Sains Malaysiana 41(11)(2012): 1367–1376

Existence of Long Memory in Ozone Time Series
(Kewujudan Ingatan-Panjang dalam Siri Masa Ozon)

MUZIRAH MUSA* & KAMARULZAMAN IBRAHIM

ABSTRACT

Long-memory is often observed in time series data. The existence of long-memory in a data set implies that the successive data points are strongly correlated i.e. they remain persistent for quite some time. A commonly used approach in modelling the time series data such as the Box and Jenkins models are no longer appropriate since the assumption of stationary is not satisfied. Thus, the scaling analysis is particularly suitable to be used for identifying the existence of long-memory as well as the extent of persistent data. In this study, an analysis was carried out on the observed daily mean per hour of ozone concentration that were available at six monitoring stations located in the urban areas of Peninsular Malaysia from 1998 to 2006. In order to investigate the existence of long-memory, a preliminary analysis was done based on plots of autocorrelation function (ACF) of the observed data. Scaling analysis involving five methods which included rescaled range, rescaled variance, dispersional, linear and bridge detrending techniques of scaled windowed variance were applied to estimate the Hurst *coefficient (H) at each station. The results revealed that the ACF plots indicated a slow decay as the number lag increased. Based on the scaling analysis, the estimated H values lay within 0.7 and 0.9, indicating the existence of long-memory in the ozone time series data. In addition, it was also found that the data were persistent for the period of up to 150 days.*

Keywords: Hurst *coefficient; long-memory; ozone; scaling analysis*

ABSTRAK

Ingatan-panjang sering diperhatikan dalam data siri masa. Kewujudan ingatan-panjang dalam set data menunjukkan bahawa titik-titik data yang berturut adalah amat berkait rapat dan berterusan dalam suatu tempoh. Satu pendekatan yang biasa digunakan dalam pemodelan data siri masa seperti model Box dan Jenkins tidak lagi sesuai berikutan andaian kepegunan tidak dipenuhi. Oleh itu, analisis penskalaan sangat sesuai untuk digunakan bagi mengenal pasti kewujudan ingatan-panjang serta sifat keberterusan data. Dalam kajian ini, analisis dijalankan ke atas data cerapan purata harian per jam siri masa ozon yang diperoleh daripada enam stesen pemantauan yang terletak di kawasan Bandar di Semenanjung Malaysia dari tahun 1998-2006. Bagi tujuan penyiasatan kewujudan ingatan-panjang, analisis awal dilakukan berdasarkan kepada plot fungsi autokorelasi (ACF) bagi data yang dicerap. Analisis penskalaan yang terdiri daripada lima kaedah penskalaan yang berbeza yang merangkupi julat penskalaan semula, varians penskalaan semula, penyerakan, teknik penyahan tren linear dan jambatan dalam penskalaan varians tertingkap (SWV) digunakan untuk menganggar pekali Hurst *(H) bagi setiap stesen. Keputusan menunjukkan bahawa plot ACF siri data menyusut dengan lambat selaras peningkatan lag. Berdasarkan analisis penskalaan, purata anggaran H terletak di antara 0.7 dan 0.9 menunjukkan kewujudan memori-panjang dalam siri masa ozon. Selain itu, didapati bahawa data tersebut berterusan sehingga ke 150 hari.*

Kata kunci: Analisis penskalaan; ingatan-panjang; ozon; pekali Hurst

INTRODUCTION

The phenomenon of long-memory which is often observed in a long time series data has received wide attention during the last few years. The first step in long-memory modelling dated back to 1950 when Hurst et al. (1965) studied the Nile river flow data and found an empirical evidence that yearly water levels exhibited extreme persistence that could not be captured by the classical ARMA models. In an economic context, Peter (1994) presented some preliminary evidence that markets were not well described by the random walk model and suggested the long-memory model as an alternative. Beran (1994) extensively discussed the statistical aspects of stationary process with long-range dependence on long-memory in his studies. Samorodnitsky (2006) reported that Mandelbrot and Ness (1968) strongly argued that long-memory process was associated with scaling and fractal behaviour involving characteristics of self-similarity and scale-invariance. In recent studies, it is shown that many natural time series such as ozone level, temperature, internet traffic, market price, blood pressure and gene expression data are characterized by self-similarity and scale-invariance (Chelani 2009; Kai et al. 2008; Li & Zhang 2007; Weng et al. 2008). In particular, self-similarity and scale-invariance are indicative of long memory.

Theoretically, there are several ways of defining long memory process. Intuitively, the presence of long-memory in a data series implies a strong correlation between the successive data points where past events affect future events. According to Beran (1994), the statistical properties of a series with long-memory can be quite different from those of a series that are iid (identical and independent). For instance, the variability properties of sample means of the assumed iid observations are far from valid in the presence of long-memory. The autocorrelation function (ACF) of a series with long-memory is decay hyperbolically to zero, following a power law and the spectral density function, *S(f)* is unbounded when the frequency *f* is near to zero. This phenomenon remains to be persistent for quite some time. Hence, the widely applied Box and Jenkins (ARIMA) models for analyzing short-memory series are no longer appropriate (Hurst et al. 1965). Therefore, long-memory analysis is pertinent for obtaining accurate information regarding the variability of observations over time for the purpose of making predictions when there is evidence of high persistent in the data.

