SCALING AND PERSISTENCE OF OZONE CONCENTRATIONS IN KLANG VALLEY, MALAYSIA
(Penskalaan dan Keberterusan Aras Kepekatan Ozon di Lembah Klang, Malaysia)

MUZIRAH MUSA1,2, ABDUL AZIZ JEMAIN2 & WAN ZAWIAH WAN ZIN2

ABSTRACT
The aim of this paper is to describe the characteristics of ozone concentration based on the value of Hurst coefficient. The Hurst coefficient, denoted as $H$, is used to explain the long-term degree of persistency in ozone concentration in Malaysia. This paper investigates the scaling properties and persistency of ozone concentrations at six selected air quality monitoring stations in Klang Valley, Malaysia. Daily mean for the hourly data of ozone concentration from 1998-2006 is considered in this study. In describing the statistical properties, several related statistics are identified and the autocorrelation functions of the observed data were plotted. The anomaly method is used to deseasonalise the seasonal variation and the natural non-stationary of the data. Next, the persistency of ozone concentration is determined by using the scaling analysis, namely Dispersional Ratio Method. This method considers two dispersion measures, that is the range and variance. The result shows that there are two different scaling regions, separated by a critical time scale, $n_c$ which is approximately 90 days. For shorter time scale ($n < n_c$), the ozone concentration persists and the concentration is high on higher time scale ($n > n_c$) where the $H$ values are between $0.5 < H < 1$ at all monitoring stations. These findings provide information on scaling behaviour and persistence in ozone concentration in the ambient air of Klang Valley, Malaysia.

Keywords: dispersional ratio; Hurst coefficient; ozone; persistence; scaling

ABSTRAK

Kata kunci: nisbah penyerakan; pelaki Hurst; ozon; keberterusan; penskalaan
1. Introduction

The analysis of ozone (O$_3$) concentration behaviour, particularly in terms of statistical properties and persistency, is of great importance in many areas such as air quality monitoring and warning system management. Ozone, identified as one of the greenhouse gases, has been the subject of intense research in recent years. It has been proved to be a serious air pollution problem for many countries all over the world (Pudasainee et al. 2010; Chelani 2009; Atkinson-Palombo et al. 2006; Varotsos et al. 2005; Lal et al. 2000). Ozone formation generally results from human activities and is known as secondary pollutant formed by photochemical reaction involving nitric oxide (NO$_2$) and volatile organic compounds (VOCs) in sunlight. Its high concentration is becoming a matter of concern due to its adverse effects on human health such as respiratory problem. In considering the harmful effect of overexposure to O$_3$, assessing and understanding of O$_3$ formation mechanism is considered as the most pervasive scientific topic and is critically important for pollution control measures. Understanding the behaviour of O$_3$ provides information to prepare for adaptation and mitigation since the rapid growth of urbanisation and industrialisation has led to the increasing problem of air pollution everywhere in the world including Malaysia.

Problems in analyzing the O$_3$ mainly arise due to the complex mechanism involved in ozone production. Ozone concentration observed over time are often recorded as time series and are characterised by a large number of large fluctuations. Thus, the autocorrelation or serial dependence often exist and it is difficult to interpret (Lee et al. 2006). This dependence is usually known as persistence and refers to ‘memory’ or internal correlations within a time series. A series is persistent if its successive values are positively correlated, whereas a series is anti-persistent if these values are negatively correlated.

Persistence analysis will be able to give a better understanding of the system governing the ozone formation (Chelani 2009) and has been proved useful for obtaining accurate information regarding time variability. It is also able to assist in prediction as high concentration is usually followed by further high concentration if similar behaviour persists. Persistence in time series has been explored in a number of studies such as, Chelani (2009) who investigated the seasonal effects of monthly hourly ground level ozone concentration at two sites in Delhi. Weng et al. (2008) who explored the underlying structure of ozone recorded at the Kaohsiung metropolitan region in Southern Taiwan. Kai et al. (2008) analysed three pollution indexes (SO$_2$, NO$_2$ and PM$_{10}$) and daily pollution indexes (APIs) of Shanghai in China. Windsor and Toumi (2001) sought to determine whether the UK pollution time series is likely to be persistent, and Vorotsos et al. (2005) examined the persistence in air pollution time series from Athens, Greece and Baltimore, Maryland. Of these, their result showed that similar persistence exists. This analysis involved the determination of Hurst coefficient, $H$ which is the most classical persistence indicator. The degree of persistence depends on the extend for $H$ close to 1.

