International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-10, August 2019

Development of Ozone Prediction Model in Urban Area

Samsuri Abdullah, Najihah Husna Ahmad Nasir, Marzuki Ismail, Ali Najah Ahmed, Mohammad Nor Khasbi Jarkoni

Abstract**:** ***One of the main challenges for countries in tropical area such as Malaysia is the high concentration of ozone (O_3) caused by elevated levels of anthropogenic and natural ozone precursors. In this study, variation of O_3 concentrations in urban area (Klang) was investigated using data covering three-year period (2012–2015) on hourly basis. Result shows that the diurnal cycle of ozone concentration has a mid-day peak (1400hrs) while lower concentration occurs at night time (2100hrs) as it titrates nitrogen dioxide (NO_2). There exists statistically significant difference ($p<0.05$) of O_3 concentration at study areas. Moderate Spearman correlation coefficient was evaluated between O_3 and NO_2 ($r=0.45$, $p<0.05$). Multiple linear regression (MLR) model was developed and signifies that nitrogen oxides (NO), relative humidity (RH), NO_2, carbon monoxide (CO), wind speed (WS), temperature (T) and sulphur dioxide (SO_2) are the significant predictors for O_3 concentration. This study suggests that the emission of O_3 precursors, particularly NOx from motor vehicles, needs to be controlled to reduce the incidence of high O_3 levels in Malaysia.***

Keywords**:** ***Klang, multiple linear regression, ozone, spearman correlation, urban area.***

I. INTRODUCTION

Ground-level ozone (O_3) is currently notified as one of the major pollutants that being inhaled in several cities and countries all over the world, including Malaysia. Factors including the deposition of photochemistry, process of transporting the pollutants locally and globally creating different variations of O_3 at different locations [1]. The variations of O_3 concentrations also being influence by meteorological parameters [2]. Nitrogen oxides (NO_x) and volatile organic compound (VOC) are known to be impacted by meteorological parameters such as sunlight, cloud cover and relative humidity [3]. In troposphere, O_3 played an important element for the chemical composition and tropospheric chemistry processes. O_3 is created by high energetic ultraviolet (UV) radiation and chemical reaction, hence, it directly affecting human health [4]-[5] such as lung tissue damage, climate problem, and vegetation [6]. Multiple Linear Regression (MLR) is widely used for prediction based on several predictors [7]-[8]. MLR analysis is simple and easy computation lead to better and widely used for understanding underlying influencing factors of O_3 variation by mathematical modeling [9]. Future O_3 concentration is important for prediction so as the relevant authorities can propose suitable actions for improving the air quality at specific location and can be used for the precautionary measures. This study emphasized on the development of ozone prediction model at urban area.

Manuscript published on 30 August 2019.
***Correspondence Author(s)**
Samsuri Abdullah*, Faculty of Ocean Engineering Technology and Informatics, University Malaysia Terengganu, 21030, Kuala Nerus, Terengganu, Malaysia; Air Quality and Environment Research Group, University Malaysia Terengganu, 21030, Kuala Nerus, Terengganu, Malaysia. Email: samsuri@umt.edu.my
Najihah Husna Ahmad Nasir, Faculty of Ocean Engineering Technology and Informatics, University Malaysia Terengganu, 21030, Kuala Nerus, Terengganu, Malaysia. Email: najihahhusna21@gmail.com
Marzuki Ismail, Faculty of Science and Marine Environment, University Malaysia Terengganu, 21030, Kuala Nerus, Terengganu, Malaysia; Air Quality and Environment Research Group, University Malaysia Terengganu, 21030, Kuala Nerus, Terengganu, Malaysia. Email: marzuki@umt.edu.my
Ali Najah Ahmed, Faculty of Engineering, Universiti Tenaga Nasional, 43650 Bangi, Selangor, Malaysia; Institute of Engineering Infrastructures, Universiti Tenaga Nasional, 43650 Bangi, Selangor, Malaysia. Email: Mahfoodh@uniten.edu.my
Mohammad Nor Khasbi Jarkoni, Faculty of Ocean Engineering Technology and Informatics, University Malaysia Terengganu, 21030, Kuala Nerus, Terengganu, Malaysia. Email: khasbijarkoni@gmail.com

II. MATERIALS AND METHODS

An urban area site was chosen in this study, perfectly located at Klang, Selangor (Lat 3° 0' 35.9784" N; Long 101° 24' 30.1464" E) in the West of Peninsular Malaysia (Figure 1). It was recorded that the study area has a population of 6.6 million in 2010 and a high number of motor vehicles with rapid industrialization and urbanization [10].

