Use of Chaotic and Time Series Analysis on Surface Ozone Study at the Tropical Region, Chennai, Tamilnadu

1. Introduction:
High concentration of surface ozone related to air pollution is becoming a matter of concern due to its adverse effects on human health, vegetation and buildings. Chennai the capital of Tamil Nadu is experiencing serious air pollution problem due to increasing number of vehicles, construction activities and power plants. Apart from the emissions of NO\textsubscript{x} and particulate matter, surface ozone concentrations are also increasing in ambient air of Chennai. Due to these increases in pollutant levels, tropospheric \textsubscript{O3} levels may increase significantly in future in this part of the world (Sahu, 2005). Ozone (\textsubscript{O3}) production efficiency is maximum over the Indian region followed by Japan and China, which is explained on the basis of increase in OH and peroxy radicals (Berntsen et al., 1996).

Surface ozone production varies along with the diurnal variation of temperature.(Samuel selvaraj, 2013). In recent years, there has been a considerable concern about the possibility of climatic changes. Alteration in our climate is governed by a complex system of atmospheric and oceanic processes and their interactions. Atmospheric processes also result in increase in surface-level ultraviolet radiation and changes in temperature and rainfall pattern. Human activities on the other hand are responsible for changes in ecosystem due to increased emissions rate of \textsubscript{CO2} and other greenhouse gases (Rai, 2010). Based on the data , it was demonstrated that all-India mean annual temperature had been rising at 0.05 °C/decade, with maximum temperature at +0.07 °C/decade and minimum temperature at +0.02 °C/decade. As a result, the diurnal temperature range shows an increase of 0.05 °C/decade. However, in northern India, the average
temperature is falling at the rate of -0.38 °C unlikely to rise in all-India average temperature (i.e. at +0.42 °C/Century) (Rai, 2010 Kothawale, 2005).

The time series consists of a set sequential numeric data taken at equally spaced intervals, usually over a period of time or space. The first step in time series analysis is to draw time series plot which provides a preliminary understanding of time behavior of the series (Marzuki Ismail et al. 2011). Time series analysis is a useful tool for better understanding of cause and effect relationship of environmental pollution. The main aim of time series analysis is to describe movement history of a particular variable in time. Many authors have tried to detect changing behavior of air pollution through time using different techniques (Hies et al. 2003). The character of ozone observations over time is often studied in terms of its periodicities, existence of any long-term trends, scaling and persistence properties. To reveal the periodicities in the ozone time series, the traditional techniques such as autocorrelation function and Fourier spectral analysis are widely used. The existence of cycles and trends is however, usually not linear and its determination is quite tedious due to interaction of several variables.

A variety of investigation tools yield information about the characterization like periodicities, persistence, and trends of the data sets for SO2, NOx, and O3. Further insight into the extension and importance of memory effects and the long-term structure of our atmospheric time series is gained by the calculation of the Hurst exponent H. This measure quantifies the extension of periods with systematic deviations from the overall mean, which is also called the persistence of the time series (Otto Klemm 1999). The phenomenon of persistence is highly relevant to time series. Persistence is the tendency for the magnitude of an event to be dependent on the magnitude of previous event(s), a memory effect, e.g., tendency for low ozone concentration to follow low ozone concentration and for high ozone concentration to follow high ozone concentration (Deepesh Machiwal., 2006).

Lyapunov exponent of a dynamical system is a quantity that characterizes the rate of separation of infinitesimally close trajectories. Positive is usually taken as an indication that the system is chaotic. Natural phenomena occur in a complex and often unpredictable way. Modeling of such phenomena is a challenging task because the variety of behavior may indicate varying unknown underlying processes. If we focus on recent models and measures of natural phenomena, one of the most important tools for eliciting the chaotic trends is the “Lyapunov exponent” which is a useful measure of the stability of a dynamic system. (Sumathi, 2012). Chaos theory studies the behavior of dynamical system that are highly sensitive to initial conditions, an effect which is popularly referred to as the butterfly effect. The system are deterministic, meaning that their future behavior is fully determined by their initial conditions, with no random element involved. In other words the deterministic natures of these systems do not make them predictable. This behavior is known as deterministic chaos. Observations of chaotic behavior in nature include changes in weather, the dynamics of satellites in the solar system, the dynamics of the action potentials in neurons and molecular vibrations. (Sumathi, 2012) indicated that the properties of the chaotic attractor can be described by a set of Lyapunov exponents.

