4.1. Two unidirectionally coupled Hénon maps

First, we apply the correntropy coefficient in detecting nonlinear interdependence of two unidirectionally coupled Hénon maps **X** and **Y** defined as

$$\mathbf{X}: \begin{cases} x_{n+1} = 1.4 - x_n^2 + b_x u_n \\ u_{n+1} = x_n, \end{cases}$$

$$\mathbf{Y}: \begin{cases} y_{n+1} = 1.4 - [Cx_n + (1-C)y_n]y_n + b_y v_n \\ v_{n+1} = y_n. \end{cases}$$

Notice that system **X** drives system **Y** with a nonlinear coupling strength C. C ranges from 0 to 1 with 0 being no coupling and 1 being complete coupling. Parameters b_x and b_y are both set to 0.3 as canonical values for the Hénonmap when analysing identical systems, and to 0.3 and 0.1 respectively for nonidentical ones. For each coupling strength, we discard the first 10 000 iterated time series as transient and obtain the next 500 data points for experiments. The correntropy coefficient η is calculated between the first component of system **X**, x_1, and the first component of system **Y**, y_1. The following simulations address different aspects of a nonlinear interdependence measure.

4.1.1. Variation of correntropy coefficient with coupling strength

First in Fig. 1, we plot the averaged correntropy coefficient η as a function of coupling strength C for identical and nonidentical maps over 10 realizations of different initial conditions. The error bars denote the standard deviation over the different realizations. Fig. 1(a) shows the identical map where the kernel size used in Gaussian kernel is set to be 0.001 according to the Silverman's rule (9). The correntropy coefficient $\eta = 1$ for $C \geq 0.7$ in Fig. 1(a) indicating perfect

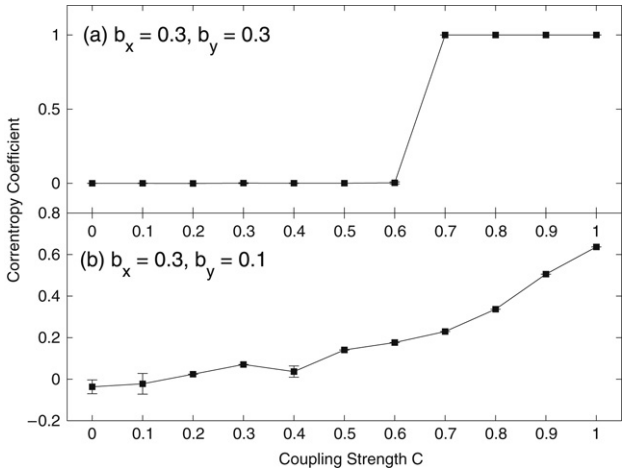

Fig. 1. Averaged correntropy coefficient for unidirectionally coupled identical (a) and nonidentical (b) Hénonmaps over 10 realizations of different initial conditions.

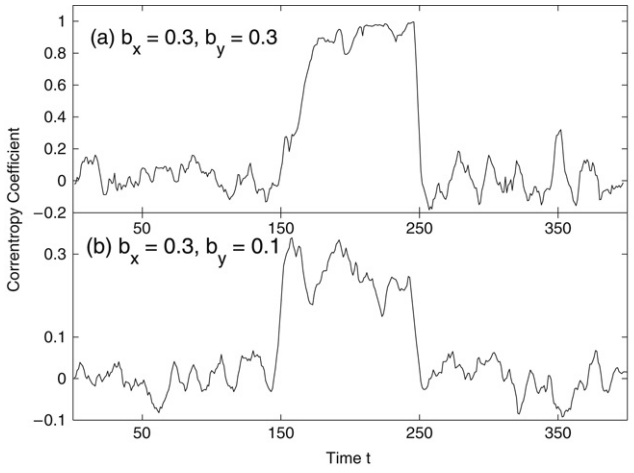

Fig. 2. Time dependent of correntropy coefficient for unidirectionally coupled identical (a) and nonidentical (b) Hénonmaps.

synchronization occurs between two coupled systems. The critical threshold $C = 0.7$ corresponds to the point when the maximum Lyapunov exponent of the response system becomes negative and identical synchronization between the systems takes place. On the other hand, the correntropy coefficient $\eta = 0$ for $C < 0.7$ suggesting no synchronization even though two systems are weakly coupled. Similar results have been reported using other nonlinear interdependence measurement in Quiroga et al. (2000) and Schmitz (2000).

