Detrended Cross-Correlation Analysis: A New Method for Analyzing Two Nonstationary Time Series

Boris Podobnik*

Department of Physics, Faculty of Civil Engineering, University of Rijeka, Rijeka, Croatia
and Zagreb School of Economics and Management, Zagreb, Croatia

H. Eugene Stanley

Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

Here we propose a new method, detrended cross-correlation analysis, which is a generalization of detrended fluctuation analysis and is based on detrended covariance. This method is designed to investigate power-law cross correlations between different simultaneously recorded time series in the presence of nonstationarity. We illustrate the method by selected examples from physics, physiology, and finance.

There are a number of situations where different signals exhibit cross correlation. In seismology, the degree of cross correlation among noise signals taken at different antennas of detector arrays is an alert signaling earthquakes and volcanic eruptions [1]. In finance, risk is estimated on the basis of cross-correlation matrices .for different assets and investment portfolios [2]. In nanodevices for quantum information processing, electronic entanglement motivates the computation of current noise cross correlations, to see whether the sign of this signal would be reversed compared to the standard devices [3].

Consider two time series $\{y_i\}$ and $\{y_i'\}$, where $i = 1, 2, \ldots, N$. Each series can be represented as a random walk of k steps, and we can define $R_k \equiv y_1 + y_2 + \ldots + y_k$ and $R_k' \equiv y_1' + y_2' + \ldots + y_k'$, where $k \leq N$. Series $\{y_i\}$ has a mean $\mu \equiv \bar{y}_i \equiv (1/N)\sum_{i=1}^{N} y_i$ and a variance $\sigma^2 \equiv \overline{(y_i - \mu)^2}$, while series $\{y_i'\}$ has a mean $\mu' \equiv \bar{y}_i'$ and a variance $\sigma'^2 \equiv \overline{(y_i' - \mu')^2}$. We assume that the autocorrelation functions $A(n) \equiv \overline{(y_k - \mu)(y_{k+n} - \mu)}/\sigma^2$ and $A'(n) \equiv \overline{(y_k' - \mu')(y_{k+n}' - \mu')}/\sigma'^2$ scale as power laws $A(n) \sim n^{-\gamma}$ and $A'(n) \sim n^{-\gamma'}$, with $0 < \gamma, \gamma' < 1$. We further assume that for the cross-correlation function $X(n) \equiv \overline{(y_k - \mu)(y_{k+n}' - \mu')}/(\sigma\sigma')$ between the time series $\{y_i\}$ and $\{y_i'\}$

$$X(n) \sim n^{-\gamma_\times}, \qquad (1)$$

with $0 < \gamma_\times < 1$, where, e.g., since $X(n)$ is in general not even, $X(n) = Cn^{-\gamma_\times}$ for positive $n \gg 1$ and $X(n) = D(-n)^{-\gamma_\times}$ for negative n ($-n \gg 1$), where C and D are constants. However, this definition in Eq. assumes stationarity of both time series, and one can question its applicability to real-world data typically characterized by a high degree of nonstationarity.

Currently there is no method to quantify the cross-correlations exponent $\gamma_\times$ between two correlated time

series in the presence of nonstationarity [4]. Here we propose such a method, and we illustrate the method by selected examples from physics, physiology and finance. To this end, for two stationary processes y_k and y_k' we calculate the expected covariance

$$\langle (R_n - \langle R_n \rangle)(R_n' - \langle R_n' \rangle) \rangle = nX(0) + \sum_{k=1}^{n-1}[n - k]$$
$$\times [X(k) + X(-k)], \qquad (2)$$

where $X(0) = \langle (y_k - \mu)(y_k' - \mu') \rangle /\sigma\sigma'$. The sums with $X(k)$ of Eq. (2) can be approximated by integrals: $\sum_{k=1}^{n-1} X(k) \approx \sum_{k=1}^{n} k^{-\gamma_\times} \approx \int_1^n dx x^{-\gamma_\times} \propto n^{1-\gamma_\times}$, and $\sum_{k=1}^{n-1} kX(k) \approx \sum_{k=1}^{n} k^{1-\gamma_\times} \approx \int_1^n dx x^{1-\gamma_\times} \propto n^{2-\gamma_\times}$, and similar for sums with $X(-k)$. Asymptotically Eq. (2) scales as

$$\langle (R_n - \langle R_n \rangle)(R_n' - \langle R_n' \rangle) \rangle \sim n^{2\lambda}, \qquad (3)$$

where the scaling exponents λ and $\gamma_\times$—respectively related to the covariance and the cross-correlation function—are not independent, since $\lambda \equiv 1 - 0.5\gamma_\times$ [5]. For self-similar processes $\{y_i\} = \{y_i'\}$, the covariance of Eqs. (2) and (3) becomes the variance that for $n \gg 1$ scales as n^{2H}, so $\lambda = H$, where H is the Hurst exponent.

