International Journal of Natural and Engineering Sciences 9 (1): 22-26, 2015
ISSN: 1307-1149, E-ISSN: 2146-0086, www.nobel.gen.tr

Mathematical Comparison of Long memory in the Surface Ozone Concentration using Various Scaling Analyses

Tamil S. SELVI[1*] Samuel R. SELVARAJ[2]
[1]Condensed Matter Research Department of Physics, KCG College of Technology, Chennai, Tamil Nadu, India
Research Scholar, Bharathiyar University, Coimbatore, Tamil Nadu, India
[2]Department of Physics, Presidency College, Chennai, Tamil Nadu, India

*Corresponding author:
Email: tssekar9773@gmail.com

Received: January 18, 2015
Accepted: March 01, 2015

Abstract

One of the main challenges for countries located in tropical areas, is the high concentration of ozone caused by elevated levels of anthropogenic and natural ozone precursors. The aim of this paper is to describe the characteristics of ozone concentration in Chennai, Tamil Nadu, India, based on the value of Hurst exponent which classifies the time series. Daily mean of the hourly data of ozone concentration from June 2011 to August 2012 is considered in this work. In this work, there are six scaling analysis, both time and frequency domain methods are used. The result of the scaling analysis show, the estimated H values lay within 0.6 and 0.8, indicating the existence of long-memory in the ozone time series data. It was also found that the data were persistent for the period of 224 days. The existence of long-memory in a data set implies that the successive data points are highly correlated, i.e. they remain persistent for quite some time.

Keywords: Hurst exponent, long memory, time and frequency domain, scaling analysis, persistent

INTRODUCTION

Ozone plays an important role in a chemistry of the earth's atmosphere, even though it is a minor constituent in terms of abundance. Ozone absorbs IR radiation and thus is one of the greenhouse gases in the troposphere [4],[9],[17]. The ozone gas can be found in two different regions of the Earth's atmosphere.The majority of the ozone (about 97%) is concentrated in the stratosphere at an altitude of 15to 55 km. The ozone layer is a layer in Earth's atmosphere which contains relatively high concentrations of ozone (O_3). This layer also absorbs 93-99% of the sun's high frequency ultraviolet light, which is potentially damaging to life on earth [1]. Normally the ozone concentration lies in the range of about 300 to 350 Dobson Units (D.U) [3]. Every 0.01 millimeter thickness of the layer is equal to one Dobson Unit [14]. Increased penetration of solar UV-B radiation is likely to have a profound impact on human health with potential risks of eye diseases, skin cancer and infectious diseases [15]. Experiments on animals show that UV exposure decreases the immune response to skin cancers, infectious agents and other antigens. In forests and grasslands increased UV-B radiation is likely to result in changes in species composition (mutation) thus altering the bio-diversity in different ecosystems [20].

Major agricultural operations are normally undertaken during Southwest and Northeast Monsoon season of Tamil Nadu. The effort devoted to study time series reflects the fact that nonlinearities convey information about the structure of the series under study [16]. There are several tests for investigating nonlinear behaviors of time series. We use the Hurst Exponent method to investigate ozone data. 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 [5], [8] [11], [24]. Self-similarity and long-range dependence (LRD) have become very important concepts in analyzing ozone data over the past years. There are two important questions related to long-range dependence that have not received as much attention: a) how can we calculate it accurately, b) what does it really mean for time series analysis and modeling?. In this work, we are interested to focus on the first question, since it is a necessary step to answer the second question. Hurst exponent can distinguish between stochastic and deterministic data. To find out whether or not a system is deterministic or random, we must determine how much "memory" (how past events affect future events) a dynamic system has. We will do this by estimating the Hurst Exponent.

Long range dependence can be thought of in two ways:[22]

In the time domain it manifests as a high degree of correlation between distantly split data points. In the frequency domain it determines the significant level of power at frequencies near zero. In this work, six methods are used for the estimation of the Hurst Exponent (H). In no particular order they are: re-scaled range, autocorrelation function, absolute moment method, aggregated variance analysis, wavelet and Periodogram method. The method originally developed by Hurst was the re-scaled range method.

The paper is organized as follows. The first part of this paper describes the data quality and location of the study area. This is followed by the explanation of the procedures and methodologies used in assessing persistency of ozone concentration. Finally, the results are presented along with discussion.

