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)

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

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).

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

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

the level of $\rho(n,0,R,R')$ with an increase in time lags. This process also implies that for $\tau > \tau_{\max}$, the level of the time-lagged cross-correlations, caused by strong time-lagged effects, becomes weaker. Consequently, time series $\{y_e(i)\}$ is attributed to a time-lagged effect and its length is limited.

From the viewpoint of the "lag" [1], an impulse in time series $\{x(i)\}$ at time $i=0$, results in continuous changes in time series $\{y_e(i)\}$, and time series $\{y_e(i)\}$ lags behind time series $\{x(i)\}$ at time $i=0$. On the contrary, in the viewpoint of the "lead" [1], an impulse of time series $\{x(i)\}$ at time $i=\tau_{\max}$, causes continuous changes in time series $\{y_e(i)\}$ in reverse temporal direction.

2.3. Relationship between maximum time lag, time lags and time scales

Hereafter $\tau^{+}_{\max}$ is denoted as positive maximum time lag and $\tau^{-}_{\max}$ is denoted as negative maximum time lag (absolute value).

When $0 < \tau + n < \tau^{+}_{\max}$, the local trend $\overline{R'_{k+\tau,i+\tau}}$ is closely related to time series $\{y_e(i)\}$, and the detrended walk $(R'_{k+\tau,i+\tau} - \overline{R'_{k+\tau,i+\tau}})$ is attributed to the structural characteristic of time series $\{y_e(i)\}$. At this time, $\rho(n,\tau,R,R')$ primarily discloses the level of time-lagged cross-correlations between time series $\{x(i)\}$ and $\{y_e(i)\}$. This level of time-lagged cross-correlations is denoted as R_e. In this circumstance, $\rho(n,\tau,R,R')$ primarily reflects the impact, which comes from a change in time series $\{x(i)\}$ at time $i=0$, on time series $\{y(i)\}$. When time lag τ is very small, but time scale n is comparatively large, $\rho(n,\tau,R,R')$ primarily discloses the level of time-lagged cross-correlations between time series $\{x(i)\}$ and the majority of time series $\{y_e(i)\}$; while when both time lag τ and time scale n are very small, $\rho(n,\tau,R,R')$ primarily reveals the level of time-lagged cross-correlations between time series $\{x(i)\}$ and the former part of time series $\{y_e(i)\}$. On the contrary, when time lag τ is comparatively large, but time scale n is very small, $\rho(n,\tau,R,R')$ primarily reveals the level of time-lagged cross-correlations between time series $\{x(i)\}$ and the latter part of time series $\{y_e(i)\}$. Moreover, for a given time lag τ, $\rho(n,\tau,R,R')$ can also reveal the structural characteristic of time series $\{y_e(i)\}$ at different time scales.

When $\tau^{+}_{\max} < \tau$, the local trend $\overline{R'_{k+\tau,i+\tau}}$ is closely related to time series $\{y(i)\}$, and the detrended walk $(R'_{k+\tau,i+\tau} - \overline{R'_{k+\tau,i+\tau}})$ is attributed to the structural characteristic of the time series $\{y(i)\}$. $\rho(n,\tau,R,R')$ primarily reveals the level of time-lagged cross-correlations between time series $\{x(i)\}$ and $\{y(i)\}$. This level of time-lagged cross-correlations is denoted as R_z. Moreover, the partial contribution to $\rho(n,\tau,R,R')$ from $\rho_{DCCA}(n)$ is comparatively large when $\tau < n$ and is comparatively small when $\tau > n$.

When $0 < \tau < \tau^{+}_{\max}$ and $\tau^{+}_{\max} < \tau + n$, both the local trend $\overline{R'_{k+\tau,i+\tau}}$ and the detrended walk $(R'_{k+\tau,i+\tau} - \overline{R'_{k+\tau,i+\tau}})$ are closely connected to time series $\{y_e(i)\}$ and $\{y(i)\}$. $\rho(n,\tau,R,R')$ is primarily attributed to R_e and R_z. When both time lag τ and time scale n are very small, the contribution to $\rho(n,\tau,R,R')$ comes mostly from R_e. On the contrary, when time lag τ is very small, but time scale n is comparatively large, the contribution to $\rho(n,\tau,R,R')$ comes mostly from R_z.

When $\tau^{-}_{\max} < \tau < 0$ and $\tau^{-}_{\max} < \tau + n < 0$, the local trend $\overline{R'_{k+\tau,i+\tau}}$ is also closely related to time series $\{y_e(i)\}$, and the detrended walk $(R'_{k+\tau,i+\tau} - \overline{R'_{k+\tau,i+\tau}})$ is attributed to the structural characteristic of the part of time series $\{y_e(i)\}$. $\rho(n,\tau,R,R')$ mainly presents the level of time-lagged cross correlations at time scales between time series $\{x(i)\}$ and $\{y_e(i)\}$. $\rho(n,\tau,R,R')$ is mainly attributed to R_e. When $\tau + n < \tau^{-}_{\max}$, the local trend $\overline{R'_{k+\tau,i+\tau}}$ is also closely connected to time series $\{y(i)\}$, and the detrended walk $(R'_{k+\tau,i+\tau} - \overline{R'_{k+\tau,i+\tau}})$ is attributed to the structural change of time series $\{y(i)\}$. $\rho(n,\tau,R,R')$ primarily reflects the level of time-lagged cross-correlation between time series $\{x(i)\}$ and $\{y(i)\}$. When $\tau < \tau^{-}_{\max}$ and $\tau^{-}_{\max} < \tau + n$, $\rho(n,\tau,R,R')$ is mainly attributed to R_e and R_z.