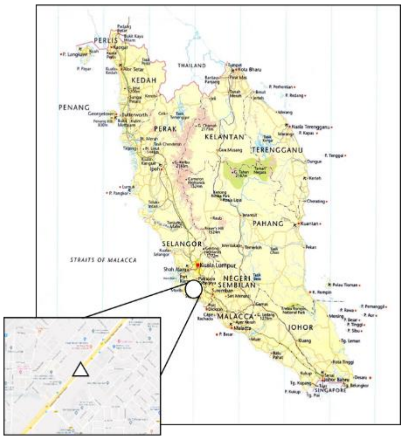

Fig. 1. Location of study area

The dataset is acquired from DOE, Malaysia for a period of year 2012-2015. Parameters included are ozone concentration, nitrogen oxides (NO_x), VOC, wind speed, temperature and relative humidity.

Due to several factors, the value in the dataset might be missing [11]. The dataset (N=10825) is divided into two parts: 70% for development and 30% for validation of the MLR model [12]. In order to reduce the bias, all missing data were removed from this study [13]. The next hours of ozone concentration were modeled using MLR in this study. This MLR technique has been verified its capability to predict the short-term ozone concentration [14], fitting the inputs in the SPSS version 25 using the stepwise method [15]. The significant predictors were evaluated using this method which the models with many significant predictors increase the R^2 value. The regression model can be written as [16]:

$$Y_i = \beta_o + \beta_i x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i \quad \text{with } i = 1, \ldots, n \tag{1}$$

Where b_i are the regression coefficients, x_i are independent variables and ε is a stochastic error associated with the regression.

The min-max technique of data normalization is applied in this study as the parameters consist of different units. Once applied, the dataset having value in the range 0 to 1 [0 1]. This technique does not introduce bias and suitable for computation carried by MLR model for better output. There are several assumptions in the development of MLR model: multi-collinearity, autocorrelation, normal distribution, and constant variance. Variable of Inflation Factor (VIF) is used to test the assumption of multi-collinearity [17]. No multi-collinearity problem if the VIF value is less than 10. Durbin-Watson Statistics is evaluated to test the assumption of autocorrelation in the developed model. The value must be within the range 0 to 4, with 2 meaning of uncorrelated. Normal distribution of the residuals of dataset and constant variance are verified graphical analysis of normal curve and scatter plot. The degree of model's development is measured by using the coefficient of determination (R^2) as follows:

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

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.

III. RESULTS AND DISCUSSION

Hourly trend of ozone concentration through boxplot is shown in Figure 2. Boxplot is essential when displaying the trend of atmospheric pollutants as more information can be gathered from one boxplot including mean, median, quartiles, minimum, and maximum concentration [18]. The diurnal trend of ozone concentration shows maximum concentration during the mid-day of around 1400 hrs, while the minimum concentration occurs during the night time around 2100 hrs due to the nitrogen dioxides titration. High ozone concentration was found during the mid-day due to the reaction of ozone precursors with the present of sunlight. During the day, from 1000hrs to 1400hrs, the presence of solar radiation induces the increment in the ozone concentration, during the nighttime, the concentration of ozone is decreasing as the absence of solar radiation for photochemical reaction. Both ozone and nitrogen dioxide are related to each other. It can be seen that the increase of nitrogen dioxide concentration will decrease the ozone concentration. Secondary pollutants such as ozone and nitrogen dioxide are formed via complex reaction. The development of ozone is induced by nitrogen dioxide, which in the daytime, the nitrogen dioxide is converted to nitrogen oxide through a photolysis process [19]. There is still existed of nitrogen dioxide concentration during the nighttime. This is due to the temperature inversion, whereby the pollutants such as nitrogen monoxide and nitrogen dioxide emitted during the day will keep under this inversion layer, making the oxides of nitrogen increase [20]. The emissions of oxides of nitrogen basically from the motor vehicles might induce the ozone concentration at study area with urban background. High skewness value of 1.219 represent the presence of outliers (Q3 + 1.5SD) and extreme values (Q3 + 3SD) (Table I). This event can be relating with high intensity of solar radiation which stimulates the ozone formation.

