Journal of Sustainability Science and Management
Volume 6 *Number* 1, June 2011: 79-88

ISSN: 1823-8556

TIME-SERIES ANALYSIS OF GROUND-LEVEL OZONE IN MUDA IRRIGATION SCHEME AREA (MADA), KEDAH.

MARZUKI ISMAIL

[1]*Department of Engineering Science, Faculty of Science and Technology, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Malaysia.*

**Corresponding author: marzuki@umt.edu.my*

Abstract: Time-series analysis and forecasting has become a major tool in many applications in air pollution and environmental management fields. The forecast of air pollution can be used as warning to the public and for decision makers to establish strategies and measures in cases of high air-pollution levels. It has long been recognised that pollutant gases cause significant impacts on crops and studies conducted revealed that surface ozone is responsible for most of the crop-yield losses from air pollutants. Hence, this study aims to fit and exemplify time-series analysis in forecasting ozone concentrations in Sungai Petani, a town situated in the Muda Irrigation Scheme Area (MADA), Kedah. In this study, Box-Jenkins methodology was used to build an Autoregressive Integrated Moving Average (ARIMA) model for monthly ozone data from 1999-2007 with a total of 108 readings. Parametric seasonally-adjusted ARIMA $(1,0,1)x(2,1,2)^{12}$ with constant model was successfully applied to predict the long-term trend of ozone concentration. The detection of a steady statistically-significant upward trend for ozone concentration in Sungai Petani is quite alarming because MADA produces 40% of the total rice production in Malaysia. This is likely due to sources of ozone precursors related to industrial activities from nearby areas and the increase in road-traffic volume.

KEYWORDS: *Surface ozone, time-series analysis, ARIMA, seasonal variation, MADA area*

Introduction

It has long been recognised that pollutant gases cause significant impacts on crops and studies conducted revealed that ground-level ozone is responsible for most of the crop-yield losses from air pollutants whereby losses from other pollutants are minimal, relative to ozone (Fuhrer *et al*., 1997; Pleijel *et al*., 1998; Ishii *et al*, 2004). Ozone is a secondary pollutant resulting from photochemical reaction of a variety of natural and anthropogenic precursors (mainly volatile organic compounds (VOCs) and oxides of nitrogen (NOx)). Despite massive and costly control efforts, countries in Europe and North America still experience severe ozone problems (Wu and Chan, 2001). People in Asia cannot escape from ozone pollution. In some large Asian cities, elevated ozone levels have been reported (Chan and Chan, 2000). Nevertheless, the long-term ozone trend, especially in Malaysia, is relatively less researched.

Development and use of statistical and other quantitative methods in the environmental sciences have been a major communication between environmental scientists and statisticians (Hertzberg and Frew, 2003). In recent years, many statistical analyses have been used to study air pollution as a common problem in urban areas (Lee, 2002). Many investigators have used probability models to explain temporal distribution of air pollutants (Bencala and Seinfield, 1979; Yee and Chen, 1997). Time-series analysis is a useful tool for better understanding of cause and effect relationships in environmental pollution (Kyriakidis and Journal, 2001). 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 behaviour of air pollution through time using different techniques (Kocak *et al,* 2000; Hies *et al,* 2003). Many others have tried to relate air pollution to human health through time-series analysis

Received: 23 Jun 2010 / Accepted: 02 February 2011

(Roberts, 2003; Touloumi *et al,* 2004). Therefore, this study aims at extending time-series analysis to give both qualitative and quantitative information about ozone concentrations in Sungai Petani and to predict future concentrations of this pollutant.

Materials and Methods

Study Area

This study was conducted in Sungai Petani (6° 11.8’ N, 100° 4.5’ E), a developing Malaysian town located in the Muda Irrigation Scheme Area (MADA), Kedah, North-west of Peninsular Malaysia (Figure 1). MADA is considered the “rice bowl” of Malaysia due to the fact that 40% of the total rice production in Malaysia comes from this area (MADA, 2010). In this area, there are dominant sources of ozone precursors related to industrial activities and road traffic.