Fig. 1(b) shows the result for unidirectionally coupled nonidentical systems. The kernel size is set to 0.4. In this case, identical synchronization is not possible and the driver has higher dimension than the response. The sharp increase of the correntropy coefficient at point $C = 0.7$ as in the identical synchronization situation can not be observed here. But the correntropy coefficient shows a consistent monotonic increase with respect to coupling strength except for the region $0.1 < C < 0.3$. The local hump in the zone $0.1 < C < 0.3$ is due to the minima of the largest subLyapunov exponent (Schiff, So, Chang, Burke, & Sauer, 1996).

4.1.2. Sensitivity of correntropy coefficient to time-dependent dynamic changes in coupling

Next we test how sensitive the correntropy coefficient is to time dependent sudden change in the dynamics of interacting systems due to coupling strength. In experiment, dynamic systems are coupled only during a single epoch and otherwise uncoupled, which basically generates nonstationarity in time series. We set the coupling strength $C = 0$ for $n \leq 10\,150$ and $n \geq 10\,250$ and $C = 0.8$ for $10\,150 < n < 10\,250$. Only 400 data samples are plotted after the first 10 000 data are discarded as transient. The sliding window used to compute the correntropy coefficient is chosen to contain 8 data samples. Kernel size is set to 0.2 for identical map and 0.3 for nonidentical map. The results are averaged over 20 independent realizations of different initial conditions ranging from 0 to 1. Fig. 2 plots the correntropy coefficient curves. In uncoupled regions, η fluctuates around 0.01 for identical maps and 0.001 for nonidentical maps. A sharp and clear increase occurs at $t = 150$ when the 0.8 coupling strength between systems $\mathbf{X}$ and $\mathbf{Y}$ is introduced, and there is a sharp and clear decrease in η falling off back to the baseline level when coupling strength reduces to zero at $t = 250$. The interval where η is noticeably higher than the baseline level matches nicely to the coupling interval. This phenomenon is observed both in identical and nonidentical Hénon maps. Therefore, although correntropy assumes stationarity in the data generation, the correntropy coefficient is potentially able to detect sudden change in the coupling between two interacting dynamic systems with a high temporal resolution, which makes this measure suitable for nonstationary data sets.

4.1.3. Robustness of correntropy coefficient against measurement noise

We analyze the robustness of correntropy coefficient when the time series are contaminated with noise. Only measurement noise is considered here which does not perturb the inherent dynamics of systems. Independent realizations of white noise are added to driver $\mathbf{X}$, response $\mathbf{Y}$ and both systems separately. The signal-to-noise (SNR) ratio is set to be 10 dB and 1dB respectively to test the performance of correntropy coefficient at different noise intensity. 500 data samples are used to calculated the correntropy coefficient, averaged over 20 realizations. Fig. 3 plots the correntropy coefficient for identical Hénonmap with white noise in response, driver and both systems. Kernel size is chosen to be 0.04 for SNR = 10 dB and 0.08 for SNR = 1 dB respectively according to the Silverman's rule (9). Note that the correntropy coefficient curves with noise become less smooth than those of noise-free ones, but the sharp increase at $C = 0.7$ is still obvious for both noise intensities. When noise level is high (SNR = 1 dB), the correntropy coefficient curve is more zigzag than that of 10 dB case, however it can still detect increases in the coupling strength. The figure also suggests that whether noise is added into the driver system, response or both systems, the performance of correntropy coefficient is very similar. Fig. 4 presents the results for nonidentical Hénon map. Kernel size is selected to 0.05 for SNR = 10 dB and 0.2 for SNR = 1 dB respectively.