The skewness value (1.219) of dataset is not reaching zero. Thus, the dataset has non-normal distribution (non-parametric), and Spearman correlation analysis is deemed suitable for the dataset [21]. The correlation analysis of ozone, precursors of ozone, and meteorological parameters is analyzed using Statistical Packages for Social Sciences (SPSS®) version 25 and tabulated in Table II. Moderate Spearman correlation coefficient was evaluated between O_3 and NO_2 at the study area (r = -0.454, p<0.01). Strong correlation was found between wind speed with temperature (r = 0.670, p<0.01), humidity (r = -0.545, p<0.01), nitrogen dioxide (r = -0.507, p<0.01), ozone (r = 0.701, p<0.01) and carbon monoxide (r= -0.593, p<0.01). Strong correlation was also evaluated between temperature and relative humidity (r = -0.874, p<0.01) and ozone (r = 0.724, p<0.01). These indicate that Klang is temperature-sensitive and it is related to photochemical reaction to O_3 formation. Relative humidity and ozone showed strong correlation (r = -0.685, p<0.01). Nitrogen oxides showed strongly correlated with ozone (r = -0.685, p<0.01) and carbon monoxide (r = 0.541, p<0.01). Nitrogen dioxides has strong correlation with carbon monoxide (r = 0.589, p<0.01). Moderate correlation is evaluated between wind speed and nitrogen oxides (r = 0.391, p<0.01). Temperature is moderately correlated with nitrogen oxides (r = -0.369, p<0.01), nitrogen dioxides (r = -0.415, p<0.05) and carbon monoxide (r = -0.398, p<0.01). It indicates that nitrogen start to decrease in value when the presence of sunlight with high of temperature produces. Ozone also show moderate correlation with nitrogen dioxides (r = -0.454, p<0.01) and carbon monoxide (r = -0.483, p<0.01).

Retrieval Number: J11270881019/19©BEIESP
DOI: 10.35940/ijitee.J1127.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

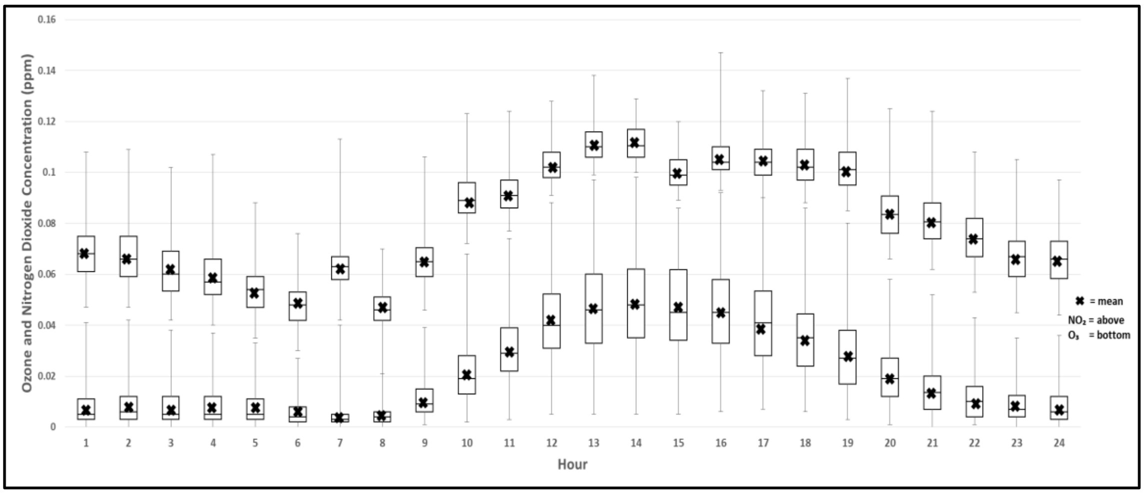

Fig. 2. Hourly trend of Ozone(O_3) titrate with Nitrogen Dioxide(NO_2)

Table- I: Descriptive Statistics for O3 concentration

Descriptive Statistics	**Klang (N=10825)**
Mean (ppm)	0.0201
Median (ppm)	0.013
Std Dev (ppm)	0.0193
Skewness	1.219
Min (ppm)	0
Max (ppm)	0.098

Table- II: Summary of Spearman Correlation Analysis (r-value) between O_3 concentration and meteorological factors

