American Journal of Environmental Sciences

Original Research Paper

Forecasting Ozone Concentrations Using Box-Jenkins ARIMA Modeling in Malaysia

[1]**Wan Rozita Wan Mahiyuddin**, [2]**Nur Izzah Jamil**, [1]**Zamtira Seman**,
[1]**Nurul Izzah Ahmad**, [1]**Nor Aini Abdullah**, [3]**Mohd Talib Latif** and [4]**Mazrura Sahani**

[1]*Institute for Medical Research, Jalan Pahang, 50588 Kuala Lumpur, Malaysia*
[2]*School of Statistics, Faculty of Science Computer and Mathematics,*
University of Technology Mara, Negeri Sembilan, Malaysia
[3]*School of Environmental and Natural Resource Sciences,*
Faculty of Science Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
[4]*Faculty of Health Sciences, Universiti Kebangsaan Malaysia,*
Jalan Raja Muda Abdul Aziz, 50300 Kuala Lumpur, Malaysia

Article history
Received: 02-03-2018
Revised: 06-04-2018
Accepted: 17-05-2018

Corresponding Author:
Wan Rozita Wan Mahiyuddin
Institute for Medical Research,
Jalan Pahang, 50588 Kuala
Lumpur, Malaysia
Tel: +603-26162666
Fax: +603-26939335
Email: rozita@imr.gov.my

Abstract: Time series analysis and forecasting has become a major tool in many applications in air pollution and environmental management fields. Among the most effective approaches for analyzing time series data is the model introduced by Box and Jenkins. In this study, we used Box-Jenkins methodology to build Autoregressive Integrated Moving Average (ARIMA) model on the average of monthly ozone data taken from three monitoring stations in Klang Valley for the period 2000 to 2010 with a total of 132 readings. Result shows that ARIMA $(1,0,0)(0,1,1)_{12}$ model was successfully applied to predict the long term trend of ozone concentrations in Klang Valley. The model performance has been evaluated on the basis of certain commonly used statistical measures. The overall model performance is found to be quite satisfactory as indicated by the values of Root Mean Squared Error, Mean Absolute Percentage Error and Normalized Bayesian Information Criteria. The finding of a statistically significant upward trend of future ozone concentrations is a concern for human health in Klang Valley since over the last decade, ozone appears as one of the main pollutant of concern in Malaysia.

Keywords: Time Series, ARIMA, Box-Jenkins, Ozone Concentration, Klang Valley

Introduction

In recent decades, air pollution has been recognized as one of the major environmental problems faced by most of the countries around the world. Air quality in Malaysia is governed by the established Malaysian Ambient Air Quality Guidelines (MAQG) of 1989 issued by Department of Environment Malaysia (DOE) using Malaysian Air Pollution Index (API). The API is an index system for classifying and reporting ambient air quality in Malaysia which are measured continuously through 52 air quality monitoring stations throughout the country. API for a given time period is calculated based on the subindex values of Sulfur Dioxide (SO_2), Nitrogen Dioxide (NO_2), Ozone (O_3), Carbon Monoxide (CO) and particulate matter less than 10 $\mu g/m^3$ (PM_{10}).

SO_2, NO_2, CO and PM_{10} are examples of primary pollutant, for which they are pollutants that emitted directly from sources. PM_{10} is inhalable material that is emitted directly from motor vehicles, power plants and other sources. It also can be formed in the atmosphere through reactions with gaseous emissions. CO is a gas emitted directly from motor vehicles and other combustion sources. SO_2 is the chemical compound produced by various industrial processes, electricity generation and fossil fuel combustion. Any material that has coal and petroleum element, either in the form of solid fuels, liquid fuels (such as gasoline, diesel and fuel oil) or natural gas contains sulfur compounds which generate SO_2 through combustion. NO_2 is one of the Nitrogen Oxides (NO_X), a group of pollutants produced from combustion processes. It is a reddish-brown gas with irritating odour produced from the combustion of

Science Publications

fossil fuels in transportation and industrial application such as waste incineration. Nitric Oxide (NO), which is emitted by motor vehicles or other combustion processes, combines with oxygen in the atmosphere producing NO_2. This gas also plays a major role in atmospheric reactions that produce O_3 (Salahudin *et al*., 2013).

O_3 is an example of secondary pollutant resulting from photochemical reaction of primary pollutants such as NOx, Organic Compounds (VOCs) and biogenic VOCs, together with certain meteorological conditions. (Ahamad *et al*., 2014; Ismail, 2011; Ismail *et al*., 2011). VOCs are emitted from various sources, including motor vehicles, chemical plants, refineries, factories, consumer and commercial products, while NOx are emitted from motor vehicles, power plants and other sources of combustion (DOE, 2006).

Various time series studies conducted world wide had shown the impact of surface O_3 on human health. Most studies reported that there were access risk on morbidity and mortality (Wan Rozita *et al*., 2013; Fischer *et al*., 2011; Nuntavarn *et al*., 2010). Surface O_3 is a crucial pollutant compared to others, because its ability to cause lung cell damage, inflammatory responses impairment of pulmonary host defenses and acute changes in lung function and chronic changes in lung cells (Folinsbee *et al*., 1992). Increased airway inflammation and deterioration in pulmonary function and gas exchange are among the health effects of O_3, as reported in laboratory studies (Mudway and Kelly, 2004; Brown *et al*., 2008).

In this past few years, higher surface O_3 levels have been reported in some Asian cities (Ismail *et al*., 2011). Recent studies in Malaysia showed the variations of surface O_3 in Klang Valley exceeded the Recommended Malaysian Air Quality Guideline (RMAQG) of 0.10ppm for the hourly level (Latif *et al*., 2012). Similar findings were reported by Ahamad and colleagues who conducted a study in 2014 which concluded that surface O_3 exceedance pattern in Klang Valley area is strongly influenced by local pollutant emission and dispersion characteristics. Geographically, Malaysia is located at the equatorial region. Thus, it poses greater risk of surface O_3 formation due to high levels of solar radiation which can promote the formation of photochemical pollutants.