In order to quantify long-range cross correlations when nonstationarities are present, we propose a modification of the above covariance analysis which we call detrended cross-correlation analysis (DCCA). We consider two long-range cross-correlated time series $\{y_i\}$ and $\{y_i'\}$ of equal length N, compute two integrated signals $R_k \equiv \sum_{i=1}^{k} y_i$ and $R_k' \equiv \sum_{i=1}^{k} y_i'$, where $k = 1, \ldots, N$. We divide the entire time series into $N - n$ overlapping boxes, each containing $n + 1$ values. For both time series, in each box that starts at i and ends at $i + n$, we define the "local trend", $\tilde{R}_{k,i}$ and $\tilde{R}_{k,i}'$ ($i \leq k \leq i + n$), to be the ordinate of a linear least-squares fit. We define the "detrended

walk'' as the difference between the original walk and the local trend. Next we calculate the covariance of the residuals in each box $f_{\mathrm{DCCA}}^2(n, i) \equiv 1/(n-1) \sum_{k=i}^{i+n} (R_k - \tilde{R}_{k,i})(R'_k - \tilde{R}'_{k,i})$. Finally, we calculate the detrended covariance by summing over all overlapping $N - n$ boxes of size n,

$$F_{\mathrm{DCCA}}^2(n) \equiv (N - n)^{-1} \sum_{i=1}^{N-n} f_{\mathrm{DCCA}}^2(n, i). \qquad (4)$$

When only one random walk is analyzed ($R_k = R'_k$), the detrended covariance $F_{\mathrm{DCCA}}^2(n)$ reduces to the detrended variance $F_{\mathrm{DFA}}^2(n)$ used in the DFA method [6].

In order to test the utility of the proposed DCCA method, power-law autocorrelated time series y_i and y'_i are generated by using a stationary linear "ARFIMA" process [7]: $y_i \equiv \sum_{j=1}^{\infty} a_j(\rho) y_{i-j} + \eta_i$ [8], where $0 < \rho < 0.5$ is a free parameter, $a_j(\rho)$ are weights defined by $a_j(\rho) \equiv \Gamma(j - \rho)/[\Gamma(-\rho)\Gamma(1 + j)]$, $\Gamma(j)$ denotes the Gamma function, and η_i denotes an independent and identically distributed (i.i.d) Gaussian variable. The parameter ρ is related to the Hurst exponent, $H = 0.5 + \rho$ [7]. We generate two time series: $\{y_i\}$ with $\rho = 0.1$ and $\{y'_i\}$ with $\rho' = 0.4$. Since both $\{y_i\}$ and $\{y'_i\}$ are generated with the same error term η_i, $X(n) \neq 0$ [9]. In Fig. 1(a) we show that each time series exhibits the power-law autocorrelations expected for ARFIMA, and that the root mean square (rms) of the detrended covariance vs. n also follows approximately a power law, consistent with the fact that $\{y_i\}$ and $\{y'_i\}$ are power-law cross correlated. For different pairs of power-law autocorrelated time series $\{y_i\}$ and $\{y'_i\}$, characterized by Hurst exponents H and H', we find the time series are also power-law cross correlated, where the exponent λ of Eq. (3) is approximately equal to the average of the Hurst exponents: $\lambda \approx (H + H')/2$.

The power-law cross correlations between $\{y_i\}$ and $\{y'_i\}$ may exist only if $A(n) \sim n^{-\gamma}$ for both processes. We generate two time series by using two uncoupled ARFIMA processes: $\{y_i\}$ with $\rho = 0.1$ and $\{y'_i\}$ with $\rho' = 0.4$. In Fig. 1(b) we find that, even though both $\{y_i\}$ and $\{y'_i\}$ are power-law autocorrelated, the detrended covariance vs n of Eq. (4) fluctuates around zero which indicates that no power-law cross-correlations are present. The same result we show for uncoupled ARFIMA processes $\{y_i\}$ and $\{y'_i\}$ defined by $\rho = 0.2$ and $\rho' = 0.3$, respectively. Generally, if the detrended covariance vs n oscillates around zero, there are no power-law cross correlations with an unique exponent, but either no cross correlations or only short-range cross correlations exist between $\{y_i\}$ and $\{y'_i\}$.

