

---

# Time Series Analysis of Surface Ozone Monitoring Records in Kemaman, Malaysia

---

Marzuki Ismail, Azrin Suroto and Nurul Ain Ismail

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/50033>

---

## 1. Introduction

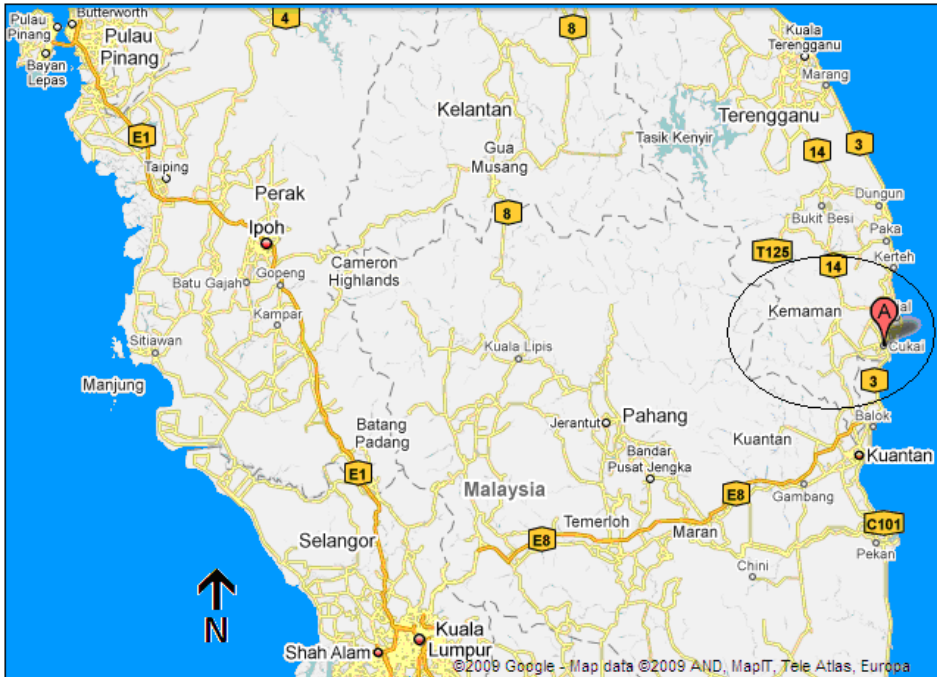
Tropospheric ozone is known as environmental air pollutants that arise from photochemical reaction among various natural and anthropogenic precursors that are volatile organic compounds (VOCs) and organic nitrogen (NO<sub>x</sub>). Accumulation of the ozone may highly happen under favorable meteorological conditions and will have an adverse effect on human health and ecosystem [1]. Chan & Chan, 2001 concluded that people in Asia also cannot escape from the adversely impact ozone pollution as there were elevated ozone level being detected. Nevertheless, the long-term ozone trend has been less researched, especially in Malaysia.

The time series analysis is one of the best tool in understanding cause and effect relationship of environmental pollution [3, 4,5]. Its applications in many studies were done to describe the past movement of particular variable with respect to time. However, there were several different techniques applied by researcher so that the change of air pollution behavior through time period can be determined [6, 7]. A study by Kuang-Jung Hsu,2003 was done by using autoregression variation (VAR) in order to establish interdependence between primary and secondary air pollutants in area of Taipei. Besides, Omidravi et al., 2008 had applied the time series analysis in their investigation in order to find the answer that relate to extreme high ozone concentrations for each season in Ishafan by using Fast Fourier Transform. Therefore, this study aims to determine qualitative and quantitative aspect of the tropospheric ozone concentrations so that prediction on future concentration of the anthropogenic air pollutant can be achieved in the study area, i.e. Kemaman, Malaysia.