General Description of Chennai City

Chennai City experiences a tropical climate. It is in the coastal area with the Bay of Bengal on the eastern side. With the improving economy and the rapid increase in the population of the urban areas in comparison to the rural areas the stress on the environment has increased notably. As the standard of living has improved, almost all the households have connected to a motorized vehicle either a scooter or a car. The increase in the population has resulted in the pollution load of big cities. The pollution levels are checked by the macroscopic parameters PM10 and NOx emissions, which is definitely showing an increase. The latitude and longitude of the city are E80^{o} 14’51” and N13^{o} 03’ 40”. The city area is about 70 square miles and the metro area range about 456 square miles.

Data collection

This work was carried out as a continuation of the previous work by the same data collection [18]. Surface ozone measurements were carried out daily and ten measurements were made on all days between 08.00 h and 17.00 h (IST) during the period from June 2011 to August 2012 in Chennai. An aeroqual series 200 ozone monitor was used to measure low and high surface ozone levels. The sensor is a gas sensitive semiconductor (GGS), which works on the principle of absorption of UV-radiation by ozone in the ambient air. The ozone sensor was calibrated as a certified UV photometer. This particular instrument was chosen for its simplicity and reliability in operation and for its ease of handling, cost-effectiveness and speed in obtaining a gas concentration level directly. An aeroqual monitor with GSS ozone sensor has been used by several workers for the measurement of atmospheric ozone and nitrogen dioxide [2][6][10][12]. This instrument was supplied by Unipro Instruments India Private Limited, Mumbai.

METHODOLOGY

The Calculation of the Hurst exponent makes use of ozone data for the Chennai city spanning from June 2011 to August 2012 as shown in Figure 1. The Hurst Exponent is a measurement that is non-deterministic in nature and measures what is observed. All the methods of estimating the Hurst exponent are broadly separated into two different domains: time domain based method and frequency domain based method. The R/S, Absolute Moment Method and the Variance Method are the time domain methods which are used in this paper. The Periodogram and Wavelet estimator are used as the Frequency domain based methods. All the computations are performed using Matlab software.

Time Domain Estimators

Rescaled range estimator (R/S)

The R/S analysis is used merely because it has been the conventional technique used for geophysical time records [7]. A time series of full length N is divided into a number of shorter time series of length n = N, N/2, N/4 ... The average rescaled range is then calculated for each value of n. For a (part) time series of length n, the rescaled range is calculated as follows: [19]

Calculate the mean and create a mean-adjusted series; then calculate the cumulative deviate series Z and compute the range R; Compute the standard deviation S and calculate the Rescaled range R (n) / S (n) and average over all the partial time series of length n. Hurst found that (R/S) scales by power-law as time increases, which shows

$$\mathbf{(R/S)\ n = C^{*}n^{H}}$$

Here C is a constant and H is called the Hurst exponent. The Hurst exponent is estimated, by sketching log(R/S) versus log n in X-Y axes. From the slope of the regression line Hurst exponent is approximated. The values of the Hurst exponent lie between 0 and 1. If the time series is merely random, then the Hurst exponent (H) equals to be 0.5. If H > 0.5, the time series cover added ‘ distance’ than a random walk, in the case of persistent motion. While if H < 0.5, the time series covers less ‘distance’ than a random walk and shows the anti-persistent behavior in time series.

The R/S method has the advantage of allowing us to determine whether the ozone time series data are long range self similar, or not. This has one important merit compared to the other methods—as it has been known and proved for about 50 years, the methods for experimenting have been well developed and applied.