2.4. Symmetry of two time series R and R′

Even if cross-correlations between time series $\{x(i)\}$ and $\{y(i)\}$ exist, the continuous changes in time series $\{y(i)\}$ in a temporal direction, resulting from an impulse in time series $\{x(i)\}$ at time $i=0$, is clearly different from the continuous changes in time series $\{x(i)\}$ in a temporal direction, resulting from an impulse in time series $\{y(i)\}$ at time $i=0$. Thus, $\rho(n,\tau,R,R')$ is not equal to $\rho(n,\tau,R',R)$ at any time scale or any time lag. In other words, there is no symmetry, but generally only asymmetry, for R and R'.

3. Application: time-lagged DCCA cross-correlation coefficient between wind speed and API

For our purpose to verify the detrended time-lagged cross-correlation analysis presented here, the time-lagged DCCA cross-correlation coefficient of wind speed and API is calculated and analyzed. The smaller the API, the better the air quality. The key four reasons for us to choose wind speed and API are: 1) Two time series of wind speed and API are nonstationary, and each series is power-law correlated; 2) There are cross-correlations between them. Generally, if there is no air pollution source surrounding the city, wind can accelerate air–pollutant diffusion to reduce API. As a result, air quality improves; 3) A time-lagged effect exists. In fact, after a strong wind blows, the air quality improves on the next day; 4) The length of the time series should be adequate.

3.1. Data source

Daily wind speed and synchronous API records data from January 2001 to December 2012 in Nanjing, China, were obtained for the analysis. The daily wind-speed data were obtained from the China Meteorological Data Sharing Service System (http://cdc.cma.gov.cn), and the daily API records data were obtained from the Ministry of Environmental Protection of the People's Republic of China (http://datacenter.mep.gov.cn). These data sets have been commonly used to study climate and air pollution changes over China in recent years [19,30]. All the records are controlled for quality.

See Fig. 1. In order to test whether time series of wind speed and API are nonstationary or not, software module of nonparametric runs test, provided by SPSS 19 (Statistical Product and Service Solutions), is used to compute the P value and to infer significance. As a result, the P value is 0.00. It is inferred that the time series of wind speed and API are nonstationary. In addition, Shen et al. reported that a cross correlation between wind speed and API by the DCCA exists [19].

3.2. Results and discussion

3.2.1. Evolution of the time-lagged DCCA cross-correlation coefficient

Fig. 2(a–f) shows that the value of $\rho(n,\tau,R,R')$ ranges between -1 and 1. This finding is in line with Eq. (5).

Fig. 2(a and c) shows that, for all studied time scales, the values of $\rho(n,\tau,R,R')$ are negative when $\tau=0$ and $\tau=1$. Furthermore, at a small time scale n, $\rho(n,1,R,R')$ is less than $\rho(n,0,R,R')$; while at a large time scale n, $\rho(n,1,R,R')$ tends to be nearly equal to $\rho(n,0,R,R')$. This confirms that air quality improves after a strong wind occurs on a subsequent day at a given small time scale. Generally, at a small time scale, air pollutants in the center of the city diffuse outside the city, leading to a decrease

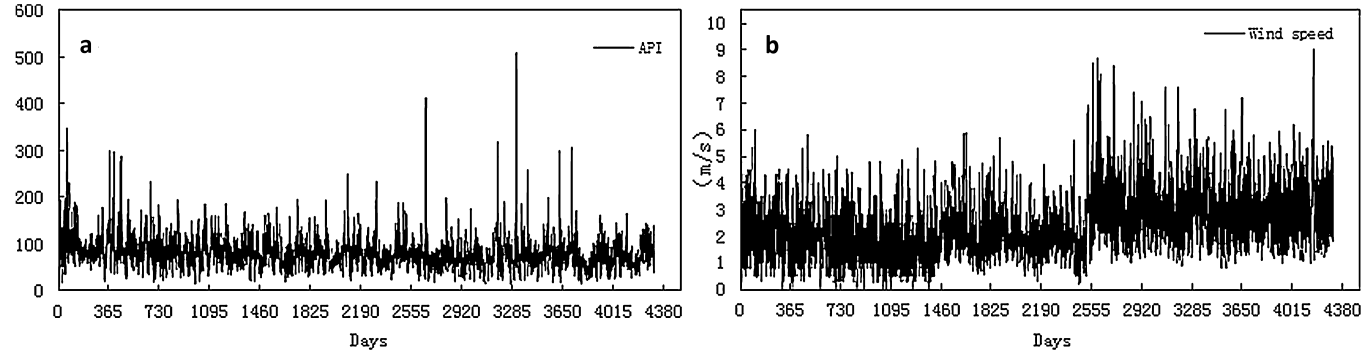

Fig. 1. (a and b) are daily original records of wind speed and API in Nanjing, China, from January 2001 to December 2012, respectively.

