Journal of Engineering Science and Technology
Vol. 16, No. 4 (2021) 3136 – 3154

DEVELOPMENT OF MODELS FOR FORE-CASTING OF SEASONAL GROUND LEVEL OZONE (O_3)

NUR N L M NAPI[1], SAMSURI ABDULLAH[1,2,*],
AMALINA A MANSOR[3], ALI N AHMED[4], MARZUKI ISMAIL[2,3]

[1]Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
[2]Institute of Tropical Biodiversity and Sustainable Development, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
[3]Faculty of Science and Marine Environment, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
[4]Institute of Energy Infrastructure (IEI), Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), Kajang, Selangor Darul Ehsan 43000, Malaysia

Abstract

Ozone (O_3) is a secondary pollutant that releases to the atmosphere through industrial and motor vehicles emission which give an adverse impact, especially on human health. The meteorological factor especially weather condition has influenced the production of O_3 concentration in the atmosphere. This study aims to develop O_3 forecasting model during different monsoon seasons. The data from the year 2012 until 2014 were acquired including O_3, nitrogen oxide (NO), nitrogen dioxide (NO_2), carbon monoxide (CO), sulphur dioxide (SO_2), wind speed (WS), ambient temperature (T) and relative humidity (RH) on an hourly basis. The Multiple Linear Regression (MLR) models were developed for the prediction of up to 3 hours in advance. Southwest Monsoon (SWM) was having a higher O_3 concentration with a mean value of 0.024 ppm while Inter Monsoon 2 (IM2) was having the lowest concentration of O_3, 0.019 ppm. The best fits MLR models for each monsoon were $O_{3,t+1}$ as compared to $O_{3,t+2}$ and $O_{3,t+3}$. The better interpretation and prediction of O_3 behaviour during monsoon conditions can help the responsible parties to plan early mitigation measures to address the air pollution problem.

Keywords: Forecasting, Meteorological, Monsoon, Multiple linear regression. Ozone.

1. Introduction

Compact urbanization and high-density of the population are demanding for better air quality because it generates a lot of air pollutants that threaten human health and our environment especially in a developing country like Malaysia [1, 2]. The emission of O_3 came through anthropogenic and biogenic emissions [3]. The emissions from anthropogenic and human activities such as motor vehicles, industrial combustion and agriculture sector caused a rise in air pollutants production that pollutes the air [4-6].

Ozone (O_3) is a predominant air pollutant and is known as greenhouse gases. It was known as a secondary pollutant that colourless and reactive oxidant gas in properties. It formed through the photo-chemical reaction which involves active primary pollutants [7, 8]. The concentration of O_3 in the troposphere and stratosphere layer was influenced by how it transports and disperses in the air with three pathways which are through natural emission, chemistry pathway and transportation pathway [9]. The rise of O_3 concentration was emitted through natural emission of O_3 from climate - sensitive such as lightning, biosphere and regional weather shifts and it also plays a part in the chemical pathway that influences the weather parameters [10, 11]. Weather conditions might influence the variation of O_3 concentrations in Malaysia via several monsoons' seasons such as the Southwest Monsoon, Northeast Monsoon and two transition periods (inter-monsoon) [12, 13]. The changes of O_3 concentrations during this time might give an adverse impact on human health, including mortality, morbidity, cardiovascular system problem, lung cancer and premature death due to O_3 exposure [14, 15]. Furthermore, the uncontrolled O_3 emission can cause global warming to our earth and affecting the crop production [16-18].

Meteorological factors played an important role in the variation of ozone concentrations. Ozone concentrations are low during low temperature since low temperature prohibit convective dispersal of ozone precursors [19]. Ozone has negative relationship with relative humidity. The diurnal pattern of relative humidity indicates high levels at midnight and in the early morning, gradually falling after sunrise contrary to the diurnal pattern of O_3 [20]. Wind allows the dispersal of the trace pollutant species along horizontal and vertical levels by promoting the mixing and transport across/ in-between the boundary layer and the upper free atmosphere. High wind speeds during the afternoon are accompanied by higher O_3 mixing ratios and low wind speeds during early morning and late night contribute to the O_3 sinking process [20]. The higher emissions of biogenic NO_x and VOCs may be the main drivers of the growing ozone pollution under the high air temperature [21]. High CO concentrations can significantly change the oxidation capacity of the atmosphere [22].

Studies have explained the use of MLR models in air quality as a tool for forecasting. MLR is used in forecasting air quality in Malaysia which it can be able to forecast with 70% accuracy [12]. Other study shows that, the air quality was forecasted with the accuracy of 78-93% in Macao using MLR, which one of the pollutants forecasted was O_3 [23]. Combination of temporal (using MLR) and spatial analysis for ozone forecasting was applied with the help of several inputs including, temperature, solar radiation, wind speed, and previous ozone values executed reliable results [24]. MLR is also used in predicting O_3 concentrations at urban traffic and background station in Oporto, Portugal based on linear and additive associations of the explanatory variables [25]. An application of MLR model shows that it worked better at suburban and rural sites than at urban sites in Hong Kong [26]. The methodologies of this model describe the relationship of dependence variable with several independent variables by a simple computation and easy to implement [12, 27].

This paper intended to investigate the trend and predict the O_3 concentration in an industrial area in Kemaman, Malaysia. The relationship of O_3 between meteorological factors and its precursors which contribute to the rise of O_3 emission in the air was also investigated. The findings will help the authority foresee the future of ozone trends in providing the mitigation and precaution plans for air quality improvement in line to achieve a green city aim that establishes under sustainable goals.

2. Materials and Methods

2.1. Site case study

Kemaman is one of the industrial parks at Terengganu that involved heavy industries activities such as oil, petrochemical, and steel production, having an area of 2,535.60 km^2 with the overall population of 201,100 people [28]. The Malaysian Department of Environment (DOE) has installed the Air Quality Monitoring Station (AQMS) at Bukit Kuang, Teluk Kalong Primary School, Kemaman (N04^o 16.260'; E103^o 25.826') which is near the city center and industrial area (Fig. 1). The meteorological condition plays a significant role in determining the fate of O_3 because of the location is experiencing four monsoon seasons; North-East Monsoon (NEM) in November until March, Inter Monsoon 1 (IM1) in April, South-West Monsoon (SWM) in June to September and Inter Monsoon 2 (IM2) in October [29].

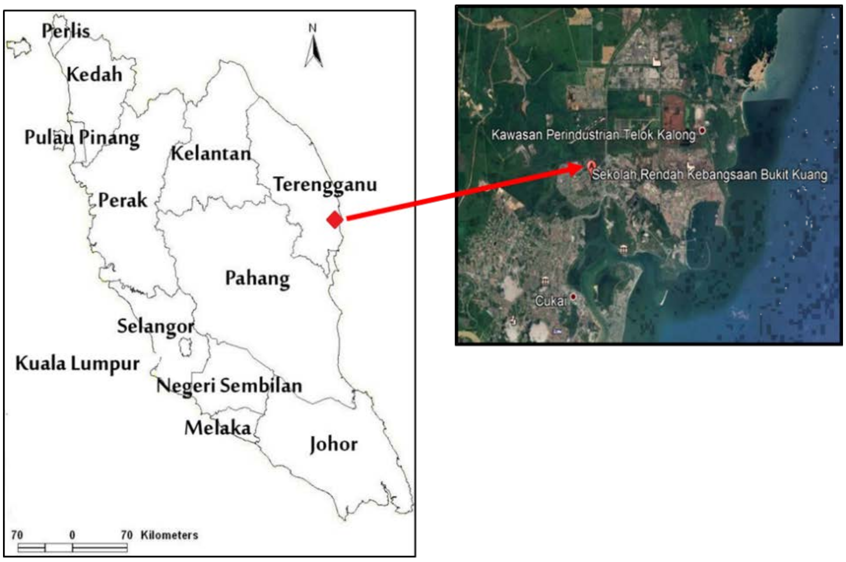


Fig. 1. Area of the study.