## 2. Material and method

This study was conducted in Kemaman (04°12'N, 103°18'E), a developing Malaysian town located in between the industrializing of Kertih Petrochemical Industrial Area in the north and industrializing and urbanizing of Gabeng Industrial Area in the South (Figure 1). In this

area, there are dominant sources of ozone precursors related to industrial activities and road traffic.



**Figure 1.** Locations of air monitoring station in Kemaman

In this study, ozone trend was examined using ozone data consisting of 144 monthly observations from January 1996 to December 2007 acquired from the Air Quality Division of ASMA for Sekolah Rendah Bukit Kuang station located in Kemaman district; one of the earliest operational stations in Malaysia. The monitoring network was installed, operated and maintained by Alam Sekitar Malaysia Sdn. Bhd. (ASMA) under concession by the Department of Environment Malaysia [10]. Tropospheric ozone concentrations data was recorded using a system based on the Beer-Lambert law for measuring low ranges of ozone in ambient air manufactured by Teledyne Technologies Incorporated (Model 400E). A 254 nm UV light signal is passed through the sample cell where it is absorbed in proportion to the amount of ozone present. Every three seconds, a switching valve alternates measurement between the sample stream and a sample that has been scrubbed of ozone. The result is a true, stable ozone measurement [11].

Time series analysis was implemented using STATGRAPHICS® statistical software package. A time series consists of a set of sequential numeric data taken at equally spaced intervals, usually over a period of time or space. This study provides statistical models for two time series methods: trend analysis and seasonal component which are both in time scale.

The seasonal decomposition was used to decompose the seasonal series into a seasonal component, a combined trend and cycle component, and a short-term variation component, i.e,

$$O_t = T_t \times S_t \times I_t \quad (1)$$

where  $O_t$  is the original ozone time series,  $T_t$  is the long term trend component,  $S_t$  is the seasonal variation, and  $I_t$  is the short-term variation component or called the error component. As the seasonality increase with the level of the series, a multiplicative model was used to estimate the seasonal index. Under this model, the trend has the same units as the original series, but the seasonal and irregular components are unitless factors, distributed around 1. As the underlying level of the series changes, the magnitude of the seasonal fluctuations varies as well. The seasonal index was the average deviation of each month's ozone value from the ozone level that was due to the other components in that month.

In trend analysis, Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model was applied to model the time series behavior in generating the forecasting trend. The methodology consisting of a four-step iterative procedure was used in this study. The first step is model identification, where the historical data are used to tentatively identify an appropriate Box-Jenkins model followed by estimation of the parameters of the tentatively identified model. Subsequently, the diagnostic checking step must be executed to check the adequacy of the identified model in order to choose the best model. A better model ought to be identified if the model is inadequate. Finally, the best model is used to establish the time series forecasting value.

In model identification (step 1), the data was examined to check for the most appropriate class of ARIMA processes through selecting the order of the consecutive and seasonal differencing required to make series stationary, as well as specifying the order of the regular and seasonal auto regressive and moving average polynomials necessary to adequately represent the time series model. The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are the most important elements of time series analysis and forecasting. The ACF measures the amount of linear dependence between observations in a time series that are separated by a lag  $k$ . The PACF plot helps to determine how many auto regressive terms are necessary to reveal one or more of the following characteristics: time lags where high correlations appear, seasonality of the series, trend either in the mean level or in the variance of the series. The general model introduced by Box and Jenkins includes autoregressive and moving average parameters as well as differencing in the formulation of the model.

The three types of parameters in the model are: the autoregressive parameters ( $p$ ), the number of differencing passes ( $d$ ) and moving average parameters ( $q$ ). Box-Jenkins model are summarized as ARIMA ( $p, d, q$ ). For example, a model described as ARIMA (1,1,1) means that this contains 1 autoregressive ( $p$ ) parameter and 1 moving average ( $q$ ) parameter

for the time series data after it was differenced once to attain stationary. In addition to the non-seasonal ARIMA (p, d, q) model, introduced above, we could identify seasonal ARIMA (P, D, Q) parameters for our data. These parameters are: seasonal autoregressive (P), seasonal differencing (D) and seasonal moving average (Q). Seasonality is defined as a pattern that repeats itself over fixed interval of time. In general, seasonality can be found by identifying a large autocorrelation coefficient or large partial autocorrelation coefficient at a seasonal lag. For example, ARIMA (1,1,1)(1,1,1)<sup>12</sup> describes a model that includes 1 autoregressive parameter, 1 moving average parameter, 1 seasonal autoregressive parameter and 1 seasonal moving average parameter. These parameters were computed after the series was differenced once at lag 1 and differenced once at lag 12.