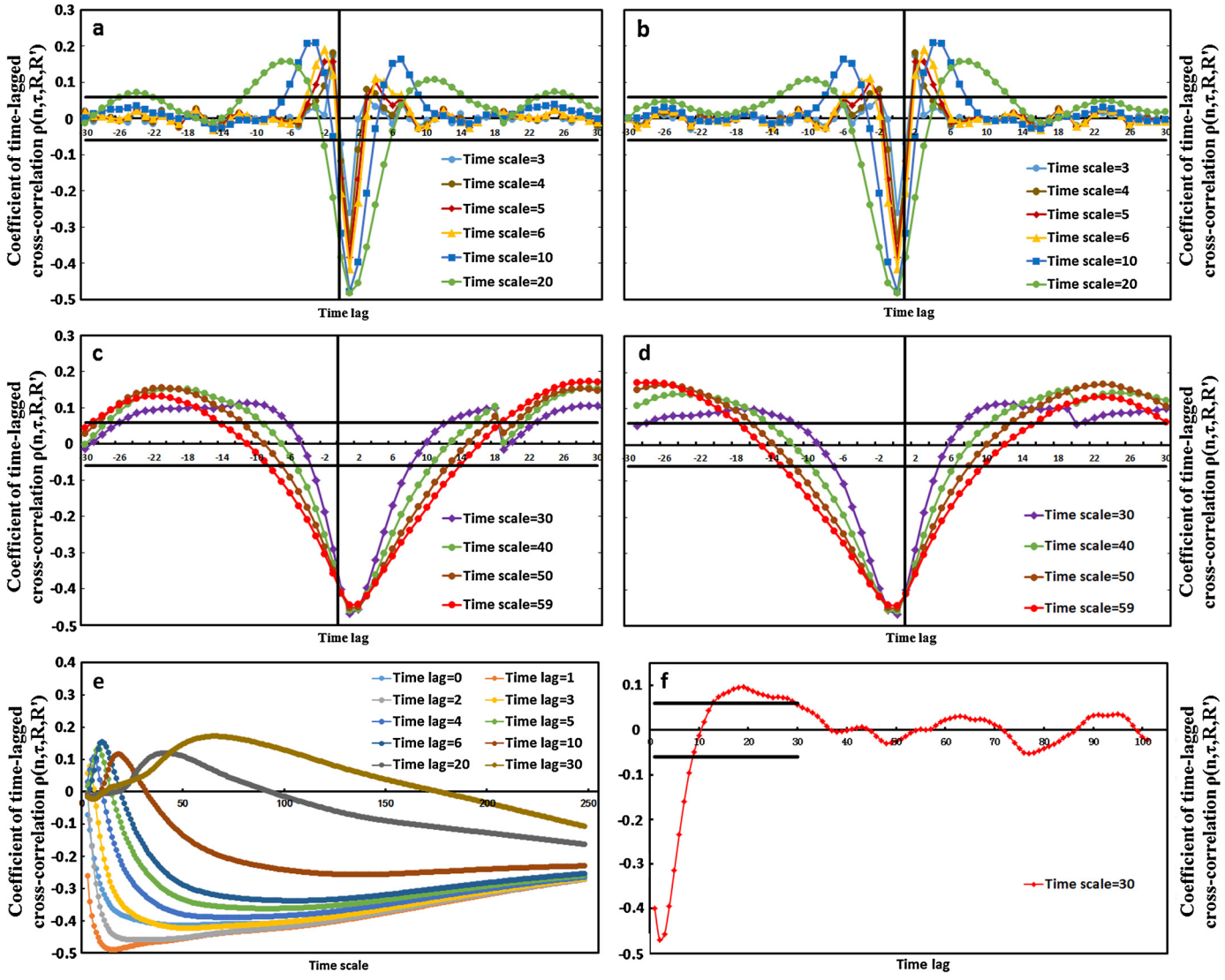

Fig. 2. (a–d) is $\rho(n,\tau,R,R')$ against time lags. (e) is $\rho(n,\tau,R,R')$ against time scales. (f) is $\rho(30,\tau,R,R')$ against time lags. In Fig. (a, c, e, f), R is the time series of wind speed and R' is the time series of API, and vice versa, in Fig. (b and d). Black coarse lines in Fig. (a–d, f) are the values of the error bar of the confidence interval of 95%. The other colored lines represent time-lagged DCCA cross-correlation coefficient. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in API on the first day, after a strong wind blows. On the second day, air quality improves due to the wind's effect on air pollution. On later days, since the wind's effect on air pollution weakens, air pollutants surrounding the city will gradually flow into the city center, causing the tardy increase in air pollution concentration. $\rho(n,\tau,R,R')$ will thus return to the level of $\rho(n,0,R,R')$ with an increase in time lags because the air pollution concentration reverts to its previous levels. Fig. 2(a and c) also shows that $\rho(n,\tau,R,R')$ varies with time lags for a fixed time scale n. $\rho(n,\tau,R,R')$ is either negative or positive. This demonstrates that a time-lagged cross-correlation may be persistent or anti-persistent at different time lags and time scales.

Fig. 2(e) shows that the value of $\rho(n,\tau,R,R')$ varies with time scales for a fixed time lag. Clearly, the variation of $\rho(n,\tau,R,R')$