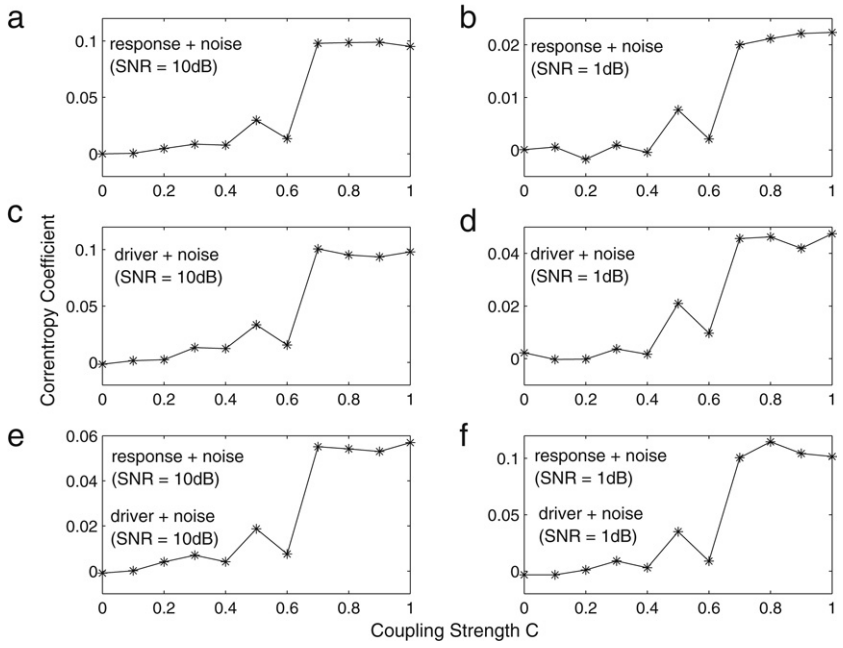

Fig. 3. Influence of different noise levels on correntropy coefficient for unidirectionally coupled identical Hénonmap with white noise in response, driver and both systems.

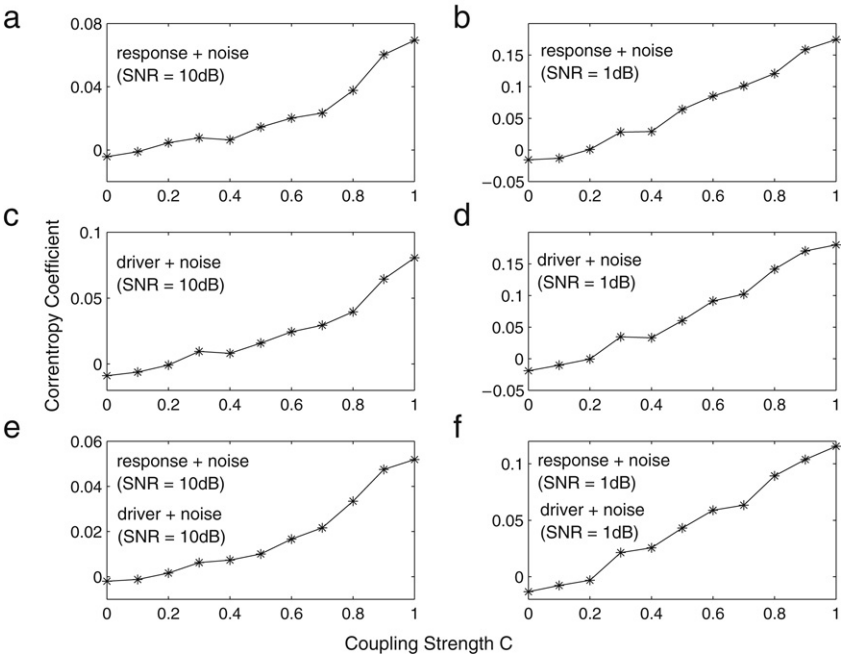

Fig. 4. Influence of different noise levels on correntropy coefficient for unidirectionally coupled nonidentical Hénonmap with white noise in response, driver and both systems.

