# SCIENTIFIC REP**O**RTS

**OPEN**

# Detecting PM2.5's Correlations between Neighboring Cities Using a Time-Lagged Cross-Correlation Coefficient

Fang Wang[1,2], Lin Wang[2] & Yuming Chen[3]

In order to investigate the time-dependent cross-correlations of fine particulate (PM2.5) series among neighboring cities in Northern China, in this paper, we propose a new cross-correlation coefficient, the time-lagged $q$-$L$ dependent height crosscorrelation coefficient (denoted by $p_q(\tau, L)$), which incorporates the time-lag factor and the fluctuation amplitude information into the analogous height cross-correlation analysis coefficient. Numerical tests are performed to illustrate that the newly proposed coefficient $\rho_q(\tau, L)$ can be used to detect cross-correlations between two series with time lags and to identify different range of fluctuations at which two series possess cross-correlations. Applying the new coefficient to analyze the time-dependent cross-correlations of PM2.5 series between Beijing and the three neighboring cities of Tianjin, Zhangjiakou, and Baoding, we find that time lags between the PM2.5 series with larger fluctuations are longer than those between PM2.5 series withsmaller fluctuations. Our analysis also shows that cross-correlations between the PM2.5 series of two neighboring cities are significant and the time lags between two PM2.5 series of neighboring cities are significantly non-zero. These findings providenew scientific support on the view that air pollution in neighboring cities can affect one another not simultaneously but with a time lag.

As a cost to rapid development of economics and progress of technology after World War II, the environmental pollution, particularly air pollution produced by industry exhaust, smoke dust, and coal combustion, has become a serious world-wide problem[1]. For instance, the Great Smog of 1952 took away 4,000 lives and led to more than 100,000 people affected by respiratory diseases, which has been regarded as the most serious environmental disaster in British history[2]. Urban air pollution directly and indirectly caused by rapid urbanization and industrialization in China in the last three decades has become more and more severe. Despite the decrease in "traditional pollutants" (e.g. NO2, SO2), fine particulate matter with a diameter $\leq 2.5\,\mu m$ (PM2.5) has become a major air pollutant that threatens human's health, including morbidity and mortality, and decreases meteorological visibility[3–13]. Naturally, as a hot societal issue, study on air pollution has attracted enormous attention from researchers in economical modeling[3–8], as well as statistical modeling[9–13]. With the advancement of modern statistic methods, it is of great importance to assess smog's trend and propagation characteristics from statistical point of view. In this context, much of existing work in the literature has been focused on studying the correlations among various air pollution indicators such as, air pollution index (API), air quality index (AQI), PM2.5 concentrations, PM10 (particulate with a diameter $\leq 10\,\mu m$) concentrations, temperature, air pressure, rainfall, relative humidity, and wind speed[9–13]. It is a common sense that smog produced at one source place can spread to its surrounding areas[6, 7]. Hence, it is more practical to explore cross-correlations of the above air pollution indicators among neighbouring cities as this helps the authority further assess the causes of local smog.

To study cross-correlations between two series, several statistical methods have been proposed in the literature. The methods include the widely used detrended cross-correlation analysis (DCCA)[14, 15], and its multifractal version (MF-DXA)[16], detrended moving average cross-correlation analysis (DMXA)[17] and MF-DMXA[18], multifractal cross-correlation analysis (MFCCA)[19], multifractal height cross-correlation analysis (MF-HXA)[20] and its

[1]College of Science/Agricultural Mathematical Modeling and Data Processing Center, Hunan Agricultural University, Changsha, P. R. China. [2]Department of Mathematics and Statistics, University of New Brunswick, Fredericton, NB, E3B 5A3, Canada. [3]Department of Mathematics, Wilfrid Laurier University, Waterloo, ON, N2L 3C5, Canada. Correspondence and requests for materials should be addressed to F.W. (email: popwang619@163.com)
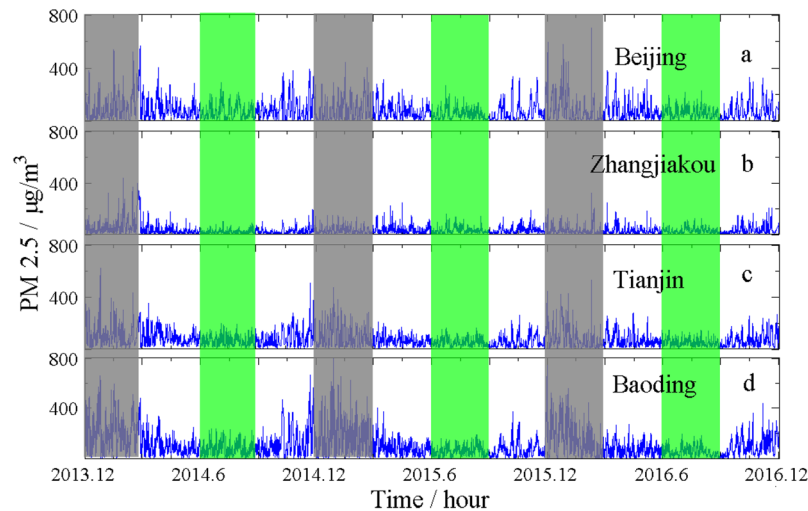
**Figure 1.** PM2.5 series of the four chosen cities in Northern China during the period of Dec. 1, 2013 and Nov. 30, 2016. The gray area refers to the winter season and the green area refers to the summer season of every year.

extension the analogous MF-HXA (AMF-HXA)[21]. Later, Podobnik et al.[22, 23] developed some statistical tests to test the presence of power-law cross-correlation between series. Moreover, DCCA with partial-correlation technique has been shown to be a powerful tool to reveal the intrinsic correlations between the two studied series interfered by another identical series[24–26]. Associated with the above mentioned methods, several cross-correlation coefficients were proposed naturally[21, 25–33]. However, the majority of the existing work focused mainly on the synchronous time series while only limited work considered the cross-correlations between two series with time lags[34–37]. In addition, the limited existing work in[34–36], was established on the basis of the detrended fluctuation analysis (DFA)[38, 39], and DCCA[14, 15]. The so-called time-lagged DCCA cross-correlation coefficient was proposed by Shen et al.[36], which can be used to detect time-dependent cross-correlations between the API and wind speed but fails to recognize the range of fluctuation amplitudes that contributed to those cross-correlations. The $q$-dependent detrended cross-correlation coefficient proposed by Kwapień et al.[33] has the flexibility to successfully detect range of cross-correlated fluctuations, but it is not applicable to series that are correlated with time lags. In this work, inspired by the paper[33], we propose a time-lagged cross-correlation coefficient that can be used to identify the range of detrended fluctuation amplitudes for signals correlated with time lags.

Another motivation of this work is to explore if there are cross-correlations for the PM2.5 series between neighbouring cities. In this study, we focus our interest on the interaction of PM2.5 series of four cities in Beijing-Tianjin-Hebei (known as the Jing-Jin-Ji area) of China, namely, Beijing (39.93°N,116.39°E), Zhangjiakou(40.81°N,114.89°E), Tianjin (39.14°N,117.21°E), and Baoding (38.89°N,115.49°E). Zhangjiakou, Tianjin, and Baoding are surrounding Beijing and are about 160 km, 115 km, and 140 km away from Beijing, respectively. All four cities have been greatly affected by heavy smog in recent years. Real-time data of PM2.5 series of these four cities from December 1, 2013 to November 30, 2016 are chosen for our consideration. The data are taken from the Ministry of Environmental Protection of the People's Republic of China (http://datacenter.mep.gov.cn) and are shown in Fig. 1.