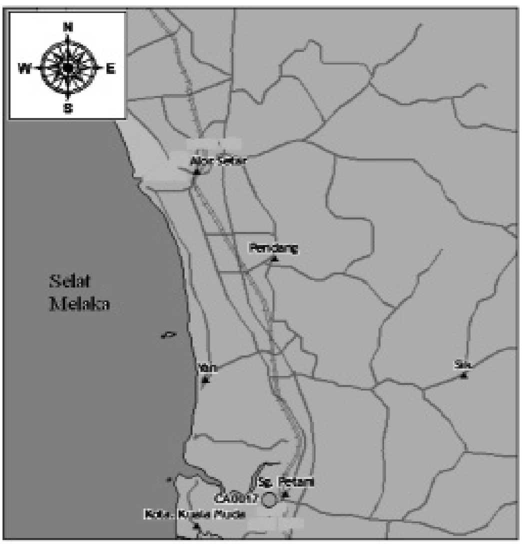

Figure 1. Location of the Study Area

Data and Monitoring Network

The monitoring network was installed, operated and maintained by Alam Sekitar Malaysia Sdn. Bhd. (ASMA) under concession by the Department of Environment Malaysia (Afroz et al, 2003). 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. In this study, the ozone trend was examined using ozone data consisting of 108 monthly observations from January 1999 to December 2007 acquired from the Air Quality Division of ASMA for Sekolah Menengah Kebangsaan Tun Ismail, Bakar Arang station (the earliest operational station in Kedah state) located in the Sungai Petani district. The time-series plot is shown in Figure 2.

Time-series Analysis

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.

Seasonal Model

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,

$$O_t = T_t \, x \, S_t \, x \, I_t \qquad \text{Equation 1}$$

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 which is called the error component. As the seasonality increases 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.

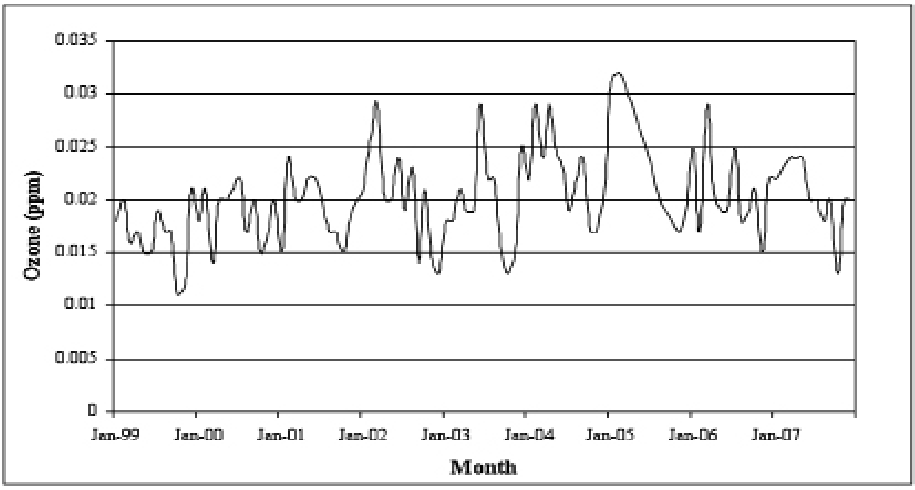

Figure 2. Original monthly ozone concentration for Sungai Petani (1999-2007)

Trend Analysis Model

In trend analysis, Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model was applied to model the time-series behaviour 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 the series stationary, as well as specifying the order of the regular and seasonal autoregressive 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 autoregressive 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 models are summarised 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 data. 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 $(1,1,1)(1,1,1)^{12}$ 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:

$$(1-f_1B)(1-a_1B^{12})(1-B)(1-B^{12})Z_t = (1-q_1B)(1-g_1B^{12})e_t$$
Equation 2

Where:

$1-f_1B$ = non-seasonal autoregressive of order 1
$1-a_1B^{12}$ = 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
$1-q_1B$ = non-seasonal moving average of order 1
$1-g_1B^{12}$ = seasonal moving average of order 1

For the seasonal model, Akaike Information Criterion (AIC) was used 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 (Hong, 1997).