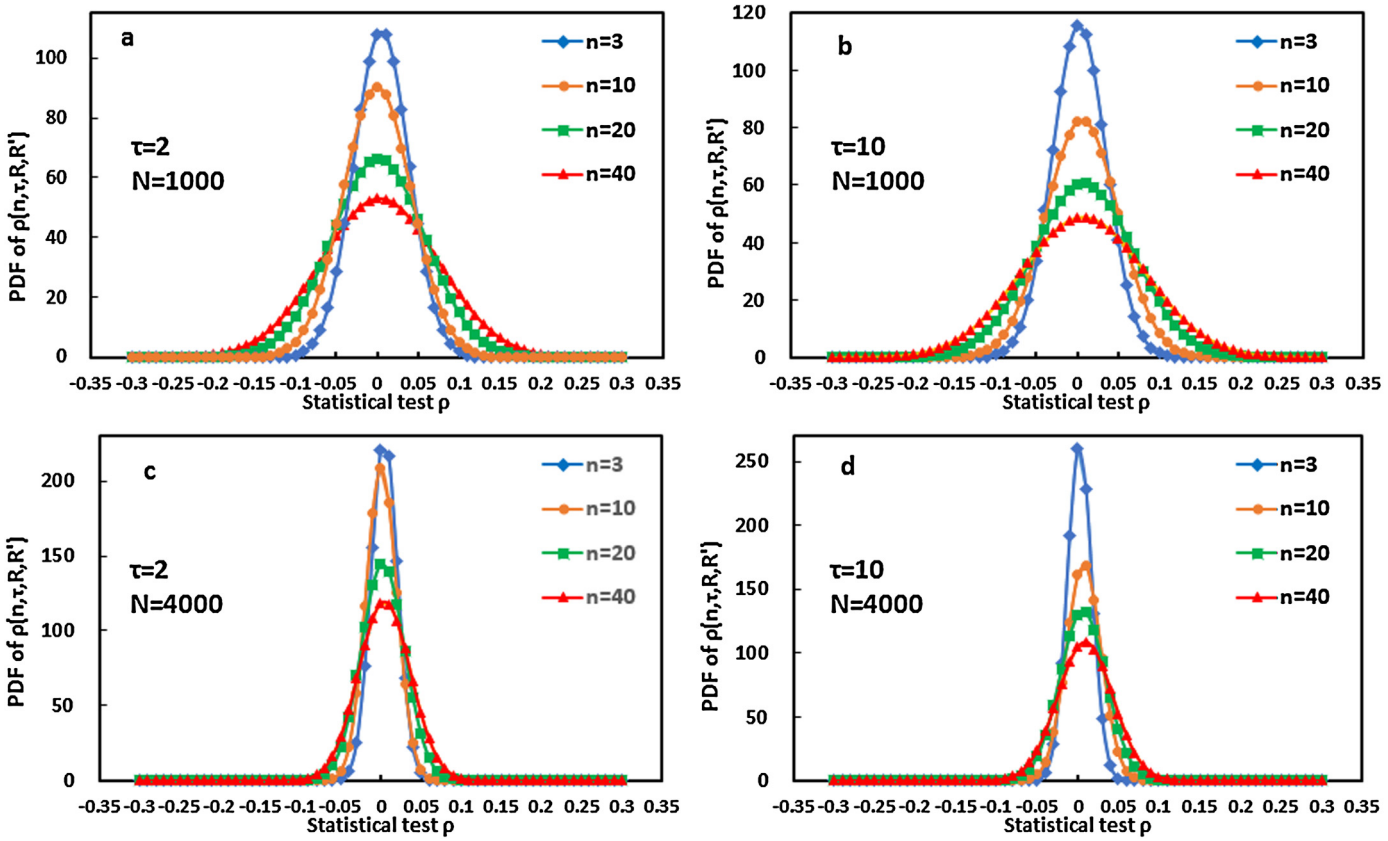

Fig. 3. Fits of probability distribution functions (PDF) of $\rho(n, \tau, R, R')$. n denotes time scale and N denotes the time series length. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

at a small time lag is different from $\rho(n, \tau, R, R')$ at a medium or large time lag.

3.2.2. Statistical test for $\rho(n, \tau, R, R')$

For our purpose to verify whether $\rho(n, \tau, R, R')$ is significant or not, a calculation of the critical numerical values [11] to determine a criterion is thus necessary.

First, it is hypothesized that each series is power-law correlated rather than identically distributed random variables (i.i.d.) [11]. Then, the DFA exponents of API and wind speed are calculated to test the presence of significant power-law cross-correlations. As a result, the DFA exponents of API and wind speed are 0.76 and 0.80, respectively. This result indicates that time series of API and wind speed are power-law correlated. Next, 1000 artificial long-term correlated time series, with DFA exponent 0.80, are generated. Because the artificial time series are generated from different white noises [31], they should have no cross-correlations with API. Finally, $\rho(n, \tau, R, R')$ of artificial data and the API is calculated. The mean coefficient of artificial data and the API is determined by the critical values. If $\rho(n, \tau, R, R')$ is outside the critical values, $\rho(n, \tau, R, R')$ is considered significant, while if $\rho(n, \tau, R, R')$ is inside the critical values, $\rho(n, \tau, R, R')$ is considered insignificant.

Fig. 3 shows that the fits of probability distribution functions (PDF) of $\rho(n, \tau, R, R')$ for different sizes of time scale and time lag, with time-series length $N = 1000$ and 4000. This result demonstrates that PDFs of $\rho(n, \tau, R, R')$ are relevant to time scale, time lag and time-series length. Single-sample *Kolmogorov–Smirnov* test is used to confirm that all studied PDFs of $\rho(n, \tau, R, R')$ are Gaussians, as expected.

Fig. 4 shows the relationship between the standard deviation of $\rho(n, \tau, R, R')$'s PDF, time-series length, time scales and time lags. Fig. 4(a) shows that the standard deviation decreases with an increase in time-series length. This result is in accordance with Ref. [12], and suggests that the larger the time-series length, the smaller the standard deviation, and that a long time series can be conducive to detrended time-lagged cross-correlation analysis. Fig. 4(b) shows that the standard deviation increases with an increase in time scale. This result is also in accordance with Ref. [12], and indicates that the larger the time-scale size, the larger the standard deviation. Fig. 4(c) shows that the relationship between the standard deviation and time lags is rather complicated. We speculate that this complicated relationship might be relevant to time-lagged effects, and further investigation should be launched.