2.2. Monitoring data

Three years of data (2012-2014) were acquired from the Malaysian Department of Environment (DOE), were then arranged, and tabulated using Microsoft Excel

2016. Parameters taken into consideration for the forecasting model are ground-level ozone (O_3, ppm), nitrogen oxide (NO, ppm), nitrogen dioxide (NO_2, ppm), carbon monoxide (CO, ppm), sulphur dioxide (SO_2), wind speed (WS, km/hr), ambient temperature (T, ^{0}C) and relative humidity (RH, %). Alam Sekitar Malaysia Sdn. Bhd. is the responsible company for installing, monitoring and operating the air quality under the Malaysian Continuous Air Quality Monitoring (CAQM) program [12] and the instrumentations for MCAQM are specified as in Table 1.

Table 1. Instrumentations for MCAQM [30, 31].

Pollutant	Instrument
O_3 (ppm)	Teledyne API Model 400/400E through UV absorption (Beer- Lambert) method with a 0.4ppb detection limit and 0.5% of precision level.
NO and NO_2 (ppm)	Model 200A NO/NO_2/NO_x analyser which applies chemiluminescence detection principles.
SO_2 (ppm)	Teledyne API Model 100A/100E using UV fluorescence method, lowest detection level of was at 0.04 ppb.
CO (ppm)	Teledyne API Model 300/300E using the non-dispersive and infrared absorption (Beer Lambert) method with 0.5% precision and 0.04 ppm of the lowest detection.
Ambient temperature (T, ^{0}C)	Met One 062 sensor.
Relative humidity (RH, %)	Met One 083D sensor.
Wind speed (WS, km/hr)	Met One 010C sensor.

The quality assurance and quality control of all instruments were programmed to have daily calibration using zero air and standard gas concentration for the monitoring data and the validation of hourly data was checked before it can be transferred to DOE [32]. The missing data due to calibration and the technical problem was deleted to produce an unbiased prediction and conservative results [33]. The data were analysed by box-and-whisker plot analysis to determine the trend and the non-parametric technique was performed by Spearman correlation analysis to investigate the relationship between the parameters [34].

2.3. Multiple Linear Regression

The statistical model named the MLR is used in this paper which relate a dependent variable with several independent variables. Stepwise MLR model has been developed in this paper, look at 95% of the confidence interval and the data set is divided respecting to 7:3 ratio for model development and validation [35, 36] by using the Statistical Package for Social Sciences (SPSS®) version 25. The MLR model's residuals were anticipated normal distribution with having 0 suggest, uncorrelated and consistent variance. The equation of the MLR model is as stated in Eq. (1).

$$y = b_0 + \sum_{i=1}^{n} b_i X_i + \varepsilon \qquad (1)$$

where b_i is the regression coefficient (independent variables), ε is the stochastic error associated with the regression.

The independent and dependent variable of the data set needs to normalize before undergoing the analysis to avoid bias due to different units of variables. The normalized data executed with data values from a 0 to 1 scale. The normalization of data is obtained by the Eq. (2) [12].

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \qquad (2)$$

where $x = (x_1 \ldots x_n)$ and z_i is normalized data.

VIF is the multi - collinearity assumption together with regression output, where the average of VIF values need below 10 which indicates no multicollinearity problem between the independent variables [37]. The VIF equation as in Eq. (3).

$$VIF = \frac{1}{(1 - R_i^2)} \tag{3}$$

where VIF_i is the variance inflation factor with ith predictor, R_i^2 is the determination in a regression of the ith predictor on all other predictors.

Durbin - Watson test is used to determine the autocorrelation ability of O_3 concentration during the current hour to predict the O_3 in the next hour which the best is in the range of zero and four [37]. The equation of DW as in Eq. (4).

$$d = \frac{\sum_{i=1}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \tag{4}$$

where n is the number of observations, $e_i = y_i - y_i$ (y_i = observed values and y_i is the predicted values).

Correlation Coefficient (R^2) is an indicator in determining the accuracy of the model for the prediction of O_3 concentrations. It also acts as an indicator to measure how good the prediction models fit the data [37]. The R^2 equation is given as in Eq. (5).

$$R^2 = \left(\frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{n.S_{pred}.S_{obs}}\right)^2 \tag{5}$$

where n = total number measurements at a particular site P_i = predicted values, O_i = observed values, $\bar{P}$ = mean of predicted values, $\bar{O}$ = mean of observed values, S_{pred} = standard deviation of predicted values and S_{obs} = standard deviation of the observed values.

3. Results and Discussion

The O_3 concentrations during the monsoon seasons were illustrated in Fig. 2 and Table 2. It is observed that SWM having the highest outlier (0.098 ppm) of O_3 concentration followed by IM1 (0.083 ppm) and NEM (0.075 ppm), meanwhile, IM2 having the lowest value of O_3 concentrations. The NEM and IM2 were having the same interquartile range value of 0.020 ppm. The descriptive statistics summary for 3-years data from 1st January 2012 until 31st December 2014 of O_3 concentration is shown in Table 1. The NEM and IM2 were having the same minimum value (0.001 ppm) while SWM and IM1 were having zero minimum value. The minimum and maximum values in this study were acceptable reading which proven by the previous study in Malaysia that having O_3 concentration from 0.001ppm until 0.174 ppm [38]. The highest mean value is 0.024 ppm (0.000-0.098 ppm) during SWM, while the lowest mean reading is 0.019 ppm (0.001- 0.072) ppm in IM2. Meanwhile, the standard deviation of SWM (0.017) is the highest compared to other monsoon seasons while the lowest standard deviation is during NEM (0.012). The SWM and NEM were having the highest median value (0.020 ppm) as compared to IM1 (0.017 ppm) and IM2 (0.015 ppm). The highest value of skewness is during IM2 (0.974) followed by IM1 (0.847), SWM (0.667) and NEM (0.573).

The Malaysia's New Ambient Air Quality Standard (NAAQS) under Department of Environment Malaysia has set a maximum limit for O_3 concentration of 0.12 ppm for 24-hour average, and the O_3 concentrations in this study were found below the standard [39, 40]. The emission of gases such as carbon monoxide, VOCs and hydrocarbon through the combustion process especially from industrial and motor vehicles were the main contributor to the rise of O_3 concentration in the atmosphere [41-43]. The meteorological factors influencing the increase of O_3 concentration especially during SWM (having highest outlier) which introduces the dry and warmer conditions that increase the ambient temperature and solar radiation, thus promote photochemical reaction to occur [44]. O_3 decreases during the IM2 as the wet season is about to occur and the maintained during the NEM (wet season). The extreme O_3 concentration is about to occur during the IM1 as the interchange of monsoon from wet to dry which induces high temperature.

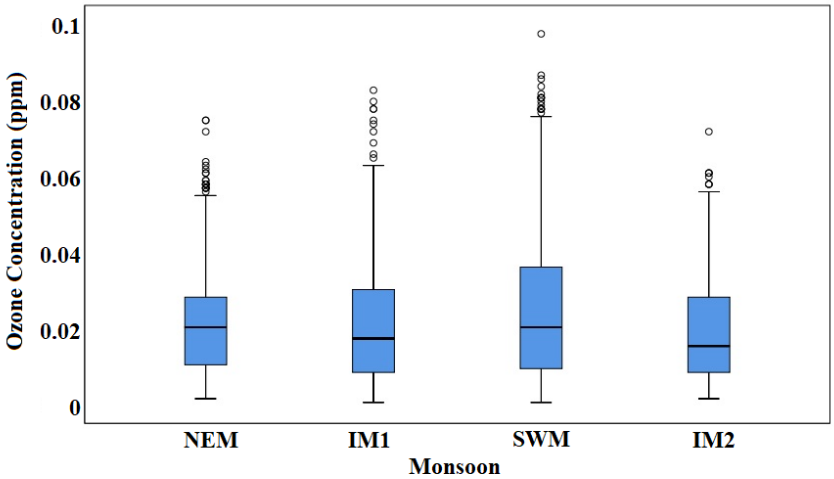


Fig. 2. The box plots analysis for Ozone (O3) concentration.