Development and the usage of statistical techniques in estimating the concentrations of air pollution are generally been made with the help of predictive air pollution models. Gaussian approach is widely used to estimate ground level air pollution concentrations (Nieuwstadt, 1980; Varma *et al*., 2014). Besides Gaussian, empirical models also had been used to estimate the ambient concentration of the level of air pollution particularly CO in a road side environment (Nunez *et al*., 1999). In order to assist in air pollution management and to reduce the health impact, forecasting of future air pollution concentrations, particularly O_3 is crucial. Autoregressive Integrated Moving Average or known as ARIMA, is a stochastic approach, suggested by Box and Jenkins, also is widely used in forecasting air pollution concentrations (Kumar *et al*., 2004; Ismail, 2011; Ismail *et al*., 2011). This method was reviewed in detail by Milionis and Davis and was found to be successfully in the context of air pollution modeling.(Wang and Guo, 2009). Therefore, in this study, we aimed to find the best forecasting model to forecast upcoming monthly surface O_3 concentration in Klang Valley using ARIMA.

Methodology

Study Area

This study was conducted in Klang Valley, which is the most industrialised and economically the fastest growing area in Malaysia. The Klang Valley, situated in the middle of the west coast of Peninsular Malaysia and has an area of about 2,832 km^2 which include Kuala Lumpur Federal Territory and its suburbs and adjoining cities and towns in the state of Selangor. Klang Valley is also known for large-scale industrial, commercial activities, densely populated areas and high volume of vehicular traffic. Based on Department of Statistics report in year 2006, the population in this area had expanded to 4.7 million (DOS, 2006). In addition and due to its characteristic as a valley, the prevailing winds in Klang Valley are generally weak resulting in stable atmospheric conditions which cause pollutants in the air to stagnate (DOE, 2006).

The exact locations of the monitoring stations, the descriptions and geographical map are given in Table 1 and Fig. 1.

The 8 h average level for O_3 was used because it was the average time recommended by the WHO for reflecting the most health-relevant exposure to O_3 (WHO, 2000). Due to that, to obtain a single value of daily pollutants readings that represented Klang Valley, the average of the O_3 concentrations were calculated daily and further monthly, across all the monitoring stations.

Study Period

We used the data series consist of 132 monthly concentrations of surface O_3 from January 2000 to December 2010, that had been recorded in Klang Valley, Malaysia from three monitoring sites located in Gombak, Petaling Jaya and Shah Alam. All the 132 observations were used in the estimation part. The data was obtained from Air Quality Division of the Department of the Environment, Malaysia, (DOE) through long-term monitoring by a private company, Alam Sekitar Sdn Bhd (ASMA).

Table 1: The exact locations and descriptions of monitoring stations in Klang Valley

Station Location	Category	Cordinates	Description
Gombak Water Service Department	Residential	N03°15.702', E101°39.103'; S3	Nearest station to Kuala Lumpur's city centre
Seri Petaling Primary School, Petaling Jaya	Industrial	N03°06.612', E101°42.274'; S1	Heavy traffic particularly during the morning rush hour
Taman Tun Dr Ismail Primary School (TTDI) Jaya, Shah Alam	Residential	N03°06.287', E101°33.368'; S2	Traffic density is lower compared with Petaling Jaya

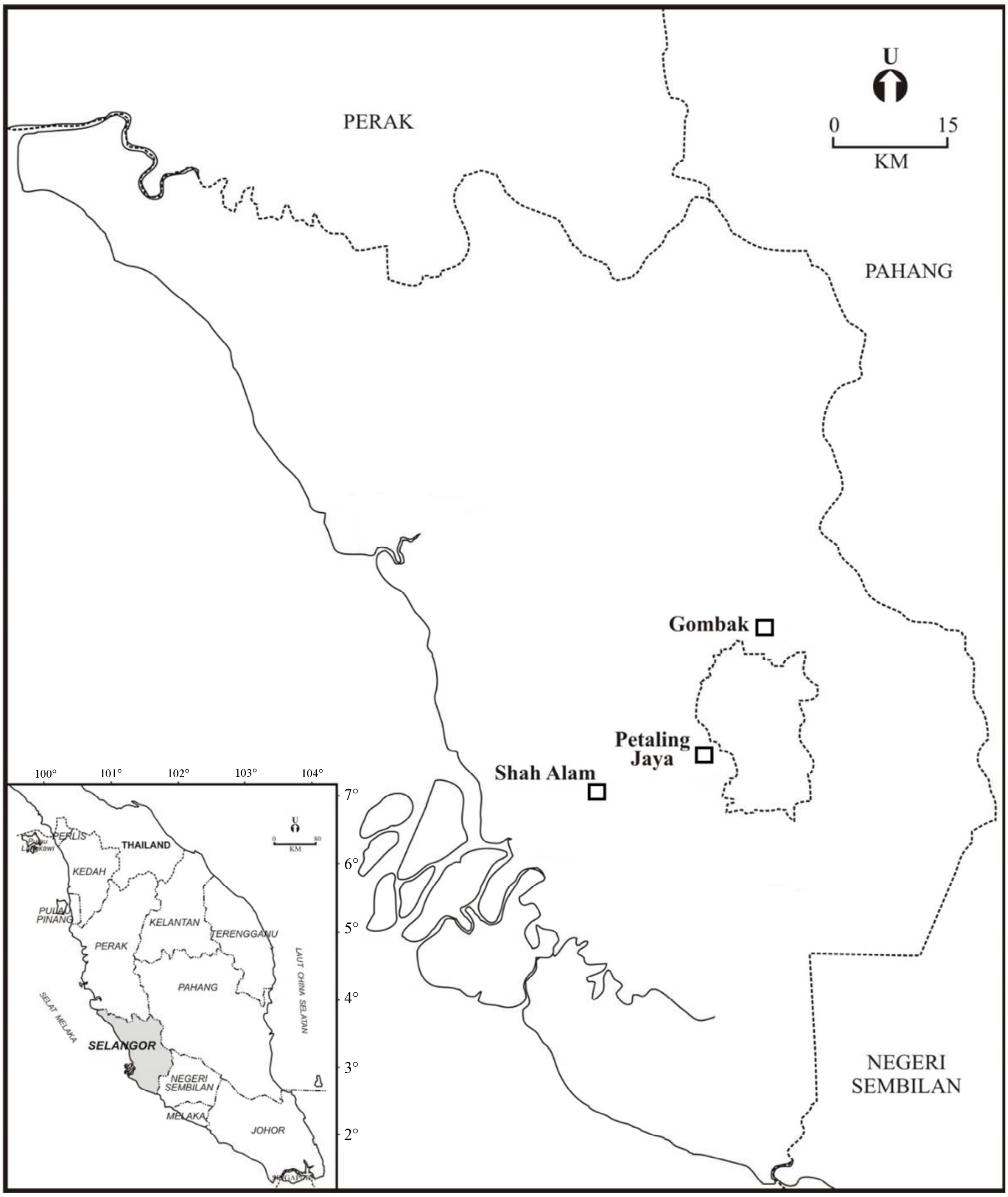


Fig. 1: A geographical map of the sampling stations located in Klang Valley, Malaysia

Time Series Analysis

The time series analysis of the data was carried out using statistical software, SPSS version 19 and Microsoft Excel 2010. The time series consists of a set of sequential numeric data at equal space intervals for a period of time.