One common way of quantifying a long-memory is via scaling analysis. As mentioned in the first paragraph, most of the studies regarding long-memory suggested that the *Hurst* coefficient (H) is a suitable indicator for measuring the extent of persistent. The presence of long-memory is inferred at certain levels of the H values ranging between 0 and 1. If $0.5 < H < 1$ this indicates that there exists a long-memory in the data series. The value of H close to 1 indicates a high degree of persistency. If $H = 0.5$, the series is random. The time series exhibit anti-persistent if $0 < H < 0.5$ (Beran 1994). The most commonly used scaling analysis methods include rescaled range analysis (R/S), rescaled variance (V/S), dispersional (disp) and detrended fluctuation analysis (DFA). These methods have occasionally been applied on environmental variables such as the acid deposition series reported in the National Deposition Program USA (Zhu & Liu 2003); ozone series reported in the polluted atmosphere of Delhi and Southern Taiwan (Chelani 2009; Weng et al. 2008) and four pollutant series-SO_2, NO_2, PM_{10} and air pollution indexes in Shanghai, China (Kai et al. 2008).

In Malaysia, many studies have been done on air pollution indexes, but none have investigated the existence of long-memory in the series. A study that was conducted by Afroz et al. (2003) provided a general review on air pollution in Malaysia and its impact on health. Meanwhile in their study, Juneng et al. (2009) investigated the spatial and temporal variability of PM_{10} concentration across Malaysia based on data from January 2000 to December 2006 observed at five monitoring stations. Based on the method of principal component analysis, their results suggested that this variability could be decomposed into four dominant modes which were characterized by the geophysical and climatic conditions of the different regions considered in their study. Recently, Azmi et al. (2010) studied the trend and status of air quality data available at three monitoring stations in the Klang Valley (Malaysia), observed from January 1997 to December 2006. Their results showed that the average concentrations of all five pollutant parameters considered were under the permissible value recommended by the Malaysian Department of Environment. Since less emphasis has been given on the study of the existence of long-memory phenomenon in the data of air pollutant series of Malaysia, the objective of this study was to investigate this phenomenon based on the ozone data observed at several monitoring stations in Peninsular Malaysia. The primarily focus in this study was to identify the existence of long-memory in the data with respect to the time domain. The scaling analysis was applied to the ozone series to determine the *Hurst* coefficient in order to examine the existence of long-memory and obtain the degree of persistency. Taking into account the persistency in the data in further analysis would help in obtaining reliable results to enhance the decision makers in implementing a policy on pollution control measure.

MATERIAL AND METHODS

OZONE DATA

The ozone concentration data was recorded as part of a Malaysian Continuous Air Quality Monitoring (CAQM) program by a private company, Alam Sekitar Sdn Bhd (ASMA) on behalf of the Air Quality Division of the Department of the Environment, Malaysia (DOE). The pollutant measurements were performed on an hourly basis and reported in part per million (ppm). For the purpose of this study, data were taken from six urban background monitoring stations located at the northern, central and southern regions of Peninsular Malaysia. The choice of the stations were made based on the most complete data series with missing values of less than 10%. The stations were categorized as industrial and residential areas which were relatively highly-populated continental region representing an urban area in Malaysia. Due to the rapid urbanization, industrialization and rapid increase of vehicular transport in congested roads, these areas were exposed to air pollution. The details of the data are shown in Table 1. Most of the stations had data recorded spanning from January 1998 to December 2006 except for station NR in which the data were recorded from December 1998 to December 2006.

For a complete data series, the missing values of less than 10% were imputed using the nearest neighbour method as proposed by Junninen et al. (2004). Then, a daily mean per hour of ozone concentration was treated as a stochastic time series to be analyzed directly in an attempt to investigate the presence of long-memory.

METHODS FOR DETECTING THE EXISTENCE OF LONG-MEMORY

As a preliminary investigation of the existence of long-memory, the autocorrelation functions (ACF) of observed

TABLE 1. Data

	Industrial Area			Residential Area		
	Code	Station Name	Data Set	Code	Station Name	Data Set
Northern region	NI	Institut Latihan Perai, Perai	Jan 1998 – Dec 2006 (9 years)	NR	Universiti Sains Malaysia, Penang	Dec 1998 – Dec 2006 (8 years)
Central region	CI	Sek. Men. Perempuan Raja Zarina, Klang	Jan 1998 – Dec 2006 (9 years)	CR	Sek. Ren. Keb.TTDI Jaya, Shah Alam	Jan 1998 – Dec 2006 (9 years)
Southern region	SI	Sek. Men. Pasir Gudang 2, Pasir Gudang	Jan 1998 – Dec 2006 (9 years)	SR	Sek. Men. Vokasional Perdagangan, Johor Bahru	Jan 1998 – Dec 2006 (9 years)