To further exemplify the potential utility of the DCCA method for analyzing real-world data, we study two time series, both of which can be considered as two outputs of a complex system: the air humidity and the air temperature [10]. We analyze absolute values of the successive differences of air humidity (denoted by $\{|y_i|\}$) and air temperature (denoted by $\{|y'_i|\}$). Figure 2 shows that each of two
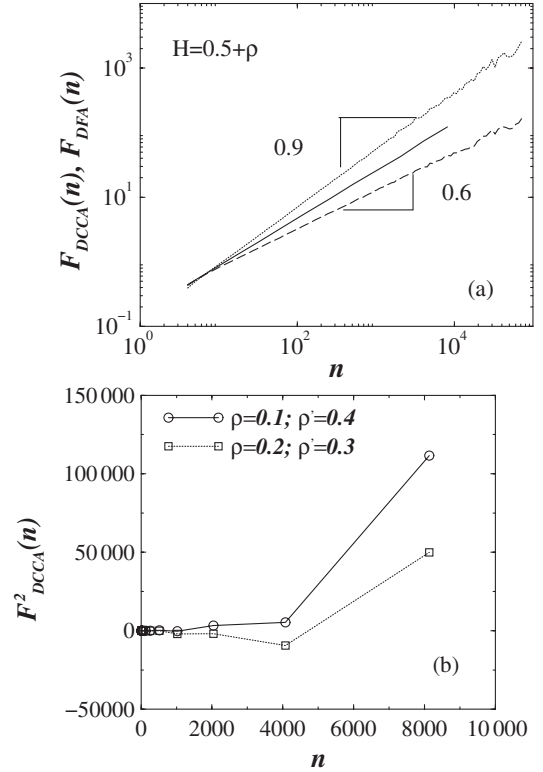

FIG. 1. (a) rms of detrended variance $F_{\mathrm{DFA}}(n)$ and detrended covariance, $F_{\mathrm{DCCA}}(n)$, where n is a scale. For two time series generated by two ARFIMA processes: $\{y_i\}$ with $\rho = 0.1$ and $\{y'_i\}$ with $\rho' = 0.4$ we show the DFA curves $F_{\mathrm{DFA}}(n)$ for both $\{y_i\}$ and $\{y'_i\}$, which can be fitted by power laws $F_{\mathrm{DFA}} \sim n^H$. Cross correlations are generated since we choose the error term to be equal for both time series: $\eta_i = \eta'_i$, where η_i corresponds to $\{y_i\}$ and η'_i corresponds to $\{y'_i\}$. When cross correlations are present, the same weights are responsible for power-law cross correlations between $\{y_i\}$ and $\{y'_i\}$. For $n \gg 1$ we find $F_{\mathrm{DCCA}}(n) \approx n^\lambda$ [see Eq. (4)], where $\lambda = 0.73$. This example illustrates the relation: $\lambda \approx (H + H')/2$. If we choose the error terms $\eta'_i = -\eta_i$, then $F_{\mathrm{DCCA}}^2(n)$ becomes negative for every n. For that case the cross-correlation function $X(n)$ becomes also negative. (b) Detrended covariance $F_{\mathrm{DCCA}}^2(n)$ of Eq. (4). We generate two pairs of two ARFIMA processes, where for each pair the time series are power-law autocorrelated, but not cross correlated, since each ARFIMA is generated by its own error term. The fluctuations, both positive and negative, indicate that two time series are not power-law cross correlated with an unique exponent, but either short-range cross correlated or not at all cross correlated.

time series $\{|y_i|\}$ and $\{|y'_i|\}$ exhibits power-law autocorrelations with similar scaling exponents. Figure 2 also shows that cross-correlations between $\{|y_i|\}$ and $\{|y'_i|\}$ exist and can be fit a power law n^λ with exponent $\lambda = 0.75$, practically equal to the exponent calculated for the temperature differences. We also analyze time series $\{y_i\}$ and $\{y'_i\}$, and we find that the DFA and DCCA analyses exhibit the correlated behavior, where $F_{\mathrm{DCCA}}^2(n)$ is negative for every n [11].