Figure 1 clearly demonstrates that there is a huge difference in the PM2.5 concentrations between the summer and the winter. We recorded hourly data of severe pollution (PM2.5 concentration $> 150\,\mu g/m^3$) of the four cities in the whole studied time period. There are 1427, 272, 1439, and 3338 hours in the three winters, respectively, while there are only 303, 2, 58, and 279 hours in the three summers, respectively. We also note in Fig. 1 that PM2.5 concentrations in the four cities follow similar trend. This indicates that cross-correlations among these series may exist.

To fully detect and quantify possible cross-correlations among the PM2.5 series mentioned above and uncover the potential time lags embedded in those cross-correlations, in this work, we propose a new cross-correlation coefficient by incorporating the time-lag factor and fluctuation information into the latest analogous height cross-correlation analysis (AHXA) coefficient $\rho(L)$ introduced by Wang et al.[21]. The new coefficient, denoted by $\rho_q(\tau, L)$, will be referred to as the time-lagged $q$-$L$ dependent height cross-correlation coefficient. Two numerical tests will be performed to assess the performance of the proposed $\rho_q(\tau, L)$. The first test is used to illustrate that the new coefficient $\rho_q(\tau, L)$ can accurately detect the cross-correlation between two series with certain time lags, while the second test is to show that the new coefficient is capable of distinguishing correlations between small and large fluctuation amplitudes. Applying the new coefficient to analyze the PM2.5 series recorded in Fig. 1 helps us understand the spreading characteristics of PM2.5 among neighboring cities in Northern China.

## Results and Discussions

### Performance of $\rho_q(\tau,L)$.
The time-lagged $q$-$L$ dependent AHXA coefficient $\rho_q(\tau,L)$ incorporates both time-lagged covariance function and the $q$-th order (co-)variance function. The former allows us to detect correlations between two series asynchronously while the latter makes the coefficient flexible to detect the range of
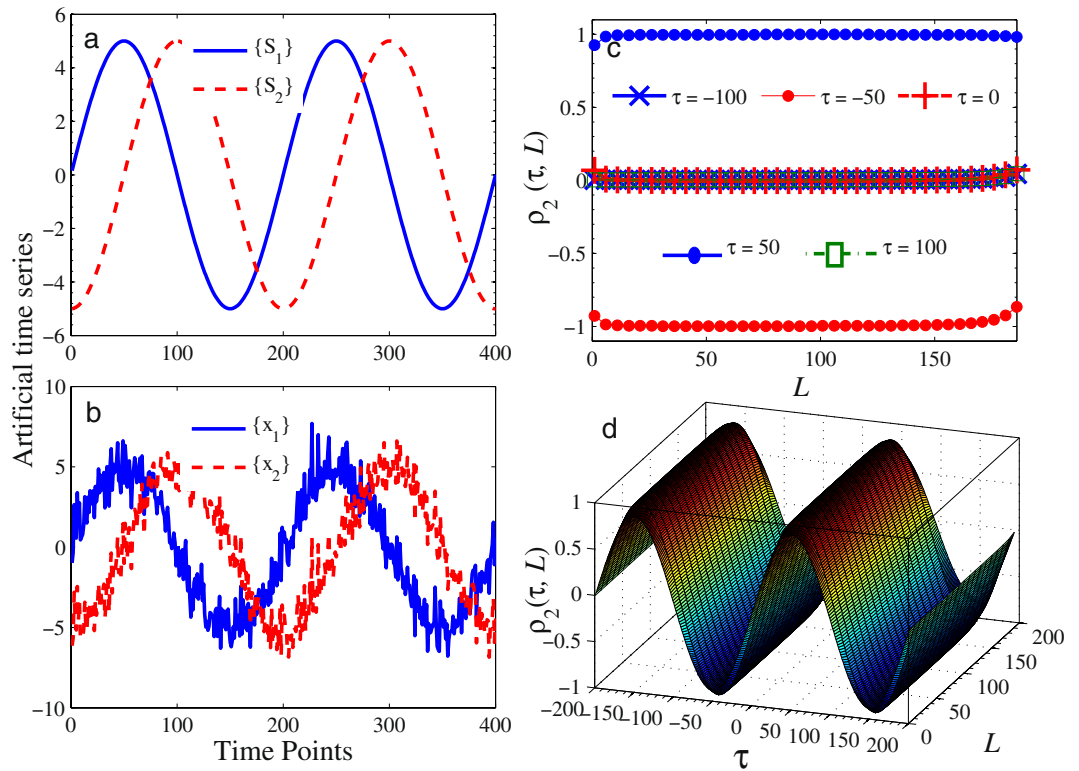
**Figure 2.** Numerical results of Test 1. (**a**) Sinusoidal signals $S_1$ and $S_2$; (**b**) Time series $\{x_1\}$ and $\{x_2\}$ generated by two sinusoidal signals with Gaussian noises; (**c**) Calculated time-lagged cross-correlation coefficients $\rho_2(\tau,L)$; (**d**) Calculated $\rho_2(\tau,L)$ surface.

cross-correlated fluctuations. In this section, we perform two numerical tests to illustrate the advantages and applicability of the proposed coefficient $\rho_q(\tau,L)$.

In our first numerical test, we generate a pair of artificial time series $x_1(t) = S_1(t) + \varepsilon(t)$ and $x_2(t) = S_2(t) + \varepsilon(t)$ for $t = 1, 2, \ldots, 10{,}000$, where $\{\varepsilon(t)\}$ is a white noise; $\{S_1\}$ and $\{S_2\}$ are two sinusoidal signals with the same period $p = 200$ and the same amplitude but different phases (they differ by one quarter of the period, as shown in Fig. 2a). Thus, $\{x_1(t)\}$ and $\{x_2(t)\}$ are two periodic sequences having the same noises (see Fig. 2b). Theoretically, the two series $\{x_1(t)\}$ and $\{x_2(t)\}$ should be perfectly correlated. But because of the difference in phases, the corresponding Pearson correlation coefficient is calculated to be 0.075 (very close to 0), which fails to detect any correlation between the two series $\{x_1(t)\}$ and $\{x_2(t)\}$. However, we can calculate the time-lagged cross-correlation coefficients by setting $\tau$ as $-100, -50, 0, 50$ and $100$, respectively. The obtained results with $q = 2$ are shown in Fig. 2c. As can be seen in Fig. 2c, when $\tau$ takes values $-100$, $0$ and $100$, the resulting coefficients are almost 0 and when $\tau = \frac{p}{4} = 50$, the two series exhibit perfect positive correlations. Perfect negative correlations are observed when $\tau = -\frac{p}{4} = -50$. Figure 2d depicts the proposed time-dependent correlation coefficients (when $q = 2$) as the time-delay $\tau$ varies from $-200$ to $200$.

Our second numerical test aims to show that the new coefficient $\rho_q(\tau,L)$ is capable of identifying different correlations at different fluctuation amplitudes. To this end, binomial multifractal series (BMFs), which can produce more different degrees of power-law for different fluctuation amplitudes, is employed in our test. Two series of $\{y_1\}$ and $\{y_2\}$ are constructed as

$$y_i = p_i^{n-n[k-1]}(1 - p_i)^{n[k-1]}, \quad k = 1, 2, \ldots, 2^n, \quad i = 1, 2. \tag{1}$$

Here parameters $p_i \in (0, 0.5)$ $(i = 1, 2)$, $n[k]$ denotes the number of digit 1 in the binary representation of the index $k$. In this test, we set $p_1 = 0.3$ and $p_2 = 0.4$. Each series is of length $2^{15}$. To illustrate that the proposed $\rho_q(\tau,L)$ can provide more information about cross-correlations at different fluctuation amplitudes, the following two experiments are carried out.

In Experiment I, we first locate the data points with small amplitudes satisfying $y_1 < 9 \times 10^{-7}$ and $y_2 < 10^{-4}$, we then replace those data points by two sinusoidal signals $G_1 = 0.0004 \sin\left(\frac{2\pi x}{200}\right)$ and $G_2 = 0.00004 \sin\left(\frac{2\pi x}{200} - \frac{\pi}{2}\right)$ accordingly. In this way, two new series $\{y'_1\}$ and $\{y'_2\}$ are generated. Since the data in the original series with small fluctuations are replaced by the two sinusoidal signals, the correlation between the new series $\{y'_1\}$ and $\{y'_2\}$ at small fluctuations is dependent on $\{G_1\}$ and $\{G_2\}$ while the correlation preserves among large fluctuations. Numerical results are presented in Fig. 3a and b. In Fig. 3a, $q = 0.5$ refers to small fluctuation amplitude. The calculated $\rho_{0.5}(\tau,L)$ surface exhibits periodicity with period of 200, the same as the two sinusoidal signals. The
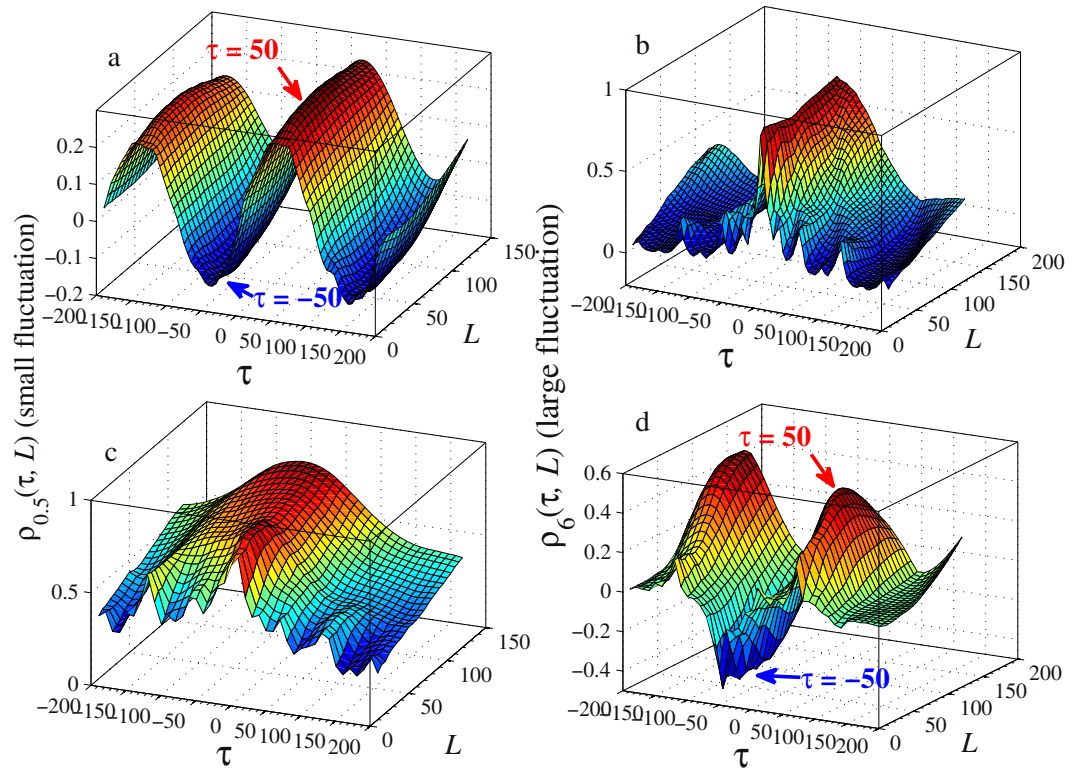
**Figure 3.** Calculated coefficients $\rho_q(\tau,L)$ for two BMFs with $p_1 = 0.3$ and $p_2 = 0.4$. (**a** and **b**) Show the $\rho_q(\tau,L)$ surface when $q = 0.5$ and 6, which describe the cross-correlation between the compound series with dominated small fluctuations and large fluctuations, respectively. (**c** and **d**) Are the opposite of (**a** and **b**), respectively.

maximum of positive correlation coefficient reaches at $\tau = 50$, which is exactly the phase difference between $\{G_1\}$ and $\{G_2\}$. On the contrary, the maximum negative correlation coefficient appears at $\tau = -50$, which means the two series are exactly half a period apart. Figure 3b illustrates the results for large fluctuation amplitudes with $q = 6$. As expected, the shape of the $\rho_6(\tau,L)$ surface is not affected by the periodic signals, thus, the maximum correlation coefficient is obtained when there is no time lag between the $\{y'_1\}$ and $\{y'_2\}$, i.e., $\tau = 0$. In Experiment II, We remove from the series $\{y_1\}$ all values which are larger than $5 \times 10^{-4}$ and from the series $\{y_2\}$ those values larger than $10^{-4}$, and replace corresponding values by $G_3 = 0.004 \sin\left(\frac{2\pi x}{200}\right)$ and $G_4 = 0.0004 \sin\left(\frac{2\pi x}{200} - \frac{\pi}{2}\right)$, respectively. Consequently, for the obtained new series, the cross-correlation at relative large fluctuations depends on the series $\{G_3\}$ and $\{G_4\}$ but the correlation preserves among the small fluctuations. Associated numerical results are presented in Fig. 3c and d. As can be seen from Fig. 3d, the maximum positive correlation coefficient appears at $\tau = 50$ and $\tau = -150$ and the maximum negative correlation coefficient occurs when $\tau = -50$. These two experiments clearly illustrate that the proposed time-lagged $q$-$L$ dependent AHXA coefficient $\rho_q(\tau,L)$ can capture not only fluctuation information but also potential time-delay in cross-correlations.

### Application of $\rho_q(\tau,L)$ to PM 2.5 series.

In this section, we apply our time-lagged cross-correlation coefficient $\rho_q(\tau,L)$ to the PM2.5 series mentioned in the introduction to illustrate the applicability of $\rho_q(\tau,L)$. Three factors, namely, the season factor in Northern China, the fluctuation amplitudes of PM2.5, and the time interval used to calculate $\rho_q(\tau,L)$, are considered to explore the time-lagged correlations. For the data reported in Fig. 1, four seasons are classified as winter (December, January, and February), spring (March, April, and May), summer (June, July, and August), and fall (September, October, and November). Two representative fluctuation amplitudes, i.e., small and large fluctuations, characterized by $q = 0.5$ and $q = 6$, respectively, are set for our consideration. The time interval $L$ is set to range from 24 hours to 720 hours with the step size 24. This means we only focus on the correlation of each series at day 1, day 2, up to day 30.

The calculated $\rho_q(\tau,L)$ surfaces of Beijing vs. Zhangjiakou for the four seasons with the two selected $q$'s are sketched in Figs 4 and 5. To help determine if the cross-correlation between PM2.5 series in the two cities is significant, threshold surfaces $\rho_q^c(\tau, L)$ (colored in gray in those sub-figures) are also constructed in Figs 4 and 5 for each season and both $q$'s. These threshold surfaces are numerically constructed via the Monte-Carlo simulations proposed by Podobnik *et al.*[23], where we used Gaussian series and repeated the calculations for 10,000 times at the 95% significance level. Note that $\rho_q(\tau, L) > \rho_q^c(\tau, L)$ means that the cross-correlation is significant at the point $(\tau,L)$ and vice versa. For the small fluctuation amplitudes of original series (refers to $q = 0.5$, Fig. 4), except for the spring, the cross-correlations vary from small time intervals to large time intervals in all other three seasons. More specifically, the $\rho_{0.5}(\tau,L)$ surfaces are above the threshold surfaces $\rho_q^c(\tau, L)$ for $L < 288\,h$ (12d) $\sim 408\,h$ (17d)
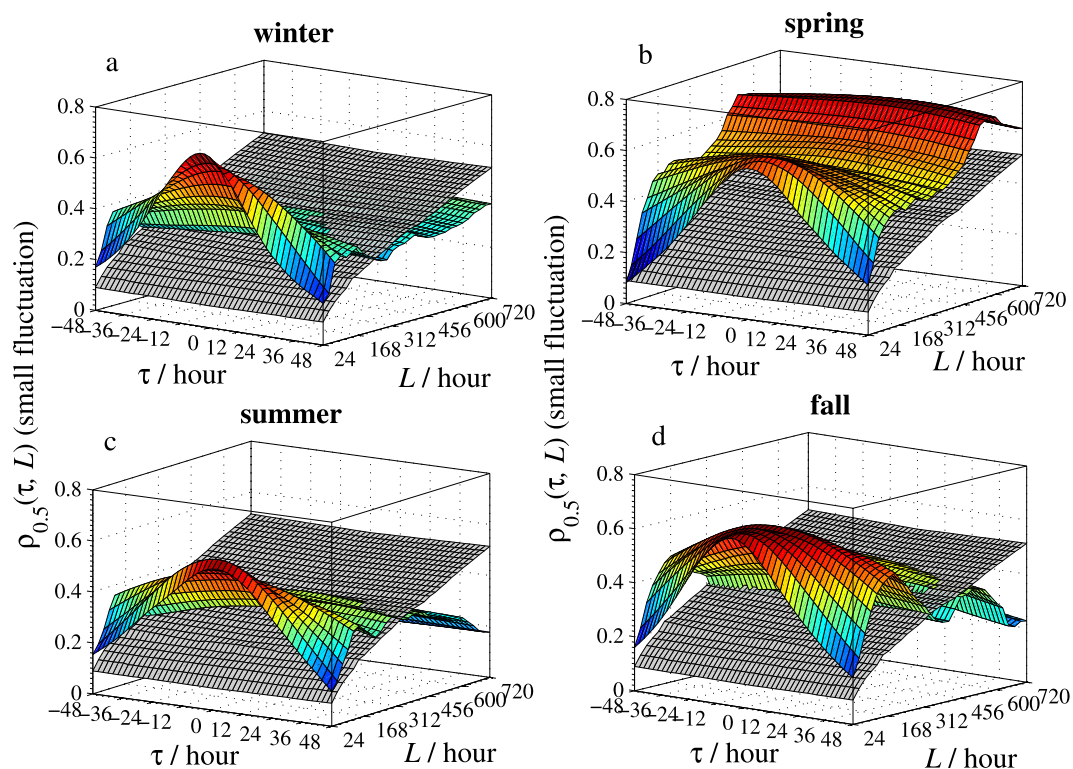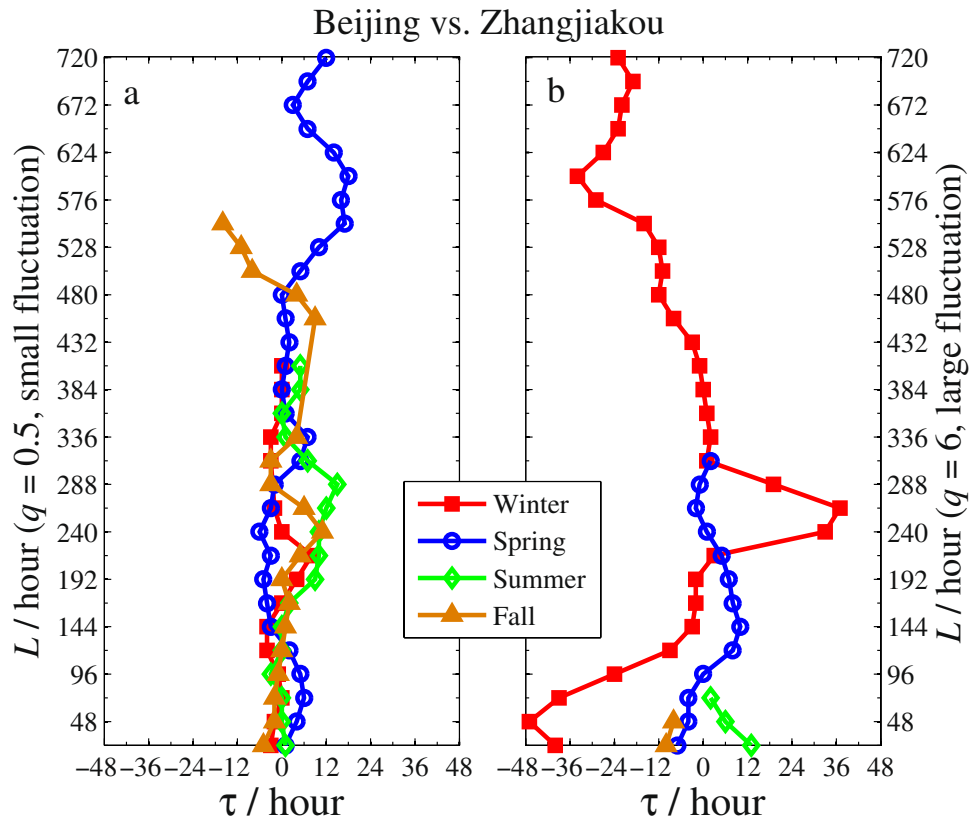
**Figure 4.** The time-lagged cross-correlation coefficient $\rho_q(\tau,L)$ with $q = 0.5$ of PM2.5 series between Beijing and Zhangjiakou. The threshold surfaces are colored in gray.
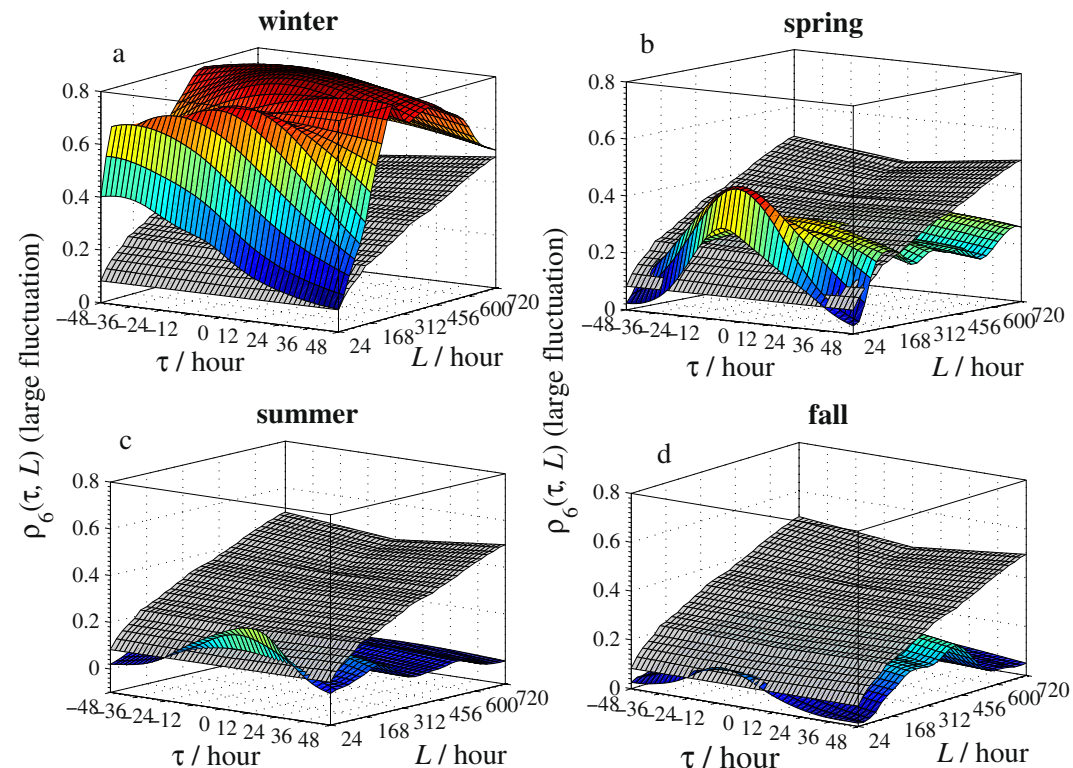


**Figure 5.** The time-lagged cross-correlation coefficient $\rho_q(\tau,L)$ with $q = 6$ of PM2.5 series between Beijing and Zhangjiakou. The threshold surfaces are colored in gray.
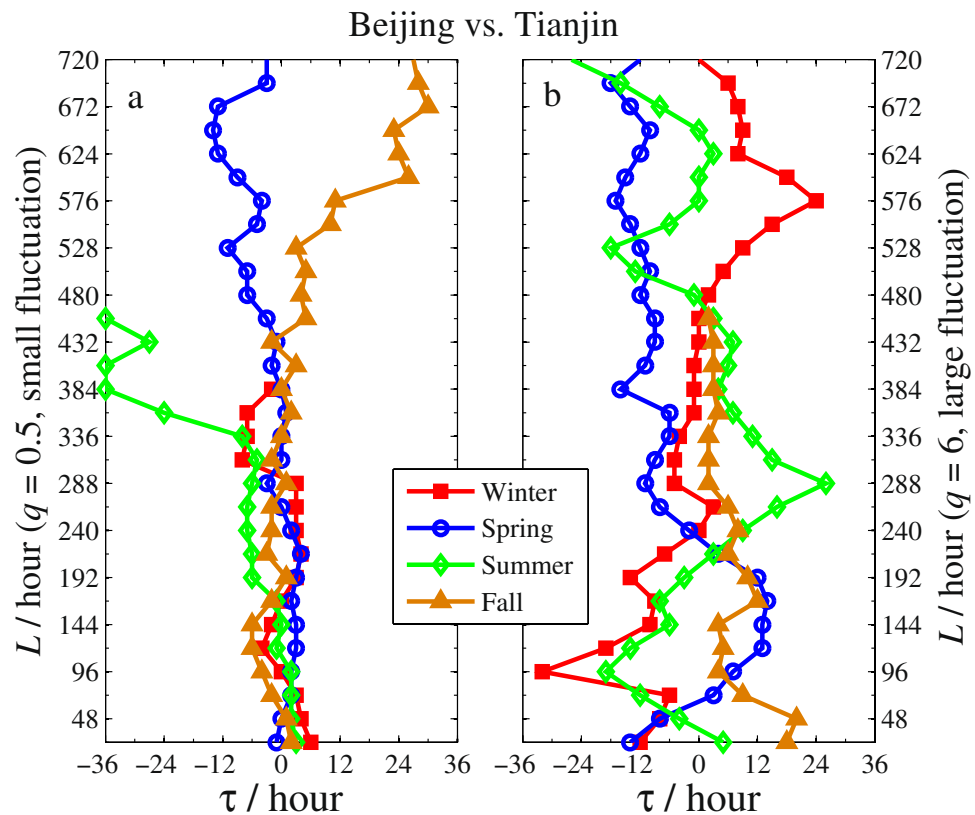
**Figure 6.** The $\tau$-$L$ curves of Beijing vs. Zhangjiakou in four seasons. (**a**) Is for the PM2.5 with small fluctuations ($q = 0.5$) and (**b**) is for the PM2.5 with large fluctuations ($q = 6$).

approximately. This implies that the long-range correlation at small fluctuation amplitudes lasts up to 12~17 days. This interesting finding explains that the PM2.5 of a relatively stable range in Beijing and Zhangjiakou can affect each other and the interaction will continue for almost 17 days in winter, summer, and fall. Meanwhile, in spring, this interaction lasts longer. However, as seen from Fig. 5, for the cross-correlation of PM2.5 with larger fluctuations (refers to $q = 6$), the spreading is most significant in winter but is almost absent in summer and fall. This indicates that the smog spread between Beijing and Zhangjiakou is most serious in winter and is least serious in summer and fall. In addition, the interaction of heavy smog between Beijing and Zhangjiakou lasts about 312 $h$ (13d) in spring, which is much less than the interaction time in winter. Similar patterns are also observed for the PM2.5 series of Beijing vs. Tianjin and Beijing vs. Baoding: the long-range correlation at small fluctuation amplitudes lasts up to 15~19 days in the winter and summer between Beijing and Tianjin while that lasts more days in spring and fall. Further, the long-range correlation at large fluctuation amplitudes can be found in all studied days in all seasons except for fall, while the correlation lasts for only 19 days in fall. For Beijing vs. Baoding: the long-range correlation at small fluctuation amplitudes lasts up to 22~29 days in the winter and fall, and all observed days in spring and summer. Meanwhile, the long-range correlation at large fluctuation amplitudes lasts longer in spring and summer than in winter and fall.

Besides the difference due to the season factor and the fluctuation amplitudes acquired from the eight sub-figures, one can also obtain information about the time lags of PM2.5 series between two neighbouring cities. For illustration, we sketch the $\tau$-$L$ curves in Figs 6–8 for Beijing vs. Zhangjiakou, Beijing vs. Tianjin, and Beijing vs. Baoding, respectively. In each figure, $\tau$ is the time lag corresponding to the maximum of $\rho_q(\tau,L)$ for each $L$. When we record $\tau$ in calculating $\rho_q(\tau,L)$, we use $x$ to stand for PM2.5 series of Beijing and $y$ for that of the other three cities. In this setting, $\tau > 0$ indicates that the largest $\rho_q(\tau,L)$ is obtained at the time point where Zhangjiakou's (or Tianjin's, or Baoding's) PM2.5 lags are behind Beijing's by $\tau$ hours. This implies that Beijing's air pollution will affect the other three cities. Conversely, $\tau < 0$ means that Beijing's air pollution will be affected by the other three cities. Here we only report these $\tau$'s obtained by the largest cross-correlation that is significant ($\rho_q(\tau, L) \geq \rho_q^c(\tau, L)$) in each figure. In addition, to eliminate the influence of the time intervals $L$, we average the $\rho_q(\tau,L)$ over every $L$ and record the maximum together with its corresponding time lag $\tau$. The results are listed in Table 1.

Based on Figs 6–8 and Table 1, we arrive at the following conclusions.

1. The time lags obtained by $\rho_q(\tau,L)$ with larger fluctuation amplitudes (Figs 6b–8b) are longer than those with smaller fluctuation amplitudes (Figs 6a–8a) and the range of the former is also much wider. This indicates that severe smog produced by PM2.5 in one city would slowly affect the air quality in its neighboring cities and the influence would last longer. For the PM2.5 of small fluctuation amplitudes, the
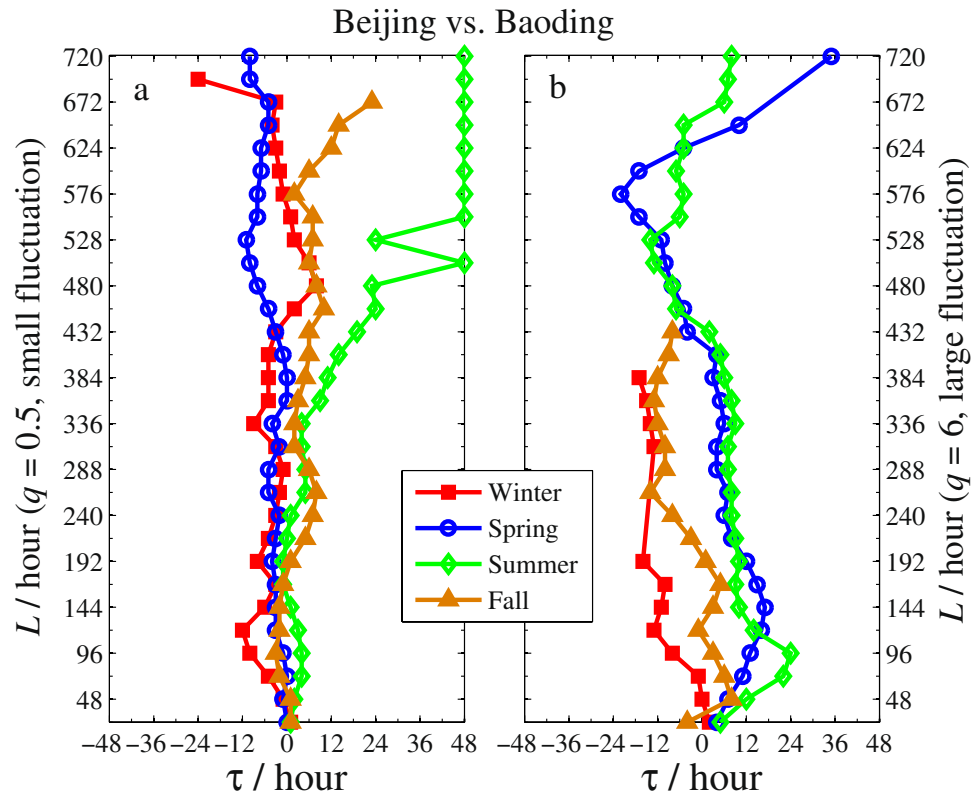
**Figure 7.** The $\tau$-$L$ curves of Beijing vs. Tianjin in four seasons (The legends are the same as in Fig. 6).

| Season | $q$ | Beijing vs. Zhangjiakou | | Beijing vs. Tianjin | | Beijing vs. Baoding | |
|---|---|---|---|---|---|---|---|
| | | $\rho$ | $\tau$ | $\rho$ | $\tau$ | $\rho$ | $\tau$ |
| winter | 0.5 | 0.3898* | −2 | 0.4533* | 2 | 0.5373* | −3 |
| | 6 | 0.7181* | −10 | 0.4822* | −7 | 0.3184* | −6 |
| spring | 0.5 | 0.6085* | 2 | 0.7211* | 0 | 0.7586* | −2 |
| | 6 | 0.2925 | −1 | 0.7456* | −4 | 0.5915* | 7 |
| summer | 0.5 | 0.3576* | 2 | 0.4725* | −1 | 0.6365* | 5 |
| | 6 | 0.0795 | 8 | 0.7511* | −1 | 0.6590* | 8 |
| fall | 0.5 | 0.4533* | −3 | 0.5913* | 1 | 0.6216* | 2 |
| | 6 | 0.1583 | −3 | 0.4353* | 9 | 0.3676* | −1 |

**Table 1.** The maximum of $\rho_q(\tau,L)$ averaged over the intervals $L$ and the corresponding time lag $\tau$ of Beijing vs. the three neighboring cities in four seasons. *Denotes 95% confidence level.

cross-correlations between Beijing and the other three neighboring cities are significant for all four seasons and smog in those cities influences one another since $\tau$ fluctuates around 0. This explains that the air quality in one city of Northern China cannot be irrelevant to that of its neighbouring cities.

2. As shown in Fig. 6b and Table 1, most time lags are less than 0 in Beijing's winters. This indicates that the air in Beijing was greatly influenced by the pollution from Zhangjiakou. In fact, due to the domination of the northwest wind from Siberia in Zhangjiakou's winter, the pollution of Zhangjiakou (located in the northwest of Beijing) can easily spread to Beijing. Furthermore, there are about 50% of time during which the wind force is above level IV (the average wind speed is greater than $30\,km/h$) in Zhangjiakou's winter, which drifts the polluted air to Beijing in less than 6 hours. This is consistent with the obtained lags for the significant cross-correlations between Zhangjiakou and Beijing. In addition, the cross-correlations of PM2.5 between Beijing and Zhangjiakou are insignificant in the other three seasons.

3. Another source of heavy air pollution in Beijing's winter is Baoding since the associated time lag $\tau < 0$ (see Fig. 8b). Nevertheless, compared to winter and fall, the smog spreads in spring and summer mainly from Beijing to Baoding since $\tau > 0$ during most time intervals and the spread is quite slow. Indeed, as seen from Table 1, the time lag corresponding to the maximum of mean $\rho_q(\tau,L)$ is 7–8 hours in spring and summer, and this number is 6 and 1 in winter and fall, respectively.

**Figure 8.** The $\tau$-$L$ curves of Beijing vs. Baoding in four seasons (The legends are the same as in Fig. 6).

4. Compared to Zhangjiakou and Baoding, heavy air pollution in Beijing and Tianjin has a greater impact on each other. This is due to the closer distance between the two cities. In the most severe smog of winter, it is clearly seen from Fig. 7b that the time lag $\tau < 0$ when the interval $L$ starts at 24 h (1d) and lasts up to 360 h (15d) while $\tau > 0$ when $L$ is beyond 15 days. This shows that the heavy smog in a winter of Tianjin would invade Beijing at a short time scale but Tianjin would be invaded by Beijing's heavy smog in a longer time scale. Overall, as shown in Table 1, except for the fall, the severe smog in the other three seasons spreads mainly from Tianjin to Beijing.

## Conclusions

In this work, by incorporating the time-lagged factor and the fluctuation amplitude information into the cross-correlation coefficient $\rho(L)$ based on analogous height cross-correlation analysis, we have proposed the time-lagged $q$-$L$ dependent AHXA coefficient $\rho_q(\tau,L)$. This newly proposed cross-correlation coefficient can be used to quantify the range of cross-correlated fluctuations with more flexibility for two time series with potential time lags. It has been shown via two tests that this new coefficient (i) can accurately capture time-lagged cross-correlations hidden between two time series; (ii) can precisely locate the time points at which the studied two series are perfectly correlated; (iii) can successfully quantify different ranges of cross-correlations between two series with different fluctuation amplitudes.

We applied this new cross-correlation coefficient to study the PM2.5 series of four neighboring cities in Northern China and found that different time lags exist in the PM2.5 series of Beijing and the other three neighboring cities at different seasons and different fluctuation amplitudes. These findings provide very useful insights on understanding the pattern of smog's propagation in Northern China and on accurately predicting future air pollution conditions. The idea can also be applied to many other fields for investigating time-dependent correlations. One particular field might be the new field of network physiology[40]. Recently, the concept of time delay stability (TDS) among key organ systems has been used to quantify physiologic interaction and transitions across distinct functional states[41, 42]. It has been found that each physiologic state is characterized by a distinct network structure with different relative contribution from individual organ systems to the global network dynamics[42]. This shows the cross-correlation is sensitive to fluctuation amplitudes and time scale. Thus the proposed method might be applied to probe complex coupling among networks of dynamical systems and to characterize the emergent functions.

## Methods

Incorporating the time lag operator into the method of analogous multifractal height cross-correlation analysis (AMF-HXA) developed by Wang et al.[21], we derive our time-lagged $q$-$L$ dependent AHXA coefficient in this

section. For two given time series $\{x_t\}$ and $\{y_t\}$, $t = 1, 2, \ldots, N$, the associated accumulated deviation series are computed by $X(t) = \sum_{i=1}^{t}[x_i - \langle x \rangle]$, $Y(t) = \sum_{i=1}^{t}[y_i - \langle y \rangle]$, where $\langle \cdot \rangle$ denotes the mean value of the whole time signal. The cross-increment of the two accumulated series with time interval $L$ is defined by

$$\Delta_L X(t)Y(t + \tau) = [X(t) - X(t + L)] \cdot [Y(t + \tau) - Y(t + \tau + L)], \tag{2}$$

where $\tau$ takes any integer within the range of the series.

In order to record the average fluctuation of cross-increment between $\{X(t)\}$ and $\{Y(t+\tau)\}$, inspired by ref. [19] and ref. [21], we employ a symbolic operator to describe the real information of $\Delta_L X(t)Y(t+\tau)$ for every $\tau$ and $L$. Hence, the $q$-th order covariance between the $\{X(t)\}$ and $\{Y(t+\tau)\}$ can be determined by

$$F_{xy}^q(L) = \langle \text{sign}(\Delta_L X(t)Y(t + \tau)) \cdot |\Delta_L X(t)Y(t + \tau)|^{q/2} \rangle, \tag{3}$$

where $t$ varies from $1 - \tau$ to $N - L$ when $\tau < 0$ and from $1$ to $N - L - \tau$ when $\tau \geq 0$. The $\text{sign}(\Delta_L X(t)Y(t+\tau))$ denotes the sign of $\Delta_L X(t)Y(t+\tau)$. The preserved sign information of $\Delta_L X(t)Y(t+\tau)$ in $F_{xy}^q(L)$ provides *i*) the fluctuation information as well as the sign information of the fluctuation function; *ii*) correct cross-correlated exponent (if correlated), and *iii*) true correlation of $\{x(t)\}$ and $\{y(t)\}$ with time-delay $\tau$ avoiding spurious correlations. The $F_{xy}^q(L)$ expresses different fluctuations through the different order $q$, which describes smaller fluctuations with smaller $q$'s. In general, exponent $q < 1$ magnifies the contribution to $F_{xy}^q(L)$ from small amplitude fluctuations and $q > 2$ magnifies the contribution from large amplitude fluctuations. Besides the $q$-th order covariance, one can also obtain the $q$-th order variance for a single signal by letting $x_t = y_t$ to arrive

$$F_{xx}^q(L) = \langle \Delta_L X(t)^q \rangle, \tag{4}$$

where $\Delta_L X(t) = |X(t + L) - X(t)|$. Thereby, if (cross-)correlations exist, there will be power-law relationships between the above $q$-th fluctuation functions and the scale $L$ and the corresponding scaling exponents can be derived from Wang *et al*.[21] and Barabasi *et al*.[43] as follows,

$$F_{xy}^q(L) \propto L^{q\lambda(q)} \text{ and } F_{xx}^q(L) \propto L^{q\gamma_{xx}(q)}. \tag{5}$$

Here $\gamma_{xx}(q)$ and $\lambda(q)$ are the so-called generalized individual and bivariate Hurst exponents, respectively. The time interval $L$ typically varies from 1 to a large positive integer $N_1$ such that a good linear fit can be obtained in the double log plot.

If the cross-correlation is absent or not significant, then the power-law relation described by Eq. (5) no longer holds and the exponent $\lambda(q)$ would be complex. To correctly quantify the level of cross-correlations, Wang *et al*.[21] proposed the so-called AHXA coefficient as follows,

$$\rho(L) = \frac{F_{xy}^2(L)}{\sqrt{F_{xx}^2(L) \cdot F_{yy}^2(L)}} = \frac{\langle \text{sign}(\Delta_L X(t)Y(t + \tau)) \cdot |\Delta_L X(t)Y(t + \tau)| \rangle}{\sqrt{\langle \Delta_L X(t)^2 \rangle \cdot \langle \Delta_L Y(t)^2 \rangle}}. \tag{6}$$

Here $\rho(L) \in [-1, 1]$ with $\rho(L) = 1$ indicating a perfect positive correlation and $\rho(L) = -1$ indicating a perfect negative correlation. However, $\rho(L)$ can only describe the cross-correlations at specific fluctuation range for two synchronous time series. To obtain more information on the amplitude of fluctuations contributed for the detected cross-correlations and to detect the range of cross-correlated fluctuations with potential time-delays, inspired by the idea of $q$-dependent DCCA coefficient proposed by Kwapień *et al*.[33], we extend $\rho(L)$ by incorporating both the time-lag factor and the fluctuation amplitude to arrive a new coefficient denoted by $\rho_q(\tau, L)$ as follows,

$$\rho_q(\tau, L) = \frac{F_{xy}^q(L)}{\sqrt{F_{xx}^q(L) \cdot F_{yy}^q(L)}} = \frac{\langle \text{sign}(\Delta_L X(t)Y(t + \tau)) \cdot |\Delta_L X(t)Y(t + \tau)|^{q/2} \rangle}{\sqrt{\langle \Delta_L X(t)^q \rangle \cdot \langle \Delta_L Y(t + \tau)^q \rangle}}. \tag{7}$$

Clearly when $q = 2$ and $\tau = 0$, $\rho_q(\tau, L)$ is exactly the AHXA coefficient $\rho(L)$. In the case where $\tau = 0$, our obtained coefficient is indeed an alternative of the $q$-dependent DCCA coefficient proposed by Kwapien *et al*. The advantage of our coefficient is that it allows us to identify the range of detrended fluctuation amplitudes that are correlated with time lags in two signals under study. Since $\rho_q(\tau, L)$ depends on three parameters, $q$, $\tau$ and $L$, we call $\rho_q(\tau, L)$ the time-lagged $q$-$L$ dependent AHXA coefficient. For each fixed $q$, we can draw a $\rho_q(\tau, L)$ surface to examine cross-correlations between two series at each time lag $\tau$ and each time interval $L$.

## References

1. Ives, M. More than 9 in 10 people breathe bad air, w.h.o. study says. https://www.nytimes.com/2016/09/28/world/air-pollution-smog-who.html?_r=0. Sep. 27, (2016).
2. Davis, D. L. A look back at the London smog of 1952 and the half century since. *Environmental health perspectives* **110**, A734 (2002).
3. Chan, C. & Yao, X. Air pollution in mega cities in china. *Atmospheric Environment* **42**, 1–42 (2008).
4. Wang, S. & Hao, J. Air quality management in china: Issues, challenges, and options. *Journal of Environmental Sciences* **24**, 2–13 (2012).
5. Han, L. J., Zhou, W. Q., Li, W. F. & Li, L. Impact of urbanization level on urban air quality: A case of fine particles (pm2.5) in chinese cities. *Environmental Pollution* **194**, 163–170 (2013).
6. Han, L. J., Zhou, W. Q. & Li, W. F. City as a major source area of fine particulate (pm2.5) in china. *Environmental Pollution* **206**, 183–187 (2015).