Periodicities in natural time series are generally due to astronomical cycles such as earth’s rotation around the sun. Due to the periodic motion of the earth around the sun and about its own axis, rainfall, like any event, is believed to have periodic components. Moreover some other researcher also suggested that the monthly rainfall sequence is considered to be a periodic-stochastic process (Rizalihadi 2002; Zakaria 2008; Duenas et al. 2005). Thus, the surface ozone has periodic and stochastic parts in nature, because ozone is influenced by climate parameters such as temperature, wind direction, speed, humidity and so on. Finally, Climate change is expected to significantly influence future O3 concentrations, as the formation, destruction and transport of O3 is strongly dependent on biogeochemical and physical processes most of which are affected by climatic factors such as temperature, rainfall and humidity. O3 will present a serious global air pollution problem by the end of the century. The results from the new scenario analysis and modeling work emphasize the important role of emissions controls for determining future O3 concentrations. They also show that by 2050, climate change may have significant impacts on ground-level O3 at the local and regional scale.
2. Study Area:
Chennai is situated on the south east coast of India and north east coast of Tamil Nadu. This area is one of the most highly populated urban sites. Chennai lies on the thermal equator and is also a coastal. The latitude and longitude of the center of the city are E 80° 14’ 51” and N 13° 03’ 40”. The geographical location of the experimental site is shown in Fig. 1 and it is located in south Chennai. The different sources of air pollution are classified under the following categories: Transport, Industries, Residential in Chennai City. This urban city can be divided into four areas, North, Central, South and West. The Northern part is primarily an industrial area comprising of petrochemical industries in the Manali area and other general industries in Ambattur. (Fig. 1). Chennai has many industrial areas.

Surface ozone was measured throughout Tamil Nadu during the year 2011 and it was found that Kanniyakumari district had the highest daily average of 17.8 ppbv (Samuel Selvaraj et. al., 2011, 2013). Moreover, the surface ozone levels studied in Chennai during 2004–2005 it was found that the hourly values varied from 1 ppbv to 50.27 ppbv. In the urban area Delhi ozone concentration in the ambient air varied from 9 to 128 ppbv at four different sites during 1989–1990. So these studies have indicated that the effects of O₃ on vegetation were quite severe in India and other parts of Asia. From our knowledge of literature survey, there was no measurements have been carried out over Chennai metropolitan area in recent years. (Samuel Selvaraj et. al., 2013). Hence through this study, the surface ozone (O₃) concentration was measured in this urban site, Chennai. We were able to understand from that the surface ozone levels reached higher levels in urban area of Chennai often. This study was conducted at Koyembedu which houses Chennai’s moffussil bus terminus and 100’s of Buses and other vehicles ply daily and hence the vehicular emission is very high. This site is surrounded by the number of Industrial areas located within a short radius.

3. MATERIAL AND Methods of Measuring Characteristics of Time Series:
A time series is the simplest form of temporal data and is a sequence of real numbers collected regularly in time, where each number represents a value. We represent a time series as an ordered set of real-valued variables. Time series can be described using a variety of qualitative terms such as seasonal, trending, noisy, non-linear, chaotic, etc. This section presents the standard statistical measures of a time
series. In addition to we have extended the scope to include a collection of special features such as long-range dependence and chaotic measures such as Hurst exponents and Lyapunov. These help to provide a rich portrait of the nature of a time series. For each of the features described below.

3.1 Measure of Persistence:

Chennai is the fourth largest Metropolitan City in India. It has been noted that the surface ozone is highly variable during the different seasons. Therefore, if its behavior could be predicted in advance, it would go a long way toward helping the agricultural and industrial activities of the region (Dhar and Rakhecha, 1983).

Hurst exponent can be obtained by using the techniques such as detrended fluctuation analysis and rescaled range analysis. Both the techniques have their advantages and disadvantages and explaining them is beyond the scope of this paper. The Hurst exponent, proposed by H. E. Hurst for use in fractal analysis (Mandelbrot and Wallis, 1968), has been applied to many research fields.