In Fig. 5(a), $\rho(n, \tau, R, R')$ of the artificial data and the API is shown. It is found that for $n = 3$, the critical value is almost 0.04 at any studied time lag, while for $n = 20$, the critical value is almost 0.06 at any studied time lag. In addition, for $n = 60$, the critical value is almost 0.08 at any studied time lag. This result means that the critical values of the artificial data versus the API remain almost unchanged at the studied time lag for a fixed time scale, and increase very slowly with an increase in time scales at a fixed time lag. Thus, for simplicity, we take the critical value 0.06 into consideration for all studied time scales and time lags in this paper.

3.3. Discussion

According to the maximum time lag of physical process discussed in Section 2.2, maximum time lag is extracted from Fig. 2(a and c). Fig. 5(b) shows the maximum time lag versus time scales. As would be expected, maximum time lag changes according to time scales. When $n = 3$, $\tau^{+}_{\max} \simeq 2.5$ days; while when $n = 5$, $\tau^{+}_{\max} \simeq 2.8$ days. In addition, when $n = 12$, $\tau^{+}_{\max} \simeq 4.0$ days. This result shows that maximum time lag increases overall with an in-

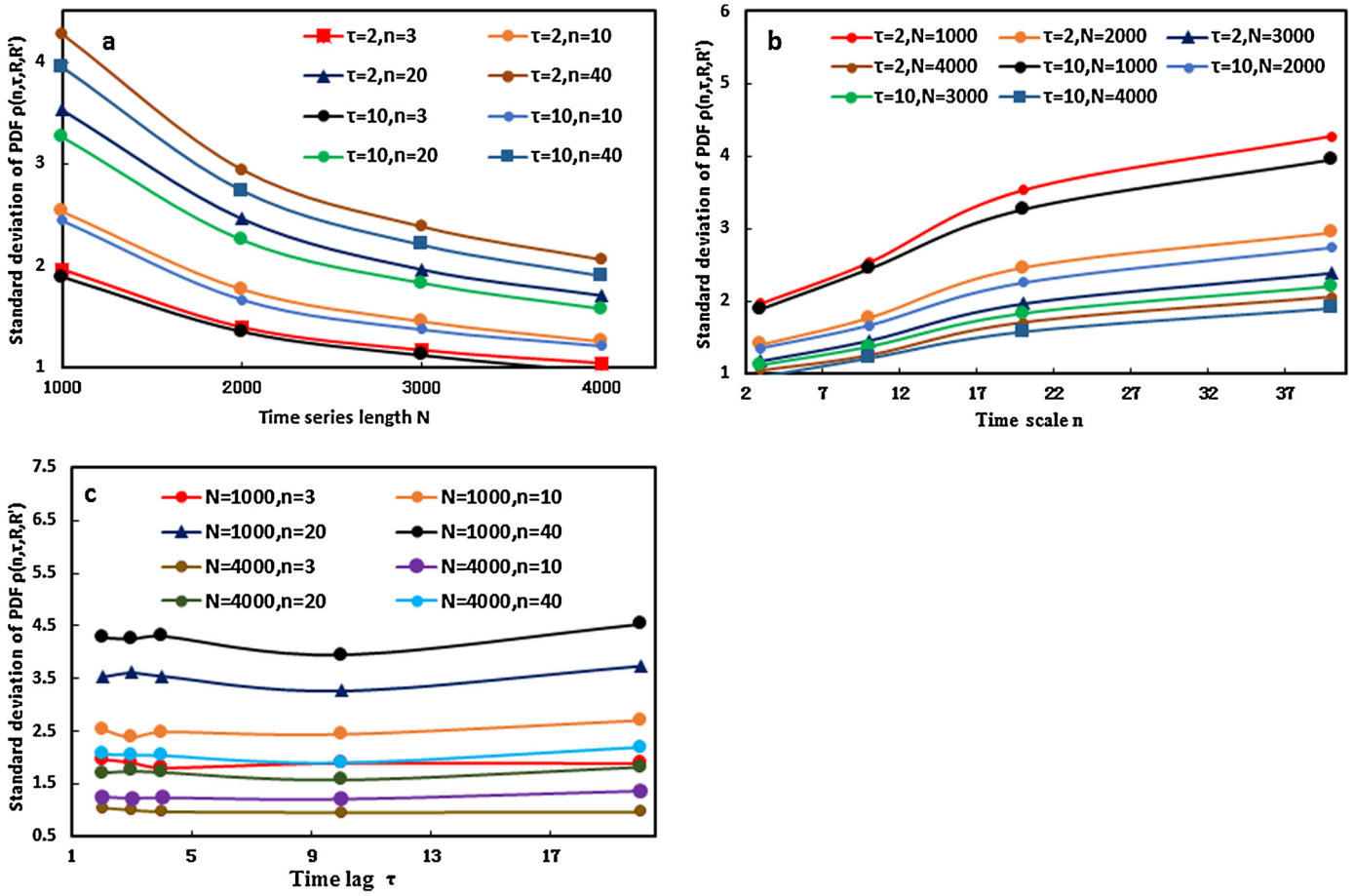

Fig. 4. (a) is the standard deviation of $\rho(n, \tau, R, R')$'s PDF versus the time-series length for time scales and time lags. Fig. (b) is the standard deviation of the $\rho(n, \tau, R, R')$'s PDF versus time scales for different time-series length and time lags. Fig. (c) is the standard deviation of the $\rho(n, \tau, R, R')$'s PDF versus time lags for different time-series length and time scales. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

