The temporal variations of PM_{10} concentration in Taipei: a fractal approach

Ding-Shun Ho[1], Lain-Chuen Juang[1], Yu-Ying Liao[1], Cheng-Cai Wang[1], Chung-Kung Lee[1*], Ting-Chu Hsu[1], Shaw-Yang Yang[2], Chung-Chin Yu[1]

Green Environment R&D Center,
[1]Department of Environmental Engineering,
[2]Department of Civil Engineering
Vanung University, Chung-Li, 320, Taiwan, ROC

Abstract

A one-year series of hourly average PM_{10} observations, which was obtained from the urban and national park air monitoring station at Taipei (Taiwan), was analyzed by descriptive statistics and fractal methods to examine the temporal structures of PM_{10} concentrations. It was found that all PM_{10} measurements exhibited the characteristic right-skewed unimodal frequency distribution and long-term memory. A monodimensional fractal analysis was performed by transferring the PM_{10} concentration time series into a useful compact form: the box-dimension (D_B)-threshold (T_h) and critical scale (C_S)-threshold (T_h) plots. Scale invariance was found in these time series and the box dimension was shown to be a decreasing function of the threshold PM_{10} level, implying multifractal characteristics, (*i.e.*, the weak and intense regions scale differently). To test this hypothesis, the PM_{10} concentration time series were transferred into multifractal spectra, i.e., the $\tau(q)$-q plots. The analysis confirmed the existence of multifractal characteristics. A simple two-scale Cantor set with unequal scales and weights was then used to fit the calculated $\tau(q)$-q plots. This model fits well with the entire spectrum of scaling exponents for the examined PM_{10} time series. The relationship between the fractal parameters and classical statistical characteristics, as well as some problems concerning the applicability of fractal methods on air pollution, are discussed.

Keywords: PM_{10}; box counting; multifractal scaling analysis; multifractal cascade model; long-range dependence; scale invariance

* Corresponding author. Fax: 886-3-4622232.
E-mail: anthony@msa.vnu.edu.tw

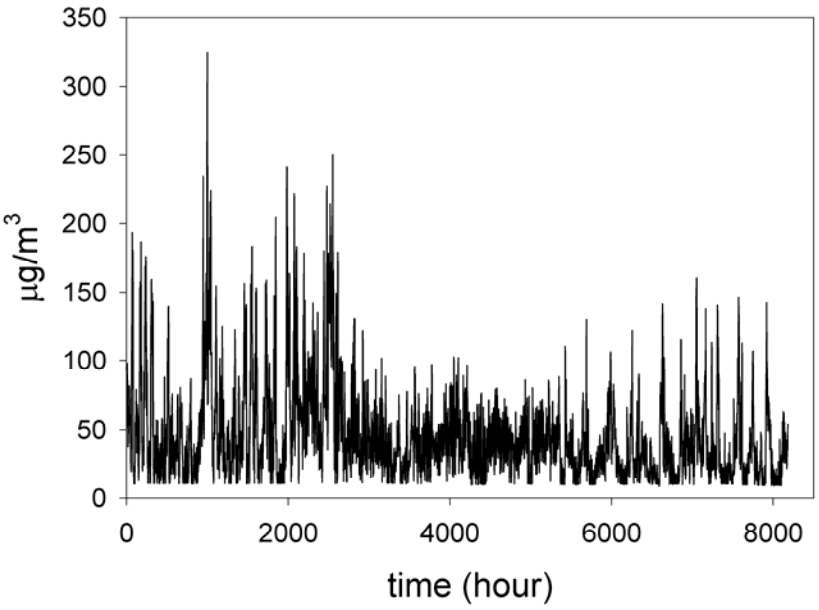

Figure 1. Time series of PM_{10} concentration in 1998, as measured at the Wan-Hwa station.

1. Introduction

PM_{10} is defined as particles with an aerodynamic diameter of approximately 10 μm or less, which means they can be inhaled. The processes of PM_{10} are complex because of the great number and variety of sources. The persistence of PM_{10} pollution is a result of industrial and societal developments, but is also influenced by meteorological factors (van der Wal and Janssen, 2000). In Taiwan, the levels of PM_{10} and other air pollutants are measured by the Taiwan Air Quality Monitoring Network (TAQMN). The Pollutant Standards Index (PSI) is used to inform the public about the current air quality and its health effects. It is found that in Taipei, two major contributors to high PSI values (poor air quality) are ozone (O_3) and PM_{10}. Because recent epidemiological studies have shown that suspended particulate matter influences respiratory health considerably (Dockery et al., 1993), it is important to examine how to effectively reduce atmospheric PM_{10} concentrations in order to decrease its adverse health effects.

The collected PM10 data are often recorded as time series and are characterized by many large fluctuations with no obvious autocorrelation (Figure 1). There are two approaches to extract air quality information from the collected data and weather parameters. One is to build an atmospheric model based on the current knowledge of fundamental chemical and physical processes, and to make predictions accordingly. The other approach starts with the statistical analysis of the collected data in order to find correlations to the atmospheric environment. However, both the accuracy and reliability of these analyses may be influenced by our fundamental knowledge of the complex structure of PM10

history at each station. Previous investigations (Lee, 2002; Lee et al., 2003a, 2003b) found that the fractal method might be an efficient tool to characterize, analyze, and compare the temporal characteristics of air pollutant concentrations. For this study, the long-range dependence of PM_{10} concentration time series measured at Taipei was first examined with the standard time series analysis. Then the clustering structure of the time series and its multiscaling characteristics were analyzed by box-counting and multifractal scaling analysis (MSA), respectively. Finally, a simple two-scale Cantor set with unequal scales and weights was used to fit with the obtained multifractal spectra. The relationship between the fractal parameters and traditional statistical parameters (coefficients of variation and skewness) is discussed. Some comments on the application of the fractal approach to the analysis of air pollution are also discussed.

2. Materials and methods

2.1 Data

In Taipei, the PM_{10} concentrations were measured with other air pollutants at seven monitoring sites of TAQMN. The PM_{10} sites were set up at traffic, urban, and national park locations, depending on their position to nearby sources. For this study, only six stations were examined because the data collected at the traffic station (Ta-Tung) were not enough to analyze. The selected sites were Shin-Lin, Chung-Shan, Wan-Hwa, Ku-Ting, and Sung-Shan (urban), and Yang-Min (national park). A previous investigation (Lee, 2002) found that most air pollutant concentration time series in Taiwan exhibited obvious annual periodicity due to systematic variations in response to seasonal and other factors, and that statistical characteristics could be extracted from the data collected over a year. Accordingly, a one-year series of hourly average values, from January to December 1998, was used in this study to examine the temporal characteristics of PM_{10}. It is worth noting that although a year consists of 8,760 hours, only about 8,400 readings for each pollutant were collected due to instrument calibration and maintenance. However, the missing observations seemed to be evenly distributed throughout the year.

Diurnal and seasonal variations play a significant role in the PM_{10} time series. Although they could have a large influence on the results of fractal analysis, using the original data to analyze the cluster structure of these time series is preferred. The reason is that any data preprocessing may strongly affect the results of fractal analysis and complicate the interpretation of the results (Klement and Kratky, 1997). Moreover, the fractal analysis made below indicates that the effect of diurnal and seasonal variations on the conclusions is insignificant. In fact, factors such as time series length may also affect the estimation of box dimension (Buczkowski et al., 1998). Thus, further investigations are still needed to examine the influence of diurnal and seasonal variations and time series length on the results of fractal analyses.