*Source: Alam Sekitar Sdn Bhd (ASMA)

data were first plotted. Then, the determination of the existence of long-memory was done by using five scaling methods: Rescaled range analysis (R/S), Rescaled variance analysis (V/S), Dispersion analysis (Disp) and the two Scaled windowed variance analysis; linear (SWVld) and bridge (SWVbd) detrending methods. Further details and the algorithm of these methods can be referred in Cajueiro and Tabak (2005) and Delignieres et al. (2006). Brief presentations of these methods and application are provided as follows:

Let $\{x_j\,;\, j = 1, 2, \ldots, N\}$ be a stochastic time series (i.e. daily mean per hour of ozone concentration in this study) with mean $\bar{x}_G = \frac{1}{N}\sum_{j=1}^{N} x_j$ and variance $s^2 = \frac{1}{N-1}\sum_{j=1}^{N}\left(x_j - \bar{x}_G\right)^2$. Then, this series is partitioned into M non-overlapping interval of equal size n, where :

$$M = N/n, \tag{1}$$

be total number of interval and:

$$\{x_{ij}\,;\, x_{ij} = x_{n(i-1)+j}\}, \tag{2}$$

representing the j^{th} element in the i^{th} interval with $i = 1, 2, \ldots, M$. For the i^{th} interval, the mean and standard deviation respectively are:

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij} \quad \text{and} \quad S_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_{ij} - \bar{x}_i\right)^2}. \tag{3}$$

Dispersion method (*DISP*). A *Dispersion* method (DISP) is based on the standard deviation of a sample mean of a given sample size. Thus, a sample mean, $\bar{x}_i$ is determined in each interval-i of equal sample size n for M possible non-overlapping interval. To describe the fluctuation of the series, the standard deviation of $\bar{x}_i$, was calculated by:

$$(\text{DISP})_n = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\bar{x}_i - \bar{x}_G\right)^2}. \tag{4}$$

These computations were repeated for different M- interval size n. According to the scaling properties, a relationship between DISP and n, can be written as (Delignieres et al. 2006) :

$$(\text{DISP})_n \propto n^{H-1}. \tag{5}$$

Rescaled Range methods (R/S). A *Rescaled Range* method (R/S) is carried out by considering the cumulative sum of deviation of the time series from its mean. As in a *DISP* method, for each interval-i size n, the cumulative sum of deviation of the time series from its mean, $y_{i,t}$ is computed by:

$$\left\{y_{i,t} = \sum_{j=1}^{t}\left(x_{ij} - \bar{x}_i\right), t = 1, 2, \ldots, n\right\}. \tag{6}$$

These generated new series $y_{i,t}$, preserved the properties such as variability of the original data.

The range, R_i which is the difference of the maximum and minimum of $y_{i,t}$, is:

$$R_i = \{\max(y_{i,t}) - \min(y_{i,t})\}_{t\in[1,n]}. \tag{7}$$

The standard deviation, S_i as in (3); and their relationship, R_i/S_i, were calculated in each i^{th}-interval and then averaged over the M possible intervals with equal size n:

$$(\text{R/S})_n = \frac{1}{M}\sum_{i=1}^{M}\frac{R_i}{S_i}. \tag{8}$$

By repeating the calculation of $(\text{R/S})_n$ for the different interval size n, a relationship between $(\text{R/S})_n$ and n can be expressed in terms of a power law (Cajueiro & Tabak 2005), hence:

$$(\text{R/S})_n \propto n^{H}. \tag{9}$$

Rescaled Variance Method (V/S). For a *Rescaled Variance* method (V/S), the algorithm is almost similar as the one for the R/S technique, but instead of using range, variance ratio V/S was used to detect the presence of long-memory process in a time series. The standard deviation of $y_{i,t}$, V_i used is given by:

$$V_i = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}\left(y_{i,t}-\bar{y}_i\right)^2} \text{ where } \bar{y}_i = \frac{1}{n}\sum_{t=1}^{n} y_{i,t}. \tag{10}$$

By repeating the calculation of (V/S) for the different interval size n, V/S is related to n in (Cajueiro & Tabak 2005):

$$(V/S)_n \propto n^H. \tag{11}$$

Scaled windowed variance methods (*SWV*). According to the *SWV* technique, after dividing the series into non-overlapping interval equal size n, the series were then detrended within each interval. Two types of detrending techniques are used: linear detrending and bridge detrending. The linear detrending (SWVld) is performed by removing the regression line within each considered interval, while bridge detrending (SWVbd) is performed by removing the line connecting the first and last points of the interval. After that, the fluctuation of the series is computed within each i^{th} interval size n. Given the standard deviation Sd for each i^{th} interval size n is :

$$(Sd)_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(r_{ij}-\bar{r}_i\right)^2}, \tag{12}$$

where r_{ij} is a deviation for j^{th} element in the i^{th} interval and $\bar{r}_i$ is its mean, respectively given as:

$$R_{ij} = x_{ij} - \hat{x}_{ij}^{(g)} \text{ and } \bar{r}_i = \frac{1}{n}\sum_{j=1}^{n} r_{ij}, \tag{13}$$

where $\hat{x}_{ij}^{(g)}$ represents the detrending technique used, as simple linear detrending, $\hat{x}_{ij}^{(ld)}$ and bridge detrending $\hat{x}_{ij}^{(bd)}$. Given that:

$$\hat{x}_{ij}^{(g)} = \begin{cases} \hat{x}_{ij}^{(ld)} = \hat{a} + \hat{b}j & ; \ \hat{a} = \bar{x}_i + \hat{b}j \ \text{ and } \\ & \hat{b} = \dfrac{\sum_{j=1}^{n} jx_{ij} - n(n+1)\bar{x}_i/2}{n(n^2-1)/12}. \\ \hat{x}_{ij}^{(bd)} = (1-w_i)x_{i1} + w_j x_{in} & ; \ w_j = \dfrac{j-1}{n} \end{cases} \tag{14}$$