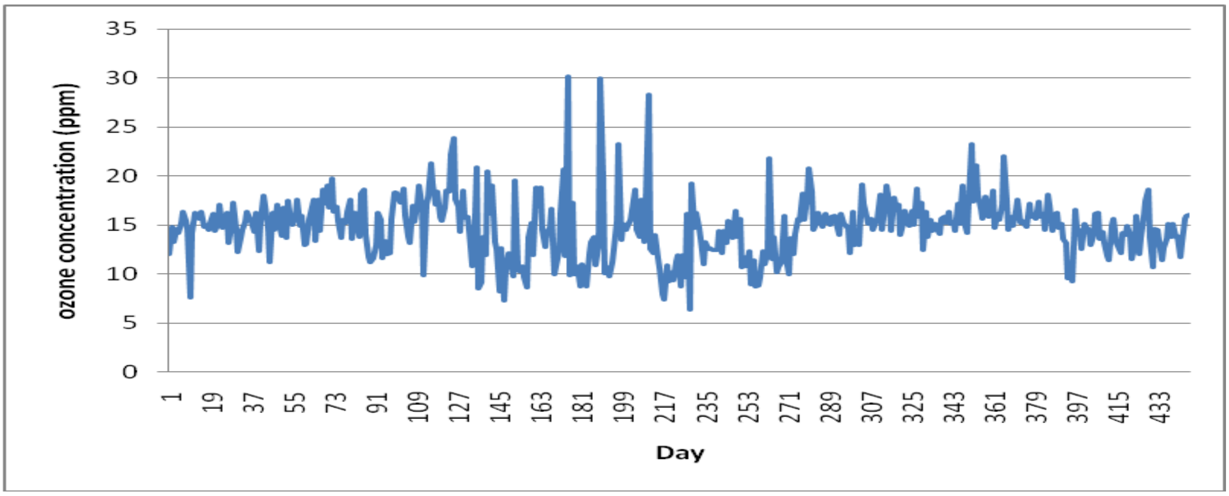

Figure 1. Time series plot for the Chennai daily ozone concentration from June 2011 to August 2012

Autocorrelation method

To estimate the Hurst exponent using the autocorrelation method, one needs to calculate a sufficient number of lags to perform the analysis. In this method, the Hurst exponent is related to the Autocorrelation function (ACF) through the slope coefficient of the estimate of the log of ACF versus the log of the frequency. We should calculate ACF of the series until the ACF is negative and use all of the positive values as a data series. A regression sketch on the natural log of the ACF values versus the natural log of the lag of the ACF values is used to estimate the Hurst exponent. The Hurst exponent is related to the slope coefficient as

$$\mathbf{H = 1 + (\alpha/2)}$$

where α is the slope of the regression. The ACF method is easier to calculate than the Rescaled range.

Absolute Moment Method for Hurst Estimation [21]

To estimate the Hurst Exponent with the absolute moment method, one starts estimation by dividing a series of length n into shorter segments of length m and then averaging the series over each segment of length m.

$$X^m(k) := \frac{1}{m}\sum_{i=(k-1)m+1}^{km} X_i, \ k = 1,2,\ldots,\frac{n}{m}$$

To get the absolute moment (AM) of the series

$$AM_n^m = \frac{1}{N/m}\sum_{k=1}^{N/m} |X^m(k) - \bar{X}|^n$$

If there is no long run dependence, then the Hurst Exponent again will be 0.5This method is generally used for n=1. If n=2 or larger it reduces to the Aggregated Variance method. In a log/log plot of the absolute moments versus m the slope (α) of the linear fit is related to the Hurst Exponent as: α = H -1 if n=1

So, **H = α + 1**

Aggregated Variance Method [21]

Estimation of Hurst Exponent with the aggregated variance method can be start with the Absolute moment method, by dividing a series of length n into shorter segments of length m and then averaging the series of each m length segment as in absolute moment method.

$$X^{(m)}(k) = \frac{1}{m}\sum_{i=(k-1)m+1}^{km} X_i, \quad k = 1,2,3,L \ ,[N/m]$$

Then the sample variance is calculated for each m length segment and the log of the variance is sketched against the log of m, as in previous methods.

$$VarX^{(m)} = \frac{1}{N/m}\left(X^{(m)}(k) - \bar{X}\right)^2$$

Once again, the slope of the linear regression of the log/log plot is related to the Hurst Exponent.

$$H = 1 - \beta/2$$

As before, if H = 0.5 then the series has no long range dependence. The disadvantage of this method is that it allows us to estimate the self-similar parameter *H* by a simple least square fit. However, this estimation of the Hurst parameter highly depends on the number of points used for calculating the slope of the line. A more robust estimation of *H* can be obtained by ensuring (i)the length of each segment is sufficiently large; and (ii) there are enough segments.