The general form of the above model describing the current value  $Z_t$  of a time series by its own past is:

$$(1 - \phi_1 B)(1 - \alpha_1 B^{12})(1 - B)(1 - B^{12}) Z_t = (1 - \theta_1 B)(1 - \gamma_1 B^{12}) \varepsilon_t \quad (2)$$

Where:

- $1 - \phi_1 B$  = non seasonal autoregressive of order 1
- $1 - \alpha_1 B^{12}$  = seasonal autoregressive of order 1
- $Z_t$  = the current value of the time series examined
- $B$  = the backward shift operator  $BZ_t = Z_{t-1}$  and  $B^{12}Z_t = Z_{t-12}$
- $1 - B$  = 1st order non-seasonal difference
- $1 - B^{12}$  = seasonal difference of order 1
- $1 - \theta_1 B$  = non seasonal moving average of order 1
- $1 - \gamma_1 B^{12}$  = seasonal moving average of order 1

For the seasonal model, we used the Akaike Information Criterion (AIC) for model selection. The AIC is a combination of two conflicting factors: the mean square error and the number of estimated parameters of a model. Generally, the model with smallest value of AIC is chosen as the best model [12].

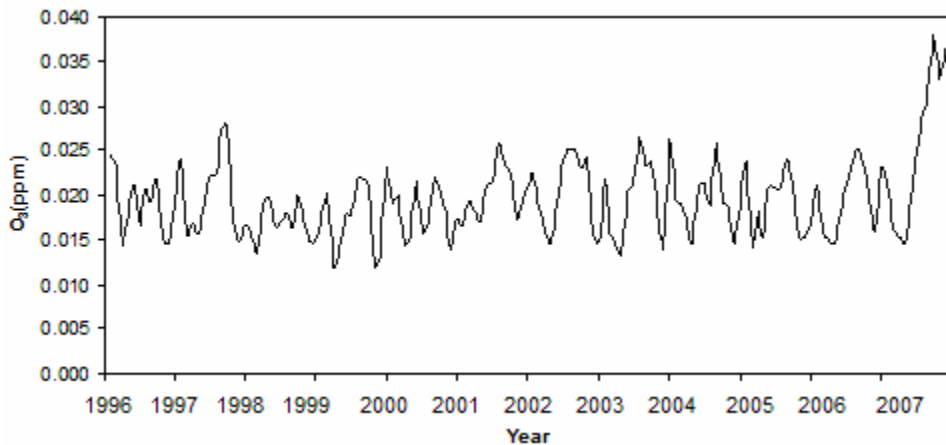
After choosing the most appropriate model, the model parameters are estimated (step 2) - the plot of the ACF and PACF of the stationary data was examined to identify what autoregressive or moving average terms are suggested. Here, values of the parameters are chosen using the least square method to make the Sum of the Squared Residuals (SSR) between the real data and the estimated values as small as possible. In most cases, nonlinear estimation method is used to estimate the above identified parameters to maximize the likelihood (probability) of the observed series given the parameter values [13].

In diagnose checking step (step 3), the residuals from the fitted model is examined against adequacy. This is usually done by correlation analysis through the residual ACF plots and the goodness-of-fit test by means of Chi-square statistics  $\chi^2$ . If the residuals are correlated, then the model should be refined as in step one above. Otherwise, the autocorrelations are white noise and the model is adequate to represent our time series.