In describing the scaling dynamic, the function $(Sd)_i$ is then averaged over all the M interval of equal size n:

$$\left(\overline{Sd}\right)_n = \frac{1}{M}\sum_{i=1}^{n}(Sd)_i. \tag{15}$$

This computation is repeated over all possible intervals with size n and a relationship between $\left(\overline{Sd}\right)_n$ and n is given by (Delignieres et al. 2006):

$$(Sd)_n \propto n^H. \tag{16}$$

Technically, the data series was first partitioned into M non-overlapping intervals with equal length size n. Then, in each interval the statistical measurement used for each method was determined and denoted as $F(n)$. For scaling dynamics, the averaged $F(n)$ over M intervals with length n is expected to obey the power law, notably:

$$F(n) \propto n^{g(H)}, \tag{17}$$

with g(H) determined by plotting the log-log plot of $F(n)$ versus n.

Finally, the agreements of all methods were observed to prove the existence of long-memory in the daily mean per hour of ozone concentration data series.

Results and Discussion

The observed ACF plots in Figure 1 shows that the daily mean per hour of ozone time series decays hyperbolically very slowly to zero and exhibits marked correlations at high lags. Theoretically, this implied the presence of long-memory in the time series. Intuitively, the indicative of long-memory implied a strong correlation with the successive data point where the corresponding autocorrelation function of daily mean per hour of ozone time series decayed more slowly than exponential decay as lag increased. This existence was proven by analyzing the daily mean per hour of ozone data using the five scaling methods mentioned earlier.

Table 2 shows the estimated H values using the five methods. As expected, all methods gave similar results in detecting the presence of long memory in all the data series at the six stations. Obviously, the estimated H values varied between 0.65 and 0.97. Additionally, based on the agreement of all methods used, the results denoted the presence of long-memory in both the industrial and residential areas. The values lay in the range of 0.69 to 0.97. However, the estimated H values which varied between 0.65 and 0.90 were observed in the residential area.

The DISP method showed similar results in all the stations with the average H values as 0.74 ± 0.02. For the R/S and V/S methods, it was noted that the V/S method produced more precise results compared to the R/S method. As can be seen, the estimated H values with the R/S method gave large values of variance compared to the V/S method. Next, by considering the trend in the ozone data series, the two tested methods, SWVld and SWVbd, gave essentially similar results. As have been presumed, it is noted that the detrending influenced the results since there was a linear trend in the ozone concentration data. Removing this trend increased the degree of persistency which was seen in the long term data. Therefore, it was important to remove the long term trend before the scaling analysis was conducted. Since most of the five methods used produced an equal H, a weighted mean $\bar{\text{H}}$ was then calculated.

Table 3 shows the weighted mean estimated H values. Clearly, the indicative value on the existence of the long-