The values of the correntropy coefficients at different coupling strength are averaged over 20 independent realizations. In both levels of noise, the correntropy coefficients consistently increase with respect to coupling strength. Also the effect of noise in response, driver or both systems does not make big differences. These results show that the correntropy coefficient is fairly robust even in the case of considerably noisy data.

4.1.4. Effect of kernel width

We have discussed the importance of the kernel width in the performance of the correntropy coefficient because it is a parametric measure in previous section. Here we demonstrate this on unidirectionally coupled identical and non-identical Hénonmaps. Fig. 5 presents a three dimensional correntropy coefficient curves of different kernel width and coupling

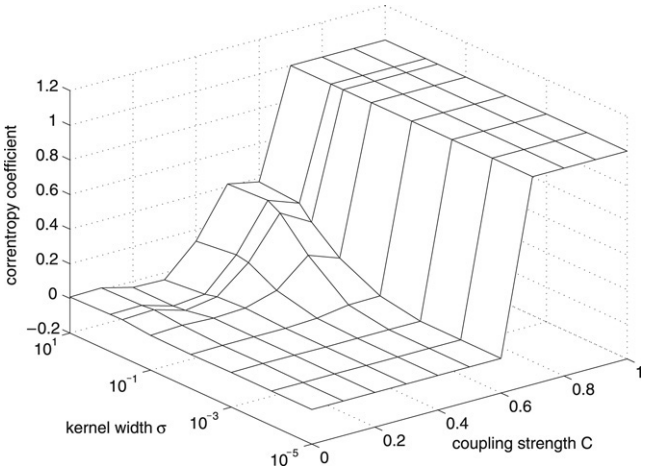

Fig. 5. Effect of different kernel width on correntropy coefficient for unidirectionally coupled identical Hénon- maps.

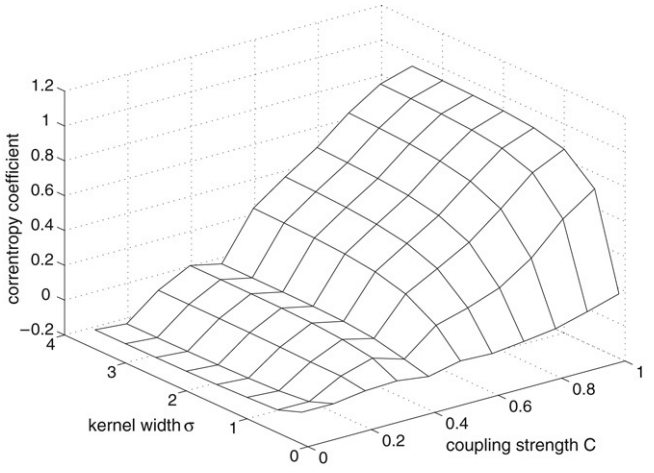

Fig. 6. Effect of different kernel width on correntropy coefficient for unidirectionally coupled nonidentical Hénon- maps.

strength for the identical Hénon map. It clearly shows that the kernel width provides a multiscale measurement for the correntropy coefficient to quantify interdependence. When the kernel width is chosen too large, $\sigma = 0.5, 1, 10$, in this case, correntropy coefficient produces similar results to those of linear correlation coefficient. On the other hand, when the kernel width approaches to the one chosen by the Silverman's rule, $\sigma = 0.001$ here, correntropy coefficient is able to characterize the nonlinear dependence between two coupled Hénonmaps. Therefore, by varying the kernel width, correntropy coefficient can measure both linear and nonlinear dependence. The results for the nonidentical Hénonmap are presented in Fig. 6. It can be seen that if kernel width is too small, the increase of the correntropy coefficient with respect to the coupling strength is not as obvious as those of suitable kernel width ($\sigma = 0.4$ here). While the kernel width is too large as per the Silverman's rule, the results of the correntropy coefficient approach to those of conventional correlation coefficient. In both figures, we see that the correntropy coefficients can either increase or decrease as the kernel width increases. These observations are consistent with our theoretical analysis in the previous section.

