Physics Letters A 379 (2015) 680–687

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Analysis of detrended time-lagged cross-correlation between two nonstationary time series

Shen Chenhua [a,b,c,*]

[a] College of Geographical Science, Nanjing Normal University, Nanjing 210046, China
[b] Jiangsu Center for Collaborative Innovation in Geographical Information Resource, Nanjing 210046, China
[c] Key Laboratory of Virtual Geographic Environment of Ministry of Education, Nanjing 210046, China

ARTICLE INFO

Article history:
Received 23 September 2014
Received in revised form 19 December 2014
Accepted 20 December 2014
Available online 24 December 2014
Communicated by C.R. Doering

Keywords:
Nonstationary time series
Time lag
Time-lagged DCCA cross-correlation coefficient
The detrended time-lagged cross-correlation analysis

ABSTRACT

A time-lagged DCCA cross-correlation coefficient is proposed with objective of quantifying the level of time-lagged cross-correlation between two nonstationary time series at time scales. This coefficient, $\rho(n, \tau, R, R')$, is defined based on a DCCA cross-correlation coefficient $\rho_{DCCA}(n)$. The implementation of this coefficient will be illustrated through selected time series of wind speed and air pollution index (API). The results indicate that both time scales and time lags are very small, $\rho(n, \tau, R, R')$ is attributed to a time-lagged effect; while when time lags are comparatively large, $\rho_{DCCA}(n)$ contributes partially to $\rho(n, \tau, R, R')$. This partial contribution is greater when $\tau < n$ and less when $\tau > n$. $\rho(n, \tau, R, R')$ is applied in meteorology. It is found that the method is reasonable and reliable. Therefore, the detrended time-lagged cross-correlation analysis can be useful to deepen and broaden our understanding of cross-correlations between nonstationary time series.

1. Introduction

Time series in the real world may contain hidden cross-correlation information due to complex interactions and diverse mechanisms. The characterization of these cross-correlations is important to an insight into their nature and mechanism [1]. For the purpose of simulation, the properties of these correlations should be studied and suitable models should be developed.

Time-lagged cross-correlation usually refers to the correlation between two time series shifted relatively in time. Time-lagged cross-correlations between time series have been studied and an analytic method has been widely applied in diverse fields [2–7]. However, nonstationary time series hamper the usage of classical statistical methods in practical application due to the stationary assumption of time series [1].

A new method, detrended cross-correlation analysis (DCCA), has been proposed to analyze power-law cross-correlations between nonstationary time series [8]. This method is a generalization of the detrended fluctuation analysis (DFA) method [9] and is based on detrended covariance. By detrending local trends, DCCA ensures that the results obtained are not affected by trend (including linear, quadratic, and even higher order trends and periodic trends) [10–12]. The DCCA can uncover more hidden correlation information than other analyses, leading to its acceptance and application in diverse fields [13–24].

However, there has to date been a few researches related to time-lagged cross-correlations. Jose et al. investigated a lagged DFA [25] for nonstationary time series based on DFA, and found that the largest correlation was at positive lags. Lin et al. studied the dynamics of the cross-correlations between stock time series based on a time delay by means of DCCA, and analyzed the time-lagged characteristics of two specific economic time series [26]. Sang et al. systematically studied the definition of time-lagged cross-correlations, and provided calculation methods and the results for two specific hydrologic time series by means of the wavelet transformation [27]. Marinho et al. found that there are several highly correlated events in geological structures, including partial and global dislocations of deposited layers [28]. Although these studies were relevant to lagged cross-correlations, there is little systematic investigation of the characteristic of time-lagged cross-correlations.

In this paper, a time-lagged DCCA cross-correlation coefficient is proposed, quantifying the level of time-lagged cross-correlations between two nonstationary time series at different time scales, based on the DCCA cross-correlation coefficient proposed by Zebende et al. [24,29]. Considering the number of fundamental variables and the possible applications, we restrict to identify and quantify the time-lagged cross-correlations between time se-

* Correspondence to: College of Geographical Science, Nanjing Normal University, Nanjing 210046, China.
E-mail address: shenandchen01@163.com.

http://dx.doi.org/10.1016/j.physleta.2014.12.036
0375-9601/

ries of wind speed and air pollution index (API) because of the simplicity of the interactional mechanism between them. Simple interactional mechanism helps us to understand the time-lagged cross-correlation deeply. Moreover, this interactional mechanism can also confirm that the method presented here is reliable and reasonable.