2.2. Standard Statistical Analysis

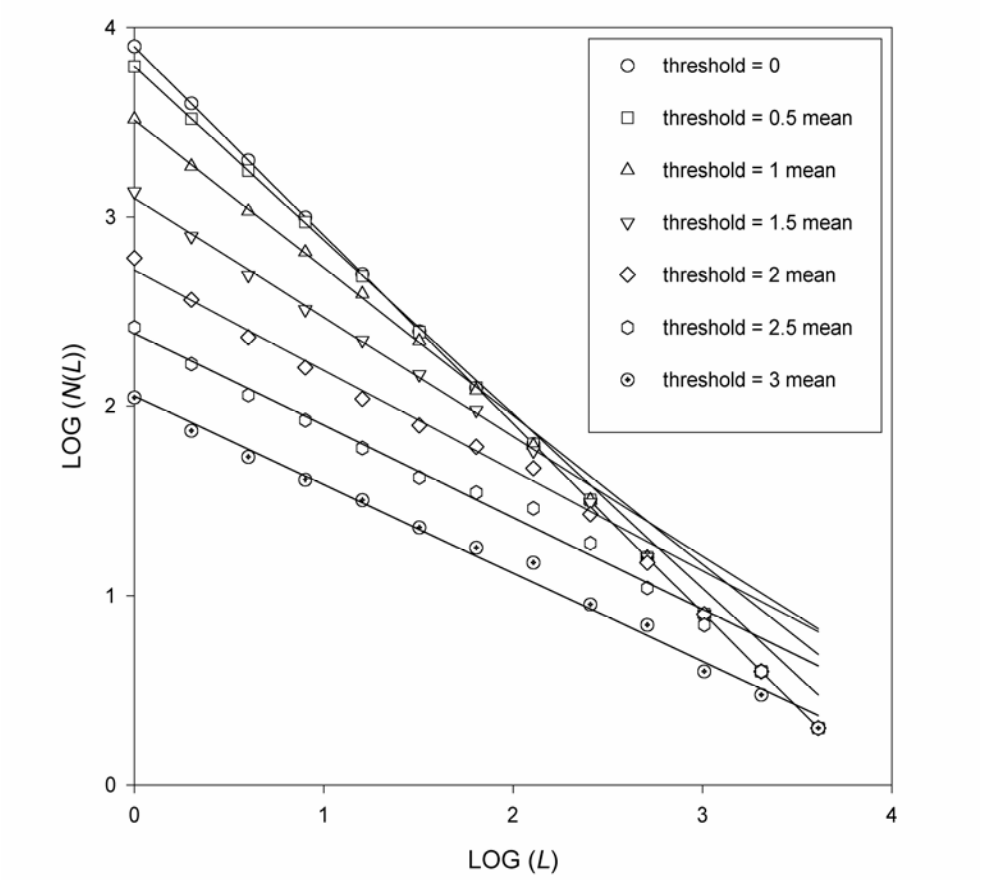

Figure 2. Box counting graphs derived from PM_{10} hourly average measurements at Shin-Lin station. Straight lines have been fitted to the sections of the graphs.

For the PM_{10} time series, we first evaluated some standard statistical parameters, such as coefficient of variation and skewness. The long-range dependency of the time series is examined based on the autocorrelation spectra.

2.3. Fractal Analysis

The scale invariance in the data set can be detected with the aid of fractal theory. Some methods have been proposed to estimate the dimension of the data set, and the dimension may be interpreted as the degree of irregularity by which the set is distributed. One method is to transform the data into a set of points whose dimension is estimated by box counting (Boxian and Lye, 1994; Lee, 2002; Lin et al., 1999, Lovejoy et al., 1987; Olsson et al., 1992; Schertzer and Lovejoy, 1987). When working with time series, another common method is to construct a phase space portrait of the process by using the correlation dimension (Olsson et al., 1993). Recently, in addition to monofractal analyses, methods suitable for analyzing multiscaling properties in time series have been suggested, such as moment scaling analysis (Klement et al., 1994; Klement and Kratky, 1997; Lee, 2002; Lee et al., 2003a, 2003b; Ho et al., 2003) and probability distribution

multiple scaling (Lovejoy and Schertzer, 1990). In this study, both the box counting technique and moment scaling analysis are adopted to examine the possible scaling characteristics in the PM_{10} time series.

In the box-counting method, the space of observation is divided into non-overlapping segments (boxes) of characteristic size L, and the number of boxes $N(L)$ needed to cover the data set is counted (Olsson et al., 1992; Olsson et al., 1993). When applying this method to a time series, the boxes represent time intervals and the space of observation is equal to the total length of the series. To convert the pollutant concentration into countable items, another step is needed. In this study, we convert the values of the PM_{10} concentration into sets of points (indicating values above threshold, T_h) by using different T_h levels. For this conversion, a zero T_h means that all hours with a registered PM_{10} concentration in the series are considered a point. If scale invariance exists in the data set, the expression $N(L) = L^{-D_B}$ will hold, with D_B as the box dimension. From the time series one can generate a plot of log $[N(L)]$ vs. log(L) and the exponents D_B can be obtained from the slope of a linear regression to the values obtained.

Figure 2 shows one typical result of applying the box-counting method, using different intensity thresholds, to the Shin-Lin monitoring stations. Some key features may be observed directly from this figure. The time scale L denotes a time interval within which PM_{10} concentration exceedances occur and the number of boxes $N(L)$ is a decreasing function of L. When T_h is 0, a linear relationship over the whole scale spectra is observed and the slope is -1. With increasing T_h, the curve is composed of two distinctly different sections: one with a slope equal to –1 and the other with $-D_B$. The value of L at the intersection of these two straight lines corresponds to the critical scale, C_S. When time scale is greater than or equal to the critical scale (C_S), the PM_{10} concentration events exceeding the threshold T_h must occur. On the other hand, the appearance of straight-lined sections with slope $-D_B$ in the log-log plots suggests the existence of scale invariance within the corresponding time scale range. This result indicates that the examined PM_{10} concentration time series can be characterized by a box dimension D_B; i.e., display scale invariance within a specific time interval. This is not surprising since the presence of fluctuations in all time scales is the origin of non-trivial scale invariance.

The temporal structure of PM_{10} concentration may depend on the threshold (T_h); the higher T_h, the more scattered the pattern, and the lower T_h, the more clustered the PM_{10} events. Since the curves are farther down and the slopes of the curves are larger (i.e., smaller D_B) as T_h increases, it is concluded that the higher the D_B the denser the time structure, and the lower the D_B the sparser the time structure. Thus, the D_B used here reveals the temporal scaling behavior of the PM_{10} concentration point set and is a measure of how the PM_{10} concentration clusters will fill the time axis they occupy. From the relationship between D_B and T_h, we can also obtain C_S as an increasing function of T_h; i.e., a larger time scale is needed to capture an occurrence of higher intensity. Basically, the implication of D_B-T_h plots is equivalent to that of C_S-T_h plots. At certain T_h, a sparser (denser) time structure may be produced with larger (smaller) C_S, and both contain all information about the PM_{10} point-processes. Thus, by analyzing D_B-T_h and C_S-T_h plots, some temporal characteristics in PM_{10} time series can be

identified.