The final stage for the modeling process (step 4) is forecasting, which gives results as three different options: - forecasted values, upper, and lower limits that provide a confidence interval of 95%. Any forecasted values within the confidence limit are satisfactory. Finally, the accuracy of the model is checked with the Mean-Square error (MS) to compare fits of different ARIMA models. A lower MS value corresponds to a better fitting model.

### 3. Results and discussion

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 Figure 2. Trend of the original series appear to be slightly increasing. Nonetheless, this needs to be tested and conformed through descriptive analysis and trend modeling.



**Figure 2.** Original monthly ozone concentration for Kemaman

In seasonality of ozone, a well-defined annual cycle was consistent with the highest ozone means occurring in August, and the lowest ozone means in November (Figure 3). Table 1 show the seasonal indices range from a low of 80.047 in November to a high of 122.058 in August. This indicates that there is a seasonal swing from 80.047% of average to 122.058% of average throughout the course of one complete cycle i.e. one year. The seasonal variation pattern in Kemaman differed from other countries, such as United States, United Kingdom, Italy, Canada, and Japan, in that the peak ozone concentration did not correspond to maximum photochemical activity in summer [14,15,16].

For the purpose of forecasting the trend in this study, the first 132 observations (January 1996 to December 2006) were used to fit the ARIMA models while the subsequent 12 observations (from January 2007 to December 2007) were kept for the post sample forecast accuracy check. Ozone concentrations data has been adjusted in the following way before the model was fit: - simple differences of order 1 and seasonal differences of order 1 were taken. The model with the lowest value (-11.8601) of the Akaike Information Criterion (AIC)

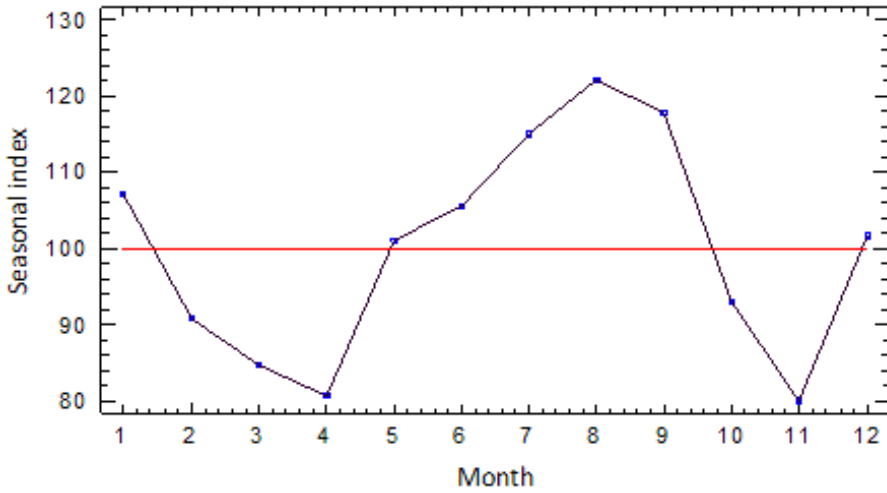


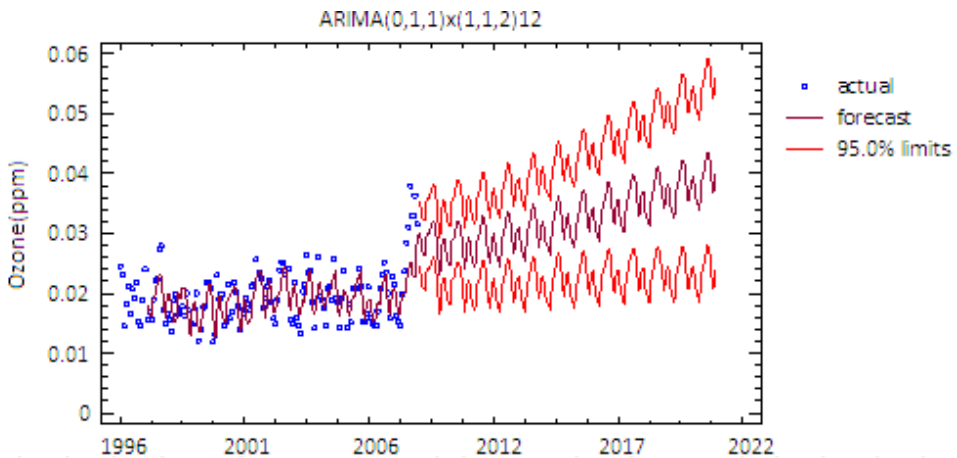
Figure 3. Annual variation of monthly ozone means