The structure of this paper is as follows. Section 2 describes methods, proposed in this paper, in detail. Section 3 uses the method to calculate a time-lagged DCCA cross-correlation coefficient of wind speed and API, and presents key findings and discussions. In Section 4, conclusions are drawn.

2. Description of the method

2.1. Time-lagged DCCA cross-correlation coefficient

Zebende proposed DCCA cross-correlation coefficient to analyze and quantify the level of cross-correlations between two time series [29]. In Ref. [29], two nonstationary time series $\{x(i)\}$ and $\{y(i)\}$ of equal length N, $i = 1, 2, \ldots, N$, were first supposed. Then, two integrated series $R_k \equiv \sum_{i=1}^{k}(x(i) - \bar{x})$ and $R'_k \equiv \sum_{i=1}^{k}(y(i) - \bar{y})$ were computed, where $\bar{x}$ and $\bar{y}$ denote the averaging over the two full time series. Next, the entire time series were divided into $N - n$ overlapping boxes [8], each containing $n + 1$ values. Finally, the covariance $f^2_{dcca}(n, i)$ of the residuals in each box of size n (time scale) that starts at i and ends at $i + n$ was calculated as in Eq. (1)

$$f^2_{dcca}(n,i) = \frac{1}{n+1}\sum_{k=i}^{i+n}(R_{k,i} - \overline{R_{k,i}})(R'_{k,i} - \overline{R'_{k,i}}) \tag{1}$$

where $\overline{R_{k,i}}$ and $\overline{R'_{k,i}}$ $(i \le k \le i+n)$ are local detrends for different time series. In the end, the detrended covariance $f^2_{dcca}(n, i)$ was summed over all overlapping $N - n$ boxes, and the DCCA cross-correlation coefficient $\rho_{DCCA}(n)$ was proposed.

As with the analysis above, if we calculate time-lagged covariance of the residuals in each box of size n [24,29] between two time series, and average time-lagged covariance over all overlapping $N - n - |\tau|$ boxes, we will implement the time-lagged DCCA cross-correlation coefficient. Hereafter τ is used to denote time lag, and it can be either positive or negative. In general, a positive τ indicates that $y(i + \tau)$ lags behind $x(i)$ while a negative τ means that $x(i + \tau)$ lags behind $y(i)$. The calculation procedure of the time-lagged DCCA cross-correlation coefficient proposed here can be summarized in the following steps:

Step 1. Determine the profiles

Similarly, two integrated series $R_k \equiv \sum_{i=1}^{k}(x(i) - \bar{x})$ and $R'_k \equiv \sum_{i=1}^{k}(y(i) - \bar{y})$, $k = 1, 2, \ldots, N$, are first computed [8], where $\{x(i)\}$ and $\{y(i)\}$ are time series of equal length N, $\bar{x}$ and $\bar{y}$ are denoted as the average over the entire time series. Of course, $R'_{k+\tau} \equiv \sum_{i=1}^{k+\tau}(y(i) - \bar{y})$, where $k + \tau = 1, 2, \ldots, N$. For convenience, the profile, $Z_k(\tau) \equiv R'_{k+\tau}$, is defined, $k = 1, 2, \ldots, N - \tau$.

Step 2. Divide the profiles

The two integrated series R_k and Z_k are divided into $N - n - |\tau|$ overlapping boxes, each containing $n + 1$ values.

Step 3. Calculate time-lagged covariance of the residuals in each box

The local trends, $\overline{R_{k,i}}$ and $\overline{R'_{k+\tau,i+\tau}}$ $(i \le k \le i+n)$ [8], are defined. For simplicity, a linear least-squares fit is used to calculate local trend for each box. For $\overline{R_{k,i}}$, in each box that starts at i and ends at $i + n$. For $\overline{R'_{k+\tau,i+\tau}}$, in each box that starts at $i + \tau$ and ends at $i + \tau + n$, where, when $\tau > 0$, $i = 1, 2, \ldots, N - n - \tau$; while when $\tau < 0$, $i = -\tau + 1, \ldots, N - n$. The detrended walk is defined as the difference between the original walk and the local trend. The time-lagged covariance of the residuals in each box is calculated [25,26,28] as in Eq. (2).

$$f^2_{dcca}(n,i,\tau,R,R') = \frac{1}{n+1}\sum_{k=i}^{i+n}(R_{k,i} - \overline{R_{k,i}})(R'_{k+\tau,i+\tau} - \overline{R'_{k+\tau,i+\tau}}) \tag{2}$$

It is to be noted that R and R' are designated as different time series.

