Journal of Environmental Science and Technology

ISSN 1994-7887

Journal of Environmental Science and Technology 8 (3): 102-112, 2015
ISSN 1994-7887 / DOI: 10.3923/jest.2015.102.112

Three Days Ahead Prediction of Daily 12 Hour Ozone (O_3) Concentrations for Urban Area in Malaysia

Muqhlisah Muhamad, Ahmad Zia Ul-Saufie and Sayang Mohd Deni
Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara, Malaysia

Corresponding Author: Ahmad Zia Ul-Saufie, Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara, Malaysia

ABSTRACT

Ground-level ozone (O_3) is a secondary pollutant and has an adverse effect on human health, agriculture and ecosystems. The aim of this study is to develop model and to predict future O_3 concentrations level in Shah Alam for next day (D+1), next two days (D+2) and next three days (D+3) using traditional method of Multiple Linear Regression (MLR) based on the concept of Ordinary Least Square estimate (OLS). This study uses daily average data of air pollutants (O_3, NO_x, NO, SO_2, NO_2, CO) and meteorological variables (WS, T, RH) that was selected from 2002 until 2013 as independent variables. The performance indicator of the models are measured by accuracy measures (Prediction accuracy, Index agreement and Coefficient of determination) and error measures (Root mean square error, Normalized absolute value). The average accuracy measures (AI, PA and R^2) show that the prediction for D+1, D+2 and D+3 is 0.4492, 0.3797 and 0.304 respectively. Meanwhile, the average error measures (RMSE, NAE) show that the prediction for D+1, D+2 and D+3 is 0.1453, 0.1374 and 0.1302, respectively.

Key words: Multiple linear regression, ordinary least square, O_3 concentrations, performance indicator

INTRODUCTION

Ground level ozone (O_3) in urban areas has become a serious air pollution problem. Based on the air quality status for 2013, 31.51% of unclean air was recorded in Shah Alam. According to Department of Environmental Malaysia (DoE), Air Pollution Index (API) was dominated by O_3 concentrations around afternoon until evening. The O_3 is a secondary pollutant. It is not originated directly from the earth's surface but it is formed by the chemical reaction under the influence of sunlight combining with nitrogen oxides (NO_x) and Volatile Organic Compound (VOCs) (DoE., 2006). VOCs are often known as nonmethane hydrocarbons (NmHC) (Ghazali *et al*., 2010). On the other hand, O_3 results in photochemical smog and as a trigger to human respiratory system problem (Sanna, 2009).

Various methods have been widely used from the previous study of O_3. One of the popular methods in prediction of O_3 concentrations level is Multiple Linear Regression (MLR). In statistical tools, analysis of regression is commonly used to analyze data. The MLR is a traditional method based on the concept of Ordinary Least Square estimate (OLS) (Ul-Saufie *et al*., 2012). MLR is used to present the relationship between dependent variable and independent variables (Chatterjee and Hadi, 2006).

Previous studies proved that MLR is a standard method and easy to be applied (Ul-Saufie *et al*., 2013). Future daily PM 10 concentrations prediction by combining regression models and feed forward back propagation models with Principle Component Analysis (PCA), in 2013. According to Barrero *et al*. (2006), the process involved in O_3 formation could be easy to understand by using MLR. Prediction of O_3 concentrations for several hours could be implemented by using MLR (Ramli *et al*., 2010). MLR is used to predict O_3 concentrations and at the same time to be used to understand the increasing and decreasing patterns of O_3 and NO_2, respectively under the influenced of weather parameters (Ghazali *et al*., 2010).

The aim of this study is to present the result of the multiple linear regression in prediction of O_3 concentrations level for the next day (D+1), the next two days (D+2) and the next three days (D+3) as the function of meteorological variables (WS, T, RH) and other pollutants concentration (NO_x, NO, SO_2, NO_2, O_3, CO).

MATERIALS AND METHODS

Study area: Monitoring station in Shah Alam is located at Taman Tun Dr. Ismail (TTDI) Jaya Primary School (N03°06.287', E101° 33.368') and nearby residential area. At the same time, this station is located at the main transportation area such as major road, highways and airport as well as surrounded by light industrial area (Azmi *et al*., 2010). Besides, Shah Alam city is located at the center of Petaling Jaya city (east) and Klang town (west) (Leh *et al*., 2014).

Monitoring record: The variables used in this study are ozone (O_3, ppm), wind speed (WS, km h^{-1}), ambient temperature (T, °C), relative humidity (RH, %), nitrogen oxide (NO_x, ppm), nitric oxide (NO, ppm), sulphur dioxide (SO_2, ppm), nitrogen dioxide (NO_2, ppm) and carbon monoxide (CO, ppm). The primary data was managed by Alam Sekitar Malaysia Sendirian Berhad (ASMA) which is the private company under supervision of Department of Environmental Malaysia (DoE).