We have used Hurst exponent method. It provides a measure for long term memory and factuality of a time series. For calculating Hurst exponent, one must estimate the dependence of the rescaled range on the time span n of observation. Various techniques have been adopted for calculating Hurst exponent. The eldest and best-known method to estimate the Hurst exponent is R/S analysis. The rescaled analysis or R/S analysis is used due to its simplicity in implementation. It was proposed by Mandelbrot and Wallis (Mandelbrot and Wallis, 1969), based on the previous work of Hurst (Hurst, 1951). The R/S analysis is used merely because it has been the conventional technique used for geophysical time records (Govindan Rangarajan and Sant, 1997). 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 (partial) time series of length n, the rescaled range is calculated as follows: (Samuel Selvaraj., 2011).

(i) Calculate the mean; (ii) Create a mean-adjusted series; (iii) Calculate the cumulative deviate series Z; (iv) Compute the range R; (v) Compute the standard deviation S; (vi) Calculate the rescaled range R (n)/ S (n) and he average of overall partial time series of length n.

Hurst found that (R/S) scales by power-law as time increases, which indicates (R/S) n = c*nH, here c is a constant and H is called the Hurst exponent. To estimate the Hurst exponent, we plot (R/S) versus n in log-log axes. The slope of the regression line approximates the Hurst exponent. The values of the Hurst exponent range between 0 and 1. Based on the Hurst exponent value H, the following classifications of time series can be realized: H = 0.5 indicates a random series; 0 < H < 0.5 indicates an anti-persistent series, which means an up value is more likely followed by a down value, and vice versa; 0.5 < H < 1 indicates a persistent series, which means the direction of the next value is more likely the same as current value (Alina Barbulescu, 2007).

The Hurst exponent is related to the Fractal dimension D of the time series curve by the formula D=2-H. The parameter H is called the Hurst exponent which takes the value between 0 and 1. If the fractal dimension D for the time series is 1.5, we again get the usual random motion. In this case, there is no correlation between amplitude changes corresponding to two successive time intervals. Therefore, no trend in amplitude can be discerned from the time series and hence the process is unpredictable. However, as the fractal dimension decreases to 1, the process becomes more and more predictable as it exhibits persistence behaviour. That is, the future trend is more and more likely to follow an established trend. As the fractal dimension increases from 1.5 to 2, the process exhibits anti-persistence. That is, a decrease in the amplitude of the process is more likely to lead to an increase in the future (Govindan Rangarajan 2004).

3.2 Measure of Chaos:

Consider two points in a space, X0 and X0 + ΔX0, each of which will generate an orbit in that space using some equation or system of equations. These orbits can be thought of as parametric functions of a variable, that is, something like time. If we use one of the orbits a reference orbit, then the separation between the two orbits will also be a function of time. Because sensitive dependence can arise only in some portions of a system (like the logistic equation), this separation is also a function of the location of the initial value and has the form ΔX(ΔX0, t). In a system with attracting fixed points or attracting periodic points, ΔX(X0, t) diminishes asymptotically with time. If a system is unstable like pins balanced on their points, then the orbits diverge exponentially for a
while, but eventually settle down. For chaotic points, the function $\Delta X(X_0, t)$ will behave erratically. It is thus useful to study the mean exponential rate of divergence of two initially close orbits using the formula:

$$\lambda = \lim_{t \to \infty} \frac{1}{t} \frac{\Delta X(X_0, t)}{\Delta X_0}$$  

(1)

$\lambda$ is useful for distinguishing among the various types of orbits. It works for discrete as well as continuous systems. The Lyapunov exponent for a set of “N” values can be found using the formula (Pietgen et al., 1992):

$$\lambda_k = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \frac{\Delta X_{k+1}}{\Delta X_k}$$  

(2)

3.3 Measure of Maximum Predictable Time Scale:

According to the maximum predictable time scale $T_f$ has a relationship with maximum Lyapunov exponent as. Much chaotic behavior in the system implies the larger Lyapunov exponent which causes in the smaller predictable time scale $T_f = 1/\lambda$ (Shalaleh Mohammadi, 2009).