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, a nonlinear estimation method is used to estimate the above identified parameters to maximise the likelihood (probability) of the observed series given the parameter values (Naill and Momani, 2009).

In diagnose checking step (step 3), the residuals from the fitted model are 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 c^2. If the residuals are correlated, then the model should be refined as in step one above. Otherwise, the residuals are white noise and the model is adequate to represent our time series.

The final stage for the modelling 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 (MSE) to compare fits of different ARIMA models. A lower MSE value corresponds to a better-fitting model.

Results and Discussion

Seasonality of Ozone

In seasonality of ozone, a well-defined annual cycle was consistent with the highest ozone means occurring in July, and the lowest ozone means in December (Figure 3). Table 1 shows the seasonal indices for each month, scaled so that an average month equals 100. The indices range from a low of 0 in December to a high of 220 in July. The seasonal variation pattern in Sg. Petani differed from other countries, such as the United States, the United Kingdom, Italy, Canada, and Japan, in that the peak ozone concentration did not correspond to maximum photochemical activity in summer (Lorenzini *et al,* 1994).

Trend of ozone

For the purpose of forecasting the trend in this study, 108 observations of monthly (January 1999 to December 2007) ozone concentrations were used to fit the ARIMA models. These data have 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 (-11.4628) of the Akaike Information Criterion (AIC) is

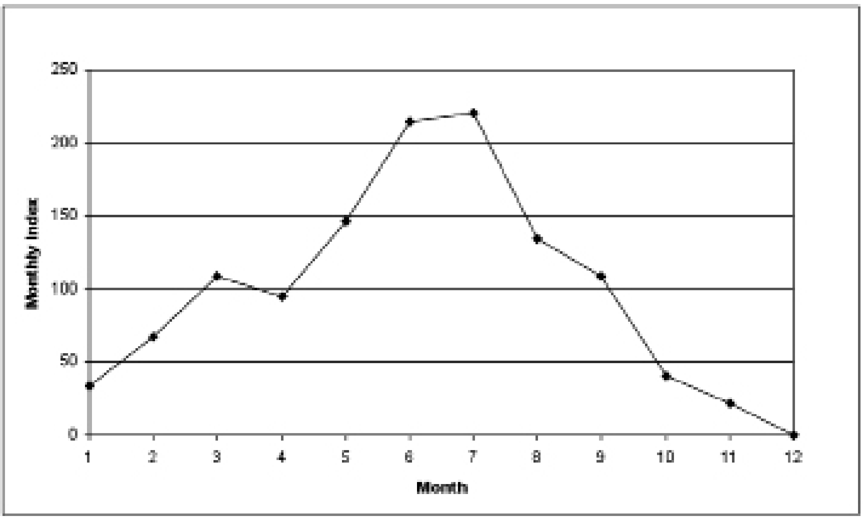

Figure 3. Annual variation of monthly ozone means.

Table 1. Seasonal Index of Ozone

Season	Index
January	33.16
February	67.09
March	108.76
April	94.48
May	145.92
Jun	214.87
July	220.02
August	134.05
September	108.97
October	40.31
November	22.19
December	0

(ARIMA) (1, 0, 1) x $(2, 1, 2)^{12}$ with constant was selected and has been used to generate the forecasts (as shown in Table 2 and 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 3, The P-value for the AR (1) term, MA (1) term, SAR (1) term, SAR (2) 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 0.00311745. Since no tests are statistically significant at the 5% significance level, the current model is adequate to represent the data and could be used to forecast the upcoming ozone concentration.