4.1.5. Ability of correntropy coefficient to quantify nonlinear coupling

To demonstrate that correntropy coefficient η is able to detect the nonlinear coupling, we use the multivariate surrogate data technique which is introduced in Prichard and Theiler (1994). Basically, in order to generate the multivariate surrogate data, first the Fourier transform is applied to each time series, then a common random number is added to each of the phases and an inverse Fourier transform is applied. The resulting time series have identical power spectra and cross-power spectra as the original time series, but any nonlinear coupling among the time series has been destroyed. In simulation, we use TISEAN package (Schreiber & Schmitz, 2000) to generate 19 realizations of the surrogate data for the time series x_n in the driver system $\mathbf{X}$ and y_n in the driver system $\mathbf{Y}$ for each different coupling strength for the unidirectionally coupled nonidentical Hénonmap. Then we compute the correntropy coefficient for both the original and the surrogate data with respect to different coupling strength. Fig. 7 plots the correntropy coefficient curve for the original data and the mean value of 19 correntropy coefficients for the surrogate data with the corresponding maximal and minimal values as error bars. To quantify the significance level, we calculate the Z-Score as $Z = \frac{|v_{\text{orig}} - \mu_{\text{surr}}|}{\sigma_{\text{surr}}}$ where v_{orig} is the correntropy coefficient value for the original data, μ_{surr} and σ_{surr} are the mean and the standard deviation for the surrogate data respectively. Table 1 presents the Z-Score values for different coupling strength. With the exception of $C = 0.2$ and 0.4, the Z-Score values are significantly larger than 1.96 which means the nonlinear coupling has been detected with a probability $p < 0.05$. These results clearly demonstrates that the correntropy coefficient is sensitive to the nonlinearity of the dependence between two coupled systems.

4.2. EEG signals

In the second experiment, we applied the correntropy coefficient to real EEG signals. The electrical potentials on the surface of the scalp of a human subject were measured and recorded with the NeuroScan EEG system (NeuroScan Inc., Compumedics, Abbotsford, Australia). A 64-channel cap was used with electrode locations according to the extended international 10/20 system and with a linked-earlobe reference. Horizontal and vertical electrooculogram (HEOG and VEOG) signals were also recorded for artifact rejection using two sets of bipolar electrodes. The data sampling rate was fixed at 1000 Hz and the online band-pass filter range was set to be maximally wide between 0.05 Hz and 200 Hz. Subjects were presented repeatedly (200 times) with unimodal auditory and visual stimuli delivered in the central visual and auditory spaces simultaneously and with the same strength to the left/right eyes and ears, as well as with simultaneous cross-modal combinations. For the purpose of this study, only the unimodal data were used. The visual stimuli consisted of 5×5 black and white checkerboards presented for 10 ms, while the auditory stimuli were 2000 Hz tones with durations

Table 1
Z-Score for the surrogate data

Coupling Strength C	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Z Score	6.9426	4.4721	1.6221	4.5845	0.7727	7.6581	9.9082	16.6999	12.2678	22.5882	19.8949

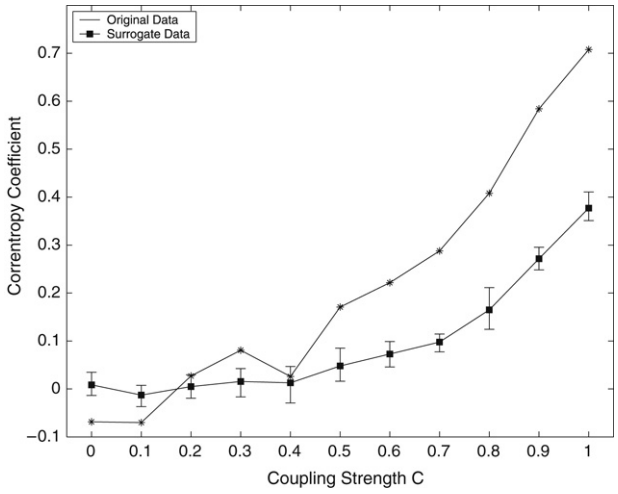

Fig. 7. Comparison of the correntropy coefficient for the original data and the surrogate data for unidirectionally coupled nonidentical Hénonmap.