3.4 Measuring Trend (Mann-Kendall Test):

There are several approaches for detecting the trend in the time series. These approaches can be either parametric or non-parametric. Parametric methods assumed the data should normally distributed and free from outliers. On the other hand, non-parametric methods are free from such assumptions. The most popularly used non-parametric tests for detecting trend in the time series is the Mann-Kendall (MK) test. It is widely used for different climatic variables (Hirsch RM, 1984). The Mann-Kendal (MK) test searches for a trend in a time series without specifying whether the trend is linear or nonlinear (Khaliq MN, 2009). It is based on the test statistics $S$, which is defined as:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sgn}(x_j - x_i)$$  

(3)

Where, $x_j$ are the sequential data values, $n$ is the length of the data set and

$$\text{sgn}(t) = \begin{cases} 
1, & t > 0 \\
0, & t = 0 \\
-1, & t < 0 
\end{cases}$$  

(4)

The value of $S$ indicate the direction of trend, A negative and positive value indicate falling, rising trend. Mann-Kendall have documented that when $n$ is greater or equal to $8$, the test statistics $S$ is approximately normally distributed with mean and variance as follows, (Raj.R.K., 2010):

$$E(S) = 0$$  

(5)

$$\text{Var}(S) = \frac{1}{18} \left[n(n-1)(2n+5) - \sum_{i=1}^{m} t_i (t_i-1)(2t_i+1)\right]$$  

(6)

where, $m$ is the number of tied groups and $t_i$ is the size of the $ith$ tie group. The standardized test statistics $Z$ is computed as follows:

$$Z_{mk} = \begin{cases} 
\frac{S-1}{\sqrt{\text{Var}(S)}}, & \text{for } S > 0 \\
0, & \text{for } S = 0 \\
\frac{S+1}{\sqrt{\text{Var}(S)}}, & \text{for } S < 0 
\end{cases}$$  

(7)

The standardized Mann-Kendall statistics $Z$ follows the standard normal distribution with zero mean and unit variance.

3.5 Measure of Periodicity:

Periodicity is one of the deterministic components in the time-series. The power spectrum is a method of analysis that was developed to handle the problem of periodicity in variations of natural events observed in time, such as in climatological and hydrological time series. Power spectrum analysis, also called generalized harmonic analysis, was derived from the principles first developed by Wiener. It is based on the premise that the time series are not necessarily composed of a finite number of oscillations, each with a discrete wavelength, but rather that they consist of virtually infinite number of small oscillations spanning a continuous distribution of wavelengths. The spectrum therefore, gives the distribution of variations in a time series over a continuous domain of all possible wavelengths.

Procedures for computing the power spectra may vary. Here, in this study, an approach described in WMO, developed by Tukey and Blackman was
employed. A detailed description of this approach can also be found in various textbooks. It can be summarized through the following steps:

(i) First, all serial correlation coefficients of normalized climatic series (eq. 1) are computed for lags from \( L = 0 \) to \( m \), where \( m \) is the maximum lag \( (m = n/3) \). The serial correlation coefficient can be computed using eq. (2).

(ii) Using the values of \( r_L \), the ‘raw’ spectral estimates, \( \hat{s}_k \), are computed using the following set of equations:

\[
\hat{s}_k = \frac{2}{m} \sum_{i=1}^{m} r_{i-k} - \frac{1}{m} r_m(-1)^k.
\]

(8)

For \( k = 1,2, \ldots, m-1 \)

Smallest is the value of \( k \) longest will be the wavelength of the spectrum, i.e. shortest wavelength is achieved at \( k = m \). The resulting values of \( S_k \) can be plotted superposed on the sample spectrum.

The statistic associated with each spectral estimate is the ratio of the magnitude of the spectral estimate to the local magnitude of the continuum (red noise continuum), found that the quantity of this ratio is distributed as Chi-square divided by the degree of freedom. The degree of freedom, \( \nu \), of each estimate of a computed spectrum is given as follows.

\[
\nu = \frac{2n-m}{2m}
\]

(9)

finally, the cycle associated in the time series is computed as follows.

\[
p = \frac{2m}{L}
\]

(10)

Here \( P' \) is the periodicity, \( m' \) is total number of time lag, and \( L' \) is the significant peak in spectral estimation. Further detail refer the full paper (Rai R.K et.al., 2010).

