Chaos in Air Pollutant Concentration (APC) Time Series

Chung-Kung Lee1*, Shu-Chen Lin2

1 Green Environment R&D Center and Department of Environmental Engineering, Vanung University, Chungli, Taiwan, ROC
2 Department of Resources and Environment, Leader University, Tainan, Taiwan, ROC

Abstract

Three chaotic indicators, namely the correlation dimension, the Lyapunov exponent, and the Kolmogorov entropy, are estimated for one-year long hourly average NO (nitrogen monoxide), CO (carbon monoxide), SO2 (sulfur dioxide), PM10 (particles with an aerodynamic diameter of approximately 10 μm or less), and NO2 (nitrogen dioxide) concentration to examine the possible chaotic characteristics in the air pollutant concentration (APC) time series. The presence of chaos in the examined APC time series is evident with the low correlation dimensions (3.42-4.71), the positive values of the largest Lyapunov exponent (0.128-0.427), and the positive Kolmogorov entropies (0.628-0.737). Since the existence of multifractal characteristics in the above time series has been confirmed in our previous investigations, the presence of chaotic behavior identified in the current study suggests the possibility of a chaotic multifractal approach for APC time series characterization. Some problems concerning the applicability of chaos analysis in air pollution are also discussed.

Keywords: Air Pollutants; Multifractal.

INTRODUCTION

Air quality changes related to human action can be investigated by long-term and large-area monitoring. The collected air quality data are often recorded as APC (air pollutant concentration) time series and are characterized by many large fluctuations without obvious autocorrelation. Based on these time series, the trends in the APC data are often investigated by statistical analysis to facilitate good air quality management. However, both the accuracy and the reliability of these statistical analyses may be strongly affected by our fundamental knowledge of the complex temporal structure of the APC history at each (monitoring) station (Horowitz and Barakat, 1979; Ho et al., 2004; Wang and Chen, 2008; Wang et al., 2008; Yang et al., 2008).
In our previous investigations (Lee, 2002; Lee et al., 2003a; Lee et al., 2003b; Lee et al., 2006a), some standard statistical methods have been adopted to examine the possible scale-invariant behavior and the clustering characteristics in the APC time series. It is found that all the examined APC time series exhibit the characteristic right-skewed unimodal frequency distribution that can be well represented by the log-normal model (Lee, 2002). Furthermore, the auto-correlation function does not decay to zero exponentially but in a slower manner, indicating the possible existence of a cluster structure (Lee, 2002; Lee et al., 2003b; Lee et al., 2006a). A mono-fractal analysis is then performed by the box counting method. Scale invariance is found in APC time series and the box dimension is shown to be a decreasing function of the threshold APC level, implying the possible presence of multifractal characteristics. To verify this hypothesis, the APC time series data is transferred into a useful compact form through the moment scaling analysis, namely, the $\tau(q)-q$ (where $\tau(q)$ is the scaling exponent of the $q$th-order moments of a given probability distribution) and $f(\alpha)-\alpha$ (the spectrum of singularities, where $f(\alpha)$ is the Legendre transform of $\tau(q)$) plots. The presence of multifractal characteristics is confirmed by the deviation from linearity in the $\tau(q)-q$ plots and the wide distribution in the $f(\alpha)-\alpha$ plots (Lee, 2002; Lee et al., 2003b; Lee et al., 2006a). It is concluded that the origin of both the pronounced right-skewness and multifractal phenomena in APC time series may be interpreted in terms of a random multiplicative process. Since multifractal characteristics indeed existed in all examined APC time series, a simple two-scale Cantor set with unequal scales and weights is presented for the APC time series (Lee et al., 2003b; Lee et al., 2006a). It is revealed that this model fits remarkably well with the entire spectrum of scaling exponents for the examined APC time series.