According to Ahamad *et al*. (2014), Ghazali *et al*. (2010) and Banan *et al*. (2013) measurements of air pollutants and meteorological variables were monitored by Teledyne Ozone Analyzer Model 400A UV Absorption (O_3), Teledyne Model 200A (NO_x, NO, NO_2), Teledyne Model 100A (SO_2), Teledyne Model 300 (CO), Met One 010C Sensor (WS), Met One 062 Sensor (T) and Met One 083D Sensor (RH). These monitoring instruments automatically record the air pollutant concentrations and meteorological variables hourly. The instruments and procedures of monitoring record is based on the method fixed by the United States Environmental Protection Agency (EPA) standard (Ghazali *et al*., 2010). Furthermore, the secondary data from 1st January 2002 until 31st December 2013 was obtained from Department of Environmental Malaysia (DoE).

In this study, the hourly concentrations for each variables were transformed into daily average concentrations. Eighty percent of monitoring records were randomly selected and twenty percent were used for validation of the models. The statistical software used in the data analysis are SPSS Version 20, MATLAB R2012a and Microsoft Excel 2013.

Variable selection: The variables selected in this study are based on previous study of O_3 concentrations level (Table 1). The formation of O_3 is a result from emission and combination of the other air pollutants through a chemical process. The main substances of O_3 formation are

Table 1: Summarization of selected variables by previous researchers

Authors	WS	T	RH	NO_x	NO	SO_2	NO_2	CO	$O_{3,D+1}$	Others
Agirre-Basurko *et al*. (2006)							X		X	
Musa *et al*. (2013)									X	
Jaioun *et al*. (2014)								X	X	PM_{10}
Wang *et al*. (2003)	X	X		X	X	X	X	X	X	WD, SR, PM_{10}
Ghazali *et al*. (2010)							X		X	UVB
Heo *et al*. (2004)	X	X	X			X	X	X	X	WD, SR
Delcloo and de Backer (2005)		X	X	X					X	
Ramli *et al*. (2010)	X	X		X					X	
Banan *et al*. (2014)	X	X	X	X	X		X	X	X	SO_2, PM_{10}, NMHC
Schlink *et al*. (2006)	X	X			X		X		X	WD, WV, SR

SO_2: Sulphur dioxides, PM_{10}: Particulate matter, WD: Wind direction, WV: Wind violation, SR: Solar radiation, NMHC: Nonmethane hydrocarbon

VOCs and NO_x. According to Department of Environment, Malaysia, 2006, VOCs are emitted from factories' chimney, motor vehicles, industrial activities, consumers and commercial products. Meanwhile, NO_x are released by motor vehicles, power plants and combustions. Most of the previous studies stated that meteorological conditions also contribute to the formation of O_3 concentrations.

Meteorological variable: Khiem *et al*. (2010) found that the low wind speed which associated with the other meteorological conditions has a high ability to contribute O_3 concentrations. Urban area has very little difference of O_3 concentrations level with rural area during high wind speeds (Husar and Renard, 1997). Temperature is also one of the main factors in the O_3 production and formation. The concentrations level of O_3 tends to increase at high temperature (Banja *et al*., 2012). Besides, relative humidity could be considered as a contributor to the O_3. The lack of photochemical process efficiency due to the high relative humidity has always been associated with low level of O_3 concentrations (Lelieveld and Crutzen, 1990).

Air pollutant variable: According to DoE. (2014), the sources of SO_2 come from power plants (50%), industrial activities (9%), motor vehicles (7%) and others (34%). The contributors of NO_2 are power plants (61%), motor vehicles (26%), industrial activities (6%) and others (7%). Meanwhile, the emissions of CO are detected from motor vehicles (95.3%), power plants (3.8%), industrial activities (0.4%) and others (0.5%). These situations increase the formation of O_3 concentrations level in Malaysia.