As a second example of real-world data, we analyze the Sleep Heart Health Study (SHHS) database which is de-
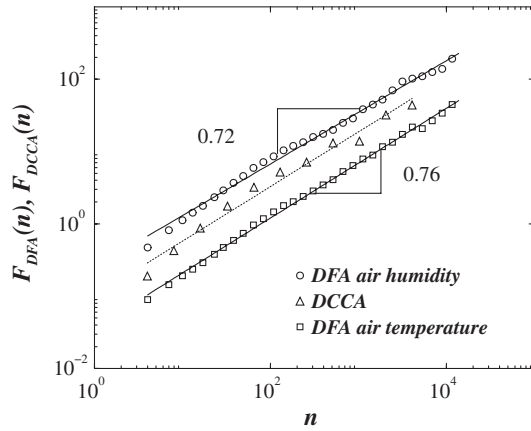
FIG. 2. Power-law autocorrelations and cross correlations in successive differences of air humidity $\{y_i\}$ and air temperature $\{y_i'\}$, recorded each 10 minutes. For time series of their absolute values, $\{|y_i|\}$ and $\{|y_i'|\}$, we find that both time series show sudden bursts of large changes. We show the rms of detrended variance $F_{\text{DFA}}(n)$ together with detrended covariance $F_{\text{DCCA}}(n)$. We find that DFA curves of $\{|y_i|\}$ and $\{|y_i'|\}$ and DCCA curve are very similar, and can be approximated with power laws $F_{\text{DFA}}(n) \sim n^H$ with scaling exponents $H = 0.72$ and $H' = 0.76$, and $F_{\text{DCCA}}(n) \sim n^\lambda$ with $\lambda = 0.75$.

signed to clarify the relationship between sleep disordered breathing and cardiovascular disease [12]. For a single patient [12,13], we analyze correlations behavior of five variables: two electroencephalography (EEG) variables, where EEG is the neuro-physiological measurement of the electrical activity of the brain recorded by electrodes commonly placed on the scalp; heart rate (HR) describing the frequency of the cardiac cycle, derived from the electrocardiogram (ECG) which records the electrical activity of the heart over time; and for both left and right eye the electrooculograms, obtained by measuring the resting potential of the retina.

We analyze the time series of two EEG variables simultaneously recorded every second, and find that each of them is short-range autocorrelated. For the absolute values of two EEG variables, the DFA curves in Fig. 3 show that each time series of the magnitudes exhibits power-law autocorrelated behavior, indicating that a large increment is more likely to be followed by a large increment. Figure 3 also shows that, besides autocorrelations, the time series of magnitudes exhibit power-law cross-correlations indicating that a large increment in one variable is more likely to be followed by large increment in the other variable. We also find that power-law magnitude cross correlations exist between the two time series of magnitudes of two EOG variables. We also find nonvanishing cross correlations between ECG time series and $C3/A2$ time series. In cross-sectional studies where many different physiological time series are recorded, an analysis based on the DCCA method should add diagnostic power to existing clinical methods employed to discriminate healthy from pathological behavior.
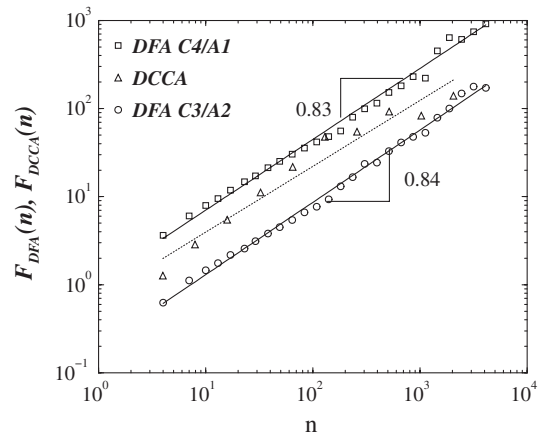

FIG. 3. Power-law autocorrelations and cross correlations between two different EEG time series: $C4/A1$ ($\{y_i\}$) and $C3/A2$ ($\{y_i'\}$) with 32 000 data points, recorded every second. For time series of their absolute values, $\{|y_i|\}$ and $\{|y_i'|\}$, we show rms of detrended variance $F_{\text{DFA}}(n)$ together with detrended covariance $F_{\text{DCCA}}(n)$. We find that $F_{\text{DFA}}(n)$ of $\{|y_i|\}$ and $\{|y_i'|\}$ and DCCA curve, $F_{\text{DCCA}}(n)$ are very similar, and can be approximated with power laws $n^\alpha(n^\lambda)$ with scaling exponents $\alpha = 0.83$ and $\alpha = 0.84$, $\lambda = 0.84$, respectively.