However, this approach provides information only about the global scaling properties of the PM_{10} concentration history. It does not take into account temporal variations of the clustering degree because the local fluctuations of the distribution are not described by a single fractal dimension. Thus, the second method employed is the recently developed multifractal scaling analysis (MSA), in which the variability of the distribution at different scales is connected through a dimension function instead of through one single dimension (Lee and Lee, 1996; Olsson et al., 1993). MSA may be used to investigate whether the probability distribution related to different intensity levels is characterized by a scaling behavior. In general, multifractal characteristics, which possess an infinite number of singularities of infinitely many types, are found on a fractal substrate, where "singularity" corresponds to a local power law behavior of the measure. However, it should be noted that it is not necessary to have a fractal structure to find multifractal phenomena (e.g., measures in time series). In addition, while this approach is able to identify fluctuations existing in distributions, it does not indicate where they occur.

Detailed information about multifractal procedures can be found in Everstz and Mandelbrot (1992), Gutfraind et al. (1991), and Halsey et al. (1986). The essence of the multifractal formalism used in this analysis is as follows: First, the normalized concentration, P_{ini}, for each hour is determined by $P_{ini} = \frac{C_{ini}}{\sum C_{ini}}$, where C_{ini} is the PM_{10} concentration at time i. The series is then divided into nonoverlapping intervals of a certain time resolution, T. Each interval is characterized by a time resolution T, and the sum of normalized concentration in the interval, a probability mass function, $P_j(T)$. Twelve time resolutions, from 2^1 to 2^{12} hour, are considered in this study. A partition function, $M_q(T)$, of order q is calculated from the $P_j(T)$ values as

$$M_q = \sum_{j=1}^{n} P_j^q(T) \qquad (1)$$

where n is the total number of the intervals of size T, and q is a real number ranging from $-\infty$ to ∞. For multifractally distributed measures, the partition function scales with the time resolution as

$$M_q \propto T^{\tau(q)}, \qquad (2)$$

where $\tau(q)$ is the mass exponent of order q. The mass exponent for each q-value can be obtained by plotting log $M_q(T)$ vs. log T. The obtained $\tau(q)$ may be regarded as a characteristic function of the fractal behavior. If $\tau(q)$ versus q is a straight line, then the data set is monofractal. If, however, $\tau(q)$ versus q is a convex function, then the data set is multifractal.

An alternative and equivalent way to study the scaling properties of PM_{10} concentration time series is by considering their spectrum of singularities. It is assumed that in each interval, the mass probability function $P_j(T)$ increases with the size T as $P_j(T) \propto T^{\alpha}$, then the singularity exponent α is a

scaling property peculiar to the interval. The index α,therefore, characterizes singularities of different strengths and is called a local fractal dimension or a singularity index. It can be determined by Legendre transformation of the τ(*q*) curve (Everstz and Mandelbrot, 1992) as

$$\frac{d}{dq}[\tau(q)] = \alpha(q). \quad (3)$$

Now, corresponding to each α, one can identify a scaling exponent (or a fractal dimension) $f(\alpha)$ if one assumes that the number of intervals of size T with the same α. $N_T(\alpha)$, is related to the T as $N_T(\alpha) \propto T^{-f(\alpha)}$. Parameter $f(\alpha)$ can be calculated (Halsey et al., 1986) as

$$\tau(q) = q\alpha(q) - f(\alpha). \quad (4)$$

The shape and the extension of the $f(\alpha)$-curve contains significant information about the distribution characteristics of the examined data set. Varying q is a trick for exploring the different regions of α. For large and positive q, we are looking for small values of α; i.e., parts of the measure in which the P_{ini} values is high. For large and negative q, we study parts of the object for which the measure is very small and corresponds to the larger value of α. On the other hand, low values of $f(\alpha)$ characterize a rare occurrence of isolated peaks in a data sample, and high values of $f(\alpha)$ characterize a more frequent and dense appearance of data values. In general, the spectrum has a concave downward curvature, with a range of α-values increasing correspondingly to the increase in the heterogeneity of the distribution. For q = 0, we can deduce $\tau(0) = -D_0$, where D_0 is the fractal dimension of the support of our measure and is equal to 1.0 because we are dealing with a one-dimensional data support (1-D time series). This turns out to be the maximum possible value of f. For a homogeneous distribution, the $\tau(q)$-q curve becomes linear and then $f = \alpha = D_0$; *i.e.*, D_0 is also the fractal dimension of all the subsets.

2.4 Multifractal Cascade Model

A simple generalized Cantor set with two rescaling parameters (l_1 and l_2) and measure parameters (p_1 and p_2) was adopted to model the multifractal spectra of PM_{10} concentration time series. It should be noted that this model with $l_1 = l_2 = \frac{1}{2}$ has been used to simulate the fully developed turbulence (Meneveau and Sreenivasan, 1987) and the air pollutant concentration data (Anh et al., 2000). In this study, however, we assume $l_1 + l_2 = 1$ (because we are dealing with a one-dimensional data support) and $p_1 + p_2 = 1$, respectively. The two-scale Cantor set is constructed on an interval E of unit length, where $E_0 = E$, E_n contains 2^n subintervals obtained by dividing each subinterval of E_{n-1} into two different length intervals. The positive measure μ on C is defined as follows. We start with the original region which has measure 1 and size 1 (i.e., $E_0 = 1$ and $\mu_0 = 1$).At the second stage, the unit mass and size are split into p_1 and p_2 as well as l_1 and l_2, respectively. This defines μ_1,which has p_1 on one interval (l_1) and p_2 on the other interval (l_2). Continued in this way, the mass on each interval of E_n will be divided randomly into the proportions p_1 and p_2 between its two subinterval in E_{n+1}.

Accordingly, a sequence $\{\mu_n\}$ can be defined, and it will converge weakly to a limiting mass distribution μon C (Anh et al., 2000). Since for each $0 \le m \le n$, a number $\binom{n}{m}$ of the 2^n intervals of E_n have mass $p_1^m(p_2)^{n-m}$, it is apparent that the $P_j(T)$ is generated by a multiplicative cascade with a binomial generator characterized by a probability p_1.