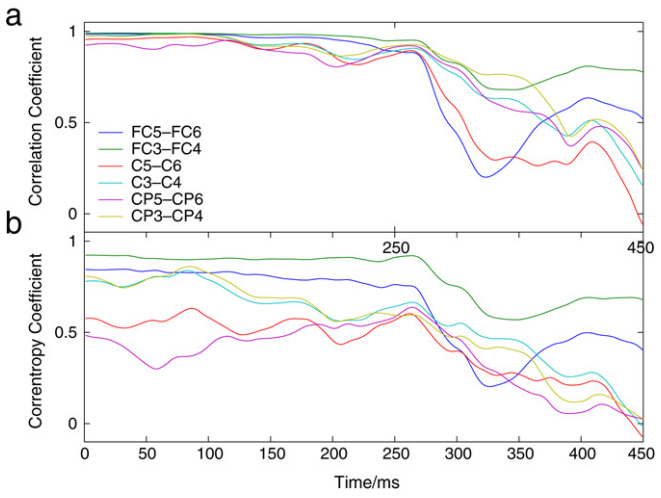

Fig. 8. Comparison of the correlation coefficient and the correntropy coefficient in characterization of synchronization among auditory cortex for audio stimuli EEG signal.

of 30 ms. The time interval between the stimuli in any of the experimental conditions was random between 1500 and 2000 ms. Following standard eye-movement artifact rejection procedures and segmentation into single epochs with alignment at the onset of the stimuli, all artifact-free epochs were averaged and normalized to zero mean and unit variance and low-pass filtered at 0–40 Hz for further analysis. Since correntropy coefficient is an amplitude-sensitive similarity measure, it is critical to normalize signals from different channels into the same dynamic range. We then applied the correntropy coefficient to the averaged data to quantify the bilateral synchronization or couplings among the corresponding sensory areas of the brain. In order to test whether the correntropy coefficient was able to detect any nonlinear couplings in the EEG signals, the results were compared to the conventional correlation coefficient. A window size of 20 ms data is used to calculate both measures corresponding to the duration of a single dipole activation in the cortex (Kotani et al., 2004). The kernel width σ in Gaussian kernel (1) used in correntropy coefficient was chosen to be 0.4 according to the Silverman's rule (9).

Fig. 8(a) and (b) show plots of the correlation and correntropy coefficients for the auditory areas of the brain as a function of time after the subject was exposed only to the audio stimuli. Several bilaterally-symmetrical pairs of electrodes were selected in the vicinity of the auditory cortex, so that both measures were computed for pairs FC5–FC6, FC3–FC4, C5–C6, C3–C4, CP5–CP6, CP3–CP4. As shown in Fig. 8(a) and (b), there are two distinct time intervals 0–270 ms and 270–450 ms in the auditory response. Both correlation and correntropy coefficients drop at 270 ms. This suggests that both measures are able to detect the changes in interhemispheric synchronization of the auditory regions. However, as the electrodes are chosen in different locations away from the auditory cortex, it is expected that during the synchronization phase (0–270 ms) the synchronization measures for different pairs should be different. Fig. 8(a) shows that the correlation coefficients for all 6 pairs are grouped together and are unable to detect the difference in activation, while Fig. 8(b) suggests that the correntropy coefficient can differentiate successfully the synchronization strength among different areas of the cortex above the left and right auditory regions. Notably, as expected from previous studies, pairs FC5–FC6 and FC3–FC4 exhibit stronger synchronization strength than the others, while most posterior pairs CP5–CP6 and C5–C6 have weaker synchronization strength. Also the synchronization patterns reveal lateral similarity in time for the pairs FC5–FC6 and FC3–FC4, for CP5–CP6 and C5–C6, and for CP3–CP4 and C3–C4. Furthermore the correntropy coefficients for pairs C5–C6, C3–C4 and CP3–CP4 peak simultaneously at 90 ms which corresponds to the first mean global field power (MGFP) peak of the EEG signal. These differences indicate that the correntropy coefficient is more sensitive and is able to extract more information as a synchronization measure than the conventional correlation coefficient.