Table 2. Summary of descriptive statistics for O_3 concentration.

Descriptive Statistics	NEM (N=4847)	IM1 (N=1700)	SWM (N=5555)	IM2 (N=576)
Mean (ppm)	0.020	0.021	0.024	0.019
Median (ppm)	0.020	0.017	0.020	0.015
Maximum (ppm)	0.075	0.083	0.098	0.072
Minimum (ppm)	0.001	0	0	0.001
Std Dev (ppm)	0.012	0.015	0.017	0.014
Skewness	0.573	0.848	0.667	0.974
Interquartile range (ppm)	0.020	0.022	0.027	0.020

The Spearman bivariate correlation analysis between O_3 with meteorological factors and O_3 precursors was tabulated in Table 3. The WS (r =0.784, $p < 0.01$), T ($r = 0.713$, $p < 0.01$) and RH ($r = -0.707$, $p < 0.01$) show a strong correlation with O_3 which WS and T were positively correlated to O_3 concentration, while the RH was negatively correlated. The higher wind speed triggers the dispersion and mixing of the atmospheric pollutant that optimized the production of O_3 by its precursor [45]. Moreover, increasing the degree Celsius of temperature will provide a lower percentage of humidity and higher solar radiation which is an optimum condition of the photochemical process to occur and increase the O_3 concentration in the atmosphere [12]. Meanwhile, the O_3 precursor such as SO_2, NO_2, NO and CO show weak and positively correlated ($r = 0.289$, $p < 0.01$), ($r = 0.210$, $p < 0.01$), ($r = 0.156$, $p < 0.01$) and ($r = 0.055$, $p < 0.01$), respectively. SO_2, NO_2, NO and CO or known as ozone precursors commonly generates through the emission of the vehicle and industries [12, 46]. These O_3 precursors were generated the ozone concentration through photolysis and oxidation process with the presence of oxidant radicals compounds such as hydroperoxyl radical (HO_2), organic peroxyl radicals (RO_2) and solar radiation that can chemically be converting the O_3 precursor to O_3 in the atmosphere [47].

Table 3. Spearman Correlation of ozone between meteorological factors and its precursors.

	O_3	WS	T	RH	NO	SO_2	NO_2	CO
O_3	1	0.784**	0.713**	-0.707**	0.156**	0.289**	0.210**	0.055**
WS		1	0.669**	-0.666**	0.282**	0.233**	0.097**	-0.047**
T			1	-0.804**	0.321**	0.362**	0.330**	0.105**
RH				1	-0.231**	-0.348**	-0.210**	0.064**
NO					1	0.248**	0.421**	0.332**
SO_2						1	0.305**	0.201**
NO_2							1	0.483**
CO								1

*Note: ** Correlation is significant at the* 0.01 *level* (2-*tailed*)

MLR is adopted in this study to analyse and predict the O_3 concentration with the independent variables of O_3 precursor and meteorological parameters (Table 4). The models were developed for predicting the next hour of $O_{3,t+1}$ concentration up to $O_{3,t+3}$ concentration, as to discern the range of significant leads for four monsoon seasons. NEM, IM1, SWM and IM2 models show that $O_{3,t+1}$ having a higher coefficient of determination of 0.829, 0.790, 0.810 and 0.778, respectively as compared to model $O_{3,t+2}$ and $O_{3,t+3}$ for each monsoon as tabulated in Table 4. The lowest R^2 for the monsoon seasons was the model $O_{3,t+3}$ which R^2 value is 0.452 (NEM), 0.285 (IM1), 0.343 (SWM) and 0.325 (IM2). It shows that the capability of the MLR model in predicting the ozone concentration is higher for the next hour and reduced for the next two hours. The capability of the MLR models for predicting ozone concentration is also varied between the seasons, thus having different MLR models for different seasons is important as each seasons have different underlying characteristics. The value of R^2 has influenced the normal distribution of residual and it negatively skewed, as illustrated in Fig. 3. The lower R^2 values were affecting the normal distribution as the extent of next hour O_3 prediction and it applicable for model $O_{3,t+2}$ and $O_{3,t+3}$ in each monsoon.

Based on the previous study in Turkey, the MLR models were performed for annual and seasonal periods to comprehend the variation of O_3 concentration which having R^2 of 0.90, 0.85 and 0.92 for an annual period, cooling season and warming season, respectively [48]. The study on the development of MLR model to determine O_3 concentration variation during the daytime, nighttime and critical conversion point (CCP) time at urban areas were conducted at Shah Alam, Malaysia which having R^2 values 0.786 (daytime), 0.531(nighttime) and 0.758 (CCP time) [30]. Based on the study in Hong Kong, having four different season which is summer, monsoon, post-monsoon and winter which influence the variation of O_3 concentration [49]. The MLR model was developed at two different stations, Sham Shui Po and Tap Mun which having the same seasonal condition with the range of R^2 between 0.54 to 0.59 for Sham Shui Po, while the range for R^2 in Tap Mun is 0.54 to 0.64 [49]. Therefore, R^2 value for $O_{3,\,t+1}$ and $O_{3,\,t+2}$ models in each monsoons season are having an acceptable range based on R^2 value based on the previous studies as tabulated in Table 5.

The range of Variance Inflation Factor (VIF) for three models is 1.170 -1.991 (NEM), 1.241- 3.466 (IM1), 1.211 - 4.497 (SWM) and 1.036 -2.300 (IM2). No multi-collinearity problem among independent variables for all models as the VIF values are less than 10. The Durbin-Watson showed that the various values for all models are 0.447-1.578 throughout all monsoon seasons. It turned into showed that the version does no longer have any first-order autocorrelation problem in which the range values are within 2 [12]. Hence, based on all models in all monsoon seasons, the best fits model for each monsoon is $O_{3,\,t+1}$ model which contains the higher value of R^2, no multi-collinearity and no first-order autocorrelation problem [50].

Table 4. Summary models for O_3 concentration forecasting on different monsoon seasons.

