Research Article

EMD Method for Minimizing the Effect of Seasonal Trends in Detrended Cross-Correlation Analysis

Keqiang Dong, You Gao, and Nianpeng Wang

1 College of Science, Civil Aviation University of China, Tianjin 300300, China
2 College of Science, Hebei University of Engineering, Handan 056038, China

Correspondence should be addressed to Keqiang Dong; hongzhangdong@163.com

Received 20 March 2013; Revised 6 September 2013; Accepted 18 September 2013

Academic Editor: Kwok-Wo Wong

Copyright © 2013 Keqiang Dong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Detrended cross-correlation analysis (DCCA) is a scaling method commonly used to estimate long-range power-law cross-correlation in nonstationary signals. Recent studies have reported signals superimposed with trends, which often lead to the complexity of the signals and the susceptibility of DCCA. This paper artificially generates long-range cross-correlated signals and systematically investigates the effect of seasonal trends. Specifically, for the crossovers raised by trends, we propose a smoothing algorithm based on empirical mode decomposition (EMD) method which decomposes underlying signals into several intrinsic mode functions (IMFs) and a residual trend. After the removal of slowly oscillating components and residual term, seasonal trends are eliminated.

1. Introduction

Recently, the existence of cross-correlations in complex systems has been extensively studied. Many studies provide strong empirical evidence for the existence of cross-correlations between time series. For example, in seismology, Campillo and Paul discussed the degree of cross-correlation among noise signals taken at different antennas of detector arrays [1]. In traffic engineering, Zebende and Filho found the existence of a cross-correlation between vehicles and passengers [2]. In finance, Shen et al. estimated the risk on the basis of cross-correlations for different assets and investment portfolios [3, 4]. In meteorology, Podobnik et al. studied the cross-correlation in successive differences of air humidity and air temperature [5, 6]. Cross-correlation functions together with autocorrelation functions are commonly used to gain insight into the dynamics of natural systems. Unfortunately, many time series, exhibiting cross-correlations, of physical, hydrological, and biological are nonstationary. Consequently, statistical properties of these systems are difficult to study due to these nonstationarities.

For determining the scaling exponent of cross-correlational time series in the presence of nonstationarities, detrended cross-correlation analysis method was developed by Podobnik et al. [5–8], and its performance was systematically tested for the effect of nonstationarities. After that, numerous methods referring to a broad range of applications [9–11] were established to investigate cross-correlational signal in the presence of nonstationarities.

However, many noisy signals in real systems exhibit trends so that the scaling phenomenon obtained by DCCA method becomes intricate to reveal the intrinsic dynamics of the real systems due to many trends changing the scaling exponent for different range of scales. Specifically, seasonal trends usually lead to a change of cross-correlations in signals. A technique, developed by Hajian and Movahed, based on singular-value decomposition (SVD) of the trajectory matrix was applied on the method specifically to eliminate possible crossovers in cross-correlation exponents [9]. Based on chaos theory and SVD, the CSVD technique was reported by Dong et al. to minimize the effect of sinusoidal trends in traffic time series and stock data [12].

In this work, the DCCA procedure is pieced together with empirical mode decomposition (EMD), which is a well-established and promising method to analyze nonlinear and nonstationary signals [13]. EMD, first proposed by Huang et al., is an empirical method to analyze data, which decomposes raw signals into several intrinsic mode functions (IMFs) and
The method was employed to cope with the crossover of scaling exponent in DFA [16, 17]. By subtracting a trend embedded in the original signal, an EMD-based DFA was proposed in [16], where in each segment EMD was applied rather than a global used form, which is a good attempt but with a disadvantage of enormous time consumption. For this reason, this paper modifies the EMD-based DFA and extends to EMD-based DCCA.

Here we first investigate the scaling characteristic of noisy signals with sinusoidal trends. And then, we obtain the scaling properties of the reconstructed data by applying the EMD-based method, and we compare the difference in the scaling results. The results indicate that sinusoidal trends embedded in original data are extracted and the intrinsic fluctuations are retrieved.

The organization of this paper is as follows. First, the DCCA method is given in Section 2. The EMD-based technique is proposed in Section 3. Section 4 is the demonstration of effectiveness of the proposed technique on artificial and aeroengine data corrupted with seasonal trend. Finally, important conclusions drawn from this study are provided in Section 5.