On the other hand, although studies conducted over the past decades on the APC time series have indicated no evidence of a deterministic behavior, it has been gradually realized that the seemingly irregular-looking dynamic behavior of air pollutant could be the result of a simple deterministic system influenced by only a few non-linear interdependent variables with sensitive dependence on initial conditions, namely, chaos. The papers by Lee et al. (1994) and Raga and Le Moyne (1996) have shown the possible presence of chaotic dynamics in the hourly average ozone concentration data. Moreover, Chen et al. (1998) and Kocak et al. (2000) have successfully performed a non-parametric short-term prediction by using the chaos theory. Recently, Sivakumar et al. (2007) also indicates the existence of nonlinear and low-dimensional deterministic behavior in the daily air quality index time series. Nonlinearity in NO2 and CO time series is detected by Kumar et al. (2008) with the Volterra–Wiener–Korenberg (VWK) series approach (Barahona and Poon, 1996). The numerical titration technique (Poon and Barahona, 2001) further reveals that the
dynamics of NO₂ and CO is indeed governed by deterministic chaos. However, none of the past studies identifies the coexistence of chaos and fractal nature in the same APC time series. If positive evidence of the coexistence of chaos and fractal behavior can be provided, the APC time series characterization can be viewed from a new perspective: the chaotic multifractal approach, as reported by Sivakumar (2001) for rainfall characterization. Therefore, it is an interesting task to examine possible presence of a chaotic nature in the APC time series that has been confirmed as having multifractal characteristics.

**ANALYSIS AND RESULTS**

In the present study, we analyze the hourly average APC data collected at the Chung-Shan air quality monitoring station, Taipei (Taiwan), from January 1998 to December 1998, to investigate the presence (or absence) of chaos and hence, the possibility of a chaotic multifractal approach for APC time series characterization. This station is located in a heavily populated area in metropolis, and is intended to provide information pertaining to human exposure. A map with the location of Chung-Shan air monitoring station is demonstrated in Fig. 1. The selected air pollutants include primary pollutants NO, CO, and SO₂, and secondary pollutants PM10 and NO₂. It is noteworthy that multifractal characteristics in the above time series have been detected with moment scaling analysis (Lee, 2002; Lee *et al.*, 2003a and 2003b). Some details of the measurement instruments used to detect the above pollutants are listed elsewhere (Lee *et al.*, 2003b). Our previous investigation (Lee, 2002) finds that most examined APC time series in Taiwan exhibit obvious annual periodicity due to the systematic variations in response to seasonal and other factors and the statistical characteristics can be extracted from the data collected over one year length. Accordingly, one-year long of hourly average values are used in this study to examine the chaos characteristics of APC time series. Although a year's time consists of 8760 hours, only about 8400 readings for each pollutant are collected due to the instrument calibration and maintenance. However, the missing observations seem to be evenly distributed throughout the year. Although the missing data may affect the quantitative results of chaos analysis, we still prefer to use the original data to make a qualitative identification of the chaos characteristics of these time series. The reason is that any data preprocessing may strongly affect the results of statistical analysis and make the interpretation of the result complex (Klement and Kratky, 1997). Moreover, the fractal analysis made in our previous investigations (Lee, 2002; Lee *et al.*, 2003a and 2003b; Lee *et al.*, 2006a) also indicated that the effect of missing data on the qualitative conclusions of fractal analysis was insignificant. In fact, other factors such as time series length and noise may also affect the estimation of chaotic indicators (Sivakumar, 2000). Therefore, further investigations to examine the
influence of time series length and noise on the results of chaos analyses are still needed.