Diagnostic Checking

Residuals from the model were tested using Ljung-Box test for white noise as shown in Table 4. A model is good if the residuals don't show any correlation or pattern, that is, if they are just plain white noise. In this test, the P-value for up to lag 12 and lag 24 is 0.279 and 0.779, respectively. So the hypothesis that the autocorrelations of the residuals are 0 cannot be rejected. This means that the residuals are not correlated, indicating that model is appropriate, i.e., fits the series well, and there is no autocorrelation left to explain. What is left after fitting the model is just noise. The plots of residual ACF (Figure 5) and PACF (Figure 6) also indicate that the residuals are white noise and not autocorrelated, confirming the Ljung-Box test result. Furthermore, as shown in Figure 7 of normal probability plot, residuals of the model are normal.

Based on the prediction for ozone concentration (Table 2 and Figure 4), there is a statistically-significant upward trend at Sg. Petani station. The detection of a steady statistically-significant upward trend for ozone concentration in Sg. Petani is quite alarming. This is likely due

Table 2. Model predicted plot of ozone concentration.

Period	*Forecast*	*Lower 95.0% Limit*	*Upper 95.0% Limit*
Jan-08	0.0181072	0.0119129	0.0243015
Feb-08	0.0277397	0.0213733	0.0341062
Mar-08	0.0210273	0.0145651	0.0274895
Apr-08	0.0234402	0.0169241	0.0299562
May-08	0.0246501	0.0181037	0.0311965
Jun-08	0.0279449	0.0213812	0.0345085
Jul-08	0.0223731	0.0157997	0.0289465
Aug-08	0.0243139	0.0177349	0.0308928
Sep-08	0.0208771	0.014295	0.0274592
Oct-08	0.0172932	0.0107093	0.0238771
Nov-08	0.0208955	0.0143106	0.0274804
Dec-08	0.0233333	0.0167478	0.0299188
Jan-09	0.0224769	0.0158884	0.0290653
Feb-09	0.0303359	0.0237475	0.0369244
Mar-09	0.0254633	0.0188748	0.0320518
Apr-09	0.0256559	0.0190674	0.0322444
May-09	0.0254939	0.0189054	0.0320824
Jun-09	0.0310437	0.0244552	0.0376322
Jul-09	0.0253392	0.0187506	0.0319277
Aug-09	0.0267862	0.0201977	0.0333748
Sep-09	0.0218227	0.0152342	0.0284112
Oct-09	0.0218779	0.0152894	0.0284664
Nov-09	0.0198739	0.0132854	0.0264625
Dec-09	0.0250352	0.0184467	0.0316237
Jan-10	0.0275905	0.020117	0.035064
Feb-10	0.0290892	0.0215689	0.0366094
Mar-10	0.0303419	0.0227952	0.0378887
Apr-10	0.0276994	0.0201376	0.0352611
May-10	0.0259743	0.0184041	0.0335446
Jun-10	0.0284686	0.0208935	0.0360437
Jul-10	0.0262005	0.0186226	0.0337783
Aug-10	0.0249004	0.017321	0.0324798
Sep-10	0.0223772	0.0147969	0.0299575
Oct-10	0.0226503	0.0150695	0.0302311
Nov-10	0.0191862	0.0116051	0.0267673
Dec-10	0.0245815	0.0170003	0.0321628
Jan-11	0.0286854	0.0209323	0.0364385
Feb-11	0.0272179	0.0194543	0.0349816
Mar-11	0.0312546	0.0234849	0.0390242
Apr-11	0.0281838	0.0204108	0.0359569