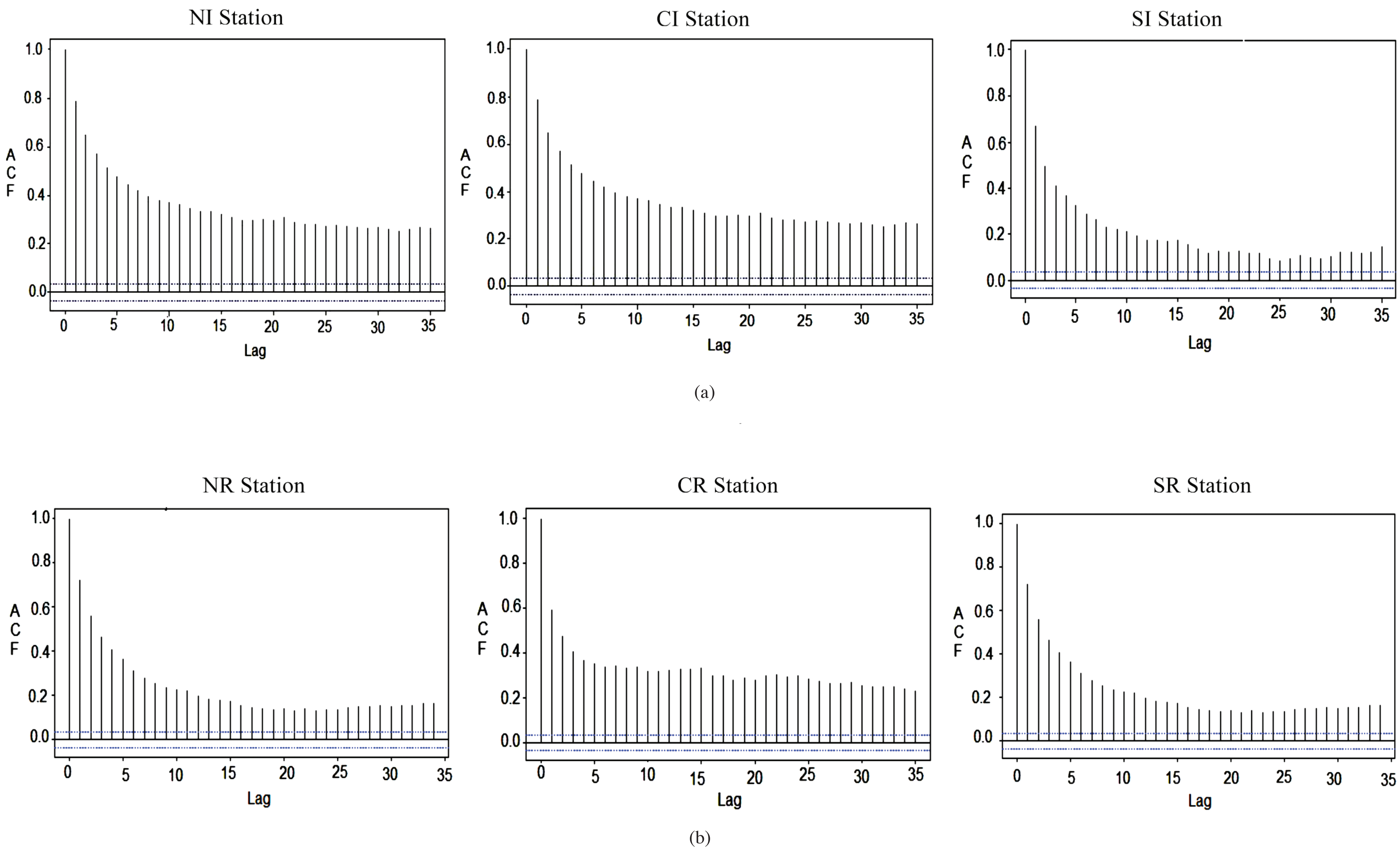


FIGURE 1. The ACF plots for (a) Industrial areas, (b) Residential areas of Northern, Central and Southern regions

TABLE 2. Estimated *Hurst* Coefficient

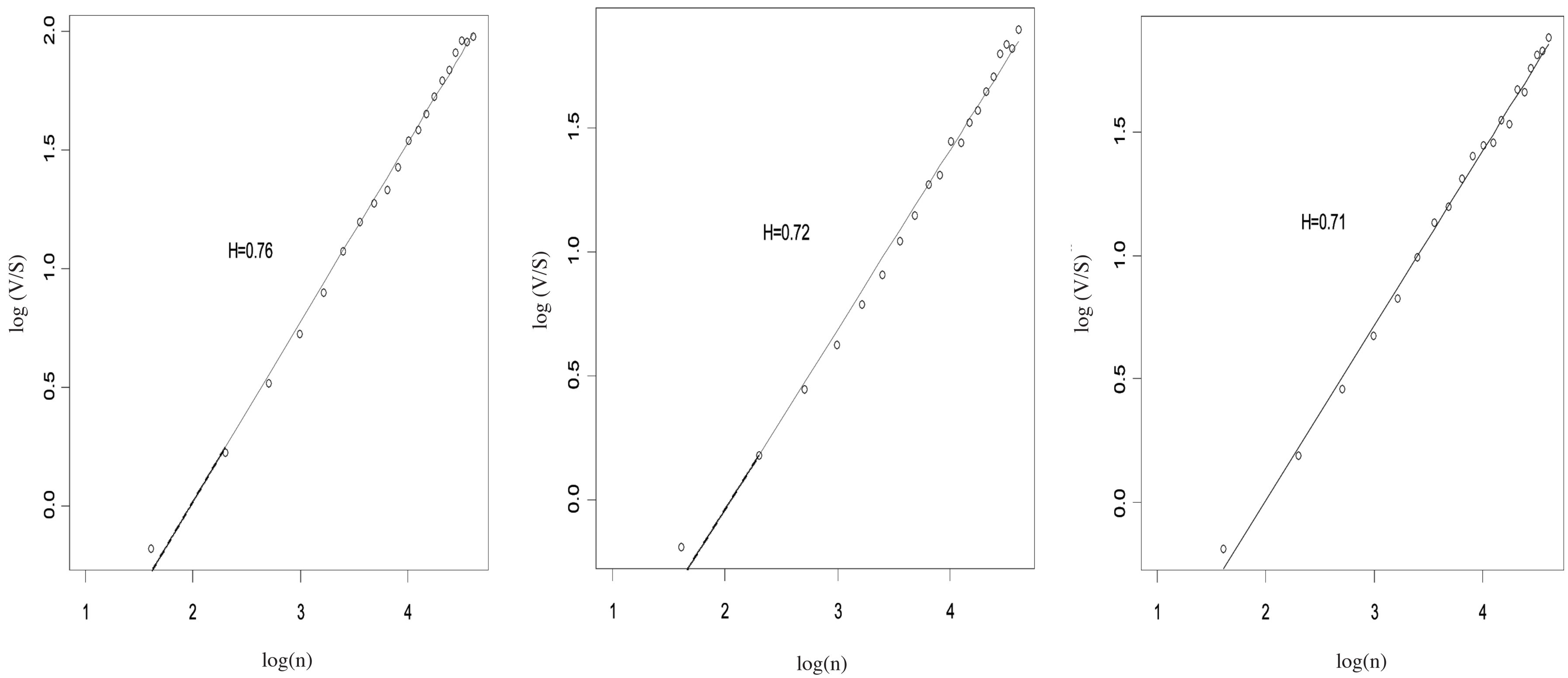
Methods	Stations					
	Northern		Central		Southern	
	NI	NR	CI	CR	SI	SR
	H (s.d)	*H (s.d)*	*H (s.d)*	*H (s.d)*	*H (s.d)*	*H (s.d)*
DISP	0.75 (0.02)	0.76 (0.02)	0.79 (0.02)	0.79 (0.02)	0.69 (0.03)	0.65 (0.03)
R/S	0.93 (0.09)	0.80 (0.08)	0.85 (0.10)	0.79 (0.10)	0.91 (0.10)	0.83 (0.08)
V/S	0.76 (0.02)	0.71 (0.02)	0.72 (0.03)	0.70 (0.04)	0.71 (0.02)	0.68 (0.02)
SWVld	0.97 (0.02)	0.90 (0.03)	0.90 (0.02)	0.84 (0.03)	0.89 (0.03)	0.84 (0.03)
SWVbd	0.93 (0.02)	0.87 (0.02)	0.88 (0.02)	0.83 (0.04)	0.87 (0.02)	0.82 (0.03)