2. DCCA Method

Many physical and biological time series are often nonstationary or have means, variances, and covariances that change over time. Nonstationary behaviors can be trends, cycles, random walks, or combinations of the three.

DCCA is a new method to quantify the cross-correlations between two nonstationary time series. This method is an extension of detrended fluctuation analysis (DFA) method. The DCCA procedure consists of four steps. Both methods are based on random walk theory. For two nonstationary time series \( x_i \) and \( y_i \), \( i = 1, 2, \ldots, N \), the DCCA method is given as follows.

**Step 1.** Compute the profiles of underlying time series using

\[
X (i) = \sum_{k=1}^{i} (x_k - \langle x \rangle),
\]

\[
Y (i) = \sum_{k=1}^{i} (y_k - \langle y \rangle),
\]

where \( \langle x \rangle = (1/N) \sum_{j=1}^{N} x_j \) and \( \langle y \rangle = (1/N) \sum_{j=1}^{N} y_j \) are the mean respectively.

**Step 2.** Cut the profiles \( X \) and \( Y \) into \( Ns = [N/s] \) non-overlapping segments of equal length \( s \), respectively. In each segment \( v \), we calculate the local trend by a least-square fit of the data and obtain the difference between the original time series and the fits.

**Step 3.** Calculate the covariance of the residuals in each segment as follows:

\[
f^2_{DCCA} (s, v) = \frac{1}{s} \sum_{k=1}^{s} (X - \bar{X}_{k,v}) (Y - \bar{Y}_{k,v}),
\]

where \( \bar{X}_{k,v} \) and \( \bar{Y}_{k,v} \) are the fitting polynomials in segment \( v \), respectively. Then the average over all segments to obtain the fluctuation function is as follows:

\[
f_{DCCA} (s) = \left( \frac{1}{2Ns} \sum_{v=1}^{2N} f^2_{DCCA} (s, v) \right)^{1/2}.
\]

**Step 4.** Determine the power-law relationship between the detrended cross-correlation function \( f_{DCCA} (s) \) and \( s \) as follows:

\[
f_{DCCA} (s) \sim s^{\lambda},
\]

where the scaling exponent \( \lambda \) represents the degrees of the cross-correlation between the two time series \( \{x_i\} \) and \( \{y_i\} \). If, however, the detrended cross-correlation oscillates around zero as a function of the time scale \( s \), there are no long-range cross-correlations. For time series \( x_i = y_i \), the standard DFA procedure is retrieved.

3. EMD-Based Method

The empirical mode decomposition (EMD) first proposed by Huang et al. is designed for the time-frequency analysis of real-world signals [13]. By applying it, the signal can be decomposed into a number of oscillatory modes named intrinsic mode functions (IMFs). The IMFs are obtained directly from the data with no a priori assumptions regarding the data nature, making EMD suitable for the analysis of nonlinear and nonstationary signals. And then, these IMFs provide meaningful instantaneous frequency estimates through Hilbert–Huang transform. The time-frequency analysis via EMD has found a wide range of applications in signal processing and related fields [18–23].

Specifically, for EMD method, if \( \{x(k)\} \) denotes an input, then the output is a set of IMFs, denoted by \( \{d_j (k)\}_{j=1}^M \), such that

\[
x (k) = \sum_{j=1}^{M} d_j (k) + r (k),
\]

where the residual \( r(k) \) is a monotonic function. The IMFs are achieved by sifting out rapidly oscillating components (dominant modes) from the data by iteratively subtracting slowly oscillating components (less dominant modes). These slowly oscillating components are defined as the local trend of a signal. The sifting process is stopped when all the less dominant modes are extracted in this paper.

The procedure used for the extraction of IMFs from a signal \( \{x(k)\} \) is as follows.

**Step 1.** Identify all the extrema (maxima and minima) of the series \( \{x(k)\} \).

**Step 2.** Obtain the signal envelope passing through the minima \( \{e_{\min}(k)\} \) and maxima \( \{e_{\max}(k)\} \) via cubic spline interpolation, respectively.
Step 3. Compute the local mean series by point averaging of the two envelopes

$$m(k) = \frac{\{e_{\min}(k) + e_{\max}(k)\}}{2}. \quad (6)$$

Step 4. Subtract the mean from the signal to obtain an IMF candidate \(s(k) = x(k) - m(k)\).