Regression analysis: Regression analysis that was used in this study is Multiple Linear Regression (MLR) based on traditional approaches of Ordinary Least Square estimate (OLS). MLR is an extension from a simple linear regression. In MLR, there are one dependent variable (response variable) and several independent variables (explanatory variables/predictors). Chatterjee and Hadi (2006) defined the general equation of MLR as follows:

$$y_i = \beta_0+\beta_1 x_{i1}+\ldots\ldots+\beta_p x_{ip}+\varepsilon, \quad i = 1,2,\ldots.n \tag{1}$$

Where:

y = Dependent variable (response variable)

x = Independent variable (predictor)

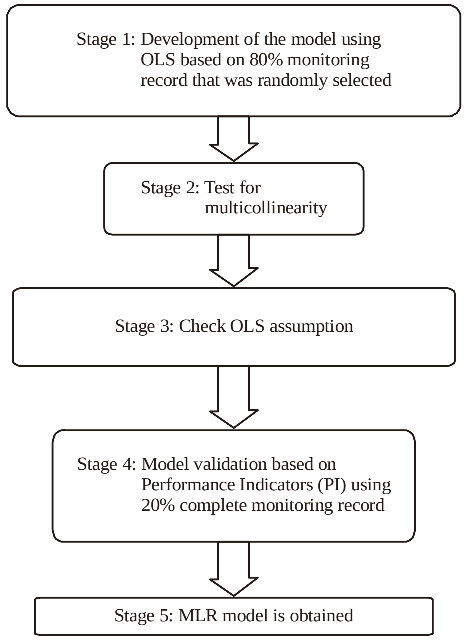

Fig. 1: Procedure for development of multiple linear regression model

Table 2: Ordinary least square assumption

Assumption	Checked by
Residuals follow a normal distribution	Graph of histogram/box plot
Residual has constant variance (homoscedasticity)	Scatter plot (the spread of point)
Residuals are uncorrelated with the independent variables	Durbin Watson test statistics (no autocorrelation present in error if close to 2)
The residual has zero mean	Graph of histogram

Source: Chatterjee and Hadi (2006), Residual = error = ε

p = Represent values of the predictors for ith unit
β_0 = Regression constant
β_p = Regression coefficient

Montgomery *et al*. (2012), found that the method of least square is used to estimate parameters and MLR mostly was used as an empirical model. There are a few stages required to obtain MLR model (Fig. 1).

Stage 1, 80% of the data for each variables was randomly selected by MATLAB R2012a. Stage 2; the Variance Inflation Factor (VIF) was used to check the multicollinearity test. The model is considered to be free of multicollinearity problem if the value of VIF is less than 10 (Field, 2005). During Stage 3, check the OLS assumptions (Table 2). Then, the models are validated by performance indicator (RMSE, NAE, PA, IA, R^2) using 20% complete monitoring record in stage 4. Finally in stage 5, the MLR model was obtained.

Performance indicator: Performance indicators are used to evaluate the performance models for next day (D+1), next two days (D+2) and next three days (D+3) predictions. The performance models (Table 3) are consists of accuracy measures (PA, IA and R^2) and error measures (RMSE and NAE).

Table 3: Performance indicators

Performances Indicator (PI)	Formulae	Notes
Normalized Absolute Error (NAE)	$\frac{\sum_{i=1}^{N}\lvert P_i - O_i \rvert}{\sum_{i=1}^{N} O_i}$	Close to 0, model is appropriate
Root Mean Square Error (RMSE)	$\sqrt{\left(\frac{1}{N-1}\right)\sum_{i=1}^{N}(P_i - O_i)^2}$	Close to 0, model is appropriate
Index of Agreement (IA)	$1-\left[\frac{\sum_{i=1}^{N}(P-O_i)^2}{\sum_{i=1}^{N}(\lvert P_i-\bar{O}\rvert+\lvert O_i-\bar{O}}\right.$	Close to 1, model is appropriate
Prediction Accuracy (PA)	$\frac{\sum_{i=1}^{N}(P_i-\bar{P})^2}{\sum_{i=1}^{N}(O_i-\bar{O})^2}$	Close to 1, model is appropriate
Coefficient of determination (R^2)	$\left[\frac{\sum_{i=1}^{N}(P_i-\bar{P})(O_i-\bar{O})}{N.S\,(pred).S\,(obs)}\right]^2$	Close to 1, model is appropriate

Source: Gervasi (2008), N: No. of sample daily measurement of a particular station, p_i: Predicted value, O_i: Observed values, $\bar{P}$: Mean of the predicted values, $\bar{O}$: Mean of the predicted values of one set daily monitoring record

RESULTS AND DISCUSSION

Descriptive statistic is used to describe a situation (Bluman, 2009). Shah Alam is located at TTDI Jaya Primary School. The mean average of O_3 concentration for Shah Alam is 0.032 ppm and the monitoring record is assumed to be moderately skewed with the value of 0.717. The maximum amount of O_3 concentration recorded was 0.097 ppm (Table 4 and Fig. 2). This is due to open burning and smokes from vehicles (DoE., 2004). According to Department of Environment, Malaysia, 2011, the unhealthy days from year 2001 to 2012 in Klang Valley was mainly due to the high concentration level of O_3. Shah Alam was recorded as having the highest number of unhealthy days except for year 2005, 2010, 2011 and 2012 (Fig. 3).