Step 4. Average over all boxes to obtain the detrended time-lagged covariance function $F^2_{DCCA}(n, \tau, R, R')$

$F^2_{DCCA}(n, \tau, R, R')$ is summed over all overlapping $N - n - |\tau|$ boxes of size n as in Eq. (3) or Eq. (4).

$$F^2_{DCCA}(n,\tau,R,R') = \frac{1}{N-n-\tau}\sum_{i=1}^{N-n-\tau} f^2_{dcca}(n,i,\tau,R,R') \quad \text{for } \tau >= 0 \tag{3}$$

$$F^2_{DCCA}(n,\tau,R,R') = \frac{1}{N-n+\tau}\sum_{i=-\tau+1}^{N-n} f^2_{dcca}(n,i,\tau,R,R') \quad \text{for } \tau < 0 \tag{4}$$

Step 5. Compute time-lagged DCCA cross-correlation coefficient

If $R_k = R'_k$, the detrended time-lagged covariance reduces to the detrended time-lagged variance [9], thus, $F_{DFA}(n, \tau, R) = \sqrt{F^2_{DCCA}(n, \tau, R, R)}$ and $F_{DFA}(n, \tau, R') = \sqrt{F^2_{DCCA}(n, \tau, R', R')}$. As the DCCA cross-correlation coefficient [29], time-lagged DCCA cross-correlation coefficient is defined as in Eq. (5).

$$\rho(n,\tau,R,R') = \frac{F^2_{DCCA}(n,\tau,R,R')}{F_{DFA}(n,\tau,R)F_{DFA}(n,\tau,R')} \tag{5}$$

From Eq. (5), $\rho(n, \tau, R, R')$ depends on the parameter n to account for time scale and on the time lag τ to account for any delayed causality. The value of $\rho(n, \tau, R, R')$ ranges from -1 to 1 according to Cauchy inequality. Clearly, when $\tau = 0$, the procedure is identical in all aspects to the procedure of the DCCA cross-correlation coefficient, and the $\rho(n, \tau, R, R')$ reduces to $\rho_{DCCA}(n)$. Furthermore, when $n \gg \tau$, $\rho(n, \tau, R, R')$ will reduce to $\rho_{DCCA}(n)$ as well. Since there is no symmetry for time series R and R', $\rho(n, \tau, R, R')$ is not usually equal to $\rho(n, \tau, R', R)$ for any time scale n and any time lag τ. $\rho(n, \tau, R, R')$ can quantify the levels of time-lagged cross-correlations between two nonstationary time series. For $\rho(n, \tau, R, R') > 0$, the larger the $\rho(n, \tau, R, R')$, the higher the time-lagged cross-correlation level; while for $\rho(n, \tau, R, R') < 0$, the smaller the $\rho(n, \tau, R, R')$, the higher the time-lagged anti-cross-correlation level. The time-lagged cross-correlation analysis can be seen as a generalization of existing detrended cross-correlation analysis schemes.

2.2. Analysis of detrended time-lagged cross-correlations

It is supposed that a cross correlation between two time series $\{x(i)\}$ and $\{y(i)\}$ of equal length N exists. Due to an impulse in time series $\{x(i)\}$ at time $i = 0$, dramatic changes in time series $\{x(i)\}$ at time $i = 0$ could lead to continuous changes in time series $\{y(i)\}$ in a temporal direction. As always, since the time-lagged effect is limited, drastic changes in time series $\{x(i)\}$ at time $i = 0$ will make the values of time series $\{y(i)\}$ become the values of time series $\{y_e(i)\}$ with $\tau < \tau_{\max}$; while the values of time series $\{y_e(i)\}$ will remain nearly unchanged with $\{y(i)\}$ with $\tau > \tau_{\max}$, where $\tau_{\max}$ is denoted as the maximum time lag, suggesting that for a given time scale n, $\rho(n, \tau, R, R')$ returns to