Monsoon	Model	R^2	Range of VIF	Durbin-Watson
NEM	$O_{3,\,t+1} = 0.023 + 0.889\,O_3 + 0.204\,NO - 0.052\,NO_2 - 0.054\,WS - 0.019\,T$	0.829	1.170 - 1.956	1.578
	$O_{3,\,t+2} = 0.061 + 0.788\,O_3 + 0.345\,NO - 0.146\,NO_2 - 0.61\,T + 0.058\,WS + 0.053\,CO$	0.625	1.180 - 1.991	0.773
	$O_{3,\,t+3} = 0.111 + 0.703\,O_3 - 0.117\,T - 0.234\,NO_2 + 0.391\,NO + 0.094\,CO + 0.034\,WS$	0.452	1.180 - 1.991	0.512
IM1	$O_{3,\,t+1} = 0.125 + 0.782\,O_3 - 0.091\,H + 0.187\,NO - 0.139\,NO_2$	0.790	1.241- 3.443	1.338
	$O_{3,\,t+2} = 0.207 + 0.565\,O_3 - 0.117\,H + 0.355\,NO - 0.265\,NO_2$	0.511	1.242 - 3.443	0.668
	$O_{3,\,t+3} = 0.227 + 0.382\,O_3 - 0.094\,H + 0.434\,NO - 0.408\,NO_2 + 0.163\,CO$	0.285	1.334 - 3.466	0.447
SWM	$O_{3,\,t+1} = 0.724\,O_3 + 0.239\,T + 0.206\,NO - 0.119\,NO_2 - 0.025\,WS + 0.054\,CO - 0.035$	0.810	1.322- 4.420	1.353
	$O_{3,\,t+2} = 0.466\,O_3 + 0.365\,T + 0.384\,NO - 0.247\,NO_2 - 0.047\,WS + 0.174\,CO - 0.026$	0.561	1.322 - 4.419	0.661
	$O_{3,\,t+3} = 0.019 + 0.376\,T + 0.525\,NO - 0.362\,NO_2 + 0.258\,O_3 + 0.294\,CO - 0.061\,WS + 0.116\,SO$	0.343	1.211 - 4.497	0.452
IM2	$O_{3,\,t+1} = 0.183 + 0.702\,O_3 - 0.155\,H + 0.197\,NO$	0.778	1.036 - 2.620	1.429
	$O_{3,\,t+2} = 0.327 - 0.253\,H + 0.398\,O_3 + 0.354\,NO$	0.517	1.036 - 2.619	0.699
	$O_{3,\,t+3} = 0.389 - 2.74\,H + 0.531\,CO - 0.257\,NO_2 + 0.158\,WS$	0.325	1.855 - 2.300	0.493

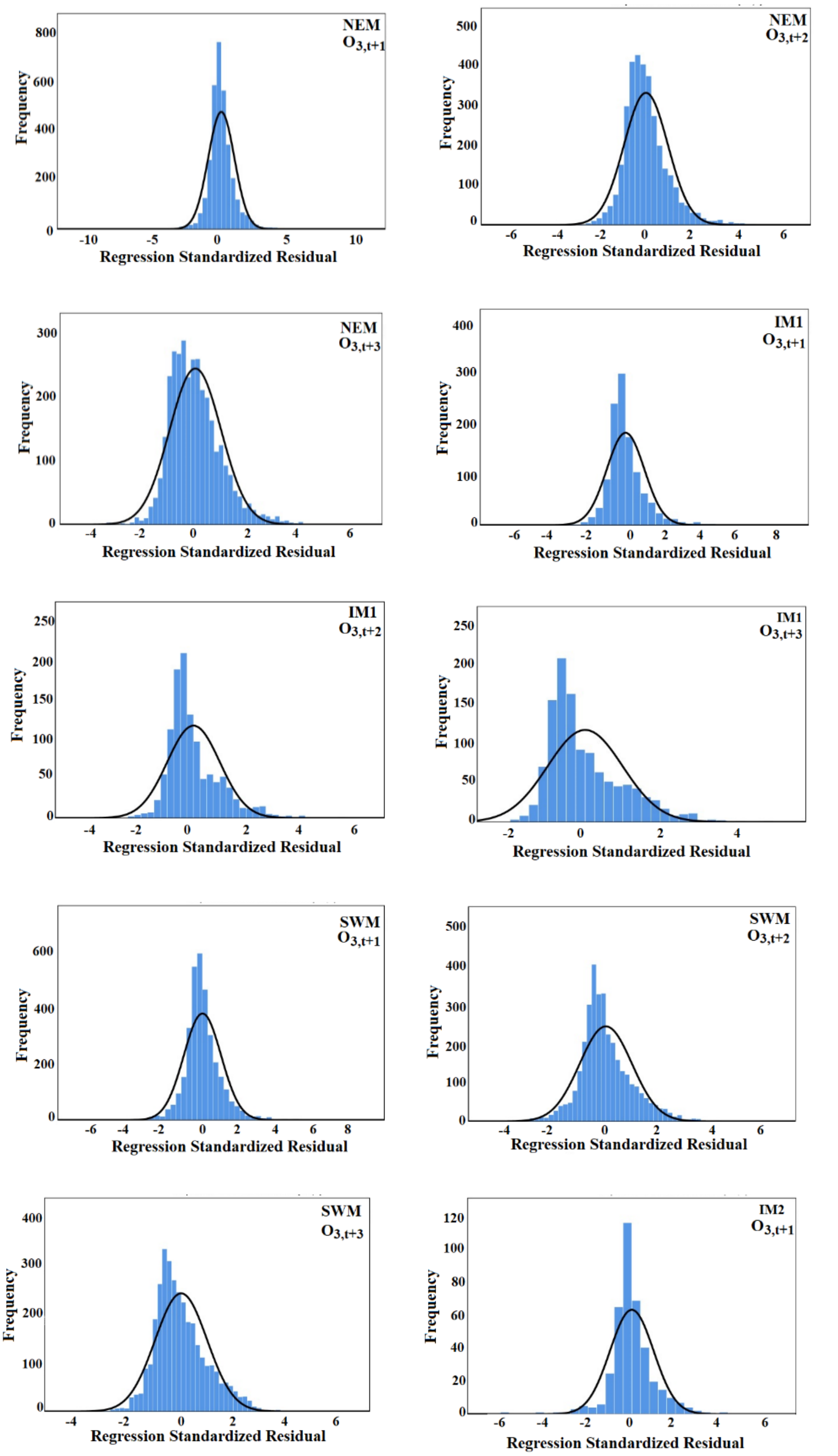
Table 5. Comparison R^2 of ozone's MLR models with similar studies.

Source	Country	Pollutant	R^2
Özbay et al. [48]	Turkey	O_3	0.85 - 0.920
Awang et al. [38]	Malaysia	O_3	0.531 - 0.786
Tong et al. [41]	Hong Kong	O_3	0.540 - 0.640

During the NEM, the significant predictor for $O_{3,t+1}$ is ozone (O_3), nitrogen oxide (NO), nitrogen dioxide (NO_2), wind speed (WS) and temperature (T). The prediction of $O_{3,t+1}$ concentration was increased by 0.889 unit when O_3 variable goes up by one unit, 0.204 unit increasing one unit of NO, 0.054 unit for the decrease in one unit of WS and 0.019 unit decreasing in one-unit T. Meanwhile, there are four predictors that influence the $O_{3,t+1}$ concentration during Inter monsoon 1. The $O_{3,t+1}$ concentration increased by 0.782 unit of one-unit O_3, decreased 0.091 unit by one unit of H, increased 0.187 unit when NO goes up by one unit and decreasing 0.139 unit as one unit of NO_2. The $O_{3,t+1}$ concentration was increasing 0.724 units, 0.239 units, 0.206 unit and 0.054 unit for one unit of O_3, T, NO, and CO. Meanwhile, $O_{3,t+1}$ concentration was decreasing 0.119 units and 0.025 units of one-unit NO_2 and WS respectively. Hence, the O_3, T, NO, CO, NO_2, and WS were identified as significant predictors during Southwest monsoon. During Inter monsoon 2, three significant predictors influence $O_{3,t+1}$ concentration. The increasing one unit of O_3 and NO will increase 0.702 and 0.197 units of $O_{3,t+1}$. The decreasing of one unit of H was decline 0.155 units of $O_{3,t+1}$ concentration. Based on this study, it shows that different ozone precursors and meteorological factors were influencing $O_{3,t+1}$ concentration according to monsoon seasons condition. O_3 and NO concentration having a positive influence on $O_{3,t+1}$ in each monsoon. In contrast, temperature and CO showed a positive relationship with $O_{3,t+1}$ concentration during SWM. The NO_2 concentration showed a negative influence during NEM, IM1 and SWM. Meanwhile, the WS give negative influence during NEM and SWM. The temperature is negatively influencing the $O_{3,t+1}$ concentration during NEM. Whilst, the humidity was found a negative relationship on $O_{3,t+1}$ concentration during IM1 and IM2.