This study applied two components of time series study. The first component was determination of seasonality through seasonal decomposition method and followed by determination of the best model by using Box -Jenkins Autoregressive Integrated Moving Average (ARIMA) model.

The seasonal decomposition was applied to decompose the seasonal variation in the series into a combination of Seasonal component (S_t), trend (T_t) , cycle component (C_t) and irregular (I_t) or short-term variation. These components are assumed to be related in a multiplicative manner or additive form, as shown below:

$$Y_t = T_t \times S_t \times I_t \quad \textit{multiplicative} \tag{1}$$

$$Y_t = T_t + S_t + I_t \quad \textit{additive} \tag{2}$$

where, Y_t is the original series of surface O_3. The T_t, has the same unit as Y_t but not S_t, C_t and I_t. As the underlying level of the series change, the magnitude of the seasonal variations also change. The S_t was the average deviation of each month's surface O_3 value from the overall average of surface O_3 level that was due to other components in that particular month.

In trend analysis, ARIMA approach was applied to determine the forecast trend. There are three main stages in building ARIMA model based on Box-Jenkins procedure. The first stage is model identification, second stage is model estimation and third stage is model application. These stages of building an ARIMA model are described in Fig. 2.

ARIMA is a general class of time series model that comprises of several techniques, which are differencing, Autoregressive models (AR) and Moving Average models (MA). An AR model of the order p is the one in which the current observation, x_t, is regressed on previous observations, x_{t-1}, x_{t-2},….x_{t-p} of the same time series. This is expressed by the equation below:

$$x_t = \xi + \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \ldots\ldots\ldots\ldots \Phi_p x_{t-p} + \varepsilon_t \tag{3}$$

where, the Φ_1, Φ_2,…..Φ_p are the regression coefficients.

ξ is the constant term and ε_t is random error.

Similarly to MA model of order q can be defined by:

$$x_t = \mu + e_t + \Theta_1 e_{t-1} - \Theta_2 e_{t-2} - \ldots\ldots\ldots\ldots \Theta_q e_{t-q} \tag{4}$$

Both models AR(p) and MA(q) can be combined to form ARMA (p,q) and can be written as:

$$\Phi_P(B) x_t = \xi + \theta_q(B) e_t \tag{5}$$

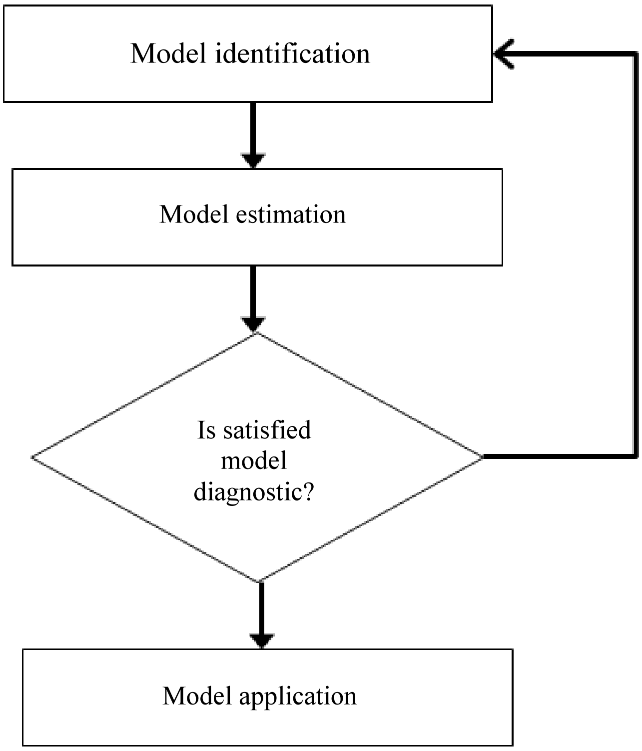


Fig. 2: Main stages of building an ARIMA model

This ARMA (p, q) model is for stationary time series, which means than it fluctuates randomly around some fixed values, either a constant or a mean. A series that does not follow this rule is called as ‘non-stationary time series’. Such series can be made stationary by taking successive differences of the data and the process is called as differencing. The process is done when the previous observation x_{t-1} is subtracted from the current observation $x_{t.}$ The number of how much differencing is needed in a non-stationary time series is denoted by the value ‘d’ in an ARIMA (p, d, q) model.

The first step in the application of the Box-Jenkins methodology is to identify the most appropriate class of ARIMA to be applied in the data series. Common statistics used to identify the model class is Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) coefficients. ARIMA models are classified as ARIMA (p, d, q) (P, D, $Q)_S$. Non-seasonal ARIMA (p, d, q) model contains three types of parameter; autoregressive parameters (p), number of differences (d) and moving average parameters (q). In addition, seasonal ARIMA (P, D, Q) or known as SARIMA, contains three types of parameter; seasonal Autoregressive (P), seasonal Differencing (D) and seasonal moving average (Q). Period or seasonality of data denotes as ‘s’.

The usual step of estimating the models and performing the necessary diagnostic testing procedures is to select the best model for forecasting. This is the crucial aspect of forecasting exercise since model that fits well may not necessary forecast well.

Ljung-Box test will be implemented in order to check for the presence of autocorrelation among the residuals by calculating the chi-square value of the error terms. Such test procedure is commonly known as portmanteau test. The model is adequate to represent our time series if the errors are white noise or uncorrelated. Graphical analysis such as plots of residual ACF, plots of residual PACF and normal probability plot should corroborate to the portmanteau test. The method with the smallest value of error measures (MSE, RMSE and MAPE) will be selected as the most appropriate ARIMA model.

Finally, goodness of fit of the model will be confirmed if it fulfills all the criteria. Therefore, the best model is now ready to be used to generate forecast values which give results of forecast values together with upper and lower limits with 95% confidence interval. Any forecast value within the range are considered acceptable. If the model fails to produce reliable forecast values or fails to explain the phenomena being investigated, then it needs to be revised and updated.