The primary goal of this study is to explore the stochastic structure of O$_3$ with an emphasis on characterizing their variability and persistency behaviour. In view of this purpose, this study is conducted to investigate the statistical properties and persistency of ozone concentration observed at six monitoring station in Klang Valley, Malaysia. This area is hypothesised as exposed to significant air pollution resulting from high population density and its location in an industrial area as well as from traffic emissions. In order to describe and investigate the variability of ozone concentration, the statistical descriptive analysis will be used. This study also aims to integrate the usage of scaling analysis by using a Dispersional Ratio Method approach in revealing additional information regarding the scaling and persistence of ozone concentration.
during the study period. This approach is proposed since the conventional analysis such as autocorrelation function and Fourier spectral analysis are said to be insufficient.

The paper is organised as follows. The first part of this paper describes the data quality, location as well as the statistics used in this study. 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.

2. Material and Method

2.1. Data and locations

The ozone concentration data was recorded as part of Malaysian Continuous Air Quality Monitoring (CAQM) program by a private company, Alam Sekitar Sdn Bhd (ASMA 2006) on behalf of the Air Quality Division of the Department of the Environment, Malaysia (DOE). The pollutants measurements are performed on hourly basis and reported in unit of part per million \(\text{ppm}\). For the purpose of study, data was taken from six stations situated in Klang Valley, Malaysia. The Klang valley is an area in Malaysia, comprises of Kuala Lumpur and its suburb and adjoining cities and towns in the state of Selangor. These sampling stations are surrounded by industries, residential and commercial areas and consequently, congested traffics. This area was expected to be affected by the pollutants, discharged continuously from the human activities. The data in detail are shown in Table 1. Most of the stations have a data record spanning from January 1998 to December 2006, except period apart from station S2 and station S4 which data are from December 1998 to December 2006 and April 2003 to December 2006, respectively.

<table>
<thead>
<tr>
<th>Code</th>
<th>Station</th>
<th>State</th>
<th>Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Jabatan Bekalan Air, Daerah Gombak</td>
<td>Selangor</td>
<td>Jan 1998 – Dec 2006 (9 years)</td>
</tr>
<tr>
<td>S3</td>
<td>Victoria Institution</td>
<td>Kuala Lumpur</td>
<td>Jan 1998 – Dec 2003 (6 years)</td>
</tr>
<tr>
<td>S5</td>
<td>Country Height, Kajang</td>
<td>Selangor</td>
<td>Jan 1998 – Dec 2006 (9 years)</td>
</tr>
</tbody>
</table>

The missing values were treated based on Azmi et al. (2010) whereby the imputed values are interpolated using the nearest neighbour value available the data set. To overcome the natural non-stationarity in the data, an anomaly method, proposed by Windsor and Toumi (2001), was applied to remove the seasonal trend. The anomaly was calculated by subtracting the hourly average for that day of the year from the observed values. Next, the daily mean hourly ozone concentration could be obtained and used for further analysis.
2.2. Method

The time series that of daily mean hourly ozone concentrations is plotted to provided visualisation of the variability and persistency of ozone concentration at the six stations. From Figure 1, it can be observed that ozone time series exhibit distinct diurnal and seasonal pattern.

![Figure 1: Time series plot of ozone concentration at six stations](image)

The standard statistical characteristics of the daily mean hourly ozone concentration are computed for each station. Next, the autocorrelation function (ACF) for ozone concentration are examined. Autocorrelation is also called lagged correlation or serial correlation, which refers to the correlation of a time series data with its own past and future values. This dependence is usually known as persistence in the terminology of the atmospheric sciences. Persistence is thus defined as the existence of positive statistical dependence among successive occurrences of a given event. The short-range correlations described by the ACF that shows exponentially decay with certain decay time, which it is decaying fast to zero. On the other hand, for long-range correlations, the ACF shows a very slow hyperbolically decay as time increased. Nevertheless, the direct calculation of ACF is usually not appropriate due to noise superimposed on the collected of data and due to underlying trends of unknown origin (Varotsos et al. 2005). In order to study the persistency in time series, the Dispersion Ratio Method, D/S is used. Then, Hurst parameter, $H$ is then used as an indicator to explain the degree of persistency in ozone concentrations time series. The Hurst parameter, $H$ can be obtained by following the steps as below.