May-11	0.0263156	0.0185406	0.0340906
Jun-11	0.0254432	0.0176671	0.0332193
Jul-11	0.0254636	0.0176869	0.0332403
Aug-11	0.022711	0.0149339	0.0304881
Sep-11	0.0227074	0.0149301	0.0304847
Oct-11	0.0208477	0.0130703	0.0286251
Nov-11	0.0199495	0.0121721	0.027727
Dec-11	0.0237941	0.0160166	0.0315716
Jan-12	0.0267603	0.0188035	0.034717
Feb-12	0.0273637	0.0193974	0.0353299
Mar-12	0.0293598	0.0213881	0.0373315
Apr-12	0.0277854	0.0198107	0.0357602
May-12	0.0268018	0.0188253	0.0347783
Jun-12	0.0256788	0.0177014	0.0336563
Jul-12	0.0250205	0.0170425	0.0329986
Aug-12	0.0230522	0.0150739	0.0310306
Sep-12	0.0231405	0.0151619	0.031119
Oct-12	0.0197767	0.0117981	0.0277554
Nov-12	0.0214364	0.0134576	0.0294151
Dec-12	0.0241401	0.0161614	0.0321189
Jan-13	0.0254093	0.017015	0.0338036
Feb-13	0.0291937	0.0207759	0.0376114
Mar-13	0.0281421	0.019711	0.0365731
Apr-13	0.027795	0.0193564	0.0362336
May-13	0.0274682	0.0190253	0.0359111
Jun-13	0.0282874	0.0198421	0.0367328
Jul-13	0.0257643	0.0173176	0.034211
Aug-13	0.0252065	0.0167589	0.033654
Sep-13	0.0237797	0.0153317	0.0322276
Oct-13	0.0207835	0.0123352	0.0292317
Nov-13	0.0224809	0.0140325	0.0309293
Dec-13	0.0254216	0.0169732	0.03387
Jan-14	0.026313	0.0177878	0.0348381
Feb-14	0.0308441	0.0223142	0.039374
Mar-14	0.0291123	0.0205797	0.0376449
Apr-14	0.0286176	0.0200835	0.0371518
May-14	0.0281596	0.0196246	0.0366947
Jun-14	0.0303811	0.0218456	0.0389167
Jul-14	0.0270858	0.01855	0.0356217
Aug-14	0.0269427	0.0184067	0.0354787
Sep-14	0.0244868	0.0159507	0.0330229
Oct-14	0.0226659	0.0141297	0.031202
Nov-14	0.0227898	0.0142537	0.031326
Dec-14	0.0265951	0.0180589	0.0351313

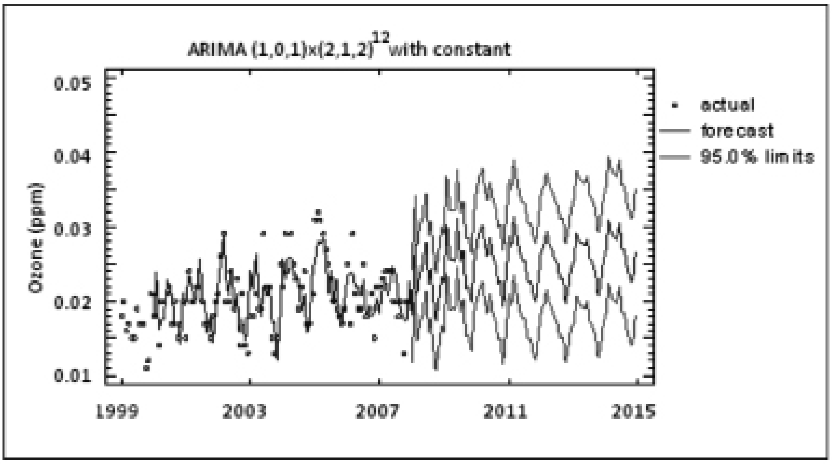

Figure 4. Model predicted plot of ozone concentration with actual and 95% confidence band.

Table 3. ARIMA Model Summary.

Parameter	*Estimate*	*Stnd. Error*	*t*	*P-value*
AR(1)	0.753937	0.19179	3.93105	0.000167
MA(1)	0.516572	0.246432	2.09621	0.038902
SAR(1)	0.6414	0.0623069	10.2942	0.000000
SAR(2)	-0.62465	0.0504516	-12.3812	0.000000
SMA(1)	1.68401	0.0473854	35.5387	0.000000
SMA(2)	-0.766594	0.041257	-18.581	0.000000
Mean	0.000577368	0.000173008	3.33724	0.001236
Constant	0.000139689			

Estimated white noise variance = 0.00000971852 with 89 degrees of freedom
Estimated white noise standard deviation = 0.00311745
Number of iterations: 16

Table 4. Modified Box-Pierce (Ljung-Box) Chi-Square statistic.