Factor	**W.Speed**	**Temp**	**RH**	**NO**	**SO_2**	**NO_2**	**O_3**	**CO**
W. Speed	1							
Temp	.670**	1						
Humidity	-.545**	-.874**	1					
NO	-.391**	-.369**	.336**	1				
SO_2	-.101**	.147**	-.088**	.286**	1			
NO_2	-.507**	-.415**	.342**	.413**	.221**	1		
O_3	.701**	.724**	-.685**	-.574**	-.030**	-.454**	1	
CO	-.593**	-.398**	.399**	.541**	.254**	.589**	-.483**	1

Weak correlation indicated that site selection parameter is not sensitive between both parameters. Weak correlations were recorded in sulphur dioxide with wind speed (r = -0.101, p<0.01), temperature (r = 0.147, p<0.01), relative humidity (r = -0.088, p<0.01), nitrogen oxides (r = 0.286, p<0.01), nitrogen dioxides (r = 0.221, p<0.01), ozone (r = -0.030, p<0.01) and carbon monoxide (r = 0.254, p<0.01).

The analysis of the air quality and meteorological dataset is continued by applying Multiple Linear Regression (MLR) model. The predicted coefficient of determination is R^2=0.810.

Retrieval Number: J11270881019/19©BEIESP
DOI: 10.35940/ijitee.J1127.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The multiple linear regression models were developed and the models summary is depicted in Table III. It was found that, the significant predictors were $O_{3,t\text{-}1}$, relative humidity, wind speed, sulphur dioxide, nitrogen oxides, nitrogen dioxides and temperature. Ozone concentrations increased by 0.681 unit when $O_{3\ t\text{-}1}$ variable goes up by one unit, 0.118 unit is decreasing one unit of relative humidity, decreased of one unit of sulphur dioxide as much as 0.106, while 0.018 decreased in nitrogen dioxides in one unit, 0.027 unit for the decreased in one unit of nitrogen oxides, 0.079 unit when wind speed increased by one unit and 0.068 change in O_3 concentrations when there is an increase in one unit of temperature. $O_{3,t\text{-}1}$, temperature and wind speed have positive influence on O_3 concentrations, while relative humidity, nitrogen dioxides, carbon monoxide, nitrogen oxides, and sulphur dioxide have negative influence on O_3 concentrations. High temperature in Malaysia (tropical) induces the formation of ozone in the atmosphere [22]. Significant factor of temperature basically helps in circulate the wind pattern, thus dilute the ozone at a particular location.

Multi-collinearity is checked by VIF values. No multi-collinearity problem spotted as the range is lower than 10. No autocorrelation problem as the Durbin Watson statistics value has value in the range of 0 and 4. Figure 3 and 4 show the normally distributed residuals and constant variance respected to zero mean, respectively. The developed MLR model is validated through the scatter plot as shown in Figure 5. Upper and lower line showing the upper bound and lower bound of 95% confidence interval of dataset. Most of the data are within the best fitted line in the middle, and the accuracy is acceptable ($R^2 = 0.797$).

Table- III: Summary Model for Ozone Forecasting at Study Area

Model	R^2	Range VIF
$O_3 = 0.088 + 0.681O_{3,\ t\text{-}1} + 0.068T + 0.079WS - 0.118RH - 0.106SO_2 - 0.027NO - 0.018NO_2$	0.810	1.136 – 6.354