TABLE 3. Weighted mean estimated *Hurst* Coefficient

	Stations					
	Northern		Central		Southern	
	NI	NR	CI	CR	SI	SR
Weighted mean, $\bar{H}$	0.85	0.80	0.84	0.80	0.79	0.73
Variance	0.21	0.20	0.22	0.27	0.22	0.19
Grand Mean, $\hat{H}$	**0.80**					

memory was 0.80 which meant that a long-memory phenomenon existed in the industrial and residential areas. The estimated $\bar{H}$ for the industrial area was on average 5% higher than the estimated $\bar{H}$ for the residential area. The existence of long-memory may have been due to the area being affected by the pollutants that are discharged continuously from the heavy volume of traffic and 24 h operating factories. A study conducted by Varotsos et al. (2005) used the detrended fluctuation analysis (*DFA*) to identify the long-memory of the hourly observations of ozone, NO_x, PM_{10} and $PM_{2.5}$ in Greece and Maryland. Their results showed that the *Hurst* value for the daytime ozone was found to range between 0.87 to 0.89 and 0.79 and 0.81 at night. While those for NO_x ranged from 0.71 and 0.73. The *Hurst* value for PM_{10} and $PM_{2.5}$ were found to vary from 0.88 to 0.93 and 1.19 to 1.21, respectively. Weng et al. (2008) concluded in their study that the daily maximum hourly ozone time series in Taiwan was identified as persistent and long-memory with a *Hurst* value of 0.75, while Kai et al. (2008) analyzed SO_2, NO_2, PM_{10} pollution indexes and daily air pollution indexes (API) of Shanghai using the R/S, DFA and spectrum. The results showed that the *Hurst* values obtained were 0.81 for SO_2, 0.84 to 0.92 for NO_2 and 0.78 to 0.81 for $PM_{10,}$ while for the API it was found to range from 0.78 to 0.83.

Next, a scaling behaviour was also described by the log-log plot of the time scale and statistical measurement used. The log-log plots of two out of five methods used in the industrial area stations are displayed in Figures 2 and 3. As can be seen, the slope of the straight line fitted that represented H indicated that the correlations existed at all the time scales. The very close values of H showed that the scaling behaviour was very similar in all stations. Specifically, for the time interval from 5 up to 150 days, the daily mean per hour of ozone concentrations behaved persistently. In other words, the fluctuations of the daily mean per hour of ozone concentrations from small time interval to larger time interval followed the power law fashion with H varying between $0.5 < H < 1$. In addition, the daily mean per hour of ozone concentration data can be characterized as self-similarity or scale-invariance. The scaling analysis offered a visually clear way to display the long-range dependence of data series or the existence of long-memory phenomenon. Windsor and Toumi (2001) applied the R/S, Sigma-T and Kurtosis analysis to the average hourly ozone, PM_{10} and $PM_{2.5}$ series in the UK with a mean *Hurst* coefficient estimator of 0.77, 0.80, 0.77, respectively. Their study proved the long-memory effect of the ozone series with a persistent duration of up to 400 days. The Rescaled range analysis, R/S on the hourly ozone data series in Delhi, conducted



(a) (b) (c)

FIGURE 2. Log-log plot of daily mean per hour of ozone concentration using *V/S* method in (a) NI, (b) CI and (c) SI stations