We also compared both measures when applied to the visual cortical areas. The measures are presented in Fig. 9 as a function of time when the subject is exposed only to visual stimuli. Again, a window size of 20 ms data is used to compute both the correlation and the correntropy coefficients, and the kernel width σ is again set to 0.4 as in the previous case. We also chose bilaterally symmetrical pairs of electrodes O1–O2, PO7–PO8, PO5–PO6, P7–P8, P5–P6 and P3–P4. In Fig. 9(b)

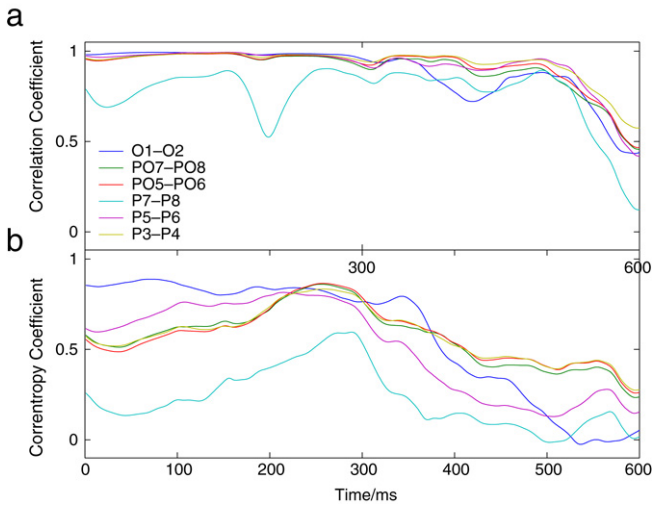

Fig. 9. Comparison of correlation coefficient and correntropy coefficient in characterization of synchronization among occipital cortex for visual stimulus EEG signal.

the correntropy coefficients for all pairs except for O1–O2 show similar synchronization patterns. The correntropy coefficient increases at first, then reaches a peak around 275 ms, after which it drops to lower levels. The maximum values of the correntropy coefficients around 275 ms correspond to the peak P1 in the visual evoked potential (Quian Quiroga, Arnhold, & Grassberger, 2001). As expected the synchronization between occipital channels O1 and O2 has the maximum strength and stays high until it decreases around 350 ms. Thus the correntropy coefficient shows that the extra-striate visual networks become increasingly recruited and synchronized until about 275 ms after the stimulus onset, while the primary visual cortex is highly synchronous for a longer period of time, until about 350 ms after onset. The channels pair P7 and P8 exhibits the weakest synchronization strength since they are located the farthest away from the primary visual cortex compared to the other electrode pairs. On the other hand, the correlation coefficients for most channel pairs display the same level of synchronization until the sharp decrease at around 500 ms (except for P7–P8). The synchronization between P7 and P8 has irregular patterns with a local minimum around 200 ms. This comparison clearly demonstrates that also in this case the correntropy coefficient measure outperforms the correlation coefficient in the quantification of the EEG signal coupling between the bilateral occipital regions of the brain in response to visual stimuli.

5. Conclusions

In this paper, we propose the correntropy coefficient as a novel nonlinear interdependence measure. Due to a positive definite kernel function, the correntropy coefficient implicitly maps the original random variables or time series into an infinite dimensional reproducing kernel Hilbert space which is uniquely induced by the centred cross-correntropy function and essentially computes the cosine of the angle between the two transformed vectors. Orthogonality in RKHS $\mathcal{H}_U$ corresponds to independence between original random variables. Comparisons between the correntropy coefficient and the conventional correlation coefficient on simulated two unidirectionally coupled Hénon maps time series and EEG signals collected from sensory tasks clearly illustrate that the correntropy coefficient is able to extract more information than the correlation coefficient in quantification of synchronization between interacting dynamic systems. Correntropy is still easy to evaluate directly from data and so it is simpler to apply than other nonlinear techniques. However, correntropy introduces an extra parameter for the analysis, which is the kernel size. The kernel size affects the mapping to the nonlinear space and so it needs to be properly selected for the application. In many applications the simple Silverman's rule sets the parameter in a range that provides the desired result; however, more sophisticated techniques as cross-validation may have to be applied for more systematic results. This is a current research topic in our laboratory (and generally in machine learning).