Month	Seasonal Index
January	107.199
February	90.8259
March	84.7179
April	80.7204
May	101.135
June	105.618
July	115.073
August	122.058
September	117.771
October	93.0941
November	80.0473
December	101.741

Table 1. Seasonal Index of Ozone

is (ARIMA)  $(0, 1, 1) \times (1, 1, 2)^{12}$  was selected and has been used to generate the forecasts (Figure 4). This model assumes that the best forecast for future data is given by a parametric model relating the most recent data value to previous data values and previous noise. As shown in Table 2, The P-value for the MA (1) term, SAR (1) term, SMA (1) term and SMA (2) term, respectively are less than 0.05, so they are significantly different from 0. Meanwhile, the estimated standard deviation of the input white noise equals 0.00277984. Since no tests are statistically significant at the 95% or higher confidence level, the current model is adequate to represent the data and could be used to forecast the upcoming ozone concentration. Therefore, we can assume that the best model for ground level ozone in Kemaman is the mathematical expression:

$$Z(t) = a(t) + 0.53a(t - 12) - 0.82(t - 1) - 1.67a(t - 12) + 0.73a(t - 24) + 0.82(1.67)a(t - 13) - 0.82(0.73)a(t - 25) \tag{3}$$



**Figure 4.** Model predicted plot of ozone concentration with actual and 95% confidence band

<i>Parameter</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>T</i>	<i>P-value</i>
MA(1)	0.818786	0.0478133	17.1246	0.000000
SAR(1)	0.531745	0.146213	3.63678	0.000400
SMA(1)	1.67374	0.092474	18.0996	0.000000
SMA(2)	-0.728689	0.081741	-8.91461	0.000000

**Table 2.** ARIMA  $(0, 1, 1) \times (1, 1, 2)^{12}$  model parameter characteristics

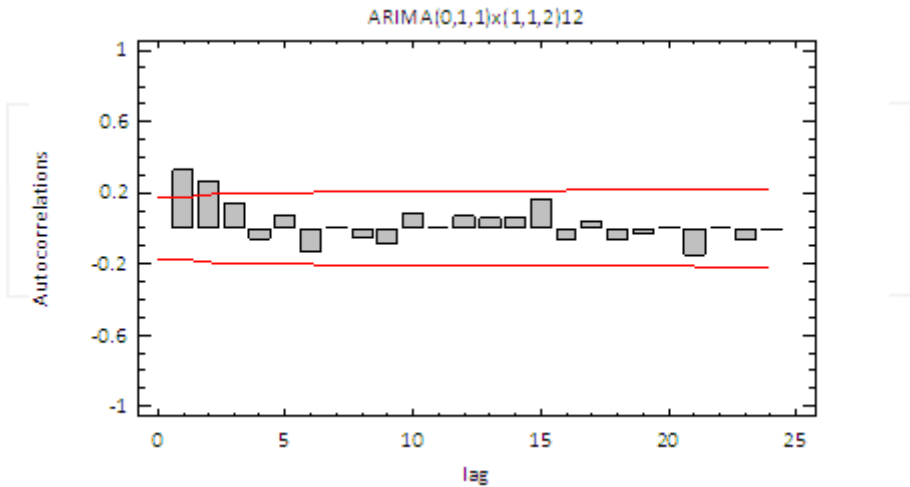
Model*	RMSE	MAE	MAPE	ME	MPE	AIC
(A)	0.00337	0.00249	13.526	0.000004	-1.5017	-11.2253
(B)	0.00271	0.00206	11.086	0.000002	-1.8498	-11.6431
(C)	0.00269	0.00201	10.786	0.000002	-1.8157	-11.6409
(H)	0.00267	0.00198	10.707	0.000003	-1.7185	-11.6712
(I)	0.00271	0.00201	10.870	-0.000050	-1.9817	-11.6423
(J)	0.00270	0.00199	10.671	0.000206	-0.5469	-11.6286
(M)	0.00258	0.00206	11.250	0.000031	-1.5638	-11.8601
(N)	0.00257	0.00204	11.192	-0.00009	-2.0636	-11.8478
(O)	0.00258	0.00206	11.298	-0.000053	-1.9673	-11.8392
(P)	0.00259	0.00207	11.260		-1.7473	-11.8382
(Q)	0.00259	0.00207	11.267	0.000030	-1.5789	-11.8335