In this study, the $P_j(T)$ is assumed to be generated by such a recursive process and its scaling is described by equation (2). To test the validity of the multiplicative cascade model for simulating the PM_{10} concentration data, both the $\tau(q)$ and $f(\alpha)$ functions are needed. For this generalized two-scale Cantor set, the analytic expressions for both the τ-q and f-α curves have been obtained by Halsey et al. (1986):

$$\tau = \frac{\ln(\frac{n}{m}-1) + q\ln(\frac{p_1}{p_2})}{\ln(\frac{l_1}{l_2})}, \tag{5}$$

$$\alpha = \frac{\ln p_1 + (\frac{n}{m}-1)\ln p_2}{\ln l_1 + (\frac{n}{m}-1)\ln l_2}, \tag{6}$$

$$f = \frac{(\frac{n}{m}-1)\ln(\frac{n}{m}-1) - (\frac{n}{m})\ln(\frac{n}{m})}{\ln l_1 + (\frac{n}{m}-1)\ln l_2}. \tag{7}$$

For any given q, the τ, α, and f could be determined by eliminating $\frac{n}{m}$ with the aid of the following equation:

$$\ln(\frac{n}{m})\ln(\frac{l_1}{l_2}) - \ln(\frac{n}{m}-1)\ln l_1 = q(\ln p_1 \ln l_2 - \ln p_2 \ln l_1).$$

It should be noted that because the equations (5) to (7) are highly nonlinear, the effects of l_1 and p_1 on both f-α and τ-q curves can be obtained only by solving those equations numerically. Our previous investigation (Lee et al., 2003b) has found that larger p_1 and smaller l_1 may produce more obvious curvature at the τ-q curve. Since higher non-linearity of the $\tau(q)$ curves is translated into a wider $f(\alpha)$ dispersion, then the more pronounced multifractal characteristics, p_1 and l_1, may be used to represent and compare the distribution's heterogeneity in the examined PM_{10} concentration time series. The distribution produced from binomial process may have tails extending to the right as p_1 becomes small. Accordingly, a smaller p_1 may correspond to larger coefficient of skewness of the PM_{10} concentration time series.

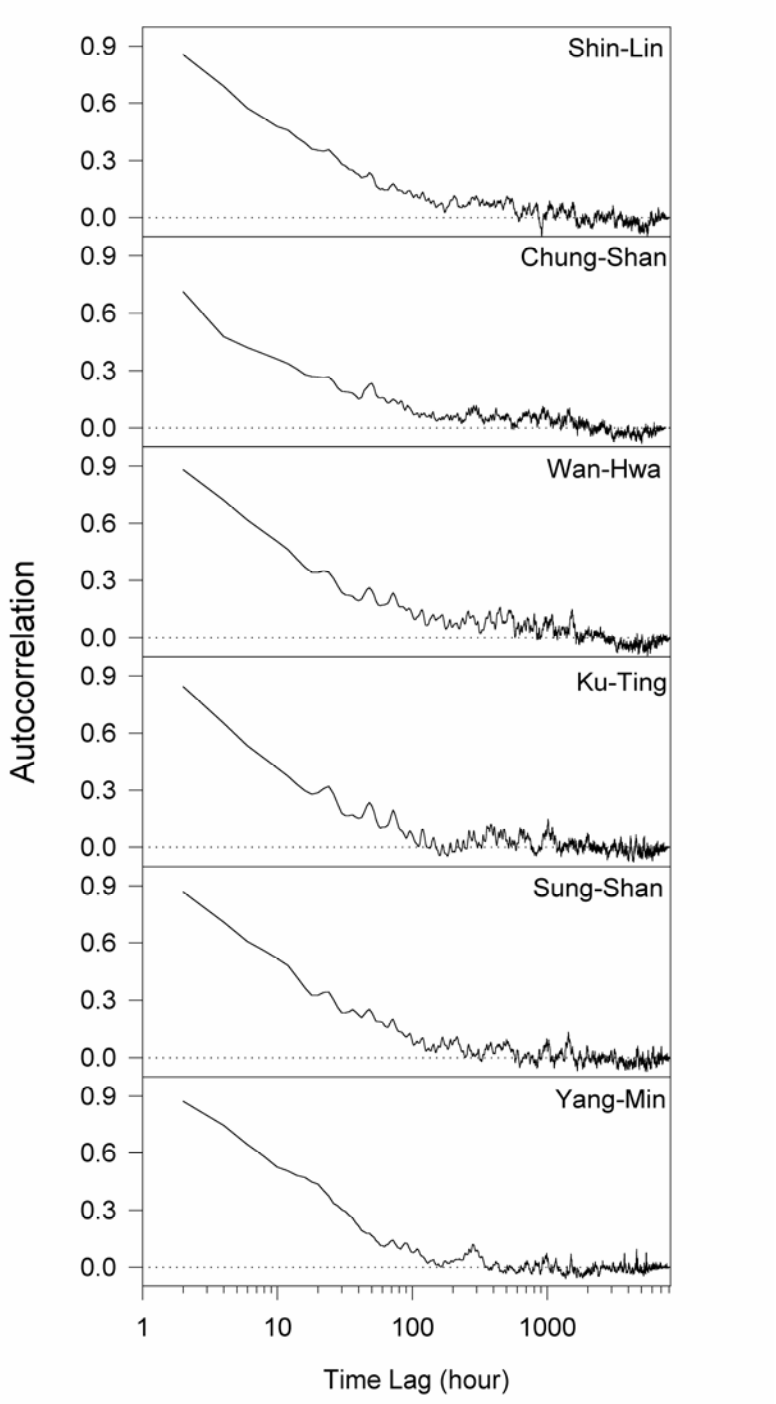

Figure 3. Autocorrelation function for all examined PM_{10} time series.

3. Results

3.1. Standard Statistical Characteristics

The standard statistical parameters estimated from the PM_{10} concentration time series are shown in Table 1. Among them, the coefficient of variation indicates the variability of PM_{10} concentration

Table 1. Basic statistical parameters and multifractal parameters for the examined PM_{10} time series.

air quality monitoring station	mean/ $\mu g/m^3$	coeff. of variation (%)	coeff. of skewness	p_1	l_1	Range of α
Shin-Lin	43.14	63.78	1.53	0.498	0.425	0.389
Chung-Shan	49.92	79.01	2.50	0.496	0.413	0.463
Wan-Hwa	45.85	74.85	2.04	0.497	0.408	0.507
Ku-Ting	43.71	70.20	1.85	0.498	0.418	0.444
Sung-Shan	42.19	63.83	1.60	0.498	0.420	0.427
Yang-Min	17.00	58.09	6.26	0.489	0.412	0.424

history. The coefficient of skewness measures the relative skewness of PM_{10} concentration frequency distribution; a positive value means that the distribution has a long tail extending to the right. As listed in Table 1, the mean values of all examined stations (except Yang-Min) are nearly equal, indicating that the spatial differences in PM_{10} concentration in Taipei are rather small. The generating sources and processes of PM_{10} are complex and the factors influencing the temporal distribution of PM_{10} may be considerable (van der Wal and Janssen, 2000). If the controlling factors (such as temperature and humidity) are distributed homogeneously in the examined monitoring stations, however, the resulting time structure of PM_{10} may correspond. The right-skewness degree of frequency distribution of the examined stations also correspond (except Yang-Min), but the concentration variability increases in the order of Yang-Min < Sung-Shan = Shin-Lin < Ku-Ting < Wan-Hwa < Chung-Shan. On the other hand, the autocorrelation spectra given in Figure 3 indicate that the autocorrelation decreases slowly in a manner that is different from an exponential decay. For Shin-Lin, Chung-Shan, Wan-Hwa, Ku-Ting, Sung-Shan, and Yang-Min station, a correlation exists up to about 600, 600, 600, 100, 300, and 300 hr, respectively. This slow decay in the autocorrelation function indicates a temporal persistence that may be related to self-similar properties in the time series.