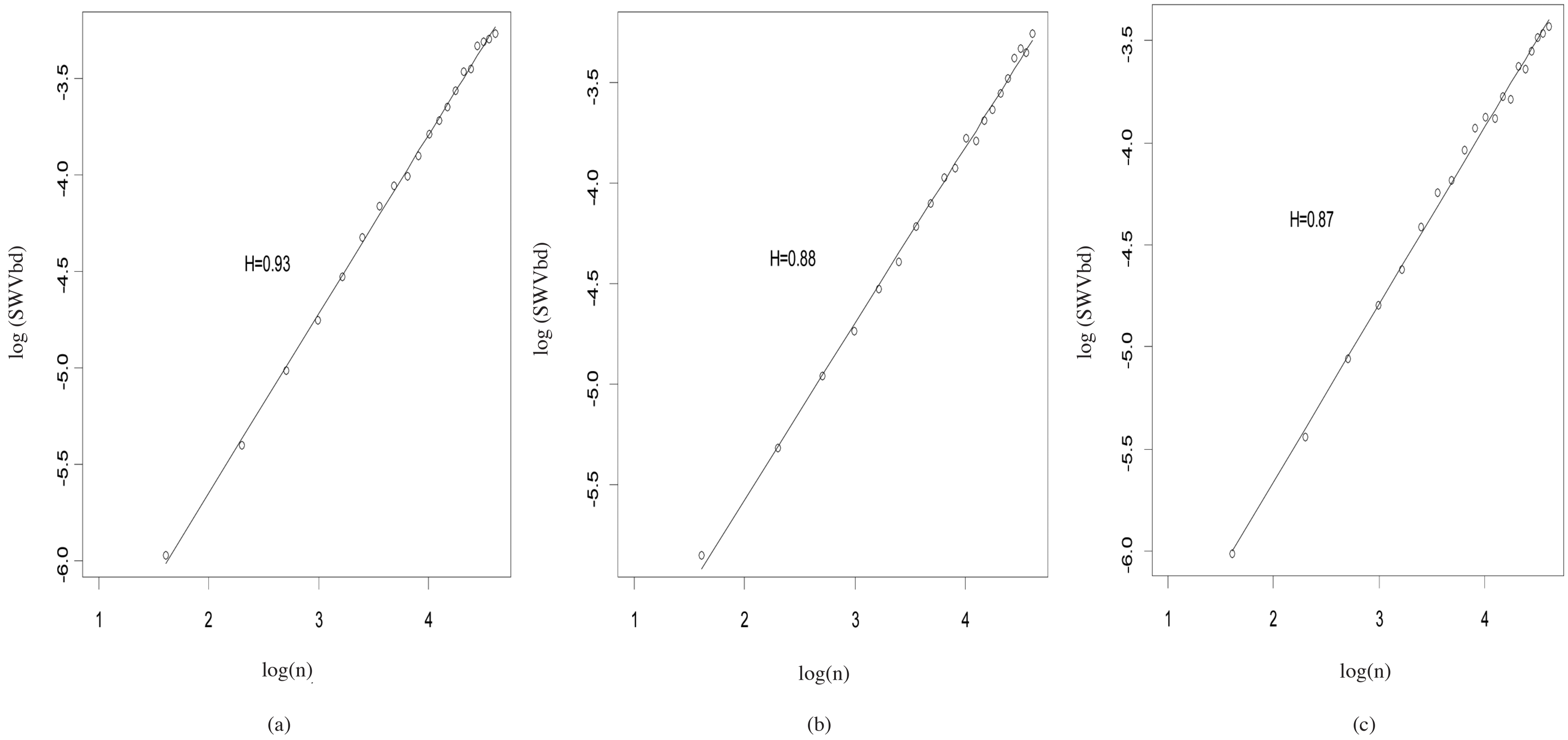


FIGURE 3. Log-log plot of daily mean per hour of ozone concentration using *SWVbd* method in (a) NI, (b) CI and (c) SI stations

by Chelani (2009) established that the *Hurst* coefficient estimator was 0.77 for the whole data series in 2006 and above 0.88 for the monthly analysis. The results showed the scaling behaviour was persistence for up to 5 days in the ozone concentration in 2006 and up to 2 days for the monthly observation. According to Koutsoyiannis (2003), a long-memory phenomenon was closely related to climate changes, since a simple scaling behaviour of climate (the daily mean per hour of ozone concentration ozone in this study) changed irregularly at all time scales. Evidently, many applied statisticians and scientists have been investigating the problem of long-memory since the pioneer work by Hurst. Therefore, a long-memory phenomenon should be considered as one of the most important characteristics in statistical analysis on air pollution modelling and forecasting.