Overall, during Southwest monsoon, the temperature has a positive influence on $O_{3,t+1}$ concentration as it introduces dry conditions which less total amount of rainfall and wet days. This dry condition promotes optimum condition for the photochemical reaction of ozone and ozone precursor to occurs which will raise the concentration of O_3 in the atmosphere [51]. The emission of NO through diesel vehicles contributes to high O_3 concentration because of the reaction of NO with oxidation hydrocarbon to form NO_2 which allowing higher ozone production. These NO predictors showed positive influences during all monsoons [52, 53]. CO is one of the combustion products from motor vehicles and industry. It was showing positive influence during SWM which involves formation and destruction reaction of O_3 by CO converting OH into the hydroperoxyl radical (HO_2) through the photolysis process [52].

Figure 3 shows the residual of predicted O_3 for all monsoon seasons, and it showed that the residuals were normally distributed. The plotted of fitted values with the predicted O_3 model's residuals were showed that the residuals are uncorrelated due to the residuals accumulated around the horizontal band and the variance is constant as presented in Fig. 4. Unfortunately, the residual started away from the horizontal band as the t increased to a second hour and third hour in each monsoon season. This condition will disturb the best fit of the models, which gives less precise O_3 concentration prediction.


NEM
$O_{3,t+1}$
NEM
$O_{3,t+2}$
NEM
$O_{3,t+3}$
IM1
$O_{3,t+1}$
IM1
$O_{3,t+2}$
IM1
$O_{3,t+3}$
SWM
$O_{3,t+1}$
SWM
$O_{3,t+2}$
SWM
$O_{3,t+3}$
IM2
$O_{3,t+1}$
Frequency
Regression Standardized Residual

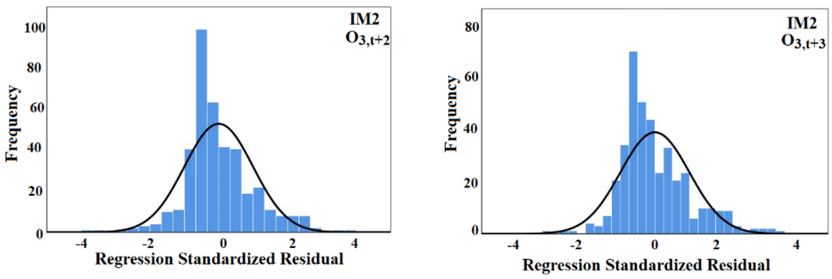
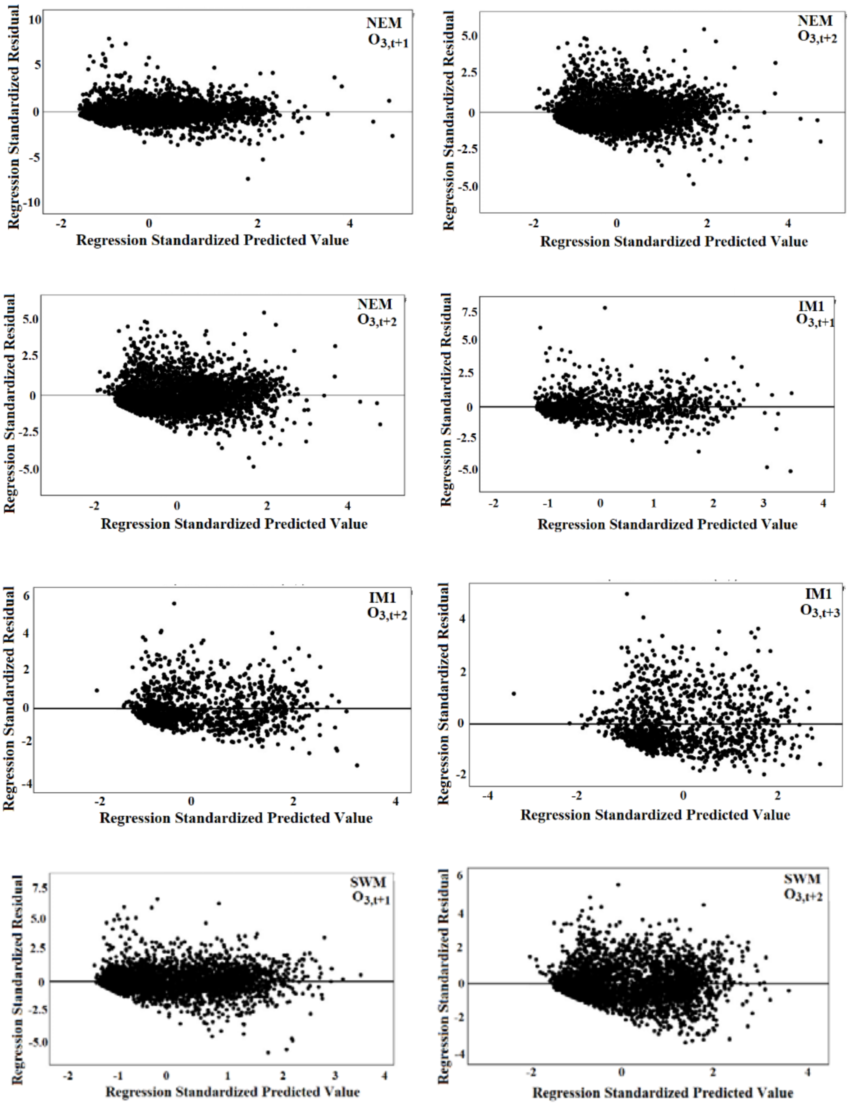


Fig. 3. Standardized residual analysis of
O_3, *t*+1, O_3, *t*+2 and O_3, *t*+3 for each monsoon seasons



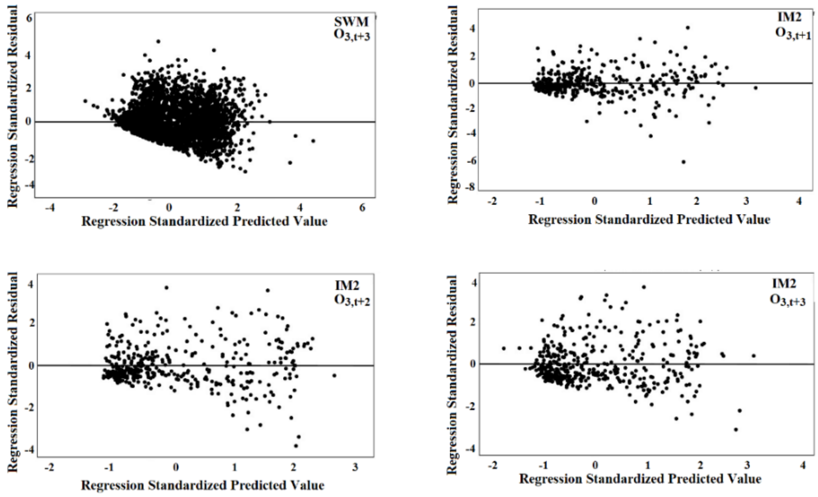
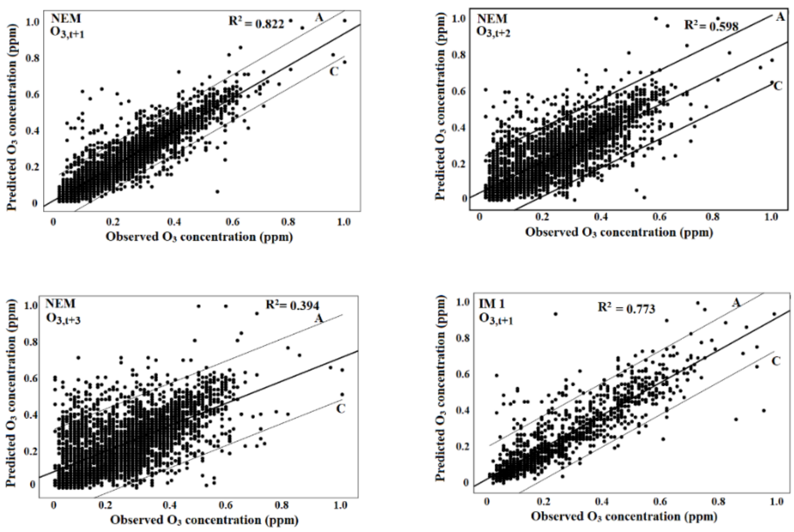