3.2. Box Dimension

Figure 4 shows both D_B-T_h and C_S-T_h plots for all examined monitoring stations. As demonstrated in Figure 4(a), the plots could be roughly divided into two groups: one is made up of Shin-Lin, Sung-Shan, and Yang-Min, and the other includes Chung-Shan, Wan-Hwa, and Ku-Ting. Under a certain T_h value, the D_B values of Shin-Lin, Sung-Shan, and Yang-Min are larger than that of other monitoring stations. Above a certain T_h value, we get the opposite result. At low T_h (0-0.5 mean), the low decrease of D_B-T_h plot for Shin-Lin, Sung-Shan, and Yang-Min indicates that at low concentration, its pattern is more discrete and the persistence of occurrence is less continuous than at other examined monitoring stations. At high T_h (>3 mean), Shin-Lin, Sung-Shan, and Yang-Min still possess a less dense and continuous pattern due to smaller D_B and the rapid decrease of D_B-T_h plots at

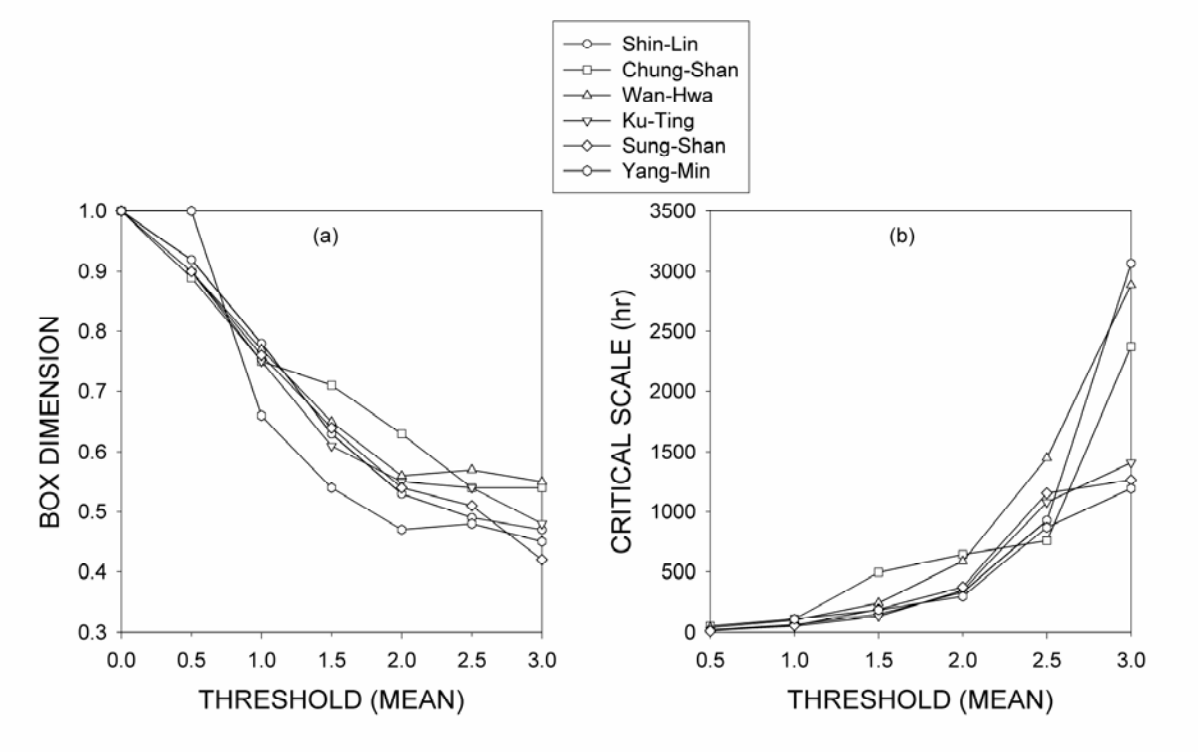

Figure 4. (a) D_B-T_h and (b) C_S-T_h plots of all examined PM_{10} time series.

intermediate T_h (1-3 mean). Thus, during a longer (or shorter) period, the high (or low) concentration events occur less readily at Shin-Lin, Sung-Shan, and Yang-Min than at other examined monitoring stations. Therefore, it may be concluded that the data of Shin-Lin, Sung-Shan, and Yang-Min are more concentrated on middle concentration regions but not at low and high ranges.

It is apparent that both D_B-T_h and C_S-T_h plots are closely related to the concentration variation of PM_{10}. Thus, it is interesting to discuss the correlation between the D_B-T_h (C_S-T_h) plot and the coefficient of variation. As mentioned earlier, when the PM_{10} data has larger D_B at low T_h and smaller D_B at high T_h, its distribution will concentrate on middle concentration regions but not at low and high ranges. In this case, it will possess a smaller temporal variation; i.e., a smaller coefficient of variation. Therefore, Shin-Lin, Sung-Shan, and Yang-Min stations may have smaller concentration variation when compared with other air monitoring stations. This result is consistent with the coefficients of variation shown in Table 1. Although both D_B-T_h and C_S-T_h plots are closely related to the coefficient of variation, it is noteworthy that the former provides a much deeper insight into data structure than the latter because it can present a more microscopic picture about the distribution of data set.

In addition to the relationship between D_B-T_h (or C_S-T_h) plots and the coefficient of variation, a closer look at the D_B-T_h plots and the corresponding coefficients of skewness in Table 1 reveals some interesting correlations. Generally, the coefficient of skewness gives a measure of the relative skewness of a distribution. For distributions that have tails extending to the right, the coefficient of skewness is positive and the distribution is called right-skewed. For a right-skewed distribution with a