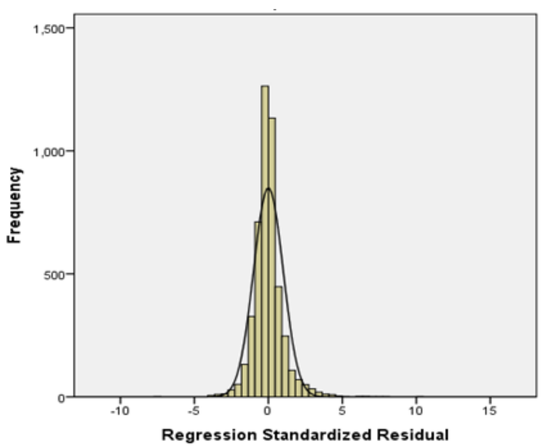

Fig. 3. Residuals of MLR Model

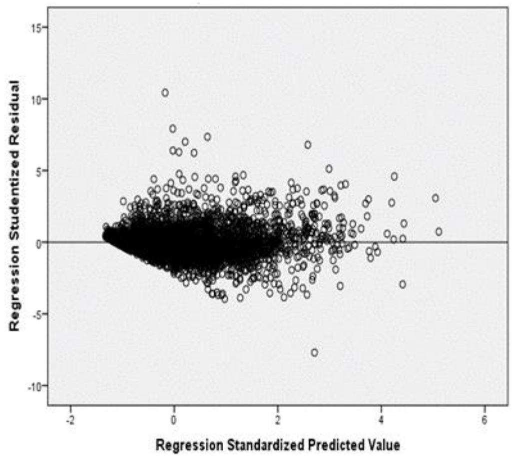

Fig. 4. Constant Variance of MLR Model

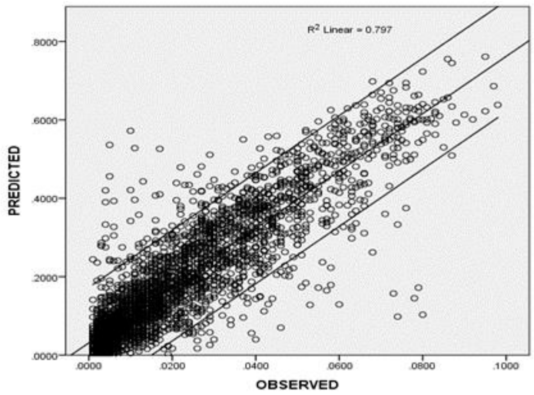

Fig. 5. Validation of MLR Model

IV. CONCLUSION

In conclusion, the result shows that the diurnal cycle of ozone concentration has a mid-day peak (1400hrs) while lower concentration occurs at night time (2100hrs) as it titrates nitrogen dioxide (NO_2). The diurnal cycle of ozone concentration reaches a peak during the middle of the day, while there was a lower ozone concentration during nighttime. This is due to the formation of ozone by photochemical reaction. Moderate Spearman correlation coefficient was evaluated between O_3 and NO_2 ($r=0.45$, $p<0.05$) which suggest that Klang area is moderately NO_2-sensitive area. Multiple linear regression (MLR) models were established and signifies that nitrogen oxides (NO), relative humidity (RH), NO_2, carbon monoxide (CO), wind speed (WS), temperature (T) and sulphur dioxide (SO_2) are the significant predictors for O_3 concentration. The predicted coefficients of determination are $R^2=0.810$. The developed MLR models are appropriate for forecasting O_3 concentrations intended for early warnings system for public health as well as for local authorities to formulate strategies in improving the air quality at Klang.

Retrieval Number: J11270881019/19©BEIESP
DOI: 10.35940/ijitee.J1127.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support received from Research Management and Innovation Center University Malaysia Terengganu. We are also would like to thanks to the Malaysian Department of Environment for the air quality data.