As a third example we analyze the daily closing values of the Dow Jones and the Nasdaq financial indices together with their corresponding trading volumes (the number of shares traded each day). For both price and trading volumes, we analyze the time series of absolute values of the differences of logarithms for successive days. In Figs. 4(a) and 4(c) we show their integrated signals $I(n) \equiv \sum_{i=1}^{n}(|y_i| - |\bar{y}_i|)$. In Figs. 4(b) and 4(d) we see that each of the four time series is power-law autocorrelated, and we also see both for absolute values of price changes ("volatility") and trading volume that the time series for Nasdaq and Dow Jones indices are power-law cross correlated.

We obtain similar results by analyzing cross-correlations between Microsoft and IBM stock prices and trading volumes, consistent with the interesting possibility that the above results may hold not only for indices, but also for individual companies. This result is especially interesting during volatile periods. Long-range cross correlations between two stocks imply that each stock separately has long memory of its own previous values and, additionally, has a long memory of previous values of the other stock.

In conclusion, we propose a new method to quantify the cross correlations between the simultaneously recorded time series. In cross-sectional studies where many different time series are recorded, the DCCA method based on local detrending together with other methods based on global oscillatory detrending may add diagnostic power to existing statistical methods. It is possible that the DCCA methods can be of use in areas where there are multiple correlated time series, such as in physiology where time series analysis is used as a clinical method to discriminate healthy from pathological behavior. In areas where periodicities are important aspect of stochastic processes, our
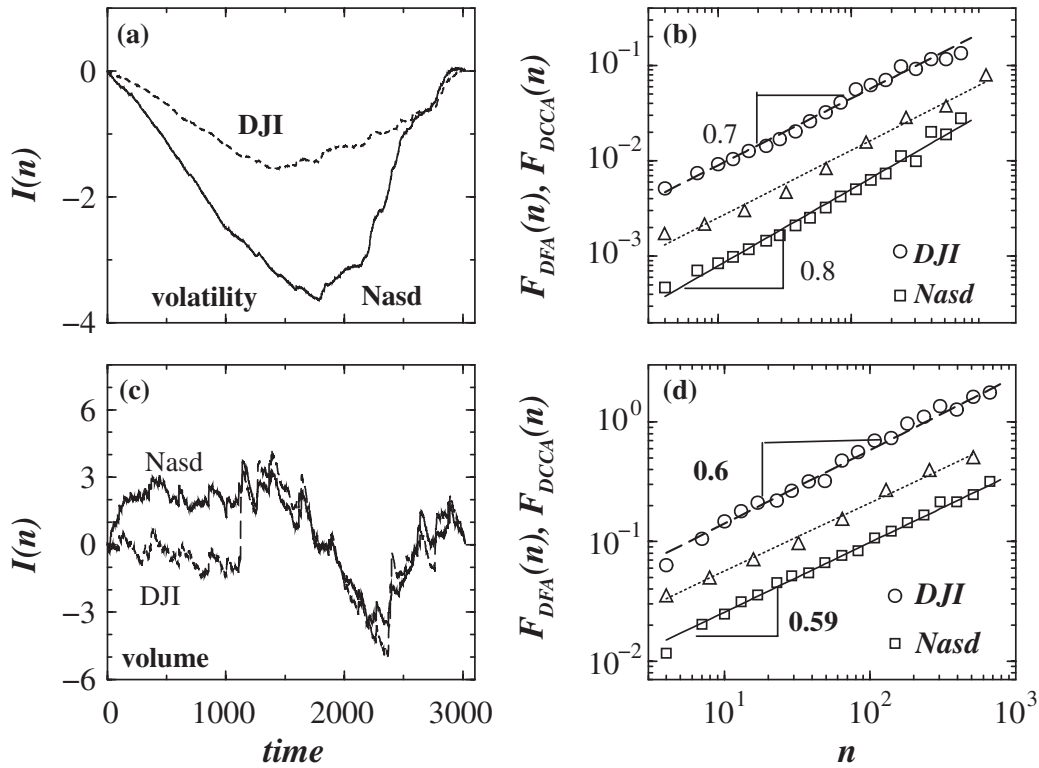
FIG. 4. Long-range autocorrelations and cross correlations in absolute values of price changes and trading volume for both Dow Jones index and Nasdaq index, recorded daily, in the period from July 1993 to November 2003. (a) Integrated profiles $I(n) \equiv \sum_{i=1}^{n}(|y_i| - |\bar{y}_i|)$ of the time series of absolute values $\{|y_i|\}$ and $\{|y_i'|\}$ of logarithmic changes in price for Dow Jones and Nasdaq, and (c) the corresponding absolute values $\{|z_i|\}$ and $\{|z_i'|\}$ for trading volume changes. (b) For price, rms of the detrended variance $F_{\text{DFA}}(n)$ curves for each $\{|z_i'|\}$ and $\{|z_i'|\}$, and also the rms of the detrended covariance, $F_{\text{DCCA}}(n)$. (d) Three curves represents the same measures but for "volume" (absolute of trading volume changes). For all DFA and DCCA curves we find both power-law autocorrelations and power-law cross correlations. Power-law cross correlations between Nasdaq and Dow Jones indices imply that current price changes of Nasdaq depend on its previous changes but also on previous price changes of Dow Jones. For trading volumes we also analyze time series $\{z_i\}$ and $\{z_i'\}$, and we find that the DFA and DCCA analyses show the anticorrelated behavior with DFA exponents $\alpha = 0.07$, $\alpha = 0.11$, and DCCA exponent $\lambda = 0.04$.