\*Models

(A) Random walk; (B) Constant mean = 0.0190056; (C) Linear trend = 0.0184806 + 0.00000789502 t  
 (H) Simple exponential smoothing with  $\alpha = 0.109$ ; (I) Brown's linear exp. smoothing with  $\alpha = 0.0572$   
 (J) Holt's linear exp. smoothing with  $\alpha = 0.1291$  and  $\beta = 0.0301$ ; (M) ARIMA(0,1,1)×(1,1,2)<sub>12</sub>  
 (N) ARIMA(1,0,1)×(1,1,2)<sub>12</sub>; (O) ARIMA(0,1,1)×(1,1,2)<sub>12</sub> with constant  
 (P) ARIMA(0,1,1)×(2,1,2)<sub>12</sub>; (Q) ARIMA(0,1,2)×(1,1,2)<sub>12</sub>

**Table 3.** Model Comparison

According to plots of residual ACF (Figure 5) and PACF (Figure 6), residuals are white noise and not-auto correlated. Furthermore, as shown in Figure 7 of normal probability plot, residuals of the model are normal.



**Figure 5.** Residual autocorrelation functions (ACF) plot



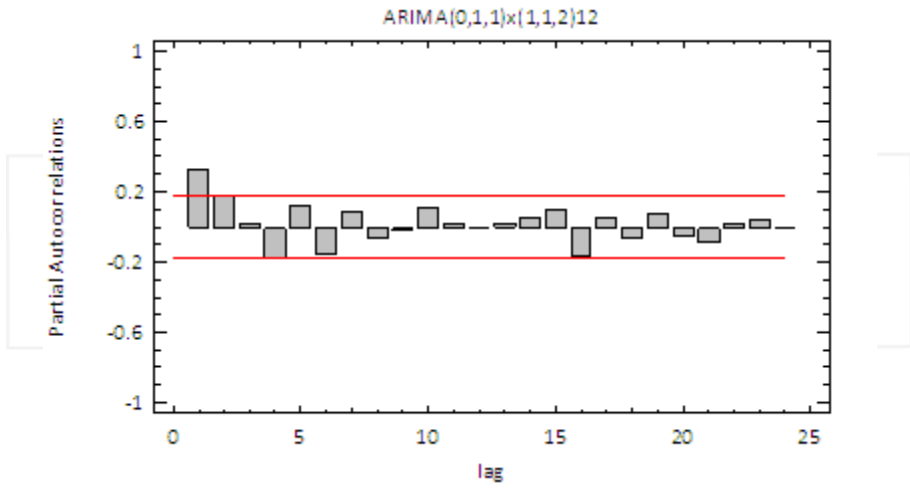


Figure 6. Residual partial autocorrelation (PACF) functions plot

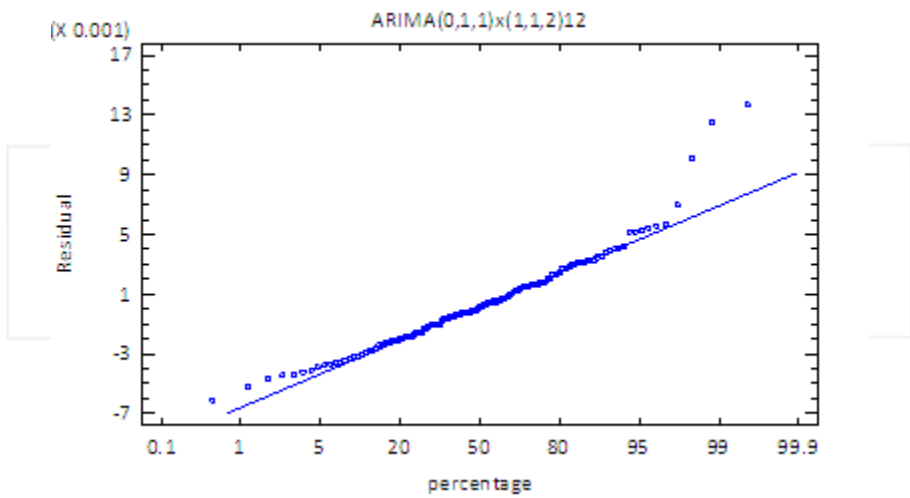


Figure 7. Residual normal probability plot

Based on the prediction for ozone concentration (Figure 4), there is a statistical significant upward trend at Kemaman station. The detection of a steady statistical significant upward trend for ozone concentration in Kemaman is quite alarming. This is likely due to sources of ozone precursors related to industrial activities from nearby areas and the increase in road traffic volume.

#### 4. Conclusion

Time series analysis is an important tool in modeling and forecasting air pollutants. Although, this piece of information was not appropriate to predict the exact monthly ozone concentration, ARIMA (0, 1, 1) × (1, 1, 2)<sup>12</sup> model give us information that can help the decision makers establish strategies, priorities and proper use of fossil fuel resources in Kemaman. This is very important because ground level ozone (O<sub>3</sub>) is formed from NO<sub>x</sub> and VOCs brought about by human activities (largely the combustion of fossil fuel). In summary, the ozone level increased steadily in Kemaman area and is predicted to exceed 40 ppb by 2019 if no effective countermeasures are introduced.

#### Author details

Marzuki Ismail\*, Azrin Suroto and Nurul Ain Ismail  
*Department of Engineering Science, Faculty of Science and Technology, Universiti Malaysia Terengganu, Kuala Terengganu, Malaysia*

#### Acknowledgement

The researchers would like to thank DOE Malaysia for providing pollutants data from 1996-2007 and the Ministry of Higher Education (MOHE) for allocating a research grant to accomplish this study.

#### 5. References

- [1] Wu, H.W.Y. & Chan, L.Y. 2001. Surface Ozone Trends in Hong Kong in 1985-1995. *Journal Of Environmental International* 26: 213-222
- [2] Chan, C.Y. & Chan, L.Y. 2000. The effect of Meteorology and Air Pollutant Transport on Ozone Episodes at a Subtropical Coastal Asian City, Hong Kong. *Journal of Geophysical Research* 105: 20707- 20719.
- [3] Kyriakidis, P.C. & Journal, A. G. 2001. Stochastic Modelling of Atmospheric Pollution: A Special Time Series Framework, Part II: Application to Monitoring Monthly Sulfate Deposition Over Europe. *Journal of Atmospheric Environment* 35: 2339-2348
- [4] Salcedo, R.L.R., Alvim, F.M., Alves, C. & Martins, F. 1999. Time Series Analysis of Air Pollution data. *Journal of Atmospheric Environment* 33 : 2361-2372.