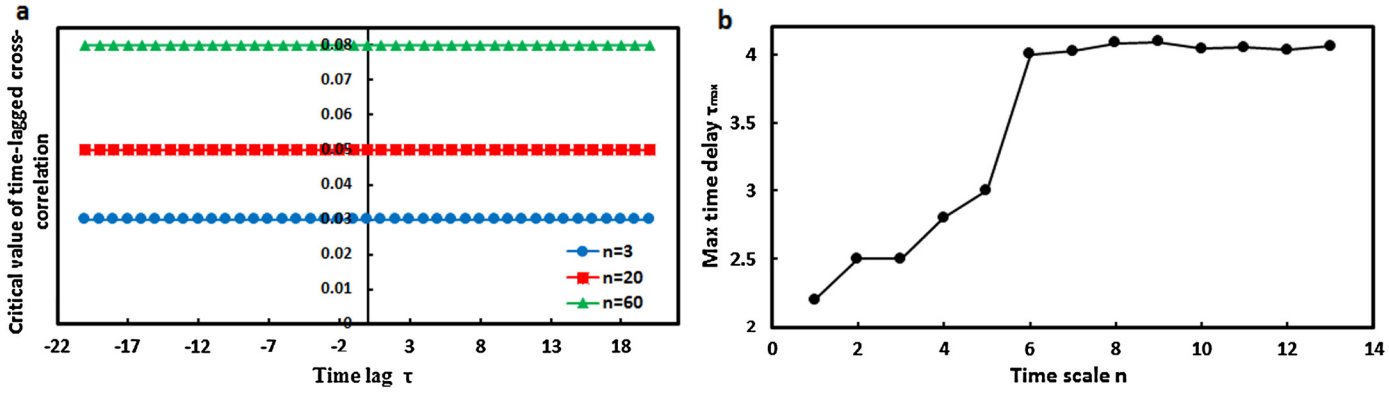

Fig. 5. (a) is critical value of time-lagged cross-correlation coefficient against time lags for different time scales. (b) is maximum time lag versus time scales. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

crease in time scales until it reaches a point where it tends nearly to be a constant.

Fig. 2(b and d) shows $\rho(n, \tau, R, R')$ versus time scales, where R is designated as the time series of API, and R' as the time series of wind speed. Comparing Fig. 2(a) to Fig. 2(b) and Fig. 2(c) to Fig. 2(d), the asymmetry about wind speed and API is found at studied time scales and time lags.

Fig. 2(e) shows that $\rho(n, \tau, R, R')$ varies with time scales in a fixed time lag, reflecting the level of cross-correlation between one time series and the other lagged time series at time scales. It is found that for a fixed time lag, the variation of $\rho(n, \tau, R, R')$ with time scales differs. The tendency of $\rho(n, \tau, R, R')$ is consistent at a large time scale. For instance, when $\tau = 1$, the value of $\rho(n, \tau, R, R')$ is negative and decreases with increased time scales. The value of $\rho(n, \tau, R, R')$ decreases first and then increase overall with increases in time scales. When $\tau = 4$, the value of $\rho(n, \tau, R, R')$ is first positive and later reaches maximum value. Afterwards, $\rho(n, \tau, R, R')$ changes from the positive value to the negative value. When time scales become comparatively large, the value of $\rho(n, \tau, R, R')$ starts to converge. $\rho(n, \tau, R, R')$ increases first and then decreases overall with increases in time scales. In addition, when $\tau = 20$, the value of $\rho(n, \tau, R, R')$ is zero at first,

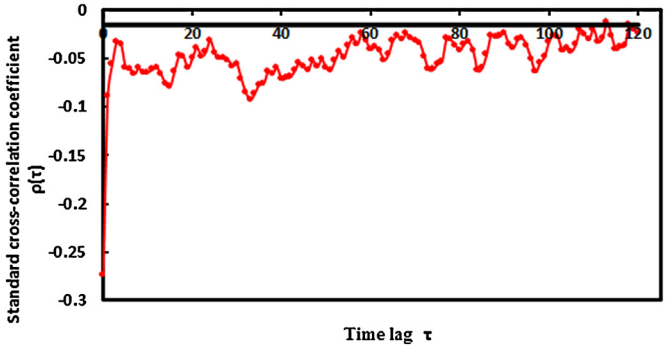

Fig. 6. Standard cross-correlation coefficients versus time lag. Black coarse line is the value of the error bar of the confidence interval of 95%. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and later becomes positive. Afterwards, $\rho(n,\tau,R,R')$ changes from the positive value to the negative value. When time scales become comparatively large, $\rho(n,\tau,R,R')$ also starts to converge. The value of $\rho(n,\tau,R,R')$ increases first and then decrease overall with increased time scales.

This situation is mainly related to R_e and R_z. When $\tau = 1$, $\rho(n,\tau,R,R')$ is related to R_e. When $n < \tau = 20$, the contribution to $\rho(n,\tau,R,R')$ from ρ_{DCCA} is minimal. This is why it is almost zero in Fig. 2(e). However, when $n > \tau = 20$, the contribution to $\rho(n,\tau,R,R')$ that derives partially from $\rho_{DCCA}(n)$ is significant. So, in Fig. 2(e), $\rho(n,\tau,R,R')$ later becomes nonzero. Moreover, due to the contribution to $\rho(n,\tau,R,R')$ from $\rho_{DCCA}(n)$, $\rho(n,\tau,R,R')$ will gradually converge to $\rho_{DCCA}(n)$ when time scales becomes very large.