REFERENCES

1. Vingarzan, R. (2004). A review of surface ozone background levels and trends. *Atmospheric Environment.* 38(21), 3431-3442.
2. Fernández-Fernández, M. I., Gallego, M. C., García, J. A., & Acero, F. J. (2011). A study of surface ozone variability over the Iberian Peninsula during the last fifty years. *Atmospheric Environment*. 45(11), 1946-1959.
3. Marathe, S. A., & 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.* 3(3): 149.
4. Moustris, K. P., Nastos, P. T., Larissi, I. K., & Paliatsos, A. G. (2012). Application of Multiple Linear Regression models and artificial neural networks on the surface ozone forecast in the greater Athens area, Greece. *Advances in Meteorology.* 2012: 1-8.
5. Abdul-Wahab, S. A., Bakheit, C. S., & Al-Alawi, S. M. (2005). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software.* 20(10): 1263-1271.
6. Ismail, M., Suroto, A. and Abdullah, S. (2015). Response of Malaysian Local Rice Cultivars Induced by Elevated Ozone Stress. *EnvironmentAsia*. 8(1): 86-93.
7. Abdullah, S., Ismail, M., Fong, S. Y. and Ahmed, A. N. (2016). Evaluation for long term PM_{10} concentration forecasting using multi linear regression (MLR) and principal component regression (PCR) models. *EnvironmentAsia.* 9(2): 101-110.
8. Abdul-Wahab, S. A., Bakheit, C. S., & Al-Alawi, S. M. (2005). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software.* 20(10): 1263-1271
9. Barrero, M. A., Grimalt, J. O., & Cantón, L. (2006). Prediction of daily ozone concentration maxima in the urban atmosphere. *Chemometrics and Intelligent Laboratory Systems.* 80(1): 67-76.
10. Leh, O. L. H., Ahmad, S., Aiyub, K., Jani, Y. M., & Hwa, T. K. (2012). Urban air environmental health indicators for Kuala Lumpur city. *Sains Malaysiana.* 41(2): 179-191.
11. Abdullah, S., Ismail, M., Ghazali, N. A. and Ahmed, A. N. (2018). Forecasting particulate matter (PM_{10}) concentration: A radial basis function neural network approach. *AIP Conference Proceedings*. 2020: 020043-1-6.
12. Abdullah, S., Ismail, M., Samat, N. N. A. and Ahmed, A. N. (2018). Modeling 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): 8226-8234.
13. Batista, G. E., & Monard, M. C. (2003). An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence.* 17(5-6): 519-533.
14. Moustris, K. P., Nastos, P. T., Larissi, I. K., & Paliatsos, A. G. (2012). Application of Multiple Linear Regression models and artificial neural networks on the surface ozone forecast in the greater Athens area, Greece. *Advances in Meteorology.* 2012: 1-8.
15. 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.
16. Kovac-Andric, E., Brana, J. & Gvozdic, V. (2009). Impact of meteorological factors on ozone concentrations modelled by time series and multivariate statistical methods. *Ecology Informatics.* 4: 117-122.
17. Ismail, M., Abdullah, S., Jaafar, A. D., Ibrahim, T. A. E. and Shukor, M. S. M. 2018. Statistical modeling approaches for PM_{10} forecasting at industrial areas of Malaysia. *AIP Conference Proceedings.* 2020: 020044-1-6.
18. Abdullah, S., Hamid, F. F. A., Ismail, M., Ahmed, A. N. and Mansor, W. N. W. (2019). Data on Indoor Air quality (IAQ) in kindergartens with different surrounding activities. *Data in Brief.* 25: 103969.
19. Ismail, M., Abdullah, S., Yuen, F. S., & Ghazali, N. A. (2016). A ten-year investigation on ozone and it precursors at Kemaman, Terengganu, Malaysia. *Environment Asia.* 9(1): 1-8.
20. Banan, N., Latif, M. T., Juneng, L., & 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.
21. Abdullah, S., Shukor, M. S. M., Shahrudin, D. and Ismail, M. (2018). The assessment of indoor air quality (IAQ) at refinery industry. *International Journal of Civil Engineering and Technology.* 9(9): 925-932.
22. Abdullah, A. M., Ismail, M., Yuen, F. S., Abdullah, S. and Elhadi, R. E. (2017). The relationship between daily maximum temperature and daily maximum ground level ozone concentration. *Polish Journal of Environmental Studies.* 26(2): 517-523.

AUTHORS PROFILE

Dr. Samsuri Abdullah is a senior lecturer in the University Malaysia Terengganu. His research interest relies on air pollution including the indoor-outdoor relationship of air pollutants and environmental noise pollution. Currently he is working on the machine learning approaches for the prediction of $PM_{2.5}$ and O_3 in Malaysia.

Ms. Najihah Husna Ahmad Nasir is a graduate in the field of Environmental Technology. She will be graduated in November 2019. Her works relies on ozone pollution and modeling.

Dr. Marzuki Ismail is an Associate Professor in University Malaysia Terengganu. His research interest relies on Air Quality Modeling and Management. He is one of the Expert Panel for the formation of $PM_{2.5}$ limit guideline in Malaysia.

Dr. Ali Najah Ahmed is a senior lecturer in the University Tenaga Nasional, Malaysia. Currently he holds a position as Head of Unit (Publication) at the Institute of Engineering Infrastructures, UNITEN. His research interest relies on hydrological modeling. He is an active researcher with many publications regarding his field.

Mr. Mohammad Nor Khasbi Jarkoni is a master graduate in Maritime Technology. Currently, he is pursuing his Doctor of Philosophy (PhD) degree in University Malaysia Terengganu. His research project relies on the emissions of engine diesel and prediction of atmospheric pollutants.