method can be used together with power spectrum in order to capture the frequency content of stochastic processes.

---

*bp@phy.hr

[1] M. Campillo and A. Paul, Science **299**, 547 (2003).

[2] L. Laloux *et al.*, Phys. Rev. Lett. **83**, 1467 (1999); V. Plerou *et al.*, *ibid.* **83**, 1471 (1999).

[3] P. Samuelsson *et al.*, Phys. Rev. Lett. **91**, 157002 (2003); A. Cottet *et al.*, *ibid.* **92**, 206801 (2004); I. Neder *et al.*, *ibid.* **98**, 036803 (2007).

[4] Recently, Jun *et al.* Phys. Rev. E **73**, 066128 (2006) proposed a detrended cross-correlation approach to quantify the correlations between positive and negative fluctuations in a single time series, and applied their approach to physiological and financial time series.

[5] The power spectrum is the Fourier transform of the autocorrelation function. Similarly, for cross correlations in the limit of small values of f, $S(f) \sim f^{-\beta_\times}$ with $X(n) \sim n^{-\gamma_\times}$, where $S(f) \approx 2\sum_{n=1}^{\infty} n^{-\gamma_\times}\cos(2\pi fn)$ we obtain $\beta_\times = 1 - \gamma_\times$ by using the Taylor expansion $(1 - y)^{-1+\gamma_\times} = 1 + \sum_{n=1}^{\infty}[(1 - \gamma_\times)\ldots(n - \gamma_\times)/\Gamma(n+1)]y^n$ and a definition of Γ function for $n \gg 1$, $\Gamma(z) \approx \Gamma(n + 1)n^z/[z(z+1)\ldots(z+n)]$.

[6] S. V. Buldyrev *et al.*, Phys. Rev. E **51**, 5084 (1995).

[7] J. Hosking, Biometrika **68**, 165 (1981); see also B. Podobnik *et al.*, Phys. Rev. E **72**, 026121 (2005).

[8] We introduce a cutoff length $\ell = 10^4$ in our numerical simulations, and we let the sum run from 1 to ℓ; i.e., we set $a_j = 0$ for $j \geq \ell$.

[9] Suppose one generates a sequence of N random numbers and then rerun the program in order to obtain a new sequence. The same sequence will be obtained and thus artificial cross correlations will be created.

[10] The data are publicly available at pdw.bic-gh.de.

[11] When $F_{\text{DCCA}}^2(n)$ is negative for every n, one may present $-F_{\text{DCCA}}^2(n)$ vs n in log-log plot.

[12] www.physionet.org/physiobank/database/shhpsgdb/.

[13] The data set is named 0000.dat.