---

\* Corresponding Author

- [5] Schwartz, J. & Marcus, A. 1990. Mortality and Air Pollution in London: A Time Series Analysis. *American Journal of Epidemiology* 131: 85-194
- [6] Hies, T., Treffeisen, R., Sebald, L. & Reimer, E. 2003. Spectral Analysis of Air Pollutants. Part I: Elemental Carbon Time Series. *Journal of Atmospheric Environment* 34: 3495-3502
- [7] Kocak, K., Saylan, L. & Sen, O. 2000. Nonlinear Time Series Prediction of O<sub>3</sub> Concentration in Istanbul. *Journal of Atmospheric Environment* 34: 1267-1271
- [8] Kuang-Jung Hsu. 2003. Time Series Analysis of the Independence Among Air Pollutants. *Atmospheric Environment*. Vol 26B, No.4, pp, 491-503,1992.
- [9] M. Omidravi, S. Hassanzadah, F. Hossicinibalam. 2008. Time Series Analysis of Ozone Data in Isfahan. *Physica A. Statistical Mechanics and Its Applications*, 387 (16-17),pp 4393-4403.
- [10] Afroz, R., Hassan, M.N., Ibrahim, N.A. 2003. Review of Air Pollution and Health Impacts in Malaysia. *Journal of Environmental Research* 92 (2): 71-77
- [11] ASMA., 2008. Alam Sekitar Malaysia Sdn Bhd. <http://www.enviromalaysia.com.my> 23/12/09.
- [12] Hong, W. 1997. A Time Series Analysis of United States Carrots Exports to Canada. Msc. Thesis, North Dakota State University.
- [13] Naill P.E. & Momani M., 2009. Time Series Analysis Model for Rainfall Data in Jordan: Case Study for Using Time Series Analysis. *American Journal of Environmental Sciences* 5 (5): 599-604
- [14] Angle, R.P. & Sandhu, H.S. 1989. Urban and Rural Ozone Concentrations in Alberta, Canada. *Journal of Atmospheric Environment* 23: 215- 221
- [15] Colbeck, I., MacKenzie, A.R. Air Quality Monographs, Air Pollution by Photochemical Oxidants, Vol.1. Amsterdam: Elsevier,1994 pp107-71, 232-326.
- [16] Lorenzini, G., Nali, C. & Panicucci, A. 1994. Surface Ozone in Pisa (Italy): A Six Year Study. *Journal of Atmospheric Environment* 28:3155-3164.
- [17] Bencala, K.E. & Seinfeld, J.H. 1979. On Frequency distribution of Air Pollutant Concentrations. *Journal of Atmospheric Environment* 10: 941-950
- [18] Gouveia, N. & Fletcher, T. 2000. Time Series Analysis of Air Pollution and Mortality : Effects by Cause, Age and Socioeconomic Status. *Journal of Epidemiology and Community Health* 54: 750-755
- [19] Hertzberg, A.M. & Frew, L. 2003. Can Public Policy be Influenced? *Environmetrics* 14 (1): 1- 10
- [20] Lee, C.K. 2002. Multiracial Characteristics in Air Pollutant Concentration Time Series. *Journal of Water Air Soil Pollution* 135: 389-409
- [21] Liu, C.M., Hung, C.Y., Shieh, S.L., & Wu, C.C. 1994. Important Meteorological Parameters For Ozone Episodes Experienced In The Taipei Basin. *Journal of Atmospheric Environment* 28: 159-173
- [22] Roberts, S. 2003. Combining data From Multiple Monitors in Air Pollution Mortality Time Series Studies. *Journal of Atmospheric Environment* 35: 2339-2348.
- [23] Touloumi, G., Atkinson, R. & Terte, A.L. 2004. Analysis of Health Outcome Time Series Data in Epidemiological Studies. *Journal of Environmetrics* 15: 101-117

- [24] Yee, E. & Chen, R. 1997. A Simple Model for Probability Density Functions of Concentration Fluctuations in Atmospheric Plumes. *Journal of Atmospheric Environment* 31: 991-1002

INTECH

INTECH