Malaysian Journal of Computing, 8 (2): 1560-1573, 2023

eISSN: 2600-8238

MODELLING TIME SERIES TROPOSPHERIC OZONE DATA AND THE PRECURSORS TO STUDY EFFECT AND RELATIONSHIP IN PETALING JAYA AND SHAH ALAM MALAYSIA

Emilya Shahira Azli[1], Nurazrin Syafiqah Mat Asri[2], Siti Nurliyani Ahmad Tajuddin[3], Norshahida Shaadan[4*], Siti Afiqah Muhamad Jamil[5] and Shamsiah Sapri[6]

[1,2,3,4,5,6] *Center of Mathematical Sciences, College of Computing, Informatics and Media, University Technology MARA (UiTM), 40450, Shah Alam, Selangor, Malaysia*

[1]emilyashahira97@gmail.com, [2]azrinsyafiqahh@gmail.com, [3]sitinurliyani23@gmail.com, [4*]shahida@tmsk.uitm.edu.my, [5]afiqahjamil@tmsk.uitm.edu.my, [6]shamsiah@tmsk.uitm.edu.my

ABSTRACT

Understanding the role of pollutant precursors is very important to provide meaningful insights for planning and designing localized air pollution control strategies. This study is conducted with the aim of describing the relationship and assessing the effect of several precursors (OX, NO, NO_2, SO_2, CO) towards O_3 using Multiple Linear Regression (MLR) at an industrial (Petaling Jaya) and an urban (Shah Alam) location in Selangor, Peninsular Malaysia. Statistical modelling and analysis were conducted based on secondary data of time-series observations of multi-variables from the period of the year 2015-2017 obtained from the Department of Environment, Malaysia. Two models were developed, evaluated and compared involving models without (default) and with a proper statistical procedure to highlight the importance of data randomization and outlier treatment in the modelling of MLR for time series air quality data set. The results have shown that the model without using a good quality data set has resulted in a false estimated MLR model. Based on the p-value, NOx and CO are the two most significant O_3 precursors in Petaling Jaya and Shah Alam indicating emission sources from vehicles contributed to O_3 changes at the sites. Stronger effect of SO_2 on O_3 with a positive relationship in the industrial site, Petaling Jaya station compared to an urban residential area, Shah Alam whereby the relationship of SO2 with O_3 is negative. The lag (AR1) variable is also shown significant. The study also concluded that the MLR model for time series observations requires good data quality (normality and non-autocorrelated) as the approach is critical to ensure a correct MLR model that satisfies the model assumptions with accurate parameter estimates. The study results contribute critical methodological knowledge to future researchers and the findings are fruitful for environmental bodies to help in managing O_3 pollution at the study locations.

Keywords: *Air pollution, Modelling, Multiple Linear Regression, Ozone, Precursors.*

Received for review: 05-05-2023; Accepted: 21-09-2023; Published: 10-10-2023
DOI: 10.24191/mjoc.v8i2.21572

1. Introduction

Ozone is type of a gas which composed of three atoms of oxygen that occurs both in earth's upper and lower atmosphere. It can be beneficial as good ozone and detrimental as bad ozone.