In order to investigate the correlation between $O_{3,\ D+1}$ (for next day) and each independent variables, regression analysis was performed based on the value of correlation coefficient (R) and scatter plot. From the regression analysis for each variables (Table 5 and Fig. 4), the relationship for each variables with $O_{3,\ D+1}$ are WS (R = -0.036), T (R = 0.155), RH (R = -0.212), NO_x (0.056), NO (R = -0.018), SO_2 (R = 0.121), NO_2 (R = 0.136), O_3 (R = 0.445) and CO (R = 0.177), where WS, RH and NO have a negative correlation with $O_{3,\ D+1}$ and the rest of variables have a positive correlation. Table 6 shows that the correlation coefficient (R) among O_3 and the other predictors from the previous studies are (Ghazali *et al*., 2010), R^2 = 89.90% for Shah Alam and Gombak (Banan *et al*., 2014), for Putrajaya, NO_x (R = 0.681), NO (R = -0.537), NO_2 (R = -0.499), for Petaling Jaya, NO_x (R = 0.515), NO (R = -0.678), NO_2 (R = -0.102) and Jerantut, NO_x (R = 0.416), NO (R = -0.557) NO_2 (R = -0.079), Ramli, Ghazali *et al*. (2010), R^2 for Shah Alam and Nilai are 89.7 and 89.0%, respectively. The previous studies from the world wide show (Wang *et al*., 2003), the value of R^2 for Hong Kong is 76.16% (Agirre-Basurko *et al*., 2006), the value of R for NO_2 is 0.88 and (Banja *et al*., 2012), T (R = 0.72), RH (R = -0.40) and R^2 is 76%.

According to Field (2005) and Montgomery *et al*. (2012), the model is considered to have a problem with multicollinearity if the value of VIF is larger than 10. Since the value of VIF for variables NO_x, NO and NO_2 are larger than 10, thus the model of O_3 for next day prediction (D+1)

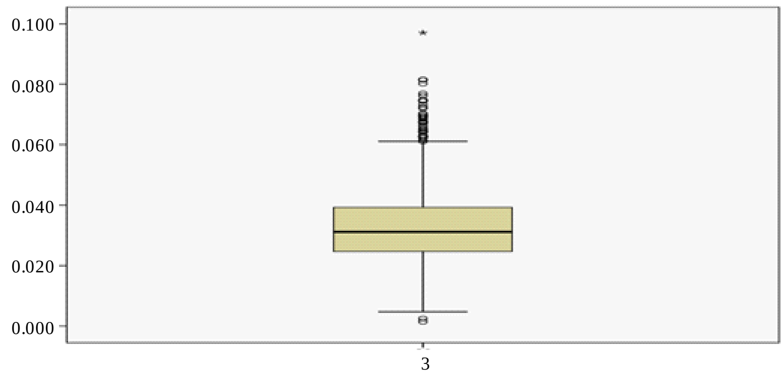

Fig. 2: Box and whisker plot for O_3 concentrations

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Banting	-	-	-	-	-	-	-	-	-	0	5	5
Shah ajam	19	67	56	88	60	47	23	44	41	32	22	11
Kajang	37	17	34	41	44	34	24	11	13	0	-	-
Choras	11	30	28	63	67	5	19	24	24	59	48	37
Klang	7	50	12	11	33	14	2	8	16	6	6	13
Petaling jaye	4	2	9	17	20	12	2	7	10	3	3	4
Kuala selanoor	8	21	0	2	13	4	0	0	10	0	0	5
Putrajaya	-	-	1	1	22	30	13	13	7	8	3	2
Batu Muda	-	-	-	1	-	-	-	-	21	17	15	25

Fig. 3: Number of unhealthy days in Klang Valley from year 2001 until 2012 (Source: DoE., 2014)

Table 4: Descriptive statistics of O_3 in Shah Alam

Parameters	O_3 (ppm)
Mean	0.032
Median	0.031
SD	0.011
Skewness	0.717
Kurtosis	1.163
Maximum	0.097

SD: Standard deviation

Table 5: Correlation coefficient between $O_{3,\ D+1}$ and each variables

Variables	Correlation coefficient (R)
WS	-0.036
T	0.155
RH	-0.212
NO_x	0.056
NO	-0.018
SO_2	0.121
NO_2	0.136
O_3	0.445
CO	0.177