Results

Figure 3 shows the time series sequence plot of O_3 concentrations in the studied area by month from January 2000 – December 2010 in the Klang Valley, Malaysia. The monthly O_3 concentrations fluctuated throughout the months with the trend line y = 0.000025 * time +0.0157 that registered a very minimal upward trend over the study period. Therefore for further modelling strategy, we considered that the concentrations of monthly O_3 fluctuate around a constant mean.

The highest peak of the O_3 concentrations was recorded on February 2005 with the reading of 0.0269ppm. The other most highest concentrations were also recorded in the months of either February or March of the years, which were March 2002 (0.0248 ppm), March 2006 (0.0242 ppm), February 2010 (0.0243 ppm) and March 2010 (0.0257 ppm). February and March are dry months with minimum rainfall. The monthly average concentrations of O_3 from 2000-2010 was 0.017ppm.

Figure 4 showed the annual cycle of seasonal index of the surface O_3. Highest seasonal index of surface O_3 occured in February and followed by March and the lowest was in July. This was consistent with the findings in Fig. 3 that highest concentrations were observed in either February or March from 2000-2010. The seasonal index was at a minimum in November, December and January, at which it began to increase, reaching a peak in February and March, before declining again. The seasonal index ranged from the lowest of 93.15 in July to the highest of 107.75 in February, indicated that there was a seasonal swing from 93.15% of average to 107.75% of average in a complete cycle of a year.

The results of ACF and PACF from surface O_3 concentrations clearly indicated that the series were stationary but contained seasonal component (Fig. 3, 4). The ACF for surface O_3 concentrations exhibited seasonality where ACF showed a wave pattern passing through zero several times and there were some peak at every dozen (Fig. 5(a)). From Fig. 5(b), the first lag on PACF also showed a spike with significant value of 0.4879. There was also significant peak at lag 12 in PACF that confirmed the presence of an annual seasonal component in the series. Therefore, the series need to be adjusted by performing seasonal differences of order 1 to station the series.

Figure 6 showed the residuals were uncorrelated by time and the residuals fluctuated around a constant mean and zero variance after performing seasonal differencing at order 1.

To ensure that a well specified model is not missed, several models will be estimated and subsequently, the best model that satisfied statistical requirements will be chosen. Therefore, Portmanteau test is carried out in order to ensure the model identified is adequate enough to represent the monthly pattern of surface O_3 concentration.

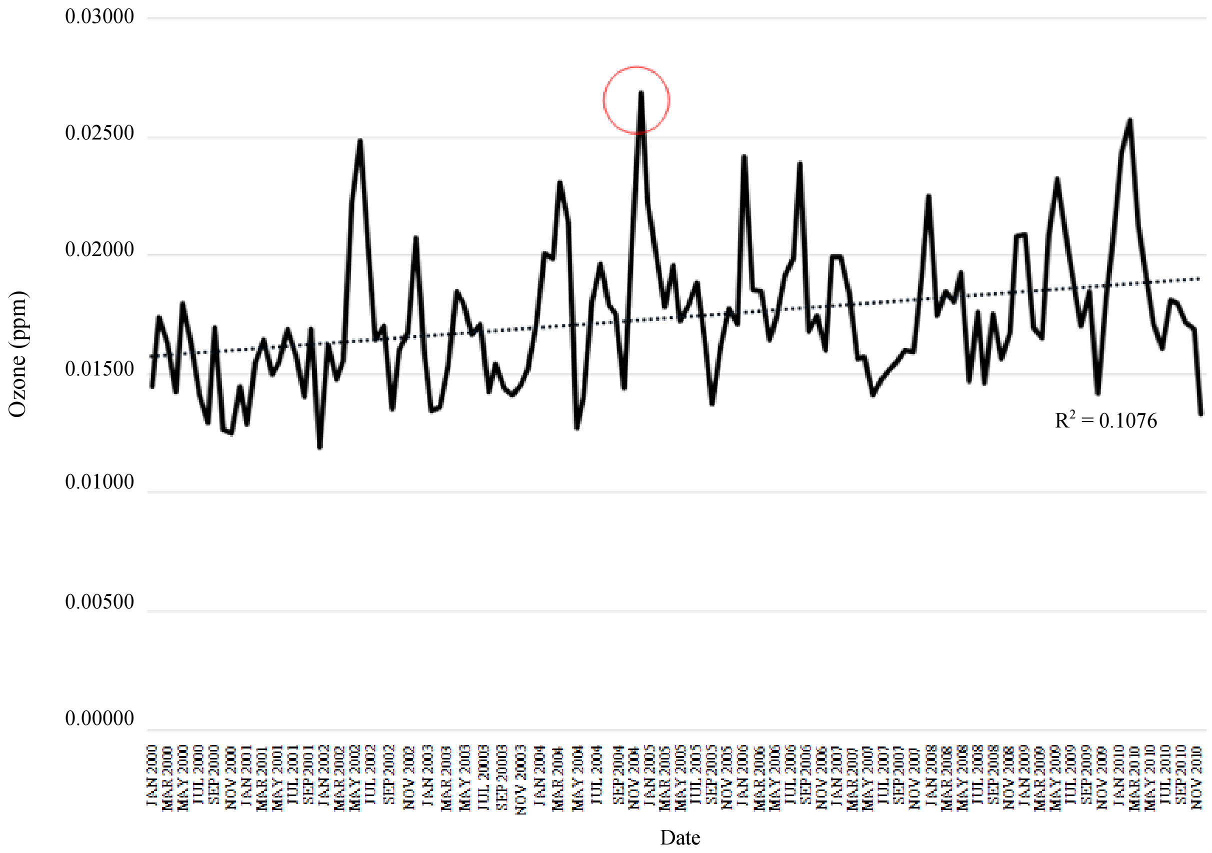


Fig. 3: Trend analysis plot of surface O_3 concentrations (Y_t)