CONCLUSION

Air pollution is one of the most important environmental problems attracting concerns of environmentalists, policy makers and the public in general. It is well known that the ozone is chiefly related to pollutant emissions but the mechanics that derives its temporal evolutions are not understood very clearly. By analyzing the scaling behaviour of the ozone concentration series, one can examine its behaviour. After applying the scaling analysis i.e. R/S, V/S, *Disp* and the two SWV methods, it was found that the ozone series exhibited a long-range dependence known as a long-memory phenomenon. Based on the results of the scaling analysis, it can be concluded that all stations displayed the existence of long-memory in their daily mean per hour Malaysian ozone data series. It was proven by the *Hurst* coefficient that the estimated values varied between $0.5 < H < 1$. Meanwhile, the results also revealed that long-memory processes occurred for time intervals of 5 up to 150 days. This might be due to the pollution discharged continuously at the industrial and most populated residential areas from the heavy volume of traffic and the 24 hours operating factories. As mentioned earlier, the information about the existence of long-memory in the data series was very important in statistical inference and meteorological modelling. Understanding and solving the environmental problems such as the air quality problem often involved certain quantitative aspects, in particular the acquisition and analysis of data. Hence, this study has highlighted new knowledge in obtaining reliable estimation and interpretation on statistical inference. Further analysis which provides sufficient information would give an effective environmental management, especially for the Malaysian meteorological purpose.

ACKNOWLEDGEMENTS

The authors are grateful to the Air Quality Division of the Department of the Environment Malaysia (DOE), Alam Sekitar Sdn Bhd (ASMA) and Assoc. Prof. Dr. Mohd Talib Latif from Universiti Kebangsaan Malaysia for providing the air quality data for this study. Special thanks also to Universiti Pendidikan Sultan Idris and Ministry of Higher Education, Malaysia for their sponsorship. This research was funded by Universiti Kebangsaan Malaysia (UKM-AP-2011-19).

REFERENCES

Afroz, R., Hassan, M.N. & Ibrahim, N.A. 2003. Review of air pollution and health impacts in Malaysia. *Environmental Research* 92(2): 71-77.

Azmi, S., Latif, M., Ismail, A., Juneng, L. & Jemain, A. 2010. Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. *Air Quality, Atmosphere & Health* 3(1): 53-64.

Beran, J. 1994. *Statistics for Long-Memory Processes*. New York: Chapman & Hall.

Cajueiro, D.O. & Tabak, B.M. 2005. The rescaled variance statistic and the determination of the Hurst exponent. *Mathematics and Computers in Simulation* 70(3): 172-179.

Chelani, A.B. 2009. Statistical persistence analysis of hourly ground level ozone concentrations in Delhi. *Atmospheric Research* 92(2): 244-250.

Delignieres, D., Ramdani, S., Lemoine, L., Torre, K., Fortes, M. & Ninot, G. 2006. Fractal analyses for 'short' time series: A re-assessment of classical methods. *Journal of Mathematical Psychology* 50(6): 525-544.

Hurst, H.E., Black, R.P. & Simaika, Y.M. 1965. *Long-term storage: An Experimental Study*. London: Constable & Co.Ltd.

Juneng, L., Latif, M.T., Tangang, F.T. & Mansor, H. 2009. Spatio-temporal characteristics of PM10 concentration across Malaysia. *Atmospheric Environment* 43(30): 4584-4594.

Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J. & Kolehmainen, M. 2004. Methods for imputation of missing values in air quality data sets. *Atmospheric Environment* 38(18): 2895-2907.

Kai, S., Chun-Qiong, L., Nan-Shan, A. & Xiao-Hong, Z. 2008. Using three methods to investigate time-scaling properties in air pollution indexes time series. *Nonlinear Analysis: Real World Applications* 9(2): 693-707.

Koutsoyiannis, D. 2003. Climate change, the Hurst phenomenon and hydrological statistics. *Hydrological Sciences Journal* 48(1): 3-24.

Li, Z. & Zhang, Y-K. 2007. Quantifying fractal dynamics of groundwater systems with detrended fluctuation analysis. *Journal of Hydrology* 336(1-2): 139-146.

Mandelbrot, B.B. & Ness, J.W.V. 1968. Fractional brownian motions, fractional noises and applications. *SIAM Review* 10(4): 422-437.

Peter, E.E. 1994. *Fractal Market Analysis: Applying Chaos Theory to Investment and Economic*. New York: John Wiley & Son, Inc.

Samorodnitsky, G. 2006. *Long Range Dependence*. New York, USA: Now Publishers Inc.

Varotsos, C., Ondov, J. & Efstathiou, M. 2005. Scaling properties of air pollution in Athens, Greece and Baltimore, Maryland. *Atmospheric Environment* 39(22): 4041-4047.

Weng, Y-C., Chang, N-B. & Lee, T.Y. 2008. Nonlinear time series analysis of ground-level ozone dynamics in Southern Taiwan. *Journal of Environmental Management* 87(3): 405-414.

Windsor, H.L. & Toumi, R. 2001. Scaling and persistence of UK pollution. *Atmospheric Environment* 35(27): 4545-4556.
Zhu, J. & Liu, Z. 2003. Long-range persistence of acid deposition. *Atmospheric Environment* 37(19): 2605-2613.

Muzirah Musa*
Department of Mathematics
Faculty of Science and Mathematics
Universiti Pendidikan Sultan Idris
35900 Tanjong Malim, Perak
Malaysia

Kamarulzaman Ibrahim
School of Mathematical Sciences
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600, Bangi, Selangor
Malaysia

*Corresponding author; email: muzirah@fsmt.upsi.edu.my

Received: 30 September 2011
Accepted: 21 May 2012