Step 5. Check the properties of \(s(k)\): if \(s(k)\) is not an IMF, replace \(x(k)\) with \(s(k)\) and repeat the procedure from Step 1 and if \(s(k)\) is an IMF, evaluate the residual \(r(k) = x(k) - s(k)\).

Step 6. Repeat the procedure from Step 1 to Step 5 by sift ing the residual signal. The sifting process ends when the residual \(r\) is monotonic.

As previously demonstrated [24], the IMFs and the residual obtained by using EMD are not always unique and could change as the stoppage criterion for the sifting process changes. Once the trend is determined, the corresponding detrending operation can be implemented. With this definition of trend, the variability of the data on various time scales can also be derived naturally [25].

As we expect to eliminate the trend from original signal \(\sum_{j=1}^{M} d_j(k)\) without slowly oscillating components \(\sum_{j=0}^{M} d_j(k) + r(k)\) will be used as signal without trend for DCCA. Slowly oscillating components may keep the major energy when signals present strong trend as we will see in Section 4.

4. EMD-Based DCCA Method for Minimizing the Effect of Sinusoidal Trends

The effectiveness of the proposed technique in minimizing the sinusoidal trends will be discussed in this section. Specifically, we consider two types of time series, artificial and aerogenous time series, in this paper.

4.1. Example 1: Artificial Time Series. In this section, we employ two-component ARFIMA method [26–28] to generate artificial long-range power-law signals \(\{u(k)\}\) and \(\{v(k)\}\) and test EMD-based method in minimizing the effects of sinusoidal trends. In this model, the series is defined by

$$u_i = \sum_{j=1}^{\infty} a_j(\rho) y_{i-j} + \eta_i, \quad (7)$$

where \(0 < \rho < 0.5\) is a free parameter, \(a_j(\rho) = \Gamma(j - \rho)/\Gamma(-\rho)\Gamma(i + j)\) are weights, \(\Gamma(j)\) is the Gamma function, and \(\eta_i\) is an independent and identically distributed (i.i.d) Gaussian variable. By introducing a parameter \(\rho\), the method can efficiently generate signals with power-law correlations characterized by \textit{a priori} known correlation exponent \(H = 0.5 + \rho\). Using the ARFIMA method, we generate two time series: \(\{u(k)\}\) with \(\rho = 0.1\) and \(\{v(k)\}\) with \(\rho = 0.4\), and the cross-correlation exponent approximately is equal to their mean value.

First, we use ARFIMA method to produce correlated signals \(\{u(k)\}\) with \(\rho = 0.1\) and \(\{v(k)\}\) with \(\rho = 0.4\) and illustrate DFA and DCCA on the signals in Figure 1. The scaling exponents \(H_i\) \((i = 1, 2)\) in DFA method correspondingly approach \(0.5 + \rho_i\) \((i = 1, 2)\), and cross-correlation exponent approximately equals to their mean value are expected.

For reliable detection of the cross-correlations, it is essential to distinguish trends from the intrinsic fluctuations in data. Generally, trends embedded in measurements are of two types: polynomial and sinusoidal trends. Although the DCCA method eliminates the polynomial trends, the sinusoidal trends remain [9].

Inspired from [29, 30], we superimpose a sinusoidal trend on the correlated signals above. A superimposed sinusoidal trend is shown as follows:

$$t(k) = A \sin\left(\frac{2\pi k}{T}\right), \quad k = 1, 2, 3, \ldots, N. \quad (8)$$

The generated data sets with sinusoidal trends, \(t_i(k)\) with the period \(T = 2^N\), will be denoted as \(x(k) = u(k) + t(k)\) and \(y(k) = v(k) + t(k)\). A conspicuous crossover (circles) is found in Figure 2, when we apply DFA on the original signals superimposed with a strong sinusoidal trend.

We, respectively, use EMD to decompose original signals with sinusoidal trends into several IMFs. The IMFs are achieved by sift ing out rapidly oscillating components from the data, by subtracting slowly oscillating components as shown in Figures 3 and 4. These slowly oscillating components are related to the sinusoidal trend in signal. We obtain the corresponding filtered data using the EMD-based algorithm and apply DFA to reconstructed data. We find that the detrended function \(F(t)\) (triangles) absent ed crossover phenomenon is similar to that of the data sets without sinusoidal trend (plus signs); see Figure 2. The results
4 Mathematical Problems in Engineering

Figure 2: DFA on signals with sinusoidal trends (circles), reconstructed data using the smoothing filter (triangles), and original signals generated by ARFIMA method (plus signs).