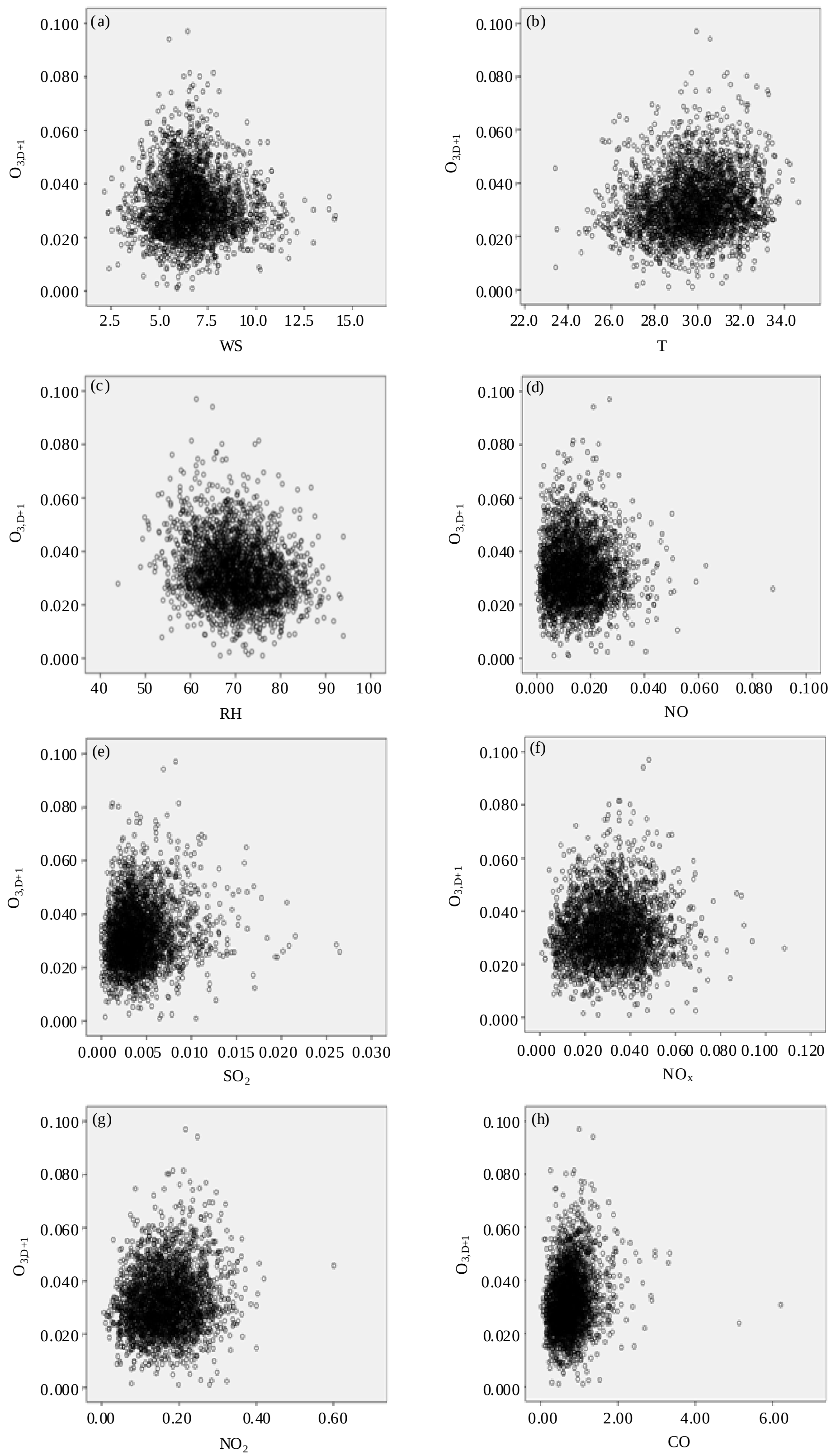

Fig. 4(a-i): Continue

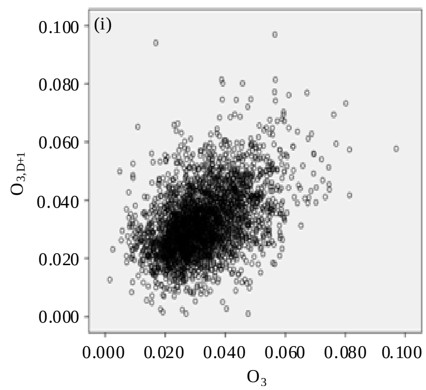

Fig. 4(a-i): Scatter plot of $O_{3,\ D+1}$ versus independent variables

Table 6: Correlation (R) and coefficient of determination (R^2) between O_3 concentrations level and its predictor variables