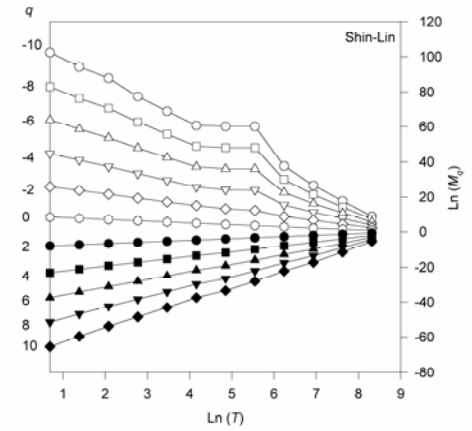

Figure 5.a

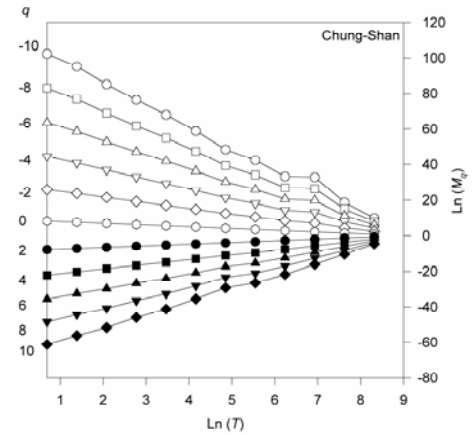

Figure 5.b

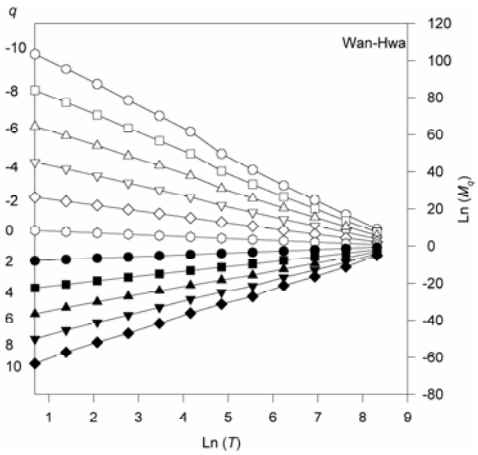

Figure 5.c

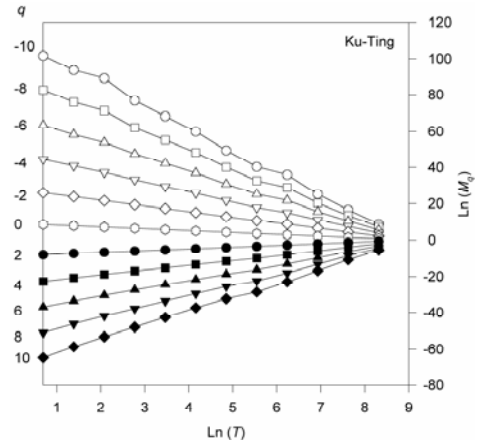

Figure 5.d

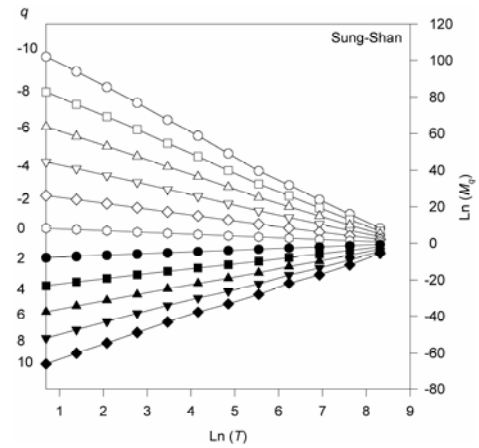

Figure 5.e

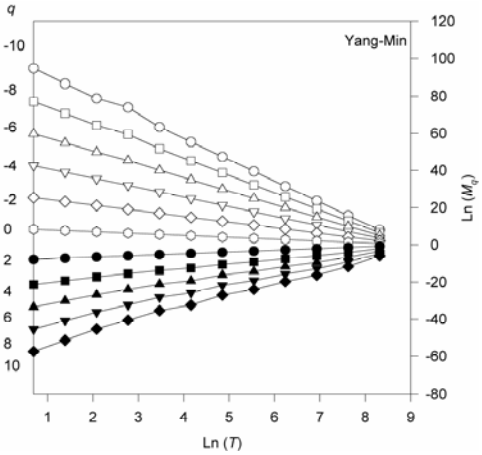

Figure 5.f

Figure 5. Qth-order moment of the normalized concentration M_q versus time scale T in a Ln-Ln scale at selected q-values for (a) Shin-Lin, (b) Chung-Shan, (c) Wan-Hwa, (d) Ku-Ting, (e) Sung-Shan, and (f) Yang-Min stations. The slope of the Ln M_q/Ln T line defines the exponent $\tau(q)$ (equation (2)).

single mode, the location of the mean is at the right side of the mode. Moreover, when the location of the mode moves more to the left, the larger is the coefficient of skewness. Accordingly, for a right-skewed distribution the D_B-T_h plot will show a sharp decrease when $T_h < 1$ mean and the larger the decreasing rate of D_B, the larger is the coefficient of skewness. As demonstrated in Figure 4(a), the D_B-T_h plots of all examined air pollutants show a pronounced decrease when $T_h < 1$ mean, indicating that the distributions of all examined PM_{10} time series are right-skewed. As shown in Table 1, the coefficient of skewness of Yang-Min station is the largest among the examined air monitoring stations, which is consistent with the decreasing rate of D_B at low threshold ($T_h < 1$ mean).

The existence of a close relationship between the D_B-T_h plots and the coefficients of variation and skewness may provide some basis for the validity of box counting technique used in this study. This also indicates that box counting is a useful approach to identify the temporal and spatial variation of PM_{10} data. Because scale invariance is closely related to the long-range dependence in the data set, it is not adequate to treat PM_{10} distribution as an independent stochastic process. This also indicates that the Poisson distribution, which assumes that the occurrence of the event is completely random in a certain time interval, is not appropriate.

Finally, it should be noted that the results of the monofractal analysis presented above show that a single dimension is insufficient to describe the scaling properties of the PM_{10} time series. It is found that the values of D_B (or C_S) decrease (or increase) as the T_h magnitude increases, indicating that different threshold intensities reveal different properties of the PM_{10} time series. Thus, a multidimensional fractal structure may be more suitable for describing the PM_{10} time series, an assumption that must be verified by using multifractal scaling analysis (Lee, 2002, Lee et al., 2003a; Ho et al., 2003).

3.3. Multifractal Scaling Analysis

Figure 5 shows plots of the qth-order moment, M_q, versus the time scale T in a log-log scale. All of these plots are close to being straight for $-10 \le q \le 10$, signifying that the studied PM_{10} time series can be regarded as multifractal measures. The observation of multifractal scaling in PM_{10} time series is encouraging because multifractal formalism has been successfully applied to systems as complex as turbulence, and it may also have great potential in modeling the complex structure of PM_{10}. Before this can be done, however, a physical interpretation must be made concerning which PM_{10} generating processes can lead to such multifractal characteristics. When applying a multiscaling approach to temporal clustering of earthquakes, multifractal characteristics are interpreted in terms of diffusive processes of stress in the Earth's crust (Godano et al., 1997). Moreover, multifractal characteristics in rainfall data have been explained with an assumption that a large-scale flux is successively broken into smaller and smaller cascades, each receiving an amount of the total flux specified by a multiplicative parameter (Olsson, 1996). On the other hand, multifractal characteristics in the stock market are interpreted with the random multiplicative process of market information (Ho et al., 2004).