Lag	12	24	36	48
Chi-Square	6.3	12.3	22.8	31.4
DF	5	17	29	41
P-Value	0.279	0.779	0.785	0.861

to sources of ozone precursors related to industrial activities from nearby areas and the increase in road-traffic volume.

Conclusion

Time-series analysis is an important tool in modelling and forecasting air pollutants. Although, this piece of information was not appropriate to predict the exact monthly ozone concentration, ARIMA (1,0,1)x $(2,1,2)^{12}$ with constant model give us information that can help the decision makers establish strategies, priorities and proper use of fossil-fuel resources in Sungai Petani. This is very important because ground-level ozone (O_3) is formed from NO_x and VOCs brought about by human activities (largely the combustion of fossil fuel). In summary, concentrations of ground-level ozone in Sungai Petani have been rising steadily, and, as ozone is harmful to vegetation, one direct effect of increasing ozone is expected to be a reduction in yield. Therefore, effective measures for NO_x and VOCs emissions reduction must be introduced immediately in MADA area to ensure that no substantial losses of rice yield occur in the area.

Acknowledgement

The researcher would like to thank DOE Malaysia for providing pollutants data from 1996-2007 to accomplish this study.

References

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.

Bencala, K. E. & Seinfield, J. H. (1979). On Frequency distribution of Air Pollutant Concentrations. *Journal of Atmospheric Environment* 10: 941-950.

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.

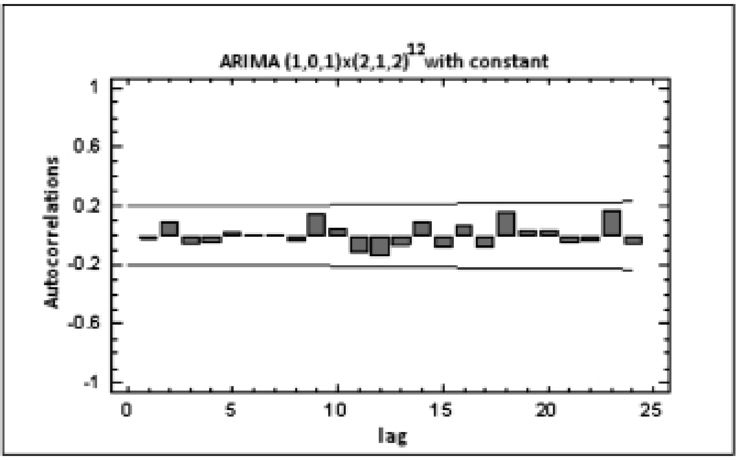

Figure 5. Residual autocorrelation functions (ACF) plot.

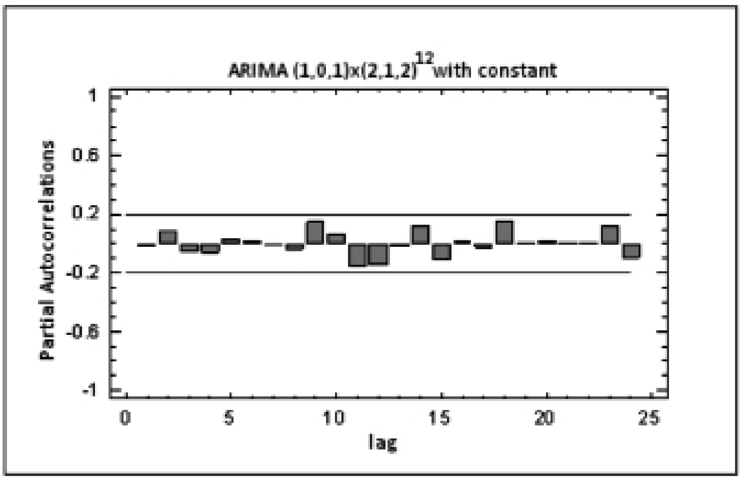

Figure 6. Residual partial autocorrelation (PACF) functions plot.