The ozone that is well known by most people as good ozone is the upper ozone called Stratospheric Ozone. It is forms naturally in the upper atmosphere and protects us from harmful ultraviolet rays (UV). Unlike Stratospheric Ozone, Tropospheric Ozone (O_3) is created through interactions between man-made emissions of Volatile Organic Compounds (VOCs) and Nitrogen Oxides (NOx) in the presence of heat and sunlight and this type is harmful if exposed beyond permissible limit. O_3 is known as a secondary pollutant, photochemical oxidant and the main component of smog (Gordon et al., 2014). The pollutant is formed by a series of reactions, under the influence of sunlight, involving volatile organic compounds (VOCs) combined with a group of air pollutants known as Nitrogen Oxides (NOx). Volatile organic compounds are emitted by automobiles and various commercial and industrial sources. NOx are by-products of burning fuel in automobiles/motor-vehicles and heavy industries. Collectively, NO_x, VOCs, methane (CH4), total non-methane hydrocarbons (TNMHCs), Carbon Monoxide (CO) and Sulphur Dioxide (SO_2) are referred to as O_3 precursors (Abdul-Wahab et al., 2005; Sharma et al., 2016). According to Cambridge dictionary, precursors is defined as "something that existed before other thing, especially if it either developed into it or had an influence on it". O_3 formation by the photochemical reactions of precursors requires at least one-hour time (Marathe, 2018). O_3 is also recognized to be a threat to human health (Carvalho et al., 2021) and have a deleterious impact on vegetation (Monks et al., 2015) as well as impacting built infrastructure (Kumar & Imam, 2013). High level of O_3 has a detrimental effect on plants, as it enters the plant leaves through stomata generates odd species which can oxidize plant tissue resulting in changes in gene expression causing impaired photosynthesis. O_3 concentrations greater than 40 ppbv may be harmful to the crop yield, biomass production, vitality and stress tolerance of forest trees (Verma et al., 2015). Long-term human exposure has an association with cause-specific CVD mortality in China, independent of particulate matter, and CVD risk factors which suggests for O3 pollution (Liu et al., 2022). Evidence also shows that O_3 is associated with respiratory morbidity and premature mortality (Gordon et al., 2014). PM10 and O_3 has been identified as the two most dominant pollutants in the Malaysia environment that has been included in the computation for Air Pollution Index (API) to measure and assess the air quality status in the country (Ahmat et al., 2019). Due to the fact that O_3 formation is also comes from the precursors which were primarily emitted form motor-vehicles engine, knowledge on the relationship between the precursors and their effect on O_3 is important. There have been many previous studies on the behavior and modeling of O_3 in various sites in Malaysia environment, which mostly uses descriptive analysis and in the modeling aspect, it is more to get the best model for prediction. Among the findings, the concentration of O_3 reaches its peak during the middle of the day while during night-time, the concentration lowers (Shaadan et al., 2018). The peak occurs at 2-3 pm (Azmi et al., 2010). The daily maximum O_3 concentration is higher in industrial area compared to urban area (Abdullah et al., 2017, Afroz et al., 2003). This indicates that the precursors of O_3 mainly NOx and VOC are produced by motor vehicles and smoke from the industrial area. Muhamad et al., (2018) stated that high O_3 concentration were caused by open burning and smokes emitted from vehicles.

Globally, the relationships among O_3 precursors, and other meteorological parameters have been examined by several studies using various statistical techniques including Principal Component Analysis (PCA) (Abdul-Wahab et al., 2005) and linear and non-linear models . There were also studies that used fuzzy logic (Mintz et al., 2005), artificial neural network (ANN) (Gao, 2018), graphical analysis and time-series analysis (Baghini et al., 2022) as well as functional data analysis to explore O_3 dynamics (Shaadan et al., 2018). Among these methods, multiple linear regression (MLR) has provided significant results in O_3 modelling (Sousa et al., 2007). In the last 5 years, the scope of the local study is about determining O_3 correlation with other pollutants and meteorological parameters using descriptive statistics by Noor et al., (2018). modelling for O_3 extreme values (Zakaria et al., 2021), modelling using MLR and PCA-MLR and comparison analysis to predict O_3 based on other pollutants and meteorological variables (Hashim et al., 2019), trend analysis of O_3 extreme (Zakaria et al., 2022), spatio-temporal variations at urban and sub-urban sites in Sarawak region of Malaysia (Mahidin et al., 2021). Noticeably, Multiple Linear Regressions (MLR) has been considered as the most common model used in the literature to explain the relationship between a particular pollutant and the associated factors. However, in the modelling context, MLR need to be carefully applied because MLR is a parametