Long-Term Sustainability and Multifractal Characteristics of Air Pollution Evolution

Wang Antao and Sun Linan

1Department of Cyber Security, Henan Police College, Zhengzhou 450046, China
2College of Geography and Environmental Science, Henan University, Kaifeng 475004, China
3College of Resources and Environment, Henan University of Economics and Law, Zhengzhou 450046, China

Correspondence should be addressed to Sun Linan; 20170171@huel.edu.cn

Received 13 January 2022; Revised 26 February 2022; Accepted 2 March 2022; Published 11 April 2022

Copyright © 2022 Wang Antao and Sun Linan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Air pollution is a complex phenomenon caused by humans causing certain substances to enter the atmosphere during production activities or natural processes. Its formation and evolution have long-term sustainability, macroscopicity, and integrity. The concentration of each pollutant is high or low, and it will last for a period of time. So, what are the macro and overall characteristics of the self-evolution of the concentration of each pollutant? This article will use the detrended fluctuation analysis (DFA) method to analyze the long-term sustainability of each pollutant concentration sequence at nine sites. At the same time, in order to describe the nonlinear characteristics of each pollutant concentration sequence in more detail, use the multifractal detrended fluctuation analysis (MF-DFA) method to analyze the internal local structure. The MF-DFA method can describe the unique mode of the pollution process during the haze period, record the detailed information of the pollutants on different time scales during the haze period, provide probability estimates for the pollutant concentration, and display the pollutant concentration. The MF-DFA method can also describe the characteristics of time series in a more detailed, precise, and comprehensive manner and quantitatively describe the long-term sustainability of time series evolution. The experimental analysis results of the MF-DFA method on the concentration of each pollutant at nine monitoring points during the haze period have achieved extraordinary results.

1. Introduction

Air pollution is a complex phenomenon in which humans discharge pollutants into an open and dissipative air system. Its formation and evolution are macroscopic and integral. With the acceleration of my country’s urbanization and industrialization, the problem of air pollution has attracted more and more people’s attention. The large amount of waste gas and impurities produced in industrial activities also seriously affect people’s physical and mental health [1]. Therefore, research on the formation and evolution of air pollution [2] is very necessary, which provides a theoretical basis for the control of air pollution.

In recent years, research on air pollution has become more and more extensive, such as the causes of air pollutants [3], the prediction of air pollutant concentrations [4], and the distribution of air pollutants [5, 6]. With the development of Internet technology, Internet technology has also been used in air pollution research [7, 8]. For example, use IoT technology to provide environmental protection departments with detailed governance information [9]. Also, the extension neural network type-1 (ENN-1) algorithm [10] improves the accuracy of air pollution assessment. Coupling coordination degree model [11], exploratory spatial data analysis [12], and spatial correlation analysis [13] are all related studies on the spatial distribution of air pollution, which analyze the temporal and spatial distribution, overall trend, and evolution law of air pollution. However, none of the above methods has meticulously described the nonlinear characteristics of each pollutant concentration sequence and analyzed the internal local structure.
In order to overcome the above-mentioned difficulties, we have proposed a layer-by-layer analysis method for DFA and MF-DFA. DFA and MF-DFA methods are often used to study the statistical characteristics of a single nonstationary time series, and it has been successfully applied to many fields such as stock finance [14], air pollution [15], heart rate dynamics [16], precipitation [17], etc. In the field of air pollution research, DFA and MF-DFA methods are also widely used; for example, Lee [18] and others have studied the multifractal characteristics of the time series of air pollutants such as O₃, CO, SO₂, and NO₂. In this article, we used the DFA method to analyze the long-term sustainability of the pollutant concentration series at nine sites. At the same time, in order to describe the nonlinear characteristics of each pollutant concentration sequence in more detail, the internal local structure is analyzed using the MF-DFA method.

The DFA method is very suitable for long-memory processes, and furthermore, extreme values can be studied. The MF-DFA method can avoid the misjudgment of the correlation and can also find the long-term continuity in the nonstationary time series. Os'Wie Cimka et al. [19] believed that the MF-DFA method can be used to explore the behavior of multifractals globally, so this method is a good multifractal analysis method, and it has been successfully used to study various nonstationary time series. Since MF-DFA is a promotion of DFA, this article takes DFA as a special case of MF-DFA and includes it in MF-DFA for the introduction.

![Image](storage.googleapis.com/download.tensorflow.org/tf2-preview/tf_image_api/imageapi.png)

### 2. The Multifractal Detrended Fluctuation Analysis (MF-DFA) Method

The steps of the MF-DFA method are as follows:

1. According to a time series of length \( N \), build a cumulative sequence:
   \[
   X(i) = \sum_{m=1}^{i} (x_m - \langle x \rangle), i = 1, 2, ..., N,
   \]
   \[
   \langle x \rangle = \frac{1}{N} \sum_{m=1}^{N} x_m.
   \]

2. Divide the cumulative sequence \( X(i) \) into \( N_t \) non-overlapping boxes of equal length \( s \), \( N_t = \text{int}(N/s) \), since \( N \) may not be an integer multiple of \( s \), there will be some remaining data at the end of the cumulative column that will not be calculated. In order to take this part of the remaining data into account when calculating, redivide this part of the remaining data. Therefore, a total of \( 2Ns \) small boxes are obtained.

3. Use the least-squares method to fit the cumulative sequence in each small box so as to calculate the local trend \( x_v(i) \) of \( 2Ns \) boxes. Then, calculate its overall elimination trend multiple covariance:

   \[
   F\_{\nu}(s) = \frac{1}{N_s} \sum_{i=1}^{N_s} \{ X[N - (v - N_s)s + i] - x_v(i) \}^2, v = N_s + 1, 2, ..., 2N_s.
   \]

4. Calculate the \( q \)-order fluctuation function:
   \[
   F(q, s) = \left\{ \frac{1}{2N_s} \sum_{v=1}^{2N_s} |F\_{\nu}(s)|^{q/2} \right\}^{1/q}, \quad q \neq 0,
   \]
   \[
   F(q, s) = \exp \left\{ \frac{1}{4N_s} \sum_{v=1}^{2N_s} \ln |F\_{\nu}(s)| \right\}, \quad q = 0.
   \]

5. If the time series \( \{x_i\} \) are related to a long-range power law, then \( F(q, s) \) and \( s \) have a power law relationship in the double logarithmic coordinates:
   \[
   F(q, s) \sim s^{\gamma(q)},
   \]

   \( h(q) \) is called the generalized Hurst exponent of order \( q \), which can be obtained by calculating the slope of the relation graph between \( F(q, s) \) and \( s \) under double logarithm. If \( h(q) \) is constant with respect to \( q \), indicating that the local structure of the time series is uniform, then the time series is monofractal. If \( h(q) \) changes with \( q \), indicating that the local structural formula of the time series is not uniform, there is a multifractal. Different \( q \) can describe the influence of different degrees of fluctuation on \( F(q, s) \). When \( q < 0 \), the size of the fluctuation function \( F(q, s) \) is greatly affected by the small fluctuation deviation \( F\_{\nu}(s) \). At this time, \( h(q) \) describes the scale behavior of small fluctuations. When \( q > 0 \), the size of the volatility function \( F(q, s) \) is greatly affected by the large volatility deviation \( F\_{\nu}(s) \). At this time, \( h(q) \) describes the scale behavior of large fluctuations.

When \( q = 2 \), MF-DFA is equivalent to the DFA method. We know that for the DFA method, \( h(2) \) can indicate whether the analyzed time series has fractal properties. When \( h(2) = 0.5 \), the sequence is a white noise sequence.
without any relationship, showing completely random characteristics, without long-term sustainability, and the value of the future moment has no relationship with the value of the past moment. When \( h(2) \neq 0.5 \), the sequence shows long-term continuity; that is, each value of the sequence “memorizes” its value for a period of time before. Furthermore, if \( h(2) > 0.5 \), the time series shows positive long-term continuity, which means that if the time series shows an upward or downward trend in a certain period of time in the past, then a certain period of time in the future will also show an upward or downward trend. The trend corresponds to the positive feedback mechanism of the system, if \( h(2) < 0.5 \), the time series has anti-long-term continuity, indicating that the time series shows more opposite trends in evolution, which corresponds to the negative feedback mechanism of the system [20].

It can be summarized as follows: if the generalized Hurst index \( h(q) > 0.5 \), it indicates that the sequence is long-term continuous; that is, if the sequence shows an upward or downward trend in the previous period, it will also show an upward or downward trend in the later period; if \( h(q) < 0.5 \), it means that the time series has antipersistence; if \( h(q) = 0.5 \), it means that the time sequence is a random walk at this time, and there is no law.

### 3. The Relationship between MF-DFA and Standard Multifractal Analysis (MFA)

The scale index \( h(q) \) of the stable and standardized sequence of the multifractal defined in formula (4) is directly related to the scale index \( \tau(q) \) defined by the standard distribution function based on the multifractal theory below [21]. Assuming that the sequence of length \( N \) is a stable standardized sequence, since there is no need to eliminate the trend, the process of eliminating the trend in step (3) of the above MF-DFA method implementation steps is not needed. Therefore, the DFA method can be replaced by the standard fluctuation analysis (FA) method, except for the definition of variance; it is the same as DFA, and it is simplified for each small box. Formula (2) in step (3) becomes as follows:

\[
F_y(s)_{FA} = |X(vs) - X((v-1)s)|^2.
\]  

(5)

Substitute this simplified formula into formula (3), then combining formula (4) to get the following:

\[
\left\{ \frac{1}{2N_s} \sum_{v=1}^{2N_s} |X(vs) - X((v-1)s)|^q \right\}^{1/q} \sim s^{h(q)}.
\]  

(6)

For the sake of simplicity, we can assume that the sequence length \( N \) is an integer multiple of the scale \( s \), and get \( N_s = N/s \), so

\[
\sum_{v=1}^{N_s} |X(vs) - X((v-1)s)|^q \sim s^{h(q)-1}.
\]  

(7)

The formula \( X(vs) - X((v-1)s) \) of formula (7) is equal to the sum of \( x_k \) in each box of length \( s \). For the standardized sequence \( x_k \), this sum is the well-known box probability \( p_x(v) \) in multifractal theory:

\[
p_x(v) \equiv \sum_{k=(v-1)s+1}^{vs} x_k = X(vs) - X((v-1)s).
\]  

(8)

The scale index \( \tau(q) \) can usually be defined by partition function \( Z_q(s) \):

\[
Z_q(s) = \sum_{v=1}^{N/s} |p_x(v)|^q \sim s^{\tau(q)},
\]  

(9)

where \( q \) is a variable in the MF-DFA method. Substituting formula (8) into (9), it can be found that equations (3)–(5) and (7) are equivalent. Therefore, we can obtain the relationship between the two multifractal indices:

\[
\tau(q) = qh(q) - 1.
\]  

(10)

Therefore, \( h(q) \) defined in formula (8) is directly related to the classical multifractal scale index \( \tau(q) \). Further, we can derive the function graph of the scale index \( \tau(q) \) and \( q \). If the time series is monofractal, then \( \tau(q) \) is a linear function of \( q \); if the time series is multifractal, then \( \tau(q) \) is a convex function of \( q \). And the larger \( \tau(q) \), the stronger the multifractal feature of the corresponding time series. Note that \( h(q) \) is different from generalized multifractal dimension \( D(q) \). The relationship between \( \tau(q) \) and \( h(q) \) is as follows:

\[
D(q) = \frac{\tau(q)}{q-1} = \frac{qh(q) - 1}{q-1}.
\]  

(11)

In this case, for a single fractal time series, when \( h(q) \) is independent of \( q \), \( D(q) \) depends on \( q \).

Another way to characterize the multifractal sequence is to find the singular spectrum [20, 21] \( f(\alpha) \) related to \( \tau(q) \) through Legendre transform:

\[
\left\{ \begin{array}{l}
\alpha = \tau'(q) \\
\hat{f}(\alpha) = q\alpha - \tau(q)
\end{array} \right.
\]  

(12)

where \( \alpha \) is the singular strength or Hurst index and \( f(\alpha) \) is the multifractal spectrum, and the value of \( f(\alpha) \) reflects the fractal dimension of the sequence subset defined by \( \alpha \). From this, the relationship between \( \alpha \) and \( f(\alpha) \) can be obtained. Using formula (10), we can directly associate the variables \( \alpha \), \( f \), \( h \) in the following way:

\[
\left\{ \begin{array}{l}
\alpha = h(q) + qh'(q) \\
f(\alpha) = q[\alpha - h(q)] + 1
\end{array} \right.
\]  

(13)

where \( \alpha \) is the singularity index, used to describe the different degrees of singularity in each interval in the time series; \( f(\alpha) \) is the multifractal spectrum, and the value of \( f(\alpha) \) reflects the fractal dimension of the singularity index \( \alpha \). In the process of analyzing the generalized binomial multifractal model in 2006, Kosielny-Bunde et al. [22] found that the \( q \)-order Hurst exponent \( h(q) \) has the following relationship:

\[
h(q) = \frac{1}{q} \ln \left( \frac{a^2 + b^q}{q} \right).
\]  

(14)

In equation (14), \( a \) and \( b \) are parameters and further we can get the width of the multifractal spectrum:
\[ \Delta \alpha = a (-\infty) (\infty) \frac{\ln b - \ln a}{\ln 2} \min_{\alpha_{\text{max}}}, \]  

where \( \alpha_{\text{min}} \) corresponds to the largest subset of the time series, \( \alpha_{\text{max}} \) corresponds to the smallest subset of the time series, and \( \Delta \alpha \) is the width of the multifractal spectrum, which represents the difference between the largest and smallest singular exponents. \( \Delta \alpha \) can be used to characterize the strength of the system’s multifractal degree; that is, the larger \( \Delta \alpha \), the more uneven the time series distribution, the greater the width of the multifractal spectrum, the more significant the intensity. \( \alpha_{\text{min}}, f(\alpha_{\text{min}}) \) reflect the nature of the smallest subset of the time series \( \alpha_{\text{min}}, f(\alpha_{\text{min}}) \) reflect the nature of the largest subset of the time series, \( f(\alpha) \max \), and corresponding \( \alpha_0 \) reflect the nature of the most probable subset. For \( \Delta f = f(\alpha)_{\text{min}} - f(\alpha)_{\text{max}}, \) where \( f_{\text{min}} \) and \( f_{\text{max}} \) are the fractal dimensions of the largest and smallest probability subsets, respectively; \( \Delta f \) is the ratio of the number of the largest and smallest probability units, that is, the ratio between the number of peaks (highest) and the number of troughs (lowest) in the time series. The multifractal spectrum is divided into left and right parts by \( \alpha = \alpha_0. \) If the left endpoint \( (\alpha_{\text{min}}, f(\alpha_{\text{min}})) \) of the multifractal spectrum is significantly lower than the right endpoint \( (\alpha_{\text{max}}, f(\alpha_{\text{max}})) \), the vertex \((\alpha_0, f(\alpha)_{\text{max}})\) moves to the right; that is, the multifractal spectrum is right hooked, the probability of the time series being at the lowest value is greater than the probability of being at the highest value, and the time series has a downward trend. If the left endpoint \( (\alpha_{\text{min}}, f(\alpha_{\text{min}})) \) of the multifractal spectrum is significantly higher than the right endpoint \( (\alpha_{\text{max}}, f(\alpha_{\text{max}})) \), the vertex \((\alpha_0, f(\alpha)_{\text{max}})\) moves to the left; that is, the multifractal spectrum is left hooked, the probability of the time series being at the lowest value is less than the probability of being at the highest value, and the time series has an upward trend.

4. Causes of Multifractals

Under normal circumstances, the multifractal characteristics of time series are mainly due to two aspects: one is the continuous influence of small and large fluctuations in the time series on different time scales, that is, the long-term sustainability; The second is the peak fat tail probability distribution characteristics of time series extreme values [23, 24]. There are two ways to determine the source of multifractal features: one is to construct alternative sequences through alternative methods; the other is to construct random sequences through random recombination methods.

We randomly process the original sequence by random recombination method, get the random sequence, and calculate the corresponding \( h_{\text{shuff}}(q) \). Because the process of random data reorganization destroys the inherent correlation of the original time series and only retains the nonlinear part of the original time series, the random series can effectively test the impact of long-term persistence on the multifractal characteristics. The replacement sequence is obtained by performing multiple phase randomization processing based on the original sequence, and the corresponding \( h_{\text{surr}}(q) \) is calculated. The replacement sequence completely removes the nonlinear characteristics of the original sequence and only retains its linear components, so it can effectively test the contribution of the peak fat tail distribution to the multifractal. We assume that the long-term persistence and the probability distribution of the spike fat tail are independent of each other, so the long-term persistence can be considered as follows:

\[ h_{\text{cor}}(q) = h_{\text{orig}}(q) - h_{\text{shuff}}(q). \]  

That is, if \( h_{\text{cor}}(q) > 0 \), it indicates that compared with the original sequence, the multifractal strength of the random recombination sequence is reduced. At this time, we can say that the long-term persistence has an effect on the multifractals, and the influence of the extreme peak fat tail distribution is considered:

\[ h_{\text{PDF}}(q) = h_{\text{orig}}(q) - h_{\text{surr}}(q). \]  

That is, if \( h_{\text{PDF}} > 0 \), the multifractal intensity of the replacement sequence is reduced compared with the original sequence. At this time, we can say that the extreme peak fat tail distribution has an impact on the multifractals. By comparing the MF-DFA results of the original time series with the rearranged time series and the replacement time series, we can analyze the causes of multifractal features:

1. If the probability distribution of extreme spikes and fat tails causes the multifractal characteristics of the time series, then the multifractal of the random recombination sequence will not affect the recombination of the sequence, and its multifractal strength is equal to the multifractal strength of its original sequence; that is, \( h_{\text{shuff}}(q) = h_{\text{orig}}(q) \). In other words, the generalized Hurst exponent \( h_{\text{surr}}(q) \) of the replacement sequence is a constant.

2. If the long-term persistence causes the multifractal characteristics of the time series, the random recombination sequence will not show the multifractal characteristics because the recombination process eliminates the inherent time correlation of the original sequence and only retains the nonlinear part of the original sequence. At this time, \( h_{\text{shuff}}(q) = 0.5 \).

3. If it is the long-term continuity of the time series and the probability distribution of extreme spikes and fat tails that together cause the formation of multifractal features, the random recombination sequence shows a weaker multifractal feature compared with the original sequence in its generalized Hurst index, at this time, \( h_{\text{shuff}}(q) < h_{\text{orig}}(q) \).

5. Analysis of Results

We collected pollution datasets from nine sites, and the evaluation indicators of pollution data were evaluated according to the indicators of PM$_{2.5}$ and SO$_2$. The nine stations are Gangli Reservoir, Water company, Henan Medical University, Management Committee, Inspection station, Middle school, Tobacco factory, Bank, and Textile factory.
5.1. Analysis of Long-Term Sustainability of Various Pollutants during Heavy Haze. First, we use the DFA method. That is, when \( q = 2 \), we will check the long-term sustainability of each pollutant. We show the relationship between \( F(q = 2, s) \) and \( s \) of each pollutant under double logarithms (Figure 1) and the Hurst index value of the slope of the fitting straight line of each pollutant log \( F(q = 2, s) \) and log \( s \) (Table 1). It can be seen from Figure 1 that the DFA analysis of the \( SO_2 \) and \( PM_{2.5} \) concentration sequences of the nine sites all show a good linear relationship. After the original concentration sequence is randomly reorganized, two species can be found. After the pollutant sequence is randomly reorganized, the \( h(2) \) value is mostly close to 0.5. It shows that the sequence after random recombination shows strong randomness, and the value of \( h(2) \) after random recombination changes greatly compared with the original sequence. The \( h(2) \) value after replacement is mostly close to the \( h_{ori}(2) \) value, indicating that the replacement sequence is very close to the original sequence; that is, the original pollutant concentration sequence is not greatly affected by the spike fat tail.

It can be seen from Table 1 that \( h_{ori}(2) > 0.5 \) and most of them are greater than 1; i.e., the pollutant time series shows strong long-term sustainability. It shows that within a certain time scale, the pollutant time series will have an impact on the current and even future trends. The observed pollutant concentration values show the same changing trend as the past values. If the pollutant concentration values in the past are large, the pollutant concentration values will be in the future for a period of time. The observed concentration of the corresponding pollutant is also larger, and vice versa. Further research found that the larger \( h_{ori}(2) \) value of each site is \( PM_{2.5} \), most of which fluctuates between 1.4 and 1.6, and the maximum value reaches 1.525, indicating that \( PM_{2.5} \) has the strongest long-term sustainability. But it also indirectly illustrates the complexity and stubbornness of \( PM_{2.5} \). This shows that the past of the haze system can continue to affect the current and future state of the system. This correlation has long-term sustainability on a certain time scale. At the same time, this is also the macroscopic manifestation of the internal dynamics of the long-term evolution of \( PM_{2.5} \) concentration. The evolution of \( PM_{2.5} \) shows irregular and nonlinear changes in time and space, with the basic characteristics of a complex system. It is precisely because of this complexity that the evolution of \( PM_{2.5} \) during the haze period is very stable and not easily destroyed. This may invalidate the red warning plan for air pollution during the haze period.

5.2. Spatial Analysis of Long-Term Sustainability of Pollutants. It can be seen from Figure 2(a) that the \( h_{ori}(2) \) value of \( SO_2 \) shows a downward trend from northwest to southeast. Higher values appear in most of the central, eastern, and southern regions. This large area, including industrial areas and residential areas, is the main functional area of Zhengzhou City. Winter is the season of high sulfur dioxide emission. Therefore, this area is also a high sulfur dioxide emission area, which is conducive to the continuity and stability of sulfur dioxide; in addition, north of winter wind is dominant, so \( SO_2 \) is easier to maintain stability and continuity in the downwind area.

It can be seen from Figure 2(b) that \( PM_{2.5} \) also show similar trends, namely, high in the east and west, lower in the northwest and central regions, and the overall \( h_{ori}(2) \) values of \( PM_{2.5} \) are both relatively large and the \( h_{ori}(2) \) value of \( PM_{2.5} \) fluctuates between 1.4 and 1.6, showing strong long-term sustainability characteristics as a whole, and the spatial difference is mainly related to the eastern and western industrial concentration factors.

5.3. Analysis of Multifractal Features of Concentration Sequences of Various Pollutants. The MF-DFA method to analyze the original sequence of the concentration of two pollutants during the haze period in Zhengzhou City and the random recombination sequence and replacement sequence after transformation are used. Figure 3 shows the concentration series of 2 pollutants at nine monitoring stations in Zhengzhou during the haze period and the relationship between the generalized Hurst index \( h(q) \) and \( q \). It can be seen from the figure that \( H \) of the concentration sequence of the two pollutants at each site is obviously not a constant, but a function of \( q \), and it keeps decreasing with the increase of \( q \), which shows the concentration of each pollutant during the heavy haze period. The time series has obvious multifractal characteristics, and its internal structure is not uniformly distributed. A single fractal model cannot accurately describe its internal essential law, indicating that the pollutant concentration has strong fractal characteristics in a small fluctuation range. After the two pollutant concentration sequences at each site are randomly reorganized and replaced in the figure, almost all \( h_{shuf}(q) \) values are close to 0.5 but not equal to 0.5. This shows that the multifractal of each pollution concentration sequence is affected by long-term sustainability during the haze period. It is also affected by the spike fat tail, but it is mainly affected by long-term persistence, and the impact of the spike fat tail is not significant.

From formula (10), we can see that \( \tau(q) \) is a function of \( q \). If the pollutant concentration sequence is a single fractal, then the image of \( \tau(q) \) with respect to \( q \) is a straight line. If the pollutant concentration sequence is multifractal, the image of \( \tau(q) \) with respect to \( q \) is an upward convex curve. Due to limited space, this article only lists the main pollutant \( PM_{2.5} \) of the haze at each site for illustration.

It can be seen from Figure 4 that the original \( PM_{2.5} \) concentration sequence \( \tau(q) \) is an upwardly convex function, and \( \tau(q) \) has a significant nonlinear change with respect to \( q \), indicating that the \( PM_{2.5} \) concentration sequence at each monitoring station has multiple fractal features. And from the convexity of the three lines, it can be seen that the convex change of the line after random recombination is more different from the convex change of the original sequence, which further shows that the long-term persistence has a greater impact on the \( PM_{2.5} \) concentration sequence. The \( q-\tau(q) \) graphs of the \( SO_2 \) pollutant also have the same performance.
Figure 1: The relationship between $F(q = 2, s)$ and $(s)$ of each pollutant under double logarithm. (a) SO$_2$, (b) PM$_{2.5}$.
It can be seen from Figure 5 that the multifractal spectrum of the nine monitoring sites is a downward opening parabola, which indicates that the PM2.5 concentration sequence during the haze period has the characteristics of multifractals. The larger the width of the multifractal spectrum, the stronger the multifractal feature.

In order to investigate the specific fractal situation of each pollutant more clearly, here we list the relevant parameter values of the multifractal spectrum $f$ and a singularity index of the two pollutants; see Table 2.

From Figure 5, it can be seen that during the haze period, the random recombination sequence corresponding to the PM$_{2.5}$ concentration of nine stations in Zhengzhou City has changed its multifractal spectrum curve $\alpha \sim f(\alpha)$ compared with the original sequence; at the same time, the multifractal spectrum curve $\alpha \sim f(\alpha)$ of the substitution sequence has also changed.

The values of $\Delta \alpha_{\text{shuff}}$ and $\Delta \alpha_{\text{surr}}$ are both greater than 0, and the value of $\Delta \alpha_{\text{shuff}}$ is obviously greater than the value of $\Delta \alpha_{\text{surr}}$, indicating that during the haze period, the multifractal characteristics of the two pollutants concentrations are affected by both the long-term persistence and the peak fat tail. But the long-term continuous impact is greater. The long-term persistence mechanism is still the main factor in the multifractal characteristics of pollutant concentration. This means that within a certain time scale, due to the long-term persistence mechanism, low-concentration pollutants may evolve into high-concentration again, leading to the appearance of more serious ash again.

In order to quantitatively describe the multifractal characteristics of large fluctuations in pollution concentration over a period of time, a quadratic function with $(\alpha_0, f(\alpha_0))$ as the vertex proposed by Shimizu in 2002 is used to fit the multifractal [25]; the formula is as follows:

$$f(\alpha) = A(\alpha - \alpha_0)^2 + B(\alpha - \alpha_0) + C. \quad (18)$$

In equation (18), the value of $\alpha$ corresponding to the peak of $f(\alpha)$ is recorded as $\alpha_0$, and $\alpha_0$ is used to characterize the degree of regularity of the potential process. The larger $\alpha_1$, the more violent the fluctuation and the more irregular. $B$ is the asymmetric coefficient. When $B = 0$, the shape of the multifractal spectrum is symmetric; when $B < 0$, the multifractal spectrum is shifted to the right, and the higher
Figure 3: The $q\sim h(q)$ graphs of the original sequence, random recombination sequence, and alternative sequence of the Hurst index of each pollutant at nine automatic monitoring stations in Zhengzhou based on the MF-DFA analysis method. (a) SO$_2$. (b) PM$_{2.5}$. 
Figure 4: The $q$-$\tau(q)$ graphs of the Hurst index of the original sequence, random recombination sequence, and alternative sequence of PM$_{2.5}$ pollutants at nine automatic monitoring stations in Zhengzhou.

Figure 5: The $\alpha$-$f(\alpha)$ graphs of the Hurst index of the original sequence, random recombination sequence, and alternative sequence of PM$_{2.5}$ pollutants at nine automatic monitoring stations in Zhengzhou.
fractal index is dominant; when $B > 0$, the multifractal spectrum is shifted to the left, at this time the lower fractal index dominates and $C$ is the constant of the equation.

It can be seen from Figure 5 and Table 2 that the left endpoint $(\alpha_{\min}, f(\alpha_{\min}))$ of the multifractal spectrum of all pollutants at each site is higher than the right endpoint $(\alpha_{\max}, f(\alpha_{\max}))$; that is, $\Delta f > 0$. This indicates that the probability of pollutant concentration being at the highest concentration is greater than the probability of being at the lowest concentration. $\Delta f$ of PM$_{2.5}$ at some monitoring stations is also less than 0. After averaging the concentrations of pollutants at the nine monitoring stations, it is found that only $\Delta f$ of PM$_{2.5}$ is less than 0 (See Figure 6), which shows that in the average state, the probability of PM$_{2.5}$ pollutant being at a lower concentration is greater than the probability of being at a higher concentration. But when $\Delta f < 0$, we know that their $\alpha - f(\alpha)$ curve is shifted to the right. At this time, $B < 0$ can be obtained from formula (18), indicating that the singular value on the left side of $\alpha_0$ has a large value range, and the pollution index larger events are more dominant. However, there are also incidents where the local pollution index drops, which is related to the change trend of PM$_{2.5}$ concentrations over time during the haze period. But overall, events with a larger pollution index are more dominant than events with a lower pollution index.

### 6. Conclusions

This article uses the MF-DFA method to analyze the long-term continuity and multifractal properties of the two pollutants PM$_{2.5}$ and SO$_2$ at nine automatic monitoring stations during a typical heavy haze period in Zhengzhou. When $q = 2$ (the MF-DFA method is equivalent to the classic DFA method at this time), first use the MF-DFA method to analyze the long-term sustainability characteristics of each pollutant. The results showed that the PM$_{2.5}$ and SO$_2$ concentration sequences of the nine sites all showed a good linear relationship, and the original concentration sequence showed strong randomness after random reorganization and replacement. This shows that long-term sustainability is related to the nature of the pollutant, but it is also affected by meteorological factors and regions (different functional areas of the city).

When $-20 \leq q \leq 20$, we used the MF-DFA method to study the fractal situation of each pollutant and analyzed the concentration series of 2 pollutants at nine monitoring stations and their transformation sequences $q \rightarrow h(q)$, $q \rightarrow \tau(q)$ and the $\alpha \sim f(\alpha)$ relationship diagram. The results show that the concentration sequence of each pollutant during the haze period has significant multifractal characteristics, and a single fractal model cannot accurately describe its inherent essential laws. Then, by randomly reorganizing and

<table>
<thead>
<tr>
<th>$q = 2$</th>
<th>Gangli reservoir</th>
<th>Water company</th>
<th>Henan medical university</th>
<th>Management committee</th>
<th>Inspection station</th>
<th>Middle school</th>
<th>Tobacco factory</th>
<th>Bank</th>
<th>Textile factory</th>
<th>Average $\Delta f_{ori}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \alpha_{ori}$</td>
<td>1.384</td>
<td>1.327</td>
<td>1.114</td>
<td>0.994</td>
<td>1.532</td>
<td>1.073</td>
<td>1.063</td>
<td>2.699</td>
<td>0.862</td>
<td>1.341</td>
</tr>
<tr>
<td>$\Delta \alpha_{shuff}$</td>
<td>0.753</td>
<td>0.742</td>
<td>0.478</td>
<td>0.489</td>
<td>0.641</td>
<td>0.811</td>
<td>0.351</td>
<td>0.536</td>
<td>0.348</td>
<td></td>
</tr>
<tr>
<td>$\Delta \alpha_{surr}$</td>
<td>1.189</td>
<td>0.893</td>
<td>0.894</td>
<td>0.852</td>
<td>1.151</td>
<td>1.174</td>
<td>0.583</td>
<td>0.687</td>
<td>0.963</td>
<td></td>
</tr>
<tr>
<td>$\Delta f_{ori}$</td>
<td>0.035</td>
<td>-0.03</td>
<td>0.072</td>
<td>0.022</td>
<td>0.041</td>
<td>-0.064</td>
<td>-0.009</td>
<td>0.038</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>$\Delta f_{shuff}$</td>
<td>0.753</td>
<td>0.742</td>
<td>0.478</td>
<td>0.489</td>
<td>0.641</td>
<td>0.811</td>
<td>0.351</td>
<td>0.536</td>
<td>0.348</td>
<td></td>
</tr>
<tr>
<td>$\Delta f_{surr}$</td>
<td>1.189</td>
<td>0.893</td>
<td>0.894</td>
<td>0.852</td>
<td>1.151</td>
<td>1.174</td>
<td>0.583</td>
<td>0.687</td>
<td>0.963</td>
<td></td>
</tr>
<tr>
<td>$\Delta f_{ori}$</td>
<td>-0.023</td>
<td>0.007</td>
<td>0.018</td>
<td>-0.15</td>
<td>-0.122</td>
<td>0.011</td>
<td>0.02</td>
<td>0.024</td>
<td>-0.015</td>
<td>-0.025</td>
</tr>
</tbody>
</table>

**Figure 6:** Hurst index $a$ of the original sequence, random recombination sequence, and alternative sequence after the average concentration of PM$_{2.5}$ pollutant at nine stations in Zhengzhou City.
replacing the two pollutant concentration sequences at each site, it is found that the multifractal of each pollution concentration sequence during the haze period is affected by both long-term persistence and spike fat tail, but it is mainly affected by long-term persistence. The spike fat tail has little effect. Finally, the fractal spectrum characteristics of pollutants at each site are analyzed, and it is found that the probability of pollutant concentration being at the highest concentration is greater than the probability of being at the lowest concentration. After averaging the pollutant concentrations at the nine sites, it was found that only PM$_{2.5}$ pollutants have a higher probability of falling at a lower concentration than at a higher concentration. This is related to the trend of PM$_{2.5}$ pollutant concentrations over time during the haze period. But overall, events with higher PM$_{2.5}$ pollution indexes are more dominant than events with lower pollution indexes.

Based on the current analysis work, we will monitor more sites and more pollutant data in the future to improve the accuracy of air pollution analysis and prediction.

**Data Availability**

The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**Acknowledgments**

This article is the phased research result of the project of “Henan Province Key R&D and Promotion Project (Science and Technology Research) (202102310342)” and “Henan Provincial Youth Natural Science Foundation (212300410103)”.

**References**