References	Site	T	RH	NO_X	NO	NO_2	R^2 (%)
Ghazali *et al.* (2010)	Shah Alam	-	-	-	-	-	89.90
Ghazali *et al.* (2010)	Gombak	-	-	-	-	-	89.90
Banan *et al.* (2013)	Putrajaya	-	-	-0.681	-0.537	-0.499	-
Banan *et al.* (2013)	Petaling Jaya	-	-	-0.515	-0.678	0.102	-
Banan *et al.* (2013)	Jerantut			-0.416	-0.557	-0.079	-
Ramli *et al.* (2010)	Shah Alam	-	-	-	-	-	89.70
Ramli *et al.* (2010)	Nilai	-	-	-	-	-	89.00
Wang *et al.* (2003)	Hong Kong	-	-	-	-	-	76.16
Agirre-Basurko *et al.* (2006)	Bilbao, Spain	-	-	-	-	0.880	-
Banja *et al.* (2012)	Tirana, Albania	0.72	-0.40	-	-	-	76.00

Table 7: Multicollinearity test of $O_{3,\ D+1}$

Variable	Regression coefficient (β_i)	VIF
Constant	0.065	
WS	0.000204	1.386
T	-0.001	3.107
RH	-0.000245	3.198
NO_x	-1.493	5679.004
NO	1.511	2291.852
SO_2	-0.010	1.299
NO_2	1.566	1428.590
O_3	0.446	1.625
CO	0.002	1.628

Table 8: Multicollinearity test of $O_{3,\ D+1}$ (Without NO_x)

Variables	Regression coefficient (β_i)	VIF
Constant	0.066	
WS	0.000201	1.386
T	-0.001	3.099
RH	-0.000248	3.188
NO	0.019	1.767
SO_2	-0.012	1.298
NO_2	0.072	2.211
O_3	0.446	1.624
CO	0.002	1.627

has multicollinearity problem (Table 7). This is due to the presence of NO_x where NO_x is a result of NO and NO_2 (NO_x = NO+NO_2) (Ghazali *et al.*, 2010). According to this problem, NO_x should not be included in this study and after the variable of NO_x was truncated, the range of VIF showed that the model was free from multicollinearity problem (Table 8).