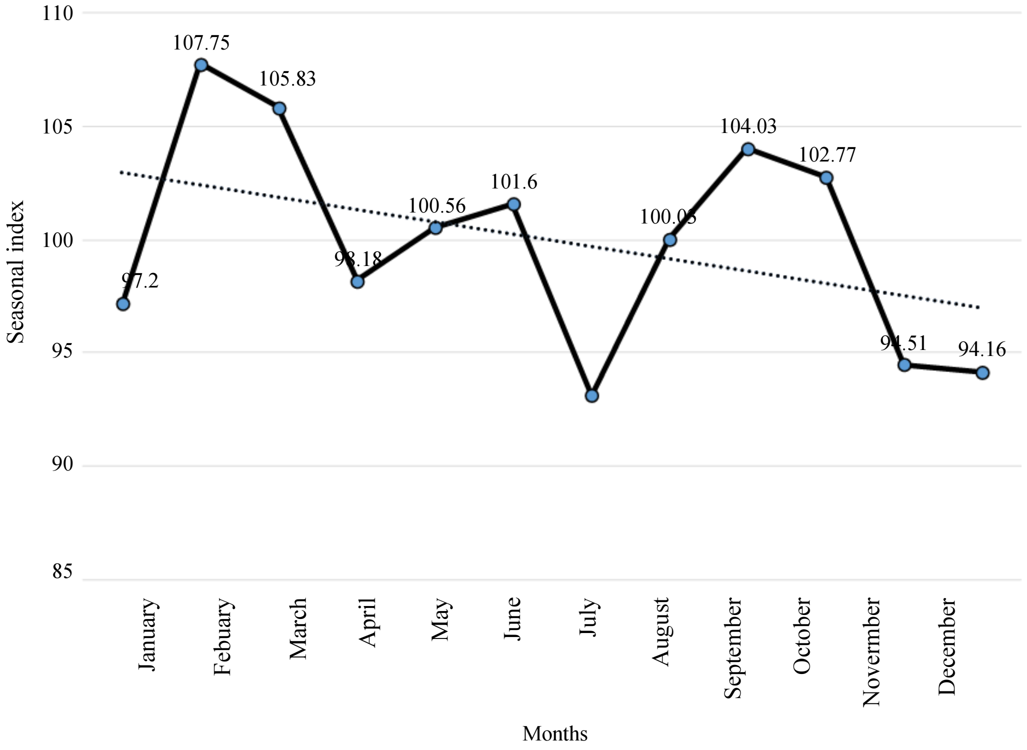
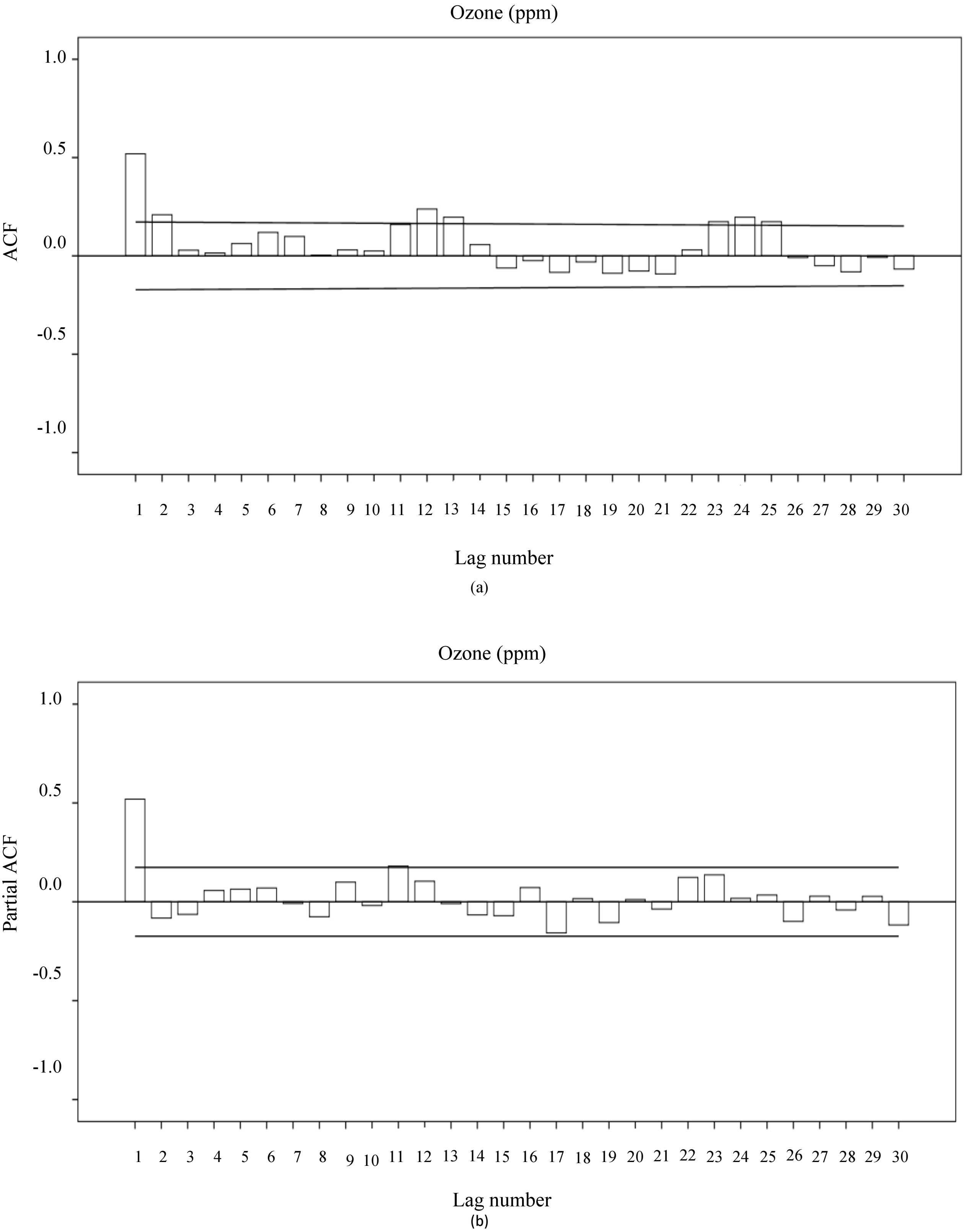


Fig. 4: Annual cycle of ozone seasonal index concentrations in Klang Valley



(a)

(b)