4. Results and Discussion:

4.1 Persistance:

The first step in time series analysis is to draw time series plot which provide a preliminary understanding of time behavior of the series as shown in Fig. 4. It shows surface ozone concentration with respect to daily time scale. The result of persistence analysis are given in the table (1), values are plotted as log \( w \) in \( x \)-axis, log \( R/S \) in \( y \)-axis as shown in fig. 2, slope is calculated for the ozone time series (fig 4). The slope for the dataset is found to be 0.4017, which is the Hurst exponent. So, the air pollutant surface ozone at this Chennai region follows anti-persistence series pattern. The fractal dimension \( D \) takes the value of 1.6 using the Hurst exponent. The fractal dimension \( D \) also exhibits anti-persistence behavior.

Fig 2. Day average of surface ozone time series for the period 2011-2012

---

655

Padma et al.
Table 1: Calculated the values of rescaled range (R/S)

<table>
<thead>
<tr>
<th>(Width) W</th>
<th>R/S</th>
<th>log W</th>
<th>log (R/S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>4.187</td>
<td>1.146128036</td>
<td>0.621902961</td>
</tr>
<tr>
<td>28</td>
<td>6.94</td>
<td>1.447158031</td>
<td>0.84135947</td>
</tr>
<tr>
<td>56</td>
<td>10.159</td>
<td>1.748188027</td>
<td>1.00685096</td>
</tr>
<tr>
<td>112</td>
<td>23.2</td>
<td>2.049218023</td>
<td>1.365487985</td>
</tr>
<tr>
<td>224</td>
<td>47.891</td>
<td>2.350248018</td>
<td>1.680253906</td>
</tr>
<tr>
<td>446</td>
<td>7.784</td>
<td>2.649334859</td>
<td>0.891202827</td>
</tr>
</tbody>
</table>

Table 2: The maximum Predictable Index

<table>
<thead>
<tr>
<th>Largest lyapunov exponent</th>
<th>Maximum predictable time scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = 0.1653$</td>
<td>6 days</td>
</tr>
</tbody>
</table>

Fig. 3: Plot for the calculation of Slope in Hurst Exponent

A Hurst exponent value $0 < H < 0.5$ will exist for a time series with anti-persistent behavior (or negative auto-correlation). Here, an increase will tend to be followed by a decrease or a decrease will be followed by an increase.

Since the Hurst exponent provides a measure for predictability, we can use this value to guide data selection before forecasting. We can identify time series with large or small Hurst exponents before we try to build a model for prediction. Furthermore, if we can focus on the periods with small or large Hurst exponents.
exponents, this will save our time and effort and hence lead to a better forecasting.

4.2 Chaos (LYAPuNOV EXPONENT):
The chaotic behavior of the atmospheric data of the surface ozone at Chennai region is calculated. From the respective error the lyapunov exponents for 446 days are calculated. The Lyapunov spectrum can be used to give an estimation of the fractal dimension of the considered dynamical system, i.e., the surface ozone time series.

- $\lambda < 0$, The orbit attracts to a stable fixed point or stable periodic orbit. Negative Lyapunov exponents are characteristic of dissipative or non-conservative systems.
- $\lambda = 0$, The orbit is a neutral fixed point (or an eventually fixed point). A Lyapunov exponent of zero indicates that the system is in some sort of steady state mode.
- $\lambda > 0$, The orbit is unstable and chaotic. Nearby points, no matter to how close, will diverge to any arbitrary separation. All neighborhoods in the phase space will eventually be visited. These points are said to be unstable.

As the variation surface ozone is periodic, an attempt is made to calculate the lyapunov exponents for the ozone time series data which is shows in Fig 3. The $\lambda$ value is positive ($\lambda >0$) which implies that the orbit is unstable and chaotic. The maximum value of lyapunov exponent for the above system is 0.1653 and implies that the system is chaotic. The lyapunov time is defined as the inverse of the maximum lyapunov exponent. For the above system, the maximum time scale obtained for the surface ozone time series is 6 days is shows in table 3. This indicates the time, after which, the system starts to behave in a chaotic manner(Shalaleh mohammadi,2009). It means that it is possible to estimate the value of surface ozone concentration up to 6 days. So, it is possible to use short time forecasting methods.