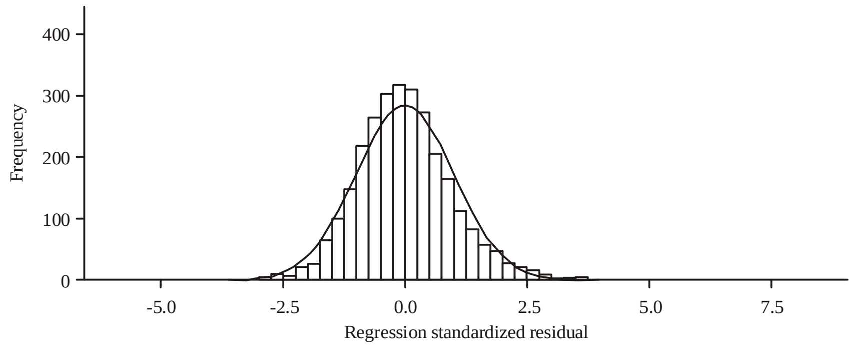

Fig. 5: Histogram and table of O_3 residual: D+1

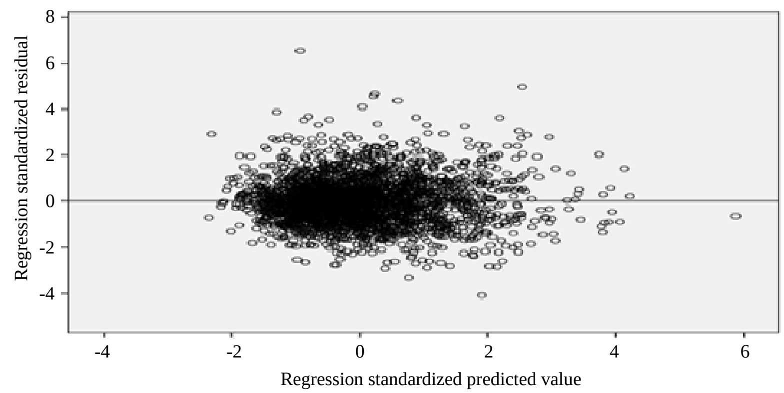

Fig. 6: Scatter plot of residual versus fitted values

Table 9: Multiple linear regression and performance indicator

Model	NAE	RMSE	IA	PA	R^2
$O_{3,\,D+1}$ = 0.066+0.000201WS-0.001T-0.000248RH+ 0.019NO-0.012SO_2+0.072NO_2+0.446O_3+0.002CO	0.2794	0.0112	0.5772	0.4535	0.3168
$O_{3,\,D+2}$ = 0.062+0.0048WS-0.001T-0.000216RH+ 0.004NO-0.054SO_2+0.047NO_2+0.367O_3+0.003CO	0.2640	0.0108	0.5216	0.4164	0.2011
$O_{3,\,D+3}$ = 0.057+0.000496WS-0.001T-0.000189RH+ 0.047NO-0.038SO_2+0.043NO_2+0.317O_3+0.002CO	0.2495	0.0109	0.4459	0.3131	0.1530

The residual of O_3 concentrations in Shah Alam for next day (D+1) shows that the graph of histogram has bell-shaped distributions which means that the residuals approximately normally distributed with zero mean of residual (Fig. 5). The assumption of the residual has a constant variance is satisfied when the scatter plot (Fig. 6) shows an equal spread and approach to regression line (homoscedasticity). Besides, the assumption of the residuals being uncorrelated with the independent variables is satisfied when the value of Durbin Watson is close to 2 (1.945).

The procedures from Table 8, Fig. 5 and 6 were repeated to obtain MLR model of Shah Alam for next two days ($O_{3,D+2}$) and next three days ($O_{3,D+3}$).

CONCLUSION

Table 9 shows the performance indicators for next day (D+1), next two days (D+2) and next three days (D+3) in Shah Alam that were obtained from the model of multiple linear regression. The average accuracy measures (AI, PA and R^2) show that the prediction for D+1 is 0.4492 followed by D+2, 0.3797 and D+3, 0.304. Besides, the average error measures (RMSE, NAE) show that the prediction for D+1, D+2 and D+3 are 0.1453, 0.1374 and 0.1302, respectively. Due to the data limitation of VOCs and UVB, the value of PA, IA and R^2 are not close to one but the model is still appropriate in prediction of O_3 concentrations level since the value of RMSE and NAE is close to zero. This is supported by the previous study from Yousef *et al*., 2008, where the best linear regression model for the air pollutant of particulate matter (PM_{10}) for dry season and wet season are 0.262 and 0.240, respectively. Therefore, these three models could be implemented for public health protection to provide early warnings to the respective populations.

ACKNOWLEDGMENT

This study was funded by Universiti Teknologi MARA, Malaysia under Grant 600-RMI/FRGS 5/3 (40/2014). Special appreciation to Department of Environmental Malaysia (DoE), Alam Sekitar Malaysia (ASMA) for providing the air quality data for this research and special thanks to Universiti Teknologi Mara, Malaysia.

REFERENCES

Agirre-Basurko, E., G. Ibarra-Berastegi and I. Madariaga, 2006. Regression and multilayer Perceptron-based models to forecast hourly O_3 and NO_2 levels in the Bilbao area. Environ. Modell. Software, 21: 430-446.

Ahamad, F., M.T. Latif, R. Tang, L. Juneng, D. Dominick and H. Juahir, 2014. Variation of surface ozone exceedance around Klang Valley, Malaysia. Atmos. Res., 139: 116-127.

Azmi, S.Z., M.T. Latif, A.S. Ismail, L. Juneng and A.A. Jemain, 2010. Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. Air Qual. Atmos. Health, 3: 53-64.

Banan, N., M.L. 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.

Banan, N., M. Latif, L. Juneng and M.F. Khan, 2014. An Application of Artificial Neural Networks for the Prediction of Surface Ozone Concentrations in Malaysia. In: From Sources to Solution, Aris, A.Z., T.H.T. Ismail, R. Harun, A.M. Abdullah and M.Y. Ishak (Eds.). Springer, Singapore, pp: 7-12.

Banja, M., D. Papanastasiou, A. Poupkou and D. Melas, 2012. Development of a Short-term ozone prediction tool in Tirana area based on meteorological variables. Atmos. Pollut. Res., 3: 32-38.

Barrero, M.A., J.O. Grimalt and L. Canton, 2006. Prediction of daily ozone concentration maxima in the urban atmosphere. Chemom. Intell. Lab. Syst., 80: 67-76.

Bluman, A., 2009. Elementary Statistics a Step by Step Approach. McGraw Hill, New York.

Chatterjee, S. and A.S. Hadi, 2006. Regression Analysis by Example. 4th Edn., John Wiley, New Jersey, Pages: 416.