7. Han, L. J., Zhou, W. Q. & Li, W. F. Increasing impact of urban fine particles (pm2.5) on areas surrounding chinese cities. *Scientific Reports* **5**, 12467 (2015).
8. Bai, X., Chen, J. & Shi, P. Landscape urbanization and economic growth in china: Positive feedbacks and sustainability dilemmas. *Environmental Science and Technology* **46**, 132–139 (2012).
9. Shen, C. H. & Li, C. An analysis of the intrinsic cross-correlations between api and meteorological elements using dpcca. *Physica A* **446**, 100–109 (2016).
10. Shi, K. Detrended cross-correlation analysis of temperature, rainfall, pm 10 and ambient dioxins in hong kong. *Atmospheric Environment* **97**, 130–135 (2014).
11. Shen, C. H. A new detrended semipartial cross-correlation analysis: Assessing the important meteorological factors affecting api. *Physics Letters A* **379**, 2962–2969 (2015).
12. Du, D. K., Lee, D. I., Kwon, B. H., Kim, K. & Park, J. K. Features of the detrended cross-correlation analysis in the time series between absorbable particulate matter and meteorological factors. *Journal of the Korean Physical Society* **63**, 10–17 (2015).
13. Zeng, M., Zhang, X. N. & Li, J. H. Dcca cross-correlation analysis of 3d wind field signals in indoor and outdoor environments. *Intelligent Control and Automation (WCICA)* 2791–2796 (2016).
14. Podobnik, B. & Stanley, H. E. Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. *Phys. Rev. Lett* **100**, 084102 (2008).
15. Podobnik, B., Horvatic, D., Petersen, A. M. & Stanley, H. E. Cross-Correlations between Volume Change and Price Change. *Proc. Natl. Acad. Sci. USA* **106**, 22079–22084 (2009).
16. Zhou, W. X. Multifractal detrended cross-correlation analysis for two nonstationary signals. *Phys. Rev. E* **77**, 066211 (2008).
17. He, L. Y. & Chen, S. P. A new approach to quantify power-law cross-correlation and its application to commodity markets. *Physica A* **390**, 3806–3814 (2011).
18. Jiang, Z. Q. & Zhou, W. Multifractal detrending moving-average cross-correlation analysis. *Phys. Rev. E* **84**, 016106 (2011).
19. Oswiecimka, P., Drozdz, S., Forczek, M., Jadach, S. & Kwapien, J. Detrended cross-correlation analysis consistently extended to multifractality. *Phys. Rev. E* **89**, 023305 (2014).
20. Kristoufek, L. Multifractal height cross-correlation analysis: A new method for analyzing long-range cross- correlations. *Europhys. Lett.* **56**, 68001 (2011).
21. F. Wang, Z. H. Y. & Wang, L. Detecting and quantifying cross-correlations by analogous multifractal height cross- correlation analysis. *Physica A* **444**, 954–962 (2016).
22. Podobnik, B. *et al.* Quantifying Cross-Correlations Using Local and Global Detrending Approaches. *Eur. Phys. J. B* **71**, 243–250 (2009).
23. Podobnik, B., Jiang, Z. Q., Zhou, W. X. & Stanley, H. E. Statistical tests for power-law cross-correlated processes. *Phys. Rev. E* **84**, 066118 (2011).
24. Qian, X. Y. *et al.* Detrended partial cross-correlation analysis of two nonstationary time series influenced by common external forces. *Phys. Rev. E* **91**, 062816 (2015).
25. Yuan, N. M. *et al.* Detrended partial-cross-correlation analysis: a new method for analyzing correlations in complex system. *Scientific Reports* **5**, 8143 (2015).
26. Yuan, N. M., Xoplaki, E., Zhu, C. & Luterbacher, J. A novel way to detect correlations on multi-time scales, with temporal evolution and for multi-variables. *Scientific Reports* **6**, 27707 (2016).
27. Zebende, G. Dcca cross-correlation coefficient, quantifying level of cross-correlation. *Physica A* **390**, 614–618 (2011).
28. Wang, F., Liao, G. P. & Zhou, X. Y. & Multifractal detrended cross-correlation analysis for power markets. *Nonlinear Dynamics* **72**, 353–363 (2013).
29. Wang, F., Liao, G. P., Li, J. H. & Zou, R. B. & Cross-Correlation Detection and Analysis for California electricity Market Based on Analogous multifractal Analysis. *Chaos* **23**, 013129 (2013).
30. Wang, F. A novel coefficient for detecting and quantifying asymmetry of California electricity market based on asymmetric detrended cross-correlation analysis. *Chaos* **26**, 063109 (2016).
31. Kristoufek, L. Detrending moving-average cross-correlation coefficient: Measuring cross-correlations between non-stationary series. *Physica A* **406**, 169–175 (2014).
32. Kristoufek, L. Measuring correlations between non-stationary series with dcca coefficient. *Physica A* **402**, 291–298 (2014).
33. Kwapien, J., Oswiecimka, P. & Drozdz, S. Detrended fluctuation analysis made flexible to detect range of cross-correlated fluctuations. *Phys. Rev. E* **92**, 052815 (2015).
34. Lin, A. J., Shang, P. J. & Zhao, X. J. The cross-correlations of stock markets based on dcca and time-delay dcca. *Nonlinear Dyn* **67**, 425–435 (2012).
35. Alvarez-Ramirez, J., Rodriguez, E. & Echeverria, J. C. Using detrended fluctuation analysis for lagged correlation analysis of nonstationary signals. *Phys. Rev. E* **79**, 057202 (2009).
36. Shen, C. H. Analysis of detrended time-lagged cross-correlation between two nonstationary time series. *Physics Letters A* **379**, 680–687 (2015).
37. Lin, A. J., Liu, K. K. L., Bartsch, R. P. & Ivanov, P. C. Delay-correlation landscape reveals characteristic time delays of brain rhythms and heart interactions. *Phil. Trans. R. Soc. A* **374**, 20150182 (2016).
38. Peng, C. K. *et al.* Mosaic organization of DNA sequences. *Phys. Rev. E* **49**, 1685–1689 (1994).
39. Kantelhardt, J. W. *et al.* Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A* **316**, 87–114 (2002).
40. Ivanov, P. C., Liu, K. K. L. & Bartsch, R. P. Focus on the emerging new fields of network physiology and network medicine. *New Journal of Physics* **18**(10), 100201 (2016).
41. Bartsch, R. P., Liu, K. K. L., Bashan, A. & Ivanov, P. C. Network physiology: how organ systems dynamically interact. *PloS One* **10**(11), e0142143 (2015).
42. Liu, K. K. L., Bartsch, R. P., Ma, Q. D. Y. & Ivanov, P. C. Major component analysis of dynamic networks of physiologic organ interactions. *Journal of Physics: Conference Series* **640**(1), 012013 (2015).
43. Barabasi, A. L., Szepfalusy, P. & Vicsek, T. Multifractal spectra of multi-affine functions. *Physica A* **178**, 17–28 (1991).

## Acknowledgements

## Author Contributions

F.W. designed the framework and performed the statistical analysis. L.W. wrote the manuscript and checked the analysis. Y.C. reviewed the analysis and provided comments on revising the manuscript.