Fig. 5: (a) ACF for O_3 concentration (b) PACF for O_3 concentration

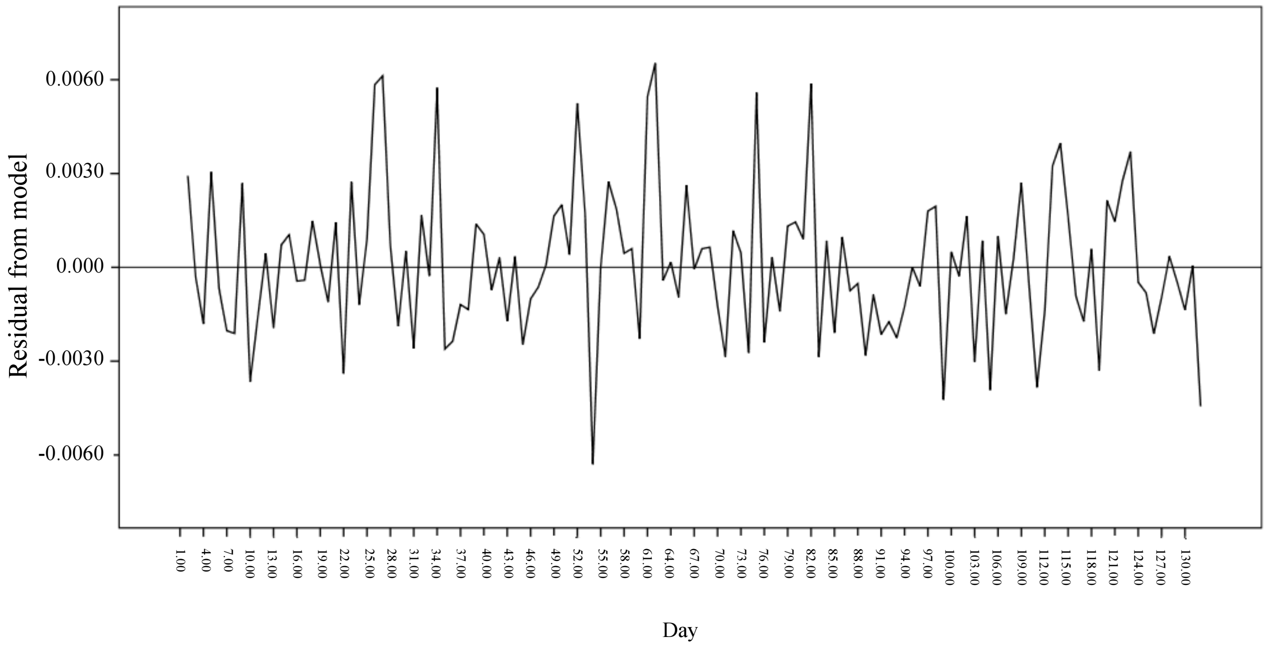


Fig. 6: Time plot of the residual (white noise) of the model after seasonal differencing

Table 2: Portmanteau for selected ARIMA model

	ARIMA $(1,0,0)(0,1,1)_{12}$	ARIMA $(1,0,1)(1,1,0)_{12}$	ARIMA $(1,0,1)(0,1,1)_{12}$	ARIMA $(0,0,1)(0,1,1)_{12}$
Ljung-Box *p-value	0.552	0.063	0.467	0.327
Conclusion	#	#	#	#
RMSE	0.002	0.003	0.002	0.003
MAPE	10.569	12.339	10.578	10.654
NBIC	-11.878	-11.560	-11.829	-11.860

* at 5% significance level
#errors are white noise

Table 2 showed the results of p-value for Ljung Box statistics. All the models were insignicant ($p>0.05$) indicating that residuals appeared to be uncorrelated and the errors were white noise. The values of Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Normalized Bayesian Information Criteria (NBIC) for all the possible models in ARIMA were also shown in Table 2. The best model was the model with the lowest value of RMSE, MAPE and NBIC. ARIMA $(1,0,0)(0,1,1)_{12}$ recorded the lowest values for RMSE and MAPE but not for NBIC. Model forecasts were found to be reasonably close to the observed values of monthly mean of surface O_3 concentrations (MAPE~10%). However, based on the principle of parsimony model, ARIMA$(1,0,0)(0,1,1)_{12}$ was chosen as the best model of Box-Jenkins Methodology and will be used to generate forecasts till year 2020.

This model ARIMA $(1,0,0)(0,1,1)_{12}$ was selected to be the best forecast for future data of surface O_3. As shown in Table 3, the p value for all the coefficients for each parameter that form the ARIMA $(1,0,0)(0,1,1)_{12}$ were less than 0.05. Indicating that the associated parameters can be judged as significantly different from zero.

Therefore, by combining both seasonal and non-seasonal model, ARIMA $(1,0,0)(0,1,1)_{12}$ the best model for surface O_3 concentrations in Klang Valley can be written in the mathematical expressions as shown below:

$$y_t = 0.368y_{t-1} + y_{t-12} + 0.368y_{t-13} + e_t - 0.939e_{t-12} \tag{6}$$

Based on the prediction for O_3 concentration (Fig. 7), O_3 registered a general upward trend over the period of 2000-2010 in Klang Valley. The O_3 concentration increased steadily in Klang Valley until 2020.

Table 3: ARIMA (1,0,0)(0,1,1)$_{12}$ model parameters

Parameter	Coefficient	SE Coefficient	t	p-value
AR1	0.368	0.087	4.236	0.000
SMA$_{12}$	0.939	0.302	3.103	0.002

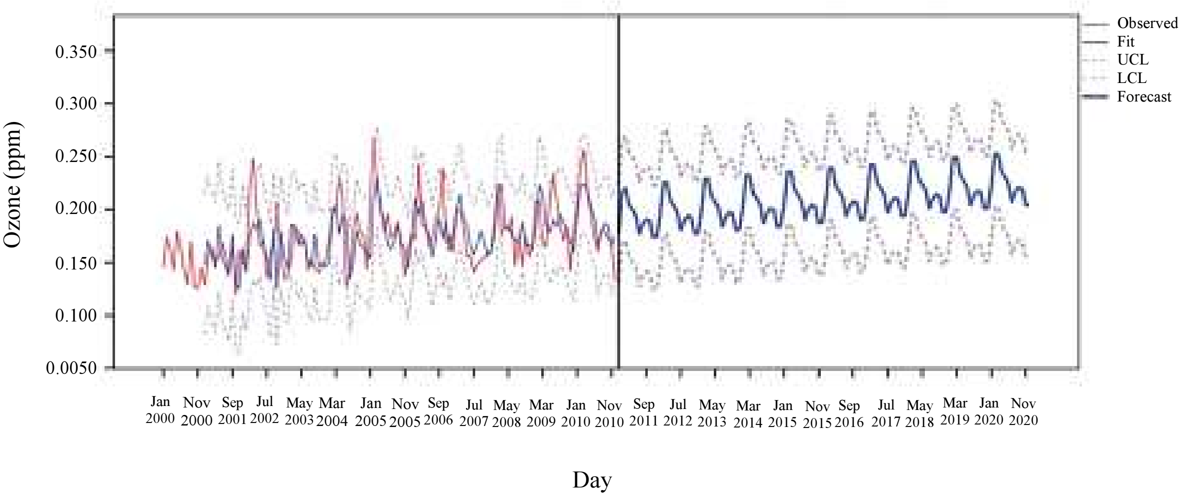


Fig. 7: Model predicted plot of O3 concentration with actual and 95% confidence interval