Fig. 4. Testing assumption of variance and uncorrelated with mean equal to zero

The predicted and observed data for the MLR model is depicted in Fig. 5. The MLR model of $O_{3,t+1}$ is the most acceptable prediction model as compared to model $O_{3,t+2}$ and $O_{3,t+3}$ in each monsoon. Most of the points in models $O_{3,t+1}$ for each monsoon were accumulated within a 95% confidence interval line and having higher values of R^2 linear as compared to models $O_{3,\,t+2}$ and $O_{3,\,t+3}$. The accuracy of models $O_{3,t+2}$ and $O_{3,t+3}$ in each monsoon become over predicted as more points disperse away from the confidence interval line and the increase of prediction hour will increase the root mean square error, mean absolute error and decrease the R^2 values which affect the performance of the models [54, 55]. Lines A and C were drawn as an upper and lower of the 95% confidence limit for MLR models. The range of R^2 linear is between 0.181 - 0.843.



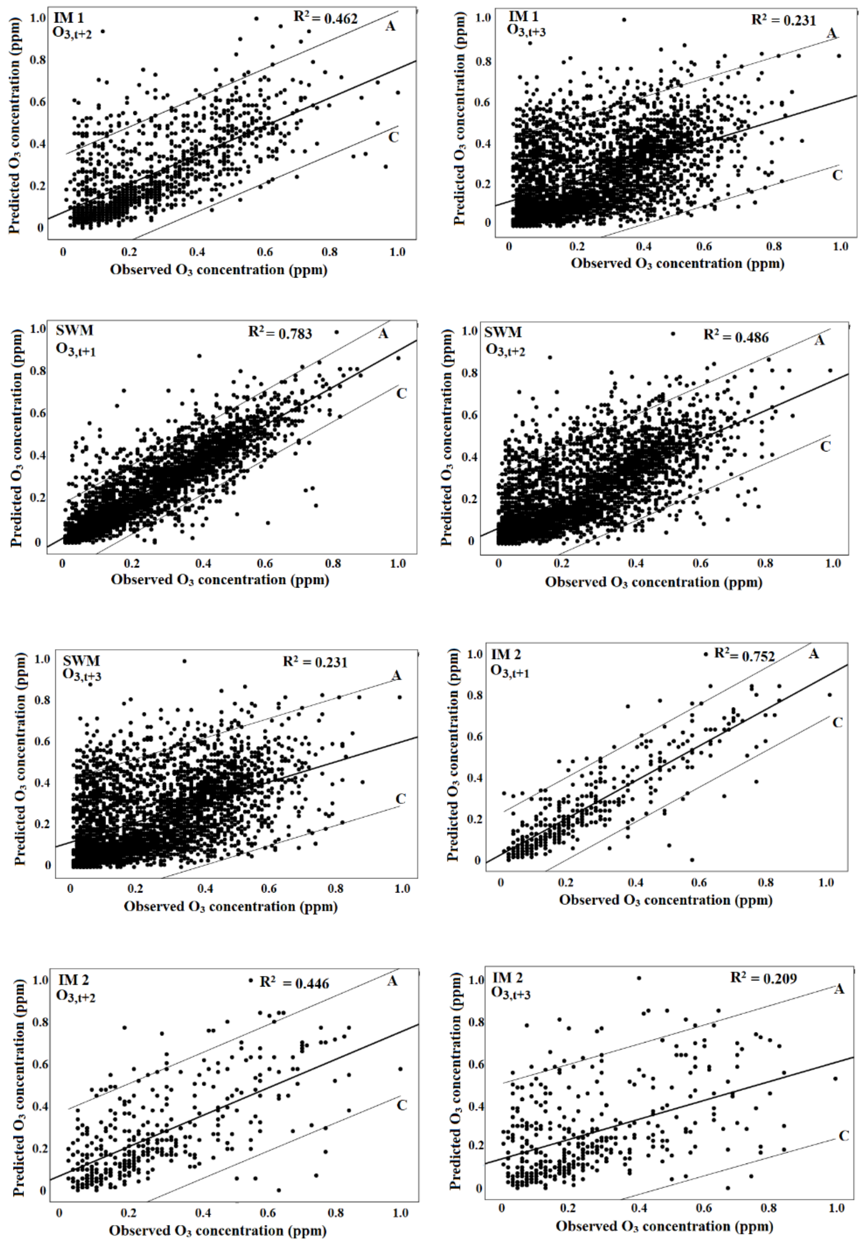


Fig. 5. Scatter plot of predicted O_3 concentration (ppm) |against observed O_3 concentration (ppm).

4. Conclusions

In conclusion, among the four monsoon seasons, we found that the concentration of O_3 during SWM is higher (mean = 0.024 ppm) and the lowest concentration of O_3 during IM2 (mean = 0.019 ppm). The SWM has the higher concentration of O_3 because of the warm and dry condition that promotes a photochemical reaction to occur. The relationship between ozone, ozone precursors and meteorological parameters was conducted via correlation analysis. Spearman correlation analysis shows that the wind speed (WS) and ambient temperature (T) having strong positively correlation (r = 0.784, p < 0.01) and (r = 0.713, p < 0.01), respectively while relative humidity (RH) was vice-versa (r = 0.707, p < 0.01) to O_3 concentration. The SO_2 (r = 0.289, p < 0.01), NO_2 (r = 0.210, p < 0.01), NO (r = 0.156, p < 0.01) and CO (r = 0.055, p < 0.01) were weakly and positively correlated with the rise of O_3 concentration. Models were developed and validated for next one to three hours for ozone concentration forecasting using stepwise MLR. The best fitted MLR model for O_3 prediction during the monsoon seasons is $O_{3,t+1}$ model, compared to $O_{3,t+2}$ and $O_{3,t+3}$ models which it showed higher R^2 values 0.829 (NEM), 0.790 (IM1), 0.810 (SWM) and 0.778 (IM2). The validation of best MLR models in each monsoon's seasons were having R^2 values of 0.811($O_{3,t+1}$; NEM), 0.754 ($O_{3,t+1}$; IM2), 0.775 ($O_{3,t+1}$; SWM) and 0.707 ($O_{3,t+1}$; IM2). The development of this model is to help the responsible bodies especially the authorities and policy makers to plan the mitigation measure for public health to improve the air quality for future projection.

Acknowledgements

This study was funded by the Fundamental Research Grant Scheme by the Malaysian Ministry of Education (FRGS/1/2019/TK10/UMT/02/4) (VOT: 59606) and Research Management & Innovation Centre, Universiti Malaysia Terengganu, as well as the Malaysian Department of Environment for providing the data.