Fig. 2(a and c) shows that $\rho(n,\tau,R,R')$ varies with time lag in a fixed time scale. For different fixed time scales, the evolution of $\rho(n,\tau,R,R')$ with time lag is basically similar. For example, in Fig. 2(a), when $n = 4$, the value of $\rho(n,\tau,R,R')$ decreases with an increase in time lag at first, and later increases with an increase in time lag. Afterwards, $\rho(n,\tau,R,R')$ decreases with an increase in time lag and converges. Although $\rho(n,\tau,R,R')$ converges to a certain fixed value with the increase in time lag, each time lag is unique at this time. For example, in Fig. 2(a), if $n = 4$, $\rho(n,\tau,R,R')$ converges to almost zero at $\tau \simeq 2$. In this situation, τ is less than n, and $\rho(n,\tau,R,R')$ is primarily attributed to R_e. While, in Fig. 2(f), if $n = 30$, $\rho(n,\tau,R,R')$ converges to almost zero at $\tau \geq 30$. In this circumstance, the contribution to $\rho(n,\tau,R,R')$ from R_z is minimal, and $\rho(n,\tau,R,R')$ tends to be almost zero with an increase in time lags.

Thus, according to Fig. 2(a and c) and Fig. 2(e), the variation of $\rho(n,\tau,R,R')$ with time lags and time scales is rather complicated. The mechanisms by which $\rho(n,\tau,R,R')$ converges to zero for $\tau \simeq 2$ and $\tau \simeq 30$ are possibly different. We speculate that for both time lags and time scales, the variation in $\rho(n,\tau,R,R')$ seems to detect different cross-correlation patterns. However, since the maximum time lag of wind speed and API is very short, the time-lagged cross-correlation pattern at small time scales is not very clear, and further study should be carried out. For our purpose to compare $\rho(n,\tau,R,R')$ with $\rho(\tau)$ calculated by standard cross-correlation analysis, the standard cross-correlation coefficient of wind speed and API, $\rho(\tau)$, is calculated. The standard cross-correlation coefficient formula is shown in Eq. (6) [1].

$$\rho(\tau) = \frac{\sum_{i=1}^{N-\tau}(x(i)-\bar{x})(y(i+\tau)-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x(i)-\bar{x})*(x(i)-\bar{x})}\sqrt{\sum_{i=1}^{N}(y(i)-\bar{y})*(y(i)-\bar{y})}} \tag{6}$$

where τ denotes time lag, $\bar{x}$ and $\bar{y}$ are the sample means.

Fig. 6 shows the standard cross-correlation coefficient of wind speed and API. On the first day, $\rho(\tau)$ is negative, and is also minimal. On the second day, $\rho(\tau)$ becomes larger. Afterwards, $\rho(\tau)$ fluctuates around certain values. This result shows that $\rho(\tau)$ does not return to $\rho(0)$. However, our results show that $\rho(n,\tau,R,R')$ could return to $\rho(n,0,R,R')$. Furthermore, according to meteorological records data in Nanjing, a strong wind can affect API for several days. Our results using the proposed method are thus judged reliable, reasonable and in keeping with real situations.

4. Conclusions

We proposed a time-lagged DCCA cross-correlation coefficient to quantify the level of time-lagged cross-correlation between two nonstationary time series at time scales. The implementation of the new method is based on a DCCA cross-correlation coefficient method. The time-lagged DCCA cross-correlation coefficient follows steps similar to the DCCA cross-correlation coefficient proposed by Zebende et al. The detrended covariances are computed as a lagged convolution. The application of the method through an example in meteorology shows that the time lags vary with time scales, and the time-lagged DCCA cross-correlation coefficient is different from the standard cross-correlation coefficient. Our findings lead us to think that the method might also be suitable to physiology, geology, ethnology, economics and other fields. In future work, we will apply the DCCA time-lagged cross correlation analytical method in diverse fields.

Conflict of interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

The authors are grateful for financial support from the National Natural Science Foundation of China (Nos. 41271189, 40971231, 40971105), and from a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

[1] C. Chatfield, The Analysis of Time Series: An Introduction, sixth ed., Chapman & Hall/CRC Texts in Statistical Science, 1996.
[2] P. Xu, B.S. Wang, W. Zhang, J.M. Lin, Y. Chen, Estimating seismic attenuation using cross correlation function, Chin. J. Geophys. 49 (6) (2006) 1738–1744 (in Chinese with English abstract).
[3] R.V.R. Goncalves, J. Zullo, L.A.S. Romani, C.R. Nascimento, A.J.M. Traina, Analysis of NDVI time series using cross correlation and forecasting methods for monitoring sugarcane fields in Brazil, Int. J. Remote Sens. 33 (15) (2012) 4653–4672.