Discussion

The results showed that the ARIMA(1,0,0)(0,1,1)$_{12}$ model of O_3 contained non-seasonal and seasonal part. The non-seasonal part was identified as AR(1) with no differencing and the seasonal part was identified as MA(1) with seasonal differencing. The AR components capture the correlation between the current values of the time series and some of its past values. AR (1) means that the current observation of O_3 concentration is correlated with its immediate past values at time t = 1. The MOVING AVERAGE (MA) component represents the influence of a random (unexplained) shocks. MA (1) for seasonal part means that a shock on the value of the O_3 concentrations series at time t is correlated with the shock or error at time t = 12. The mixture of AR and MA model with seasonality exist in the O_3 series of this study, consistent with other findings in Malaysia (Ismail, 2011; Ismail *et al*., 2011).

Other results also showed that February and March were the months that recorded high concentrations of O_3 from 2000-2010. Both of the months included in the transition period of Northeast Monsoon (NEM) from November to January and the Southwest Monsoon (SWM) from May to August (Wong *et al*., 2009). In addition, during this transition period, an increased number of sunny hours were evident. Therefore, due to the positive relationship of O_3 with temperature and high level of solar radiation and inverse relationship with rainfall (Tan *et al*., 2014), it is expected that the concentrations of O_3 were observed to peak during these months.

The low level of average monthly O_3 in ambient air compared to the MAAQG for the study period, consistent with the findings by other research in few different cities in Peninsular Malaysia (Rahman *et al*., 2012; Awang *et al*., 2015; Banan *et al*., 2013). Although lower concentration of O_3 do not necessarily indicate cleaner air, few studies showed excedeences especially in the urban area and industrial zones (Awang *et al*., 2013). Furthermore, study by Wan Rozita *et al*. (2013) revealed associations between O_3 and daily natural mortality and between O_3 and daily respiratory mortality even though the concentrations of O_3 were far below the guidelines.

The fitted series of O_3 during the study period together with the prediction series based on the selected model, showed consistent increasing trend till year 2020. One of the main challenges for countries in tropical region is the high concentrations of O_3 caused by elevated levels of anthropogenic and natural O_3 precursors, particularly NO_x, the emissions from motor vehicles. The interaction of NO and O_3 lead to the formation of NO_2 which in turn contributes to the amount of O and O_3 in the atmosphere. A study by Banan *et al*. (2013) concluded that O_3 concentrations were higher particularly in sub-urban areas, as a result of down winds, compared to those in urban areas. Meteorological factors such as sunlight, the ambient temperature, cloud cover, water vapour concentrations, humidity and wind directions, influenced the variations of O_3 in the ambient air (Rahman *et al*., 2012).

The New Ambient Air Quality Standard was established by DOE in 2014 in order to replace the older Malaysia Ambient Air Quality Guideline that has been used since 1989. The New Ambient Air Quality Standard adopts 6 air pollutants criteria that include 5 existing air pollutants which are particulate matter with the size of less than 10 micron (PM_{10}), sulfur dioxide (SO_2), carbon monoxide (CO), nitrogen dioxide (NO_2) and ground level ozone (O_3) as well as 1 additional parameter which is particulate matter with the size of less than 2.5 micron ($PM_{2.5}$).

The air pollutants concentration limit will be strengthen in stages until 2020. There are 3 interim targets set which include interim target 1 (IT-1) in 2015, interim target 2 (IT-2) in 2018 and the full implementation of the standard in 2020. Since substantial literatures has been published on the increasing level of O_3 in Klang Valley, to minimize the effects of ground-level O_3 pollution from becoming too hazardous, an action plan for early information should be developed. The action plan should includes to identify the causes, assist policy makers in formulating policies and strategies to address the issue of O_3. It will also recommend measures of prevention, monitoring and compliance activities that contribute to O_3 pollution reduction.

These new guidelines and action plan are timely and consistent with the current situation of rapid urbanization in Malaysia in order to ensure sustainable development in the country.

Conclusion

An ARIMA model of the order $(1,0,0)(0,1,1)_{12}$ was found to fit the time series of monthly mean surface O_3 concentrations in Klang Valley from 2000 to 2010. The predictions estimates for the univariable model were found to be satisfactory. The model applied was not designed to forecast episodic circumstances but to recognize their behavior and assess its time-based progression. The study demonstrates that the ARIMA modeling approach is a useful tool for analysing non-stationary data, containing ordinary or seasonal series. It also has proven could be effectively used for obtaining short-term forecasts of air quality. Univariable forecasting using routinely-data collected is still not popularly used in the area of air pollution in Malaysia. A comparative study of the ARIMA modeling approach with other relevant alternatives of forecasting such as Classification and Regression Trees and Artificial Neural Network Models in the context of air quality modeling would be highly desirable.

An accurate methodology to forecast O_3 concentration is needed for the strategy planning and control of air pollution. Studies had shown that O_3 seriously endangers human health and environment at the level below the guideline. The effective management of the control and public warning strategies for O_3 concentration can be efficiently implemented by the accurate forecast of O_3 concentration.

We hope that the outcome of the study will be used to assist the development of the action plan of O_3 in Malaysia. This study will contribute to the formulation of pertinent public policies to address the root causes of air pollution. Further conclusive studies on health risk of air pollution particularly O_3 recommended to be conducted in order to provide evidence to guide the air quality management for health protection and sustainable city.

Acknowledgement

The authors would like to gratefully acknowledge the Director General of Health, Malaysia and the Director of the Institute for Medical Research (IMR), Ministry of Health Malaysia, for the permission to publish this paper. We also would like to thank Department of the Environment (DOE), for providing the data. This study received grant from National Institute of Health (JPP-IMR 13-014; NMRR-13-630-16667).

Author's Contributions

Wan Rozita Wan Mahiyuddin: Designed the research plan and organized the study, writing the manuscript, coordinating the data analysis and data management.

Nur Izzah Jamil: Managing the database, did the data analysis and writing the final model.

Zamtira Seman: Writing the final model and coordinated the mouse work.

Nurul Izzah Ahmad, Nor Aini Abdullah, Mohd Talib Latif and Mazrura Sahani: Contributed to the writing of the manuscript.