Acknowledgments

This work was partially supported by NSF grant ECS-0601271, Graduate Alumni Fellowship from University of Florida and research scholarship from RIKEN Brain Science Institute.

References

Arnhold, J., Grassberger, P., Lehnertz, K., & Elger, C. E. (1999). A robust method for detecting interdependencies: application to intracranially recorded EEG. *Physica D*, *134*, 419–430.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, *68*(3), 337–404.

Grenander, U. (1981). *Abstract Inference*. New York: John Wiley & Sons.

Kotani, K., Kinomoto, Y., Yamada, M., Deguchi, J., Tonoike, M., Horii, K., et al. (2004). Spatiotemporal patterns of movement-related fields in stroke patients. *Neurology and Clinical Neurophysiology*, *63*, 1–4.

Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London*, *209*, 415–446.

Parzen, E. (1959). Statistical inference on time series by hilbert space methods *Tech. report 23*. Stat. Dept., Stanford Univ.

Pereda, E., Quian Quiroga, R., & Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology*, *77*, 1–37.

Pompe, B. (1993). Measuring statistical dependencies in a time series. *Journal of Statistical Physics*, *73*, 587–610.

Prichard, D., & Theiler, J. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Physics Review Letters*, *73*(7), 951–954.

Principe, J. C., Xu, D., & Fisher, J. W. (2000). Information theoretic learning. In S. Haykin (Ed.), *Unsupervised adaptive filtering* (pp. 265–319).

Quian Quiroga, R., Arnhold, J., & Grassberger, P. (2000). Learning driver-response relationships from synchronization patterns. *Physics Review E*, *61*(5), 5142–5148.

Rosenblum, M. G., Pikovsky, A. S., & Kurths, J. (1996). Phase Synchronization of Chaotic Oscillators. *Physics Review Letters*, *76*(11), 1804–1807.

Di Russo, F., Martinez, A., Serono, M. I., Pitzalis, S., & Hillyard, S. A. (2001). Cortical sources of the eearly components of the visual evoked potential. *Human Brain Mapping*, *15*, 95–111.

Santamaria, I., Pokharel, P., & Principe, J. C. (2006). Generalized correlation function: definition, properties, and application to blind equalization. *IEEE Transactions on Signal Processing*, *54*(6), 2187–2197.

Schiff, S. J., So, P., Chang, T., Burke, R. E., & Sauer, T. (1996). Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Physics Review E*, *54*, 6708–6724.

Schmitz, A. (2000). Measuring statistical dependence and coupling of subsystem. *Physics Review E*, *62*, 7508–7511.

Schreiber, T., & Schmitz, A. (2000). Surrogate Time Series. *Physica D*, *142*, 346–382.

Shaw, J. C. (1981). An introduction to the coherence function and its use in EEG signal analysis. *Journal of Medical Engineering and Technology (London)*, *5*(6), 279–288.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. New York: Chapman and Hall.

Stam, C. J., & van Dijk, B. W. (2002). Synchronization likelihood: An unbiased measure of generalized sysnchronization in multivariate data sets. *Physica D*, *163*, 236–251.

Xu, J. -W., Bakardjian, H., Cichocki, A., & Principe, J. C. (2007). A new nonlinear similarity measure for multichannel biological signals. In *Proc. intl. joint conf. on neural networks*.

Xu, J. -W., Pokharel, P. P., Paiva, A. R. C., & Principe, J. C. (2006). Nonlinear component analysis based on correntropy. In *Proc. intl. joint conf. on neural networks* (pp. 3517–3521).