4.3 Trend (Mann-Kendall Test):
For the original Mann–Kendall test, the time series must be serially independent in nature. In many real situation, the observed data are serially dependent(i.e., autocorrelated). The auto correlation in the observed data will results in interpretation of trend test results. Previous researcher stated that ‘Positive serial correlation among the observations would increase the chance of significant answer, even absence of trend’ ( Cox and Stuart,1955). If the seasonality exists in the data, by dividing the observations into separate classes according to the season and then performing the Mann-Kendall trend test on the sum of the statistics from each season, the effect of seasonality can be eliminated, This modification is called seasonal Mann-Kendall test(Hirsch RM, 1984). The standardized Mann-Kendall statistics $Z$ follows the standard normal distribution with zero mean and unit variance. Here the null hypothesis about no trend is accepted due to the ‘p’ value(0.47) is greater than the significance level of alpha (0.05) 46% in this study.

4.4 Periodicity:
The procedure for evaluating the results of power spectrum analysis mention in WMO (World Meteorological Organization, Geneva) is described below in fig. 5. The spectral estimates are shown in Fig.(5) which is a sample time series of surface ozone of Chennai region. If a time series has persistence, the spectrum changes over most of the wavelengths and the amplitude of the spectrum has a decreasing from long to short wavelengths.(i.e. corresponding to increasing order of lags). If the spectrum of a time series having persistence with necessary exponential relationship between $r_1$, $r_2$ and $r_3$ (i.e. Markov-type persistence), the appropriate null hypothesis was assumed to be a Markov red noise continuum. From our spectrum it can be stated that times series with positive lag-3 shows the low frequency variability (i.e. $L = 3$, and $P = 2m/L = 98.6$ days).

However, for presentation, power spectrum plots of surface ozone are given Fig.(5). To evaluate the power spectrum results, generally period values were computed using eq (10) (i.e. $P = 2m/L$) for all the time series. The value of period is computed with their significance level for surface ozone. From this method period is calculated as 98.6 days.
5. Conclusion:
Based on the Time series analysis, following conclusions can be drawn;

1. There is a considerable result had been drawn from ozone time series pattern in the analysis of persistence and periodicity, trend and chaotic. From our study the air pollutant surface ozone at this Chennai region follows anti-persistence series pattern. Since the Hurst exponent provides a measure for predictability, we can use this value to guide data selection before forecasting. Anti persistence is the tendency for the magnitude of an event to be independent on the magnitude of previous event(s), a memory effect, e.g., tendency for low ozone concentration to follow high ozone concentration and that for high ozone concentration to follow low ozone concentration. Furthermore, we can focus on the short time prediction model to predict the ozone level in this region.

2. As the variation surface ozone is periodic, an attempt is made to calculate the lyapunov exponents for the time series data, after calculating $\lambda$ (Fig 3). The maximum value of lyapunov exponent for the system is 0.1653 and $\lambda$ value is positive ($\lambda > 0$) which implies that the system is unstable and chaotic. For the above system, the value of lyapunov time is 6.05 This indicates the time, after which, the system starts to behave in a chaotic manner.

3. The standardized Mann-Kendall statistics $Z$ follows the standard normal distribution with zero mean and unit variance and the null hypothesis about no trend is accepted due to the ‘$p$’ value is greater than the significance level of alpha.

4. From the presentation of power spectrum plots of surface ozone it is evaluated that the power spectrum results period is calculated as 98.6 days.

6. Acknowledgments:
The authors wish to thank the Tamil Nadu Pollution Control Board, Chennai for providing valuable guidance and for checking the accuracy of the instruments. We also thank our co-researchers who have given their valuable suggestions enabling us to release this research paper and their help is gratefully acknowledged. Authors are grateful to editorial board members and team of reviewers who have helped to bring quality to this manuscript.

References:
4) Deepesh machiwal and madan k. jha(2006): Time series analysis of hydrologic data for water