Ethics

This paper has been approved to be published by Malaysia Research Ethical Committee (MREC) under the above mentioned grant. We declare that there is no conflict of interest in this paper.

References

Ahamad, F., M.T. Latif, R. Tang, L. Juneng and D. Dominick *et al*., 2014. Variation of surface ozone exceedance around Klang Valley, Malaysia. Atmospheric Res., 139: 116-127. DOI: 10.1016/j.atmosres.2014.01.003

Awang, N.R., N.A. Ramli, A.S. Yahaya and M. Elbayoumi, 2015. High nighttime ground-level ozone concentrations in Kemaman: NO and NO_2 concnetraions attributions. Aerosol Air Qual. Res., 15: 1357-1366. DOI: 10.4209/aaqr.2015.01.0031

Awang, N.R., N.A. Ramli, N.I. Mohammed and A.S. Yahaya, 2013. Time series evaluation of ozone concentrations in Malaysia based on location of monitoring stations. Int. J. Eng. Technol., 3: 390-394.

Banan, N., M.T. Latif, L. Juneng and F. Ahamad, 2013. Characteristics of surface ozone concentrations at stations with different backgrounds in the Malaysian Peninsula. Aerosol Air Qual. Res., 13: 1090-1106. DOI: 10.4209/aaqr.2012.09.0259

Brown, J.S., T.F. Bateson and W.F. McDonnell, 2008. Effects of exposure to 0.06 ppm ozone on FEV1 in humans: A secondary analysis of existing data. Environ. Health Perspect., 11: 1023-1026. DOI: 10.1289/ehp.11396

DOE, 2006. Environment quality report, 2005. Department of Environment, Ministry of Natural Resources and Environment, Malaysia, Kuala Lumpur.

DOS, 2006. Basic population characteristics by administrative districts. Department of Statistics, Kuala Lumpur, Malaysia.

Fischer, P.H., M. Marra, N. Janssen and F.R. Cassee, 2011. Trends in relative risk estimates for the association between air pollution and mortality in the Netherlands, 1992-2006. Environ. Res., 111: 94-100. DOI: 10.1016/j.envres.2010.09.010

Folinsbee, L.J., M.J. Hazucha, E. Seal, H.R. Kehrl and D.H. Horstman, 1992. Cumulative ozone exposure dose explains responses to different exposure regimens. Am. Rev. Respor. Dis., 145: 195-195.

Ismail, M., 2011. Time series analysis of ground-level ozone in Muda Irrigation Scheme Area (MADA), Kedah. J Sustain. Sci. Manage., 6: 79-88.

Ismail, M., M.Z. Ibrahim, T.A. Ibrahim and A.M. Abdullah, 2011. Time series analysis of surface ozone monitoring records in Kemaman, Malaysia, Sains Malaysiana. 40 : 411-417

Kumar, K., A.K. Yadav, M.P. Singh and V.K. Jain, 2004. Forecasting daily maximum surface ozone concentrations in Brunei Darussalam-An ARIMA modeling approach. J. Air Waste Manage. Assoc., 54: 809-814. DOI: 10.1080/10473289.2004.10470949

Latif, M.T., L.S. Huey and L. Juneng, 2012. Variations of surface ozone concentration across the Klang Valley, Malaysia. Atmosphere Environ., 61: 434-445. DOI: 10.1016/j.atmosenv.2012.07.062

Mudway, I.S. and F.J. Kelly, 2004. An investigation of inhaled ozone dose and the magnitude of airway inflammation in healthy adults. Am. J. Resp. Crit. Care, 169: 1089-1095. DOI: 10.1164/rccm.200309-1325PP

Nieuwstadt, F.T.M., 1980. Prediction of air pollution frequency distribution-part II. The Gaussian plume model. Atmosphere Environ., 14: 259-265. DOI: 10.1016/0004-6981(80)90286-3

Nunez, C.M., G.H. Ramsey, M.A. Bahner and C.A. Clayton, 1999. An empirical model to predict styrene emissions from fiber-reinforced plastics fabrication processes. J. Air Waste Manage. Assoc., 49 :1168-1178. DOI: 10.1080/10473289.1999.10463800

Nuntavarn, V.V., N. Vajanapoom and B. Ostro, 2010. Estimating the effects of air pollution on mortality in Bangkok, Thailand. Health Effects Inst. Res. Report, 154; 231-266.

Rahman, S.R.A., S.N.S. Ismail, M.F. Ramli, M.T. Latif and E.Z Abinin *et al*., 2012. The assessment of ambient air pollution trend in Klang Valley, Malaysia. World Environ., 5: 1-11.

Salahudin, S.N., M.M. Abdullah and N.A. Newaz, 2013. Emissions: Sources, policies and development in Malaysia. Int. J. Educ. Res., 1: 1-12

Tan, K.C., H.S. Lim and M.Z.M. Jafri, 2014. Analysis of total ozone in Peninsular Malaysia retrieved from SCIAMACHY. Atmosphere. Pollut. Res., 5: 42-51. DOI: 10.5094/APR.2014.006

Varma, S.A.K., M. Srimurali and S.V.K. Varma, 2014. Prediction of ground level concentrations of air pollutants using Gaussian model, Rayalaseema Thermal Power Project, Kadapa, A.P., India. Energy Environ. Eng., 2: 91-97.

Wan Rozita, W.M., S. Mazrura, A. Rasimah, C.M. Wong and T.Q. Thach *et al*., 2013. Short-term effects of air pollution on mortality in Klang Valley, Malaysia. Atmosphere Environ., 65: 69-79. DOI: 10.1016/j.atmosenv.2012.10.019

Wang, W. and Y. Guo, 2009. Air pollution PM2.5 data analysis in Los Angeles Long Beach with seasonal ARIMA Model. Proceedings of the International Conference on Energy and Environment Technology, Oct. 16-18, IEEE Xplore Press, Guilin, Guangxi, China, pp: 7-10. DOI: 10.1109/ICEET.2009.468

WHO, 2000. Air Quality Guidelines For Europe. 1st Edn., World Health Organization, Regional Office for Europe, Copenhagen, Denmark, ISBN-10: 9289013583, pp: 273.

Wong, C.L., R. Venneker, S. Uhlenbrook, A.B.M. Jamil and Y. Zhou, 2009. Variability of rainfall in Peninsular Malaysia. Hydrol. Earth Syst. Sci., 6; 5471-5503. DOI: 10.5194/hessd-6-5471-2009