There are a large number of methods available in the literature to identify the existence of chaos in a time series, among them the correlation dimension (Grassberger and Procaccia, 1983a and 1983b), the Lyapunov exponent (Wolf et al., 1985), and the Kolmogorov entropy (Grassberger and Procaccia, 1983c) methods have been widely employed. Thus, in the present study these three methods are applied to the examination of the presence of chaos in the APC time series. The algorithms of these methods use the phase-space reconstruction of the time series. In general, a dynamic system can be described by a phase-space diagram whose trajectories describe the evolution of the dynamical system from some known initial states through time. In dissipative systems, in which the energy is not conserved, the trajectories eventually converge to some subspace regardless of the initial conditions. This subspace is called the attractor of the system and has a topological dimension less than or equal to the Euclidean dimension $m$ of the phase-space it lies in. If the dynamic system is very sensitive to the initial conditions, the attractor would have a non-integer dimension. Such an attractor is called ‘strange attractor’, and the system that contains a strange attractor is called a chaotic dynamic system. A method for reconstructing a phase-space from a time series with time delays is initiated by Packard et al. (1980) and put on a firm mathematical basis by Takens (1981). According to Takens’ time-delay embedding theorem (Takens, 1981), if $x(t)$ is a scalar time series in discrete time that are obtained from a continuous time multidimensional deterministic system with an attractor contained in a manifold of dimension $d$, there exists an embedding dimension $m \leq 2d + 1$ such that the vectors with time-delayed coordinate

$$X_t = \{x_t, x_{t+\tau}, x_{t+2\tau}, \ldots, x_{t+(m-1)\tau}\}$$

(where $t = 1, 2, \ldots, N - (m - 1)\tau/\Delta t$; $\tau$ is a delay time taken to be a suitable multiple of the sampling time $\Delta t$) will trace out a trajectory that represents a smooth coordinate transformation of the original trajectory of the system. Therefore, the trajectory of the delay vectors will have the same topological dimension as the underlying attractor of the dynamical system.

**Correlation dimension**

For an $m$-dimensional phase-space, the correlation function $C(r)$ is defined by
Grassberger and Procaccia (1983a, b) as

$$C(r) = \lim_{N \to \infty} \frac{2}{N(N-1)} \sum_{i \neq j}^{N} H(r - |X_i - X_j|)$$

(1)

where $H$ is the Heaviside step function, with $H(u) = 1$ for $u > 0$ and $H(u) = 0$ for $u \leq 0$, where $u = r - |X_i - X_j|$, $r$ is the radius of the sphere centered on $X_i$ or $X_j$, and $N$ is the number of data points. If the attractor for the time series data exists, then, for positive values of $r$, $C(r)$ is related to the radius $r$ by the following relation:

$$C(r) \approx a r^\nu$$

(2)

where $a$ is a constant and $\nu$ is the correlation exponent or the slope of the log $C(r)$ versus log $r$ plot. If correlation exponent is saturated with an increase in the embedding dimension $m$, then the system is generally considered to exhibit chaos. The saturation value of the correlation exponent is defined as the correlation dimension of the attractor, and the nearest integer above the saturation value provides the minimum number of the embedding dimensions of the phase-space required to model the dynamics of the attractor. For random processes, $\nu$ varies linearly with the increasing embedding dimension without arriving at a saturation value.

One typical plot for the relationship between the correlation function $C(r)$ and the radius $r$ on log-log scale with the embedding dimension $m$ from 2 to 20 is shown in Fig. 2 for SO$_2$. For each $m$, this figure indicates a clear scaling region that allows fairly accurate estimation of the correlation exponents. The dependence of the correlation exponents on the embedding dimensions for all examined APC time series is shown in Fig. 3. As demonstrated in Fig. 3, the correlation exponent increases with the increasing embedding dimension up to a certain value, and then saturates beyond that value, which may be taken to be an indication of deterministic dynamics. The saturation values of the correlation exponent (or correlation dimension) for NO, CO, SO$_2$, PM$_{10}$, and NO$_2$ are estimated as $3.42 \pm 0.19$, $4.71 \pm 0.17$, $3.98 \pm 0.20$, $4.32 \pm 0.25$, and $4.25 \pm 0.25$, respectively. The low correlation dimensions indicate that these time series exhibit low-dimensional chaotic behavior. As the nearest integer above the correlation dimension value generally provides the number of dominant variables influencing the dynamics of the underlying system, the correlation dimensions for the five time series indicate that the minimum number of variables essential to modeling the dynamics of NO, CO, SO$_2$, PM$_{10}$, and NO$_2$ process are 4, 5, 4, 5, and 5, respectively. However, it is difficult to give comments on the most probable variables in each case, because the concentrations of an air pollutant observed in a city often are influenced by hundreds or thousands of sources in the area, atmospheric variables, dilution and chemical reactions in the atmosphere, interaction with biological
systems, and other phenomena.