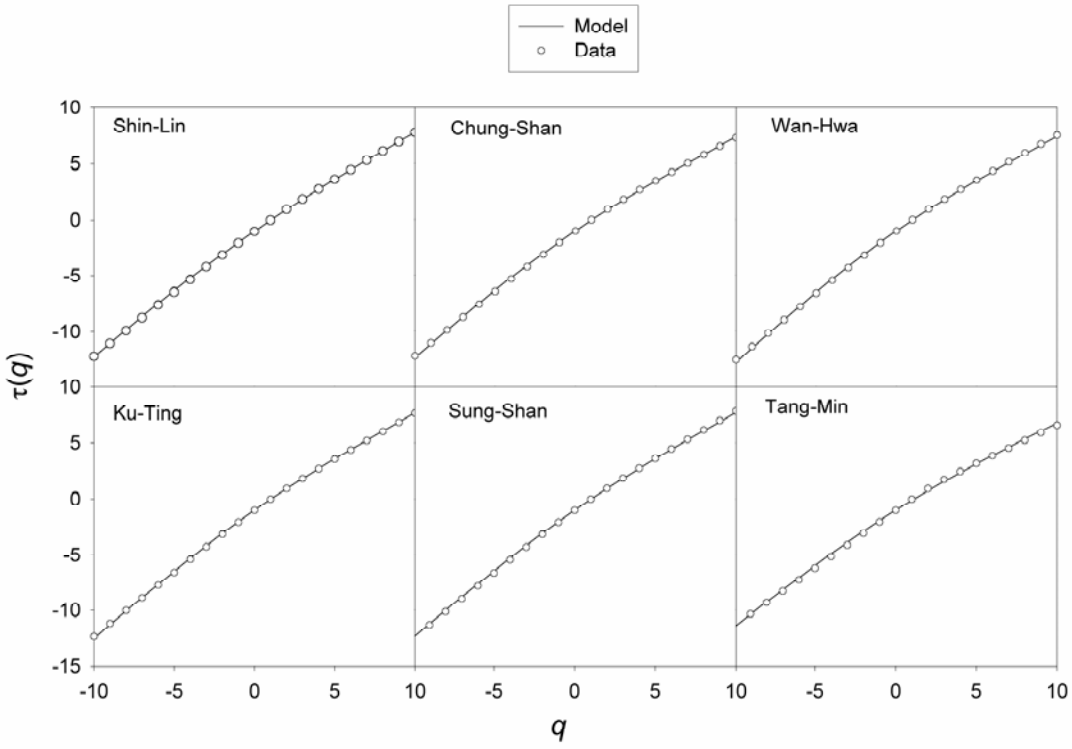

Figure 6. Fitting model to the experimental $\tau(q)$ curves on $-10 \le q \le 10$ for all examined PM_{10} time series.

It should be noted that the stochastic processes proposed for such systems to generate multifractal characteristics are closely related to the heart of turbulence, namely, the multiplicative cascade process. The only difference is the characteristic physical quantity that accompanies the stochastic processes. For earthquakes, rainfall, the stock market, and turbulence, the corresponding characteristic quantity is stress, water, market information, and energy, respectively. Accordingly, the multifractal characteristics in PM_{10} time series may be interpreted with the aid of random multiplicative process of PM_{10} concentration (Lee, 2002; Lee et al., 2003b).

Since multifractal characteristics indeed exist in all examined PM_{10} time series and can be viewed as a result of random multiplicative process, we next fit the multifractal cascade model to those experimental $\tau(q)$ curves. The values of p_1 and l_1 can be estimated by comparing the experimental $\tau(q)$ curves with the curves computed from model for a range of values of p_1 and l_1. In Figure 6, we show a comparison between the measured $\tau(q)$ curve and equation (6) for all examined monitoring stations. The agreement is remarkable and the estimated p_1 and l_1 parameters are shown in Table 1. It is worth mentioning that if we use equal scales, (*i.e.*, $l_1 = l_2 = 0.5$), no choice of p_1 would have been satisfactory. Since both p_1 and l_1 are determined, the corresponding $f(\alpha)$ curves can be obtained with the aid of equations (7) and (8). As demonstrated in Figure 7 and Table 1, the α range, then the multifractal characteristics (or the distribution's heterogeneity), increases in the order of Shin-Lin < Yang-Min < Sung-Shan < Ku-Ting < Chung-Shan < Wan-Hwa.

An important question is this: What is the physical meaning of p_1 and l_1 and can they provide a

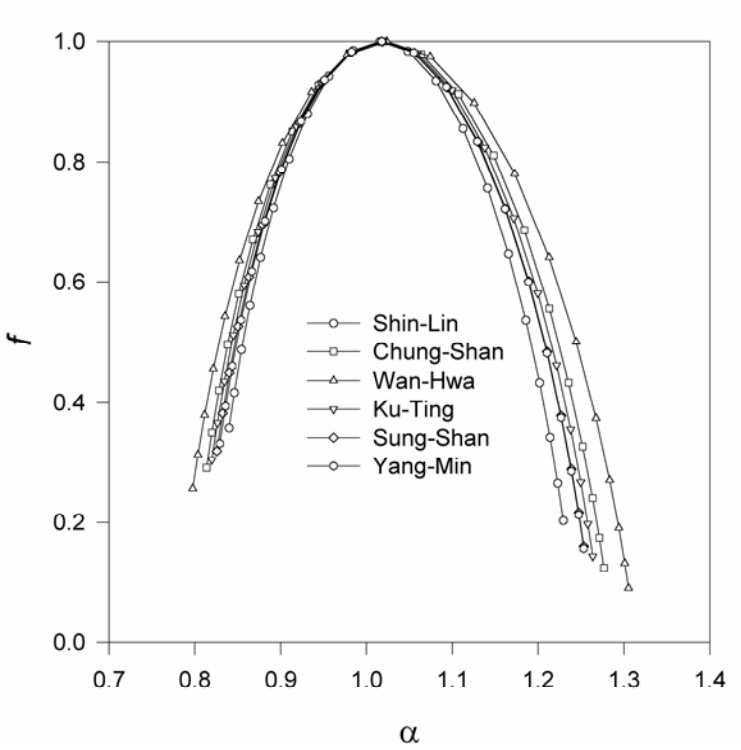

Figure 7. The plots of f vs. α for the examined PM_{10} time series.

useful means for comparison and classification of some statistical characteristics in different PM_{10} time series? It is well known that the distribution produced from binomial process may have tails extending to the right as p_1 becomes small. Accordingly, a smaller p_1 may correspond to larger coefficient of skewness. As demonstrated in Table 1, when $l_1 < \frac{1}{2}$, an inverse relationship is indeed found between p_1 and the coefficient of skewness.

Another interesting feature is the relationship between multifractal characteristics and the corresponding values of p_1 and l_1. As mentioned earlier, larger p_1 and smaller l_1 may correspond to stronger multifractal characteristics. Accordingly, as shown in Table 1 and Figure 7, the multifractal property (or the distribution's heterogeneity) of Chung-Shan, Ku-Ting, and Wan-Hwa may be more obvious than that of Shin-Lin, Sung-Shan, and Yang-Min. This result is similar to the results of Kravchenko et al. (1999), in that a significant correlation is observed between the coefficients of variation and the multifractal characteristics (or the range of α). However, it is noteworthy that on multifractal analysis, αdispersion strength represents the range between the high and low PM_{10} values. On the other hand, the number of high and low concentration data are related to the left and right parts of the $f(\alpha)$ spectrum, respectively. Since the variation of data distribution is determined by both the data values and their corresponding number, the range of α itself is insufficient to determine the relative variability when comparing different data sets. This result also indicates that the multifractal approach provides a much deeper insight into data structure than the coefficient of variation because

it can provide a more microscopic picture about the distribution of data set.

4. Discussion and Conclusions

Some standard statistical methods have been used to investigate the clustering properties of PM_{10} time series in Taipei. The autocorrelation of all PM_{10} time series does not decay to zero exponentially but in a slower manner. Multifractal analysis indicates that the PM_{10} time series could be viewed as multifractal measures that may be the result of a random multiplicative process. A simple two-scale Cantor set with unequal scales and weights was then presented for the PM_{10} time series. This model fits well with the entire spectrum of scaling exponents for the examined PM_{10} time series. The validity of the fractal approach is supported by the existence of a close relationship between the practical implications of D_B-T_h plots (or l_1 and p_1) and the coefficient of variation and skewness.

The results obtained from this analysis are encouraging and some practical implications can be addressed. First, the existence of scale-invariant properties suggests that scale shifts are theoretically possible. A shift from the time scales defined by the data collection systems to any scale needed for a particular forecasting or estimation problem would be feasible. If the variability or uncertainty associated with these processes could be identified, prediction based on scale-invariant properties may be useful. Second, multifractal analysis may be used to examine the relationship between the PM_{10} and weather data by transferring the data into useful multifractal parameters, namely, the l_1 and p_1. This may be viewed as an additional recognition method when using weather data as input to forecast the PM_{10}.

However, some important comments on the limitations of fractal analysis should also be addressed. First, to make the multifractal cascade model an efficient tool for characterization, analysis, and comparison of air pollutant concentration temporal characteristics, a clear relationship between the model parameters and traditional statistical quantities is needed. To do this, however, the connection between the multifractal parameters and traditional statistical properties must first be identified because mutifractal parameters are functions of model parameters. Yet comparing the multifractal cascade model with the use of only the coefficients of variation and skewness may be unfair because statistical analysis of the PM_{10} data collected at each air quality monitoring station routinely reveals a high variation of concentration, right-skewed frequency distribution, and long-term memory. A more appropriate approach would be to compare it with the use of the aforementioned three characteristics of PM_{10} data. As shown in Figure 2, the long-range dependence does exist in the examined PM_{10} time series. It is found, however, that the relationship between the model parameters and the long-range dependence in the examined PM_{10} data is difficult to identify. To make the multifractal approach an efficient tool for analysis of air pollutant concentration data, it may be an important task in the future to reach a more concise conclusion for the connection between the model parameters and the long-range dependence of data set.

Finally, although the two-scale Cantor set performed here can be regarded as a convenient model for PM_{10} distribution in time, it is difficult to conclude that the PM10 distribution is governed exactly by a single two-scale Cantor set with p_1 and l_1 as parameters, especially if one is interested in modeling correctly the scaling properties of PM_{10} time series. Still, the two-scale Cantor set is merely one of the possible forms to generate multiplicative cascades. To achieve more confidence when we adopt the multifractal cascade model to simulate and predict the air pollutant concentration data, a further comparison among the different multiplicative cascade models is still needed to identify which one is the best.

Acknowledgement

The work is supported by National Science Council (Taiwan, ROC) grant NSC90-2211-E238-001 and the data are supplied by the Environmental Protection Administration of ROC.

References

Anh V. V., Lam K. C., Leung Y. and Tieng Q. M. (2000), Multifractal Analysis of Hong Kong Air Quality Data. Environmetrics 11: 139-149.

Boxian W. and Lye L.M. (1994), Identification of Temporal Scaling Behaviour of Flood: a Study of Fractals. Fractals 2: 283-286.

Buczkowski S., Hildgen P. and Cartilier L. (1998), Measurements of Fractal Dimension by Box-counting: a Critical Analysis of Data Scatter. Physica A 252: 23-34.

Dockery D. W., Pope C. A., Xu X., Spengler J. D., Ware J. H., Fay M. E., Ferris B. G. and Speizer F. E. (1993), An Association Between Air Pollution and Mortality in Six U.S. Cities. New Engl. J. Mde. 329: 1753-1759.

Ho D. S., Lee C. K., Wang C. C. and Chuang M. (2004), Scaling Characteristics in the Taiwan Stock Market. Physica A 332: 448-460.

Evertsz C. J. G. and Mandelbrot B. B. (1992), Multifractal Measures (Appendix B). In Chaos and fractals, Peitgen, H. O. et al. (eds.); Springer-Verlag, New York, 922-953.

Godano C., Alonzo, M. L. and Vilardo G. (1997), Multifractal Approach to Time Clustering of Earthquakes. Application to Mt. Vesuvio Seismicity. Pure Appl. Geophys. 149:375-390.

Gutfraind R., Sheintuch M. and Avnir D. (1991), Fractal and Multifractal Analysis of the Sensitivity of Catalytic Reactions to Catalyst Structure. J. Chem. Phys. 95: 6100-6111.

Halsey T. C., Jensen M. H., Kadanoff L. P., Procaccia I. and Shraiman B. I. (1986), Fractal Measures and Their Singularities: the Characterization of Strange Sets. Phys. Rev. A 33: 1141-1151.

Klement S. and Kratky K. W. (1997), Multifractal Analysis of Airborne Particle Count Data: the Influence of Data Preprocessing. Aerosol Sci. Technol. 26: 12-20.

Klement S., Nittman J., Kratky K. W. and Actio W. P. (1994), Multifractal Analysis of Airborne

Microcontamination Particles. Aerosol Sci. Technol. 20: 100-110.

Kravchenko A. N., Boast C. W. and Bullock D. G. (1999), Multifractal Analysis of Soil Spatial Variability. Agron. J. 91: 1033-1041.

Lee C. K. (2002), Multifractal Characteristics in Air Pollutant Concentration Time Series. Water, Air & Soil Pollution 135: 389-409.

Lee C. K., Ho D. S., Yu C. C. and Wang C. C. (2003a), Fractal Analysis of Temporal Variation of Air Pollutant Concentration by Box Counting. Environ. Modelling & Software 18: 243-251.

Lee C. K., Ho D. S., Yu C. C., Wang C. C. and Hsiao T. H. (2003b), Simple Multifractal Cascade Model for the Air Pollutant Concentration Time Series. Environmetrics 14: 255-269.

Lee C. K. and Lee S. L. (1996), Heterogeneity of Surfaces and Materials, As Reflected in Multifractal Analysis. Heterogen. Chem. Rev. 3: 269-302.

Lin S. C., Liu C. L. and Lee T. Y. (1999), Fractality of Rainfall: Identification of Temporal Scaling Law. Fractals 7: 123-131.

Lovejoy S. and Schertzer D. (1990), Multifractals, University Classes and Satellite and Radar Measurements of Clouds and Rain Fields. J. Geophys. Res. 95: 2021-2034.

Lovejoy S., Schertzer D. and Tsonis A. A. (1987), Functional Box-counting and Multiple Elliptical Dimensions of Rain. Science 235: 1036-1038.

Meneveau C. and Sreenivasan K. R. (1987), Simple Multifractal Cascade Model for Fully Developed Turbulence. Phys. Rev. Lett. 59: 1424-1427.

Olsson J. (1996), Validity and Applicability of a Scale-independent, Multifractal Relationship for Rainfall. Atmos. Res. 42: 53-65.

Olsson J., Niemczynowicz J. and Berndtsson R. (1993), Fractal Analysis of High-resolution Rainfall Time Series. J. Geophys. Res. 98: 23265-23274.

Olsson J., Niemczynowicz J., Berndtsson R. and Larson M. (1992), An Analysis of the Rainfall Time Structure by Box Counting-Some Practical Implications. J. Hydrol. 137: 261-277.

Schertzer D. and Lovejoy S. (1987), Physical Modeling and Analysis of Rain and Clouds by Anisotropic Scaling Multiplicative Processes. J. Geophys. Res. 92: 9693-9714.

van der Wal J. T. and Janssen L. H. J. M. (2000), Analysis of Spatial and Temporal Variations of PM10 Concentrations in the Netherlands Using Kalman Filtering. Atmos. Environ. 34: 3675-3687.

Received for review, November 02, 2003
Accepted, February 19, 2004