Frequency Domain Estimators

Periodogram Estimator

The slope of the log of the Periodogram (I) versus the log of the frequency over the entire domain from 0 to πgives the Hurst exponent H = (1- α) / 2 where α is the slope of the regression and the Fourier coefficients a_k and b_k can be calculated by the following equations [23].

$$a_k = \frac{2}{n}\sum_{t=1}^{n} Z_t \cos(w_k t), k = 1,2,\ldots,\frac{n-1}{2} \text{ if } n \text{ is odd}$$

and

$$b_k = \frac{2}{n}\sum_{t=1}^{n} Z_t \sin(w_k t), \quad k = 1,2,\ldots,\frac{n-1}{2}$$

The Fourier coefficients are used to calculate the Periodogram (I) where the regression of log(I) versus log (frequency) is used to estimate the Hurst exponent. The slope of the regression is used to calculate the Hurst exponent. The Periodogram method resulted to be the simplest as it can be implemented by a sequence of eight vectorial products and an FFT operator.

Wavelet-based estimator

Another method that helps us eliminate the effect of trends in the signals we study is wavelet analysis. The generalized form of a wavelet transform is, as described by [13]

$$W(t,a) = \frac{1}{a^{1/2}}\int_{-\infty}^{\infty} g\left(\frac{t'-t}{a}\right) f(t')dt'$$

This is a more general transform than the Fourier transform, because this is not only getting a picture of the frequency distribution in the signal, but also shows how the frequencies differ with time. In this expression “a” is the scale parameter. By selecting a small value for “ a” one can get the information of how much the signal variability on short time scales (high frequencies), and with the larger values one can study how much the signal varies on longer time scales (low frequencies). It is common to study the set of scales where “a” is increased by a factor of two each time. The function g (t’) is called as the mother wavelet.

$$g(t') = \left(\frac{1}{2\pi}\right)^{1/2}(1 - t'^2)e^{-t'^2/2}$$

If we calculate the variance of W and plot it as a function of a for a time series and taking logarithm on both sides

Log (W) = β log (a) + constant

Hence we obtain a straight line in a log-log plot, with slope β . From this we get an estimate of the Hurst exponent by the relation

$$H = (\beta + 1) / 2$$

The main difference is that wavelets are well localized in both time and frequency domain. In wavelet theory, it is often possible to obtain a good approximation of the given function f by using only a few coefficients which are the greatest achievement in comparing to the Fourier transform.

RESULTS AND DISCUSSION

The observed time series plot in Figure 1 shows that the daily measurement of ozone time series from June 2011 to August 2012. The existence of long memory was proven by analyzing the daily measurement of ozone data using the time and frequency based domain scaling methods mentioned earlier. Table 1 and 2 shows the estimated Hurst values using the time and frequency based domain scaling methods. As expected, all methods gave similar results in detecting the presence of long memory in the data series in Chennai. Obviously, the estimated Hurst values varied between 0.6 and 0.8. Additionally, based on the agreement of all methods used, the results denoted the presence of long-memory for the ozone data in Chennai. To calculate the time interval of long term memory another statistics used in R/S analysis is v-variable, defined as

$$V_N = \frac{(R/S)}{\sqrt{N}}$$

V-Statistics are used to estimate the length of persistent behavior with the value of the length of the cycle when the curve progresses to its height. The results support that long-memory process occurred at time gap of 224 days. Therefore, a long-memory phenomenon should be considered as one of the most important characteristics in statistical analysis on ozone data modeling and forecasting.

Table 1 shows the estimated Hurst values using time domain methods

TIME DOMAIN METHOD	HURST VALUE
R/S	0.7493
ACF	0.7450
AGG. MOMENT	0.6025
AGG. VARINACE	0.7469

Table 2 shows the estimated Hurst values using frequency domain methods

FREQUENCY DOMAIN METHOD	HURST VALUE
PERIODOGRAM	0.7472
WAVELET	0.8042

CONCLUSION

Air pollution is one of the most important environmental problems attracting importance of environmentalists, policy makers and the public in common. By resolving the scaling behavior, of the ozone concentration series its behavior can be identified. After applying the scaling analyses, it was established that the ozone series exhibits a long-range dependence. From the results of the scaling analyses, it can be identified that the Chennai station displayed the existence of long-memory in their daily mean concentration of ozone data series. It was proved that the estimated value of the Hurst exponent lies between 0.5 < H< 1. Meanwhile, the results also revealed that the long-memory processes occurred in the time gap of 224 days. This might be due to the pollution discharge continued at the industrial and most populated residential areas from the heavy volume of traffic and the 24 hours operating factories. The information about the existence of long-memory in the data series was very important in statistical inference and meteorological modeling. Understanding and solving the environmental problems such as the air quality problem often involved certain quantitative aspects, specifically the acquisition and data analysis. Hence, this study has pointed out new knowledge in obtaining reliable estimation and interpretation on statistical inference. Further analysis, provides sufficient information would give an effective environmental management, especially for the Chennai meteorological purpose.