Nomenclatures

R^2	Correlation coefficient
$O_{3,t+1}$	Next hour of predicted ozone
$O_{3,t+2}$	Next two hours of predicted ozone
$O_{3,t+3}$	Next three hours of predicted zone

Greek Symbols

ε	Stochastic Error

Abbreviations

AQMS	Air Quality Monitoring Station
ASMA	Alam Sekitar Malaysia
CAQM	Continuous Air Quality Monitoring
CCP	Critical conversion point
CO	Carbon monoxide
CO_2	Carbon dioxide
DOE	Department of Environment
DW	Durbin Watson
HO_2	Hydroperoxyl radical
IM1	Inter monsoon one
IM2	Inter monsoon two
MLR	Multiple linear regression
MMD	Malaysian Meteorological Department
NAAQS	New ambient air quality standard
NEM	Northeast monsoon

NEM	North east monsoon
NO	Nitrogen oxide
NO_2	Nitrogen dioxide
O_3	Ozone
RO_2	Organic peroxyl radicals
SO_2	Sulphur dioxide
SPSS	Statistical Package for Social Science
SWM	South west monsoon
T	Temperature
VIF	Variance of inflation factor
WS	Wind speed

References

1. Han, L.; Zhou, W.; and Li, W. (2018). Growing urbanization and the impact on fine particulate matter ($PM_{2.5}$) dynamics. *Sustainability*, 10(6), 1696.
2. Cho, H.S.; and Choi, M.J. (2014). Effects of compact urban development on air pollution: Empirical evidence from Korea. *Sustainability* (*Switzerland*), 6(9), 5968-5982.
3. Abdullah, S.; Ismail, M.; and Fong, S.Y. (2017). Multiple Linear Regression (MLR) models for long term PM_{10} concentration forecasting during different monsoon seasons. *Journal of Sustainability Science and Management*, 12(1), 60-69.
4. Osborne, S.; Pandey, D.; Mills, G.; Hayes, F.; Harmens, H.; Gillies, D.; Buker, B.; and Emberson, L. (2019). New insights into leaf physiological responses to ozone for use in crop Modelling. *Plants*, 8(4), 84.
5. Huang, D.; Li, Q.; Wang, X.; Li, G.; Sun, L.; He, B.; Zhang, L.; and Zhang, C. (2018). Characteristics and trends of ambient ozone and nitrogen oxides at urban, suburban, and rural sites from 2011 to 2017 in Shenzhen, China. *Sustainability* (*Switzerland*), 10(12), 4530
6. Lee, K.J.; Kahng, H.; Kim, S.B.; and Park, S.K. (2018). Improving environmental sustainability by characterizing spatial and temporal concentrations of ozone. *Sustainability* (*Switzerland*), 10(12), 4551
7. Pawlak, I.; and Jarosławski, J. (2019). Forecasting of surface ozone concentration by using artificial neural networks in rural and urban areas in central Poland. *Atmosphere*, 10(2), 52.
8. Marathe, S.A.; and Murthy, S. (2015). Seasonal variation in surface ozone concentrations, meteorology and primary pollutants in Coastal Mega City of Mumbai, India. *Journal of Climatology & Weather Forecasting*, 03(03), 1000149.
9. Young, P.J.; Naik, V.; Fiore, A.M.; Gaudel, A.; Guo, J.; Lin, M.Y.; Neu, J.; Parrish, D.D.; Rieder, H.E.; Schnell, J.L.; Tilmes, S.; Wild, O.; Zhang, L.; Brandt, J.; Delcloo, A.; Doherty, R.M.; Geels, C.; Hegglin, M.I.; Hu, L.; Im, U.; Kumar, R.; Luhar, A.; Murray, L.; Plummer, D.; Rodriguez, J.; Saiz-Lopez, A.; Schultz, M.G.; Woodhouse, M.; Zeng, G.; and Ziemke, J. (2018). Tropospheric ozone assessment report: Assessment of global-scale model performance for global and regional ozone distributions, variability, and trends. *Elementa*, 6, 0-81.

10. Lu, X.; Zhang, L.; and Shen, L. (2019). Meteorology and climate influences on tropospheric ozone: a review of natural sources, chemistry, and transport patterns. *Current Pollution Reports*, 5, 238-260.
11. Banerjee, A.; Archibald, A.T.; Maycock, A.C.; Telford, P.; Abraham, N.L.; Yang, X.; Braesicke, P.; and Pyle, J.A. (2014). Lightning NOx, a key chemistry - climate interaction: impacts of future climate change and consequences for tropospheric oxidising capacity. *Atmospheric Chemistry and Physics*, 14(18), 9871-9881.
12. Abdullah, S.; Ismail, M.; Ahmed, A.N.; and Abdullah, A.M. (2019). Forecasting particulate matter concentration using linear and non-linear approaches for air quality decision support. *Atmosphere*, 10, 667.
13. Rajab, J.M.; MatJafri, M.Z. and Lim, H.S. (2013). Combining multiple regression and principal component analysis for accurate predictions for column ozone in Peninsular Malaysia. *Atmospheric Environment*, 71, 36-43.
14. Nuvolone, D.; Petri, D.; and Voller, F. (2018). The effects of ozone on human health. *Environmental Science and Pollution Research International*, 25(9), 8074-8088.
15. Mabahwi, N.A.; Leh, O.L.H.; and Omar, D. (2015). Urban air quality and human health effects in Selangor, Malaysia. *Procedia - Social and Behavioral Sciences*, 170, 282-291.
16. Portmann, R.W.; Daniel, J.S.; and Ravishankara, A.R. (2012). Stratospheric ozone depletion due to nitrous oxide: Influences of other gases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1593), 1256-1264.
17. Tagaris, E.; Manomaiphiboon, K.; Liao, K.J.; Ruby Leung, L.; Woo, J.H.; He, S.; Amar, P.; and Russell, A.G. (2007). Impacts of global climate change and emissions on regional ozone and fine particulate matter concentrations over the United States. *Journal of Geophysical Research Atmospheres*, 112(14), D14312.
18. Ismail, M.; Suroto, S.; and Abdullah, S. (2015). Response of Malaysian local rice cultivars induced by elevated ozone stress. *Environment Asia*, 1(1), 86-93.
19. Sharma, A.; Mandal, T.K.; Sharma, S.K.; Shukla, D.K.; and Singh, S. (2017). Relationships of surface ozone with its precursors, particulate matter and meteorology over Delhi. *Journal of Atmospheric Chemistry*, 74, 451-474.
20. Liu, P.; Song, H.; Wang, T.; Wang, F.; Li, X.; Miao, C.; and Zhao, H. (2020). Effects of meteorological conditions and anthropogenic precursors on ground-level ozone concentrations in Chinese cities. *Environmental Pollution*, 262, 114366.
21. Wang, Q.; Han, Z.; Wang, T.; and Zhang, R. (2008). Impacts of biogenic emissions of VOC and NO_x on tropospheric ozone during summertime in eastern China. *Science of the Total Environment*, 395(1), 41-49.
22. Wang, Z.; Li, Y.; Chen, T.; Zhang, D.; Sun, F.; Wei, Q.; Dong, X.; Sun, R.; Huan, N.; and Pan, L. (2015). Ground-level ozone in urban Beijing over a 1-year period: Temporal variations and relationship to atmospheric oxidation. *Atmospheric Research*, 164-165, 110-117.
23. Lei, M.T.; Monjardino, J.; Mendes, L.; Goncalves, D.; and Ferreira, F. (2019). Macao air quality forecast using statistical methods. *Air Quality, Atmosphere & Health*, 12, 1049-1057.
24. Ahmadi, M.; Huang, Y.; and John, K. (2017) Application of spatio-temporal clustering for predicting ground-level ozone pollution. In: Griffith D., Chun

Y., Dean D. (eds) *Advances in Geocomputation. Advances in Geographic Information Science.* Springer, Cham.