**Fig. 2.** Log $C(r)$ versus log $r$ plots for SO$_2$ time series.

**Fig. 3.** The variation of the correlation exponent with the embedding dimensions for the examined APC time series.

**Lyapunov exponent**

The second measure of the chaotic nature in the APC time series is the Lyapunov exponent which gives the average exponential rate of divergence or convergence of the nearby orbits in the phase-space. Because the presence of a positive Lyapunov exponent implies the divergence of the nearby trajectories, a system having at least one positive Lyapunov exponent is often considered to be chaotic. In this study, the algorithm and the computer program given by Wolf et al. (1985) which gives the Lyapunov exponents are adopted. The largest Lyapunov exponent $\lambda_1$ is defined as

$$\lambda_1 = \frac{1}{N_m \Delta t} \sum_{j=1}^{M} \log_2 \frac{L'(t_j)}{L(t_j-\Delta t)}$$

(3)

where $\Delta t$ is the time interval between two successive observations, $M$ is the number of replacement steps, $N_m$ is the total number of points in the sequence $(X_t)$, $L(t_j-1)$ is the Euclidean distance between the point $\{x(t_j), x(t_j+1), ..., x(t_j+(m-1)\Delta t)\}$ and its nearest neighbor, and $L'(t_j)$ is the evolved length of $L(t_j-1)$ at a time $t_j$ (Jayawardena and Lai, 1994). When $\lambda_1 > 0$, it means that the time series has at least one positive Lyapunov exponent and it is chaotic. For $\lambda_1 \leq 0$ and $\lambda_1 = \infty$, the time series corresponds to a regular motion process (such as periodic systems) and a stochastic process, respectively. For the purpose of detecting chaos in a time series, it is not necessary to determine all the Lyapunov exponents. In the calculations based on the algorithm of Wolf et al. (1985), it is found that all the values of $\lambda_1$ are positive and finite over a range of $m$ between 1 and 10. The mean of a series of $\lambda_1$ generated in different dimensional phase-spaces from $m = 1$ to 10 is taken as an estimated largest Lyapunov exponent for each APC time series (see Table 1). It is evident that all the examined APC time series can be
Table 1. The dependence of the largest Lyapunov exponent on the embedding dimension for the examined APC time series.

<table>
<thead>
<tr>
<th>air pollutant</th>
<th>embedding dimension</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>1</td>
<td>0.577</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.545</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.521</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.517</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.521</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.521</td>
</tr>
<tr>
<td>CO</td>
<td>1</td>
<td>0.374</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.219</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.219</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.219</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.219</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.219</td>
</tr>
<tr>
<td>SO2</td>
<td>1</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.113</td>
</tr>
<tr>
<td>PM10</td>
<td>1</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.048</td>
</tr>
<tr>
<td>NO2</td>
<td>1</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.076</td>
</tr>
</tbody>
</table>

regarded as chaotic series, because the $\lambda_1$ value of each air pollutant is positive.

**Kolmogorov entropy**

The Kolmogorov entropy of a time series gives a lower bound to the sum of the positive Lyapunov exponents. Since the calculation of the Kolmogorov entropy $K$ is difficult, an approximate estimation for the value of Kolmogorov entropy is usually conducted with the aid of the second order Renyi entropy $K_2$, which may be estimated with the distance (in log-log coordinates) between successive correlation curves $C_m(r)$ and $C_{m+1}(r)$ (Grassberger and Procaccia, 1983c), i.e.,

$$K_2(m) \equiv \lim_{r \to 0} \frac{1}{\Delta r} \left\{ \log[C_m(r)] - \log[C_{m+1}(r)] \right\}$$

(4)

and

$$K_2 \equiv \lim_{m \to \infty} [K_2(m)]$$

(5)

The $K_2$ entropy and Kolmogorov entropy are thought to have the same qualitative behavior, i.e., zero, positive and finite, and infinite values corresponding to a regular system, a chaotic system, and a stochastic process, respectively.