REFERENCES

[1] Albritton, Daniel, What Should Be Done in a Science Assessment In Protecting the Ozone Layer: Lessons, Models, and Prospects, 1998.

[2] Ali, A., 2008. Factors affecting on response of broad bean and corn to air quality and soil CO2 flux rates in Egypt. Water Air Soil Pollut., 195, 311–323.

[3] Alternative Fluorocarbons Environmental Acceptability Study (AFEAS), Washington, DC, 1995.

[4] Chalita, S., Hauglystaine, D., Le Treut, H. and Muller, J. F., 1996. Radiative forcing due to increased tropospheric ozone concentrations. Atmos. Environ., 30, 1641–1646.

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

[6] Frei, M., Tanaka, J. P. and Wissuwa, M., 2008. Genotypic variation in tolerance to elevated ozone in rice; dissection of distinct genetic factors linked to tolerance mechanisms. J. Exp. Bot., 13, 3741–3752.

[7] Govindan, Rangarajan and Dhananjay, A. Sant; 1997. A climate predictability index and its applications, Geophysical Research letters; 24, 1239-1242.

[8] Kai, S.et.al; 2008. Using three methods to investigate time-scaling properties in air pollution indexes time series. Nonlinear Analysis: Real World Applications; 9(2), 693-707.

[9] Lacis, A., Wuebbles, D. J., Logan, J. A., 1990. Radiative forcing of climate in the vertical distribution of ozone. J. Geophys. Res., 95, 9971–9981.

[10] Laurinaviene, D., 2008. Ground level ozone air pollution in Vilnius City. *Environ. Res., Eng. Manage.*, 3, 21–28.

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

[12] Lee, S. and Tsai, S.-W., 2008. Passive sampling of ambient ozone by solid phase micro extraction with on-fiber derivatization. Anal. Chim. Acta, 610, 149–1526.

[13] Malamud, B.D., Turcotte, D.L., 1999. Self-affine time series I: generation and analyses. Advances in Geophysics. 40, 1–90.

[14] Margitan J. J. HO 2 in the Stratophere: 3 In-situ Observations. 1991. Geophysical Research Letters, 8 (3).

[15] Morrisette, Peter M., 1995. The Evolution of Policy Responses to Stratospheric Ozone Depletion. Natural Resources Journal, 29.

[16] Panas, E., and Ninni, V; 2000. Are oil markets chaotic? A non-linear dynamic analysis. Energy economics; 22, 549-568.

[17] Ramnathan, V., Singh, H. B., Cicerone, R. J. and Kiehl, J. T., 1985. Trace gas trends and their potential role in climate change. J. Geophys. Res., 90, 5547–5566.

[18] Samuel Selvaraj, R, Padma, K, and Milton Boaz , B., 2013. Seasonal variation of surface ozone and its association with meteorological parameters, UV-radiation, rainfall and cloud cover over Chennai, India. Current Science; 105, 676-684.

[19] Samuel Selvaraj R., Umarani P.R., Vimal Priya S.P. and Mahalakshmi N.; 2011. Fractal dimensional analysis of geomagnetic as index. International Journal of current Research, 3,146-147.

[20] Stedman, D. H., 1981. Atomic Chlorine and the Chlorine Monoxide Radical in the Stratosphere: Three in Situ Observations. Science, 198.

[21] Taqqu, . M. S ; Teverovsky, V; Willinger, W; 1995. Estimators for long range dependence: an empirical study, Fractals, 3, 785 – 788.

[22] V. Teverovsky, M. S. Taqqu, W. Willinger, On Lo's modified R/S statistic, Preprint.

[23] Wei, William, Time Series Analysis Univariate and Multivariate Methods. 2006, Pearson Addison Wesley, 2nd edition.

[24] 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), 405-414.1997.