Delcloo, A.W. and H. de Backer, 2005. Modelling planetary boundary layer ozone, using meteorological parameters at Uccle and Payerne. Atmos. Environ., 39: 5067-5077.

DoE., 2004. Malaysia environmental quality report 2003. Department of Environment, Ministry of Natural Resources and Environment, Kuala Lumpur, Malaysia.

DoE., 2006. Malaysia environmental quality report 2005. Department of Environment, Ministry of Natural Resources and Environment, Putrajaya, Malaysia.

DoE., 2014. Malaysia environmental quality report 2013. Department of Environment, Ministry of Natural Resources and Environment, Kuala Lumpur, Malaysia.

Field, A.P., 2005. Discovering Statistics Using SPSS. 2nd Edn., SAGE Publication Ltd., London.

Gervasi, O., 2008. Computational Science and its Applications. Springer, New York, USA.

Ghazali, N.A., N. Ramli, A. Yahaya, N.F. Yusof, N. Sansuddin and W. Al Madhoun, 2010. Nurul adya transformation of nitrogen dioxide into ozone and prediction of ozone concentrations using multiple linear regression. Environ. Monit. Assess., 165: 475-489.

Heo, J.S., K.H. Kim and D.S. Kim, 2004. Pattern recognition of high episodes in forecasting daily maximum ozone levels. Terrestrial Atmospheric Oceanic Sci., 15: 199-220.

Husar, R.B. and W.P. Renard, 1997. Ozone as a function of local wind direction and wind speed evidence of local and regional transport. Center for Air Pollution Impact and Trend Analysis (CAPITA), May 7, 1997.

Jaioun, K., K. Saithanu and J. Mekparyup, 2014. Multiple linear regression model to estimate ozone concentration in chonburi, Thailand. Int. J. Applied Environ. Sci., 9: 1305-1308.

Khiem, M., R. Ooka, H. Huang, H. Hayami, H. Yoshikado and Y. Kawamoto, 2010. Analysis of the relationship between changes in meteorological conditions and the variation in summer ozone levels over the Central Kanto area. Adv. Meteorol. 10.1155/2010/349248

Leh, O.L.H., S.N.A.M. Musthafa and A.R.A. Rasam, 2014. Urban environmental heath: Respiratory infection and urban factors in urban growth corridor of Petaling Jaya, Shah Alam and Klang, Malaysia. Sains Malaysiana, 43: 1405-1414.

Lelieveld, J. and P.J. Crutzen, 1990. Influences of cloud photochemical processes on tropospheric ozone. Nature, 343: 227-233.

Montgomery, D.C., E. Peck and G. Vining, 2012. Introduction to Linear Regression Analysis. 5th Edn., John Wiley, New Jersey, ISBN-13: ISBN-13: 9780470542811, Pages: 672.

Musa, M., A. Jemain and W.W. Zin, 2013. Scaling and persistence of ozone concentrations in Klang Valley. J. Qual. Measur. Anal., 9: 9-20.

Ramli, N.A., N.A. Ghazali AND A.S. Yahaya, 2010. Diurnal fluctuations of ozone concentrations and its precursors and prediction of ozone using multiple linear regressions. Malaysian J. Environ. Manage., 11: 57-69.

Sanna, E., 2009. Air Pollution and Health. Health and the Environment, New York.

Schlink, U., O. Herbarth, M. Richter, S. Dorling, G. Nunnari, G. Cawley and E. Pelikan, 2006. Statistical models to assess the health effects and to forecast ground-level ozone. Environ. Modell. Software, 21: 547-558.

Ul-Saufie, A.Z., A.S. Yahaya, N.A. Ramli and H.A. Hamid, 2012. Performance of multiple linear regression model for long-term PM_{10} concentration prediction based on gaseous and meteorological parameters. J. Applied Sci., 12: 1488-1494.

Ul-Saufie, A.Z., A.S. Yahaya, N.A. Ramli, N. Rosaida and H. Abdul Hamid, 2013. Future daily PM_{10} concentrations prediction by combining regression models and feedforward backpropagation models with Principle Component Analysis (PCA). Atmos. Environ., 77: 621-630.

Wang, W., W. Lu, X. Wang and A.Y. Leung, 2003. Prediction of maximum daily ozone level using combined neural network and statistical characteristics. Environ. Int., 29: 555-562.

Yousef, N.F., N.A. Ghazli, N.A. Ramli, A.S. Yahia, N. Sansuddin and W.A. Al Madhoun, 2008. Correlation of PM_{10} concentration and weather parameters in conjunction with haze event in Seberang Perai, Penang. Proceeding of the International Conference on Construction and Building Technology, June 16-20, 2008, Kuala Lumpur, Malaysia.