For a series of embedding dimensions, Eq. (4) is used to evaluate the quantity $K_2(m)$ for each APC time series. The dependence of the $K_2$ entropy on the embedding dimension of $m = 2$ to 19 is listed in Table 2. Here the chaotic behavior in all APC time series is evident with the positive and finite $K_2$ values.
DISCUSSION AND CONCLUSIONS

For the above results, three important comments should be addressed. Firstly, since the multifractal characteristics in the time series used in this study have been detected in our previous investigations (Lee, 2002; Lee et al., 2003a and 2003b), the results shown here provide a positive evidence for the coexistence of multifractal and chaotic behaviors in the APC time series. It is well known that multifractal behavior is frequently associated with systems where the underlying physics is governed by a random multiplicative process (Olsson and Niemczynowicz, 1996; Godano et al., 1997; Ho et al., 2004; Lee et al., 2006b). However, the existence of chaos identified in this study indicates that multifractal approaches may provide positive evidence of a multifractal nature not only in stochastic processes but also in chaotic processes. A possible implication of this may be that the APC data characterization can be viewed from a new perspective, i.e., the chaotic multifractal approach. However, to make chaotic multifractal model an efficient tool for characterization, analysis, and comparison of the APC temporal characteristics, a clear relationship between both multifractal and chaos parameters and traditional statistical quantities is needed. In general, statistical analysis of the APC data collected at each air quality monitoring station routinely reveals high variation of concentration, right-skewed frequency distribution, and long term memory. In our previous investigation (Lee et al., 2003b), the relationship between the...
coefficients of variation and skewness and multifractal parameters has been well established. However, it is found that the correlation between the multifractal parameters and the long-range dependence in the examined APC data is difficult to identify, although it is well known that the existence of multifractal characteristics is closely related to the long-range dependence in the data set. Therefore, it is an interesting and promising task in the future to make the relationship between the multifractal parameters and the long-range dependence of APC data set more transparently relevant as well as to establish the relationship between the chaos indicators and the above mentioned three traditional statistical quantities (i.e., the coefficients of both variation and skewness and the long term memory). Second, although conceptually simple, the estimation of the chaotic parameters from a time series may be significantly influenced by the size of the sample, the delay time, and the presence of noise (Sivakumar, 2000). Therefore, it is still necessary to conduct further investigations on the presence of a chaotic nature in the APC time series using more APC data and other chaos identification methods, in order to provide a more solid basis for the application of chaos theory on the APC time series characterization. Some recent developments in the field of nonlinear dynamics may provide an insight into the chaotic nature of air pollutants. For instance, Barahona and Poon (1996) have developed a Volterra–Wiener–Korenberg (VWK) series approach to detect nonlinearity in a time series. This approach is able to detect nonlinearity even when data are heavily contaminated with noise, or strong periodicity is present. Poon and Barahona (2001) developed a novel numerical titration technique to detect chaos in a non-linear time series, even if the time series is short and noisy in nature. Cao (1997) has proposed a modified form of the false nearest-neighbourhood (FNN) method for ascertaining the dimensionality of the system. This method overcomes many shortcomings of the widely used FNN method and is equally powerful when the number of data points is less (~1000 points). Recently, the chaotic nature of NO2 and CO time series are clearly detected with the aid of VWK approach, numerical titration technique, and Cao’s method by Kumar et al. (2008). Finally, the similarity of the chaos nature involved in APC time series and rainfall is interesting, although their microscopic generating processes may be different in a fundamental way. Air pollution has an external forcing (emissions of pollutants) which has a non-trivial time structure connected with human activity, and is bound to generate long term correlation or periodicity in the data. Rainfall is also correlated with periodic forcing, but of completely different origin. Therefore, the time series of rainfall and APC would have significantly different properties although the chaos characteristics in their generating processes are similar.

ACKNOWLEDGEMENTS

The work is supported by the grant
REFERENCES


Received for review, September 8, 2008
Accepted, November 10, 2008