Figure 3: The IMFs for $\rho = 0.1$, where rapidly oscillating components are from IMF1 to IMF7. We trade the superposition of rapidly oscillating components as correlated signals without trend.

in Figure 2 indicate the remarkable ability of EMD-based method in eliminating a sinusoidal trend in DFA.

Next, we apply DCCA on the original signals superimposed with sinusoidal trend. As shown in Figure 5, the scaling exponent by DCCA shows an identical result compared to DFA exponents in Figure 2. Specifically to the crossover in circles plot, we again use EMD-based method to the signals and compare the DCCA function of reconstructed data to DCCA function of original data. Similar scaling properties are illustrated in Figure 5, which indicates a good performance of EMD in elimination of a sinusoidal trend.

We note that crossovers caused by any sinusoidal trend in artificial time series can be solved by EMD. We also test the effect of real time series on the method and find the expected results in the next section.

In the following discussion, we study the effectiveness of the proposed method on correlated signals with seasonal trend. Given a seasonal trend,

$$t(k) = A \sin \left( \frac{2\pi k}{T} \right) + B \cos \left( \frac{2\pi k}{T} \right), \quad k = 1, 2, 3, \ldots, N,$$

(9)
where $A$ and $B$ are the amplitude. The generated data sets with seasonal trends will be denoted as $x(k) = u(k) + t(k)$ and $y(k) = v(k) + t(k)$. We apply DCCA on the new signals, which is affected by the trend. When the EMD-based method is applied, the DCCA function of reconstructed data is close to the DCCA function of original data.
4.2. Example 2: Aeroengine Time Series. In this section, the effectiveness of the proposed method is analyzed by applying the aeroengine series. The sinusoidal trends superimposed on the exhaust gas temperature (EGT) and engine fan speeds (NI), respectively, denoted by \( \{\nu(k)\} \) and \( \{\nu(k)\} \) will be discussed. The aircraft data includes operational flight data from on-board flight data recorders. These recorders are part of the Aircraft Condition Monitoring System (ACMS) such as Smart ACMS Recorder (SAR) and Quick Access Recorder (QAR). The QAR data are employed because the data includes an extensive list of flight parameters recorded at specific sampling intervals which are set by the manufacturer.

The DCCA of the original aeroengine data, original data with the sinusoidal trend, and reconstructed data is shown in Figure 7. It can be seen that the original characteristics of the aeroengine data are retained by the proposed smoothing technique. Obviously, the detrended cross-correlation function \( F(s) \) of filtered data is similar to cross-correlation function obtained by the data without sinusoidal trend. Similar scaling properties are illustrated compared to Figure 2, which indicates a good performance of EMD-based method in elimination of a sinusoidal trend not only in artificial time series but also in aeroengine time series.

5. Conclusion

In the paper, we consider EMD-based method to eliminate sinusoidal trends which introduce crossovers in cross-correlation exponent in DFA and DCCA. The effectiveness of the proposed EMD-based method is applied to long-range power-law cross-correlated signals which are generated by two-component ARFIMA artificially. After the removal of slowly oscillating components, sinusoidal trend is successfully eliminated. EMD shows advantage in eliminating sinusoidal trend in artificial time series.

EMD together with DCCA method can be applied to real dynamics to reliably detect long-range cross-correlation where sinusoidal trend exists. In aeroengine dynamics, we apply EMD method by decomposing underlying signals into several intrinsic mode functions and present similar results with artificial time series. We note that EMD is of great advantage when dealing with sinusoidal trend. Therefore, we do believe that our smoothing filter algorithm may provide some help to minimize the effect of the trends introduced by the sinusoidal trends and facilitate a reliable extraction of the scaling exponent.

Acknowledgments

The financial support from the funds of the Fundamental Research Funds for the Central Universities under Grant no. ZXH 2012C002, the National Natural Science Foundation of China under Grant no. U1232301, and the Natural Science Foundation of Hebei Province (E2012402013) are gratefully acknowledged.

References