**Dispensional Ratio Method, $D/S$**

This method, $D/S$ is determined based on the ratio of the selected dispersion measure, $D$ to the standard deviation, $S$ of the time series. This method was adapted based on the famous the method of rescaled range analysis, $R/S$ proposed by Hurst et al. (1965) and rescaled variance analysis, $V/S$ introduced by Cajueiro and Tabak (2005a).
The analysis begins with a series of daily mean hourly ozone concentration $x_j; j = 1,2,\ldots,N$, partitioned into $M$ non-overlapping interval of equal size $n$, where $M = N/n$ is the total number of interval and $\{x_{ij}\}$ representing the $j$th element in $i$th interval with $i=1,2,\ldots,M$. For the $i$th interval, the mean and variance are

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

While the cumulative sum of deviation of the time series is given as

$$
y_{it} = \sum_{j=1}^{t} (x_{ij} - \bar{x}_i), \quad t = 1,2,\ldots,n
$$

and will be used for further analysis. This generated newly series, $y_{it}$, preserve the original properties such as variability. Next, the determination of $D/S$ is given by

$$
D_i/S_i = \begin{cases} 
\frac{R_i}{S_i} & \text{for} \quad R_i = \left\{ \max(y_{it}) - \min(y_{it}) \right\}_{t=1\ldots n} \\
\frac{V_i}{S_i} & \text{for} \quad V_i = \sqrt{\text{var}(y_{it})} 
\end{cases} .
$$

The function of $D_i/S_i$ is then averaged over all the $i$th interval of equal size $n$ ;

$$
\left( \frac{D}{S} \right)_n = \frac{1}{M} \sum_{i=1}^{M} \left( \frac{D_i}{S_i} \right) .
$$

Values of $\left( \frac{D}{S} \right)_n$ for various $i$th interval of size $n$ are then calculated. The relationship between $\left( \frac{D}{S} \right)_n$ and $n$ can then be expressed in terms of power law, that is

$$
\left( \frac{D}{S} \right)_n \propto n^H
$$

The value of $H$ can be computed by running a regression equation of
\[
\log\left(\frac{D}{S}\right) = H \log(n) + \log(c)
\] 

(6)

where the slope of the curve, \(H\), represents the degree of persistency. The Hurst coefficient values with \(0.5 < H < 1\), characterises persistent or positively correlated time series while \(0 < H < 0.5\) indicates anti-persistent or negatively correlated time series. For \(H=0.5\) the time series are said to be independent or random.

3. Result and discussion

During the first phase of the study, the distribution of data was described using classical descriptive statistics. The detail results for six monitoring stations are shown in Table 2. For all stations under study, the mean and median are very similar. Also, the coefficients of variation are nearly equal, indicating that the spatial differences in ozone concentration are rather small. Furthermore, the mean values for all stations was far below the value suggested by Malaysian Ambient Air Quality Guidelines for the 24-hour average concentration and the maximum value are also still below the recommendation threshold limit.

The coefficients of skewness for all stations are positive and this implies that the ozone concentration are all right-skewed. However, the degree of right-skewness of S5 station is found to be weaker than that of other stations.