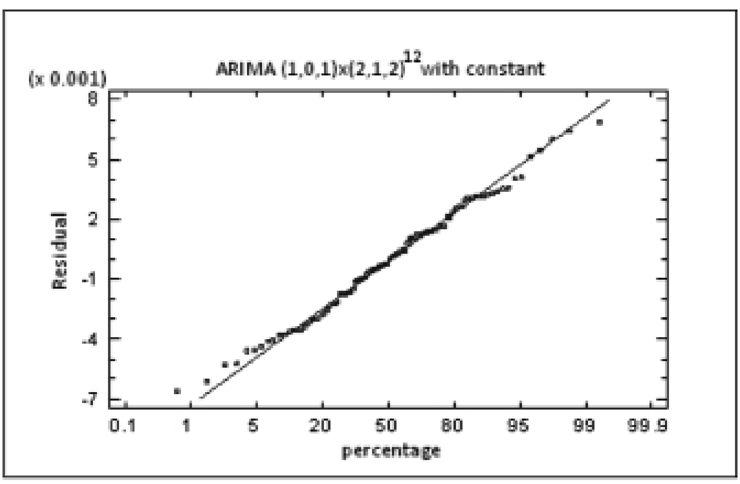

Figure 7. Residual normal probability plot.

MADA. (2010). How much MADA contributes in the total Domestic rice production? Retrieved June 9, 2010 from http://www.mada.gov.my/web/guest/soalanlazim.

J. Fuhrer, L. Skärby and M. R. Ashmore. (1997). Critical levels for ozone effects on vegetation in Europe. *Environmental Pollution* 97 (1–2): 91–106.

Hertzberg, A. M. & Frew, L. (2003). Can Public Policy be Influenced? *Environmetrics* 14 (1): 1- 10.

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.

Hong, W. (1997). A Time Series Analysis of United States Carrots Exports to Canada. Msc. Thesis, North Dakota State University.

Ishii, S., Marshall, F. M., Bell, J. N. B., & Abdullah, A. M. (2004). Impact of ambient air pollution on locally grown rice cultivars (*Oryza sativa* L.) in Malaysia. *Water, Air, and Soil Pollution, 154*, 187–201.

Kocak, K., Saylan, L. & Sen, O. (2000). Nonlinear Time Series Prediction of O_3 Concentration in Istanbul. *Journal of Atmospheric Environment* 34: 1267-1271.

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.

Lee, C. K. (2002). Multiracial Characteristics in Air Pollutant Concentration Time Series. *Journal of Water Air Soil Pollution* 135: 389-409.

Lorenzini, G., Nali, C. & Panicucci, A. (1994). Surface Ozone in Pisa (Italy): A Six Year Study. *Journal of Atmospheric Environment* 28:3155-3164.

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.

Pleijel, H., Danielsson, H., Gelang, J., Sild, E., Selldén, G. (1998). Growth stage dependence of the grain yield response to ozone in spring wheat (*Triticum aestivum* L.). *Agr. Ecosyst. Environ.* 70: 61–68.

Roberts, S. (2003). Combining data From Multiple Monitors in Air Pollution Mortality Time Series Studies. *Journal of Atmospheric Environment* 35: 2339-2348.

Touloumi, G., Atkinson, R. & Terte, A. L. (2004). Analysis of Health Outcome Time Series Data in Epidemiological Studies. *Journal of Environmetrics* 15: 101-117.

Wu, H. W. Y. & Chan, L. Y. (2001). Surface Ozone Trends in Hong Kong in 1985-1995. *Journal Of Environmental International* 26: 213-222.

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.