25. Sousa, S.I.V.; Martins, F.G.; Pereira, M.C.; and Alvim-Ferraz, M.C.M. (2006). Prediction of ozone concentrations in Oporto city with statistical approaches. *Chemosphere*, 64(7), 1141-1149.
26. Zhao, W.; Fun, S.; Guo, H.; Gao, B.; Sun, J.; and Chen, L. (2016). Assessing the impact of local meteorological variables on surface ozone in Hong Kong during 2000-2015 using quantile and multiple line regression models. *Atmospheric Environment*, 144, 182-193.
27. Abdul-Wahab, S.A.; Bakheit, C.S.; and Al-Alawi, S.M. (2005). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling and Software*, 20(10), 1263-1271.
28. Department of Statistics (DOS). (2014). Population Distribution by Local Authority Areas ad Mukims, Putrajaya: Department of Statistics Malaysia. Retrieved April 17, 2019, from https://www.dosm.gov.my
29. Malaysian Meteorological Department. (2019). Malaysia Climate. Retrieved April 17, 2019, from http://www/met.gov.my.
30. Awang, N.R.; Elbayoumi, M.; Ramli, N.A.; and Yahaya, A.S. (2016). Diurnal variations of ground-level ozone in three port cities in Malaysia. *Air Quality, Atmosphere and Health*, 9(1), 25-39.
31. Latif, M.T.; Dominick, D.; Ahamad, F.; Khan, M.F.; Juneng, L.; Hamzah, F. M.; and Nadzir, M.S.M. (2014). Long term assessment of air quality from a background station on the Malaysian Peninsula. *Science of the Total Environment*, 482-483(1), 336-348.
32. Banan, N.; Latif, M.T.; Juneng, L.; and Ahamad, F. (2013). Characteristics of surface ozone concentrations at stations with different backgrounds in the Malaysian Peninsula. *Aerosol and Air Quality Research*, 13(3), 1090-1106.
33. Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, 64(5), 402-406.
34. Wang, Z.; Lv, J.; Tan, Y.; Guo, M.; Gu, Y.; Xu, S.; and Zhou, Y. (2019). Temporospatial variations and Spearman correlation analysis of ozone concentrations to nitrogen dioxide, sulfur dioxide, particulate matters and carbon monoxide in ambient air, China. *Atmospheric Pollution Research*, 10(4), 1203-1210
35. Abdullah, S.; Ismail, M.; Samat, N.N.A.; and Ahmed, A.N. (2018). Modelling particulate matter (PM_{10}) concentration in industrialized area: A comparative study of linear and nonlinear algorithms. *ARPN Journal of Engineering and Applied Sciences*, 13(20), 8227-8235.
36. Roy, K.; and Ambure, P. (2016). The “double cross-validation” software tool for MLR QSAR model development. *Chemometrics and Intelligent Laboratory Systems*, 159, 108-126.
37. Ul-Saufie, A.Z.; Yahaya, A.S.; Ramli, N.A.; and Hamid, H.A. (2011). Comparison between multiple linear regression and feed forward back propagation neural network models for predicting PM_{10} concentration level based on gaseous and meteorological parameters. *International Journal of Applied Science and Technology*, 1(4), 42-49.

38. Awang, N.R.; Ramli, N.A.; Yahaya, A.S.; and Elbayoumi, M. (2015). Multivariate methods to predict ground level ozone during daytime, nighttime, and critical conversion time in urban areas. *Atmospheric Pollution Research*, 6(5), 726-734.
39. Department of Environment, Malaysia. (2013). Malaysia Environmental Quality Report 2014. Retrieved April 17, 2019, from https://www.doe.gov.my/portalv1/en/
40. Hashim, N.I.M.; and Noor, N.M. (2017). Variations of ground-level ozone concentration in Malaysia: A case study in West Coast of Peninsular Malaysia. *MATEC Web of Conferences*, 97, 1-6.
41. Tong, L.; Zhang, H.; Yu, J.; He, M.; Xu, N.; Zhang, J.; Qian, F.; Feng, J.; and Xiao, H. (2017). Characteristics of surface ozone and nitrogen oxides at urban, suburban and rural sites in Ningbo, China. *Atmospheric Research*, 187, 57-68.
42. Lanzafame, R.; Scandura, P.F.; Famoso, F.; Monforte, P.; and Oliveri, C. (2014). Air quality data for Catania: Analysis and investigation case study 2010-2011. *Energy Procedia*, 45(2), 681-690.
43. Hung-Lung, C.; Jiun-Horng, T.; Shih-Yu, C.; Kuo-Hsiung, L.; and Sen-Yi, M. (2007). VOC concentration profiles in an ozone non-attainment area: A case study in an urban and industrial complex metroplex in southern Taiwan. *Atmospheric Environment*, 41(9), 1848-1860.
44. Suhaila, J.; Deni, S.M.; Wan Zin, W.Z.; and Jemain, A.A. (2010). Trends in Peninsular Malaysia rainfall data during the southwest monsoon and northeast monsoon seasons: 1975-2004. *Sains Malaysiana*, 39(4), 533-542.
45. Agudelo-Castaneda, D.M.; Teixeira, E.C. and Pereira, F.N. (2014). Time-series analysis of surface ozone and nitrogen oxides concentrations in an urban area at Brazil. *Atmospheric Pollution Research*, 5(3), 411-420.
46. He, H.; and Lu, W.Z. (2012). Decomposition of pollution contributors to urban ozone levels concerning regional and local scales. *Building and Environment*, 49(1), 97-103.
47. Li, B.; Zhao, J.; and Lu, J. (2015). Numerical study of the simultaneous oxidation of NO and SO_2 by ozone. *International Journal of Environmental Research and Public Health*, 12(2), 1595-1611.
48. Özbay, B.; Keskin, G.A.; Doğruparmak, Ş.Ç.; and Ayberk, S. (2011). Multivariate methods for ground-level ozone modeling. *Atmospheric Research*, 102(1-2), 57-65.
49. Zhang, J.; and Ding, W. (2017). Prediction of air pollutants concentration based on an extreme learning machine: The case of Hong Kong. *International Journal of Environmental Research and Public Health*, 14(2), 1-19.
50. Gagné, P.; and Dayton, C.M. (2002). Early scholars: Best regression model using information criteria. *Journal of Modern Applied Statistical Methods*, 1(2), 479-488.
51. Reddy, B.S.K.; Reddy, L.S.S.; Cao, J.J.; Kumar, K.R.; Balakrishnaiah, G.; Gopal, K.R.; Reddy, R.R.; Narasimhulu, K.; Lal, S.; and Ahammed, Y.N. (2011). Simultaneous measurements of surface ozone at two sites over the southern Asia: A comparative study. *Aerosol and Air Quality Research*, 11(7), 895-902.

52. Teixeira, E.C.; de Santana, E.R.; Wiegand, F.; and Fachel, J. (2009) Measurement of surface ozone and its precursors in an urban area in South Brazil. *Atmospheric Environment*, 43(13), 2213-2220.
53. Blanchard, C.L.; and Tanenbaum, S.J. (2003). Differences between weekday and weekend air pollutant levels in Southern California. *Journal of the Air and Waste Management Association*, 53(7), 816-828.
54. Zeyu, W.; Yueren, W.; Ruochen, Z.; Srinivasan, R.S.; and Ahrentzen, S. (2018). Random forest based hourly building energy prediction. *Energy and Buildings,* 171, 11-25.
55. Zounemat-Kermani, M. (2012). Hourly predictive Levenberg-Marquardt ANN and multi linear regression models for predicting of dew point temperature. *Meteorology and Atmospheric Physics,* 117(3-4), 181-192.