Table 2: Statistics of the daily mean hourly of ozone concentrations (ppm)

<table>
<thead>
<tr>
<th>Summary</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>3285</td>
<td>3285</td>
<td>2555</td>
<td>1360</td>
<td>3285</td>
<td>3285</td>
</tr>
<tr>
<td>Mean</td>
<td>0.016</td>
<td>0.016</td>
<td>0.013</td>
<td>0.014</td>
<td>0.019</td>
<td>0.020</td>
</tr>
<tr>
<td>Median</td>
<td>0.015</td>
<td>0.014</td>
<td>0.012</td>
<td>0.013</td>
<td>0.018</td>
<td>0.019</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.006</td>
<td>0.007</td>
<td>0.006</td>
<td>0.006</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.050</td>
<td>0.056</td>
<td>0.041</td>
<td>0.039</td>
<td>0.049</td>
<td>0.054</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.002</td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>Coefficient of Variation</td>
<td>0.410</td>
<td>0.424</td>
<td>0.446</td>
<td>0.412</td>
<td>0.360</td>
<td>0.370</td>
</tr>
<tr>
<td>Coefficient of Skewness</td>
<td>0.508</td>
<td>0.532</td>
<td>0.512</td>
<td>0.519</td>
<td>0.304</td>
<td>0.457</td>
</tr>
</tbody>
</table>

*Note: standard safe level by Malaysian Ambient Air Quality Guidelines (MAAQG): 0.10ppm*

The autocorrelation, \(ACF\), of the daily mean hourly of ozone concentration are given in Figure 2 that exhibited marked correlations at high lags. Moreover, the autocorrelation function, \(ACF\) decreases slowly in the manner that certainly different from an exponential decay, where it hyperbolically decays to zero. Theoretically, this gave an initial insight on the existence of persistency within the observed ozone time series. These results are further confirmed by scaling analysis and the results are graphically shown in Figures 3 and 4.
Figure 2 shows the R/S analysis for the daily mean hourly ozone concentrations at six monitoring stations. The plots obviously show two different scaling regions separated by a point known as the critical time scale, \( n_c \). This critical time scale, \( n_c \), is about 90 days, or equivalent to 3 months. Referring to Figure 3(a) a plot for station S1 (Gombak) on a shorter time periods \((n < n_c)\), the plot can be fitted to a straight line with a Hurst coefficient, \( H \), of 0.76 which exhibits persistence. Over longer time periods \((n > n_c)\), a line with an increased slope with \( H \) is 0.99 indicating high persistence at large temporal scale. The similar results are also obtained at the other monitoring stations as shown as in Figure 3(b) – (f). Thus, this shows that when the temporal scale is larger than 90 days, the persistence becomes higher.

Figure 4 shows the results of the V/S analysis performed for the daily mean hourly ozone concentrations at six monitoring stations. In this case, it is still found the two different scaling regions with two different Hurst coefficients, \( H \) with a critical time scale, \( n_c \), of about 90 days. Figure 4(a), shows for shorter time periods \((n < n_c)\), the \( H \) value is 0.61 and 0.97 for longer time periods \((n > n_c)\). The results for the others monitoring stations are shown in Figure 4(b) – (f) respectively. The series is persistence in time span less than 90 days and shows high persistence in time span greater than 90 days which is consistent with the result of R/S analysis.

Table 3 provides the detailed results for a straight line plotted with two different scaling regions with the estimated \( H \) values, multiple R-squares and standard error of the line calculated using the two dispersional ratio methods; R/S and V/S.
Table 3: The estimated H values using D/S method

<table>
<thead>
<tr>
<th>Methods</th>
<th>R/S</th>
<th>V/S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n &lt; n_c$</td>
<td>$n &gt; n_c$</td>
</tr>
<tr>
<td>S1</td>
<td>$\hat{H}_1$</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.99</td>
</tr>
<tr>
<td>S2</td>
<td>$n &gt; n_c$</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>$n &gt; n_c$</td>
<td>0.74</td>
</tr>
<tr>
<td>S3</td>
<td>$n &gt; n_c$</td>
<td>0.74</td>
</tr>
<tr>
<td>S4</td>
<td>$n &gt; n_c$</td>
<td>0.74</td>
</tr>
<tr>
<td>S5</td>
<td>$n &gt; n_c$</td>
<td>0.74</td>
</tr>
<tr>
<td>S6</td>
<td>$n &gt; n_c$</td>
<td>0.74</td>
</tr>
</tbody>
</table>

* $n_c$: critical time scale = 90 days (3 months)

Both R/S and V/S methods confirm that persistence exists in the daily mean hourly ozone concentration series. These results verify that the Hurst coefficients obtained are reliable since both method R/S and V/S gave very close values for $H$ coefficient. The daily mean hourly ozone concentrations series obey a power-law in two scaling regions with mean Hurst coefficient, $H_1$, between 0.60 and 0.92, whereas $H_2$ is between 0.74 and 0.99. In other words, the linear relationship as shown in Figures 3 and 4 indicates that the ozone concentration fluctuations in small time intervals are related to the fluctuations in larger time intervals in a power-law fashion.