[4] H. Li, S.H. Futch, J.P. Syvertsen, Cross-correlation patterns of air and soil temperatures, rainfall and diaprepes abbreviatus root weevil in citrus, Pest Manag. Sci. 63 (11) (2007) 1116–1138.
[5] D.B. Yoon, J.H. Park, S.H. Shin, Improvement of cross-correlation technique for leak detection of a buried pipe in a tonal noisy environment, Nucl. Eng. Technol. 44 (8) (2012) 977–984.
[6] D.C. Mei, L.C. Du, C.J. Wang, The effects of time delay on stochastic resonance in a bistable system with correlated noises, J. Stat. Phys. 137 (2009) 625–638.
[7] L.C. Du, D.C. Mei, Stochastic resonance in a bistable system with global delay and two noises, Eur. Phys. J. B 85 (2012) 75, http://dx.doi.org/10.1140/epjb/e2012-21053-0.
[8] B. Podobnik, H.E. Stanley, Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series, Phys. Rev. Lett. 100 (2008) 084102.
[9] C.K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Mosaic organization of DNA nucleotides, Phys. Rev. E 49 (2) (1994) 1685–1689.
[10] B. Podobnik, I. Grosse, D. Horvatić, S. Ilic, P.Ch. Ivanov, H.E. Stanley, Quantifying cross correlations using local and global detrending approaches, Eur. Phys. J. B 71 (2) (2009) 243–250.
[11] D. Horvatic, H.E. Stanley, B. Podobnik, Detrended cross-correlation analysis for non-stationary time series with periodic trends, Europhys. Lett. 94 (1) (2011) 18007.
[12] B. Podobnik, Z.Q. Jiang, W.X. Zhou, H.E. Stanley, Statistical tests for power-law cross correlated processes, Phys. Rev. E 84 (2011) 066118.
[13] R. Balocchi, M. Varanini, A. Macerate, Quantifying different degrees of coupling in detrended cross-correlation analysis, Europhys. Lett. 101 (2013) 20011.
[14] Duncan A.J. Blythe, A rigorous and efficient asymptotic test for power-law cross-correlation, http://arxiv-web3.library.cornell.edu/abs/1309.
[15] L. Kristoufek, Measuring correlations between non-stationary series with DCCA coefficient, Physica A 402 (2014) 291–298.
[16] L. Kristoufek, Detrending moving-average cross-correlation coefficient: measuring cross correlations between non-stationary series, Physica A 406 (2014) 169–175.
[17] G.F. Zebende, A. Machado Filho, Cross-correlation between time series of vehicles and passengers, Physica A 388 (23) (2009) 4863–4866, http://dx.doi.org/10.1016/j.physa.2009.07.046.
[18] G.F. Zebende, P.A. da Silva, A. Machado Filho, Study of cross-correlation in a self-affine time series of taxi accidents, Physica A 390 (9) (2011) 1677–1683, http://dx.doi.org/10.1016/j.physa.2010.12.038.
[19] C.H. Shen, C.L. Li, Y.L. Si, A detrended cross-correlation analysis of meteorological and API data in Nanjing, China, Physica A 419 (2015) 417–428, http://dx.doi.org/10.1016/j.physa.2014.10.058.
[20] Y. Wang, Y. Wei, C. Wu, Cross-correlations between Chinese A-share and B-share markets, Physica A 389 (2010) 5468–5478.
[21] D.D. Kang, D.I. Lee, B.H. Kwon, K. Kim, J.K. Park, Features of the detrended cross-correlation analysis in the time series between absorbable particulate matter and meteorological factors, J. Korean Phys. Soc. 63 (1) (2013) 10–17.
[22] A. Király, I.M. Jánosi, Detrended fluctuation analysis of daily temperature records: geographic dependence over Australia, Meteorol. Atmos. Phys. 88 (2005) 119–128.
[23] P. Oświecimka, S. Drożdż, M. Forczek, S. Jadach, J. Kwapień, Detrended cross-correlation analysis consistently extended to multifractality, Phys. Rev. E 89 (2) (2014), http://dx.doi.org/10.1103/PhysRevE.89.023305.
[24] R.T. Vassoler, G.F. Zebende, DCCA cross-correlation coefficient apply in time series of air temperature and air relative humidity, Physica A 391 (7) (2012) 2438–2443.
[25] J. Alvarez-Ramirez, E. Rodriguez, J.C. Echeverria, Using detrended fluctuation analysis for lagged correlation analysis of nonstationary signals, Phys. Rev. E 79 (2009) 057202.
[26] A.J. Lin, P.J. Shang, X.J. Zhao, The cross-correlations of stock markets based on DCCA and time-delay DCCA, Nonlinear Dyn. 67 (1) (2012) 425–435, http://dx.doi.org/10.1007/s11071-011-9991-8.
[27] Y.F. Sang, D. Wang, J.C. Wu, Q.P. Zhu, Wavelet cross-correlation method for hydrologic time series analysis, J. Hydraul. Eng. 41 (11) (2010) 1272–1279 (in Chinese with English abstract).
[28] E.B.S. Marinho, A.M.Y.R. Sousa, R.F.S. Andrade, Using detrended cross-correlation analysis in geophysical data, Physica A 392 (9) (2013) 2195–2201, http://dx.doi.org/10.1016/j.physa.2012.12.038.
[29] G.F. Zebende, DCCA cross-correlation coefficient: quantifying level of cross-correlation, Physica A 390 (4) (2011) 614–618.
[30] N.M. Yuan, Z.T. Fu, J.Y. Mao, Different multi-fractal behaviors of diurnal temperature range over the north and the south of China, Theor. Appl. Climatol. 112 (3–4) (2013) 673–682.
[31] F.A.B.F. De Moura, M.L. Lyra, Delocalization in the 1D Anderson model with long-range correlated disorder, Phys. Rev. Lett. 18 (17) (1998) 3735–3738.