It is worth to mention that, the results obtained are similar as a study done by Weng et al. (2008) which explored the underlying structure of ozone at ground level and identified ozone persistence with Hurst coefficient above 0.75. Windsor and Toumi (2001) applied R/S analysis to average hourly ozone, PM$_{10}$ and PM$_{2.5}$ series in UK and obtained mean Hurst coefficient of 0.77, 0.808, and 0.77, respectively. Their result showed high persistence up to 400 days. Whereas, Chelani (2009), investigated the seasonal effects of monthly hourly ground level ozone concentration using rescaled range R/S find out that, for monthly observation, persistence observed up to 2 days and the difference has significant effect on the persistence property in time series. Moreover, Varotsos et al. (2005) in their study found an hourly ozone time series observed with high persistence of about 1 week to 5 years. Cajueiro and Tabak (2005b) used R/S and V/S to test the persistency of the equity returns of Australia, Hong Kong, Singapore and Taiwan. They found both results for shuffled and unshuffled returns gave similar results. As for the R/S, the Hurst coefficient values obtained are between 0.5 and 0.6, while for the V/S, the Hurst coefficient values are in between 0.4 to 0.6. Turvey (2007) also showed similar result to the analysis done on the agricultural commodity price using V/S by obtaining Hurst coefficient values between 0.4 to 0.6.
Scaling and persistence of ozone concentrations in Klang Valley, Malaysia

Figure 3: Plot log(R/S) against log(n) for the six monitoring stations
Figure 4: Plot log(V/S) against log(n) for six monitoring stations
4. Conclusion

Analyses of the scaling and persistency of daily mean hourly ozone concentrations at six monitoring stations in Klang Valley, Malaysia allows the following conclusions:

The results show that the average and maximum values of ozone concentrations at all monitoring stations are under the permissible value as recommended by the Malaysian Department of Environment. The preliminary analysis based on the autocorrelation functions proved that these series are positively correlated to each other. As such, this shows that persistence in ozone concentrations does exist in the ambient air of Klang Valley, Malaysia and this is further confirmed by Dispensatio Ratio Methods, D/S. This method has indentified persistency of ozone concentrations occurs in two scaling regions. For a shorter time scale which is a period from 5 up to 90 days, the ozone concentration is found to be persistent with $H$ value in between 0.65 to 0.80. On a larger time scale, that is, for period over 90 days, it remains persistence with high persistivity and its $H$ values are in between 0.93 and 0.94. These findings are important in helping us to understand the behaviour of ozone concentrations and hence may facilitate further studies such as studying the scaling factor with particular scaling coefficient as such study may assist in obtaining predictions on the future behaviour of the ozone system. The findings of this study is hoped to benefit researchers, decision makers and publics in monitoring and ensuring the sustainability of meteorology systems. Future research could be done involving more stations.

Acknowledgement

The authors are grateful to the Air Quality Division of the Department of the Environments (DOE), Malaysia, 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 to Universiti Pendidikan Sultan Idris, Universiti Kebangsaan Malaysia and Ministry of Higher Education, Malaysia on their sponsorship. This research is partially funded by DLP-2013-007 and UKM-AP-2011-19.

References


---

1Department of Mathematics
Faculty of Science and Mathematics
Universiti Pendidikan Sultan Idris
35900 Tanjong Malim
Perak DR, MALAYSIA
Mel-e: muzirah@fsmt.upsi.edu.my*

2School of Mathematical Sciences
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor DE, MALAYSIA
Mel-e: azizj@ukm.my, w_zawiah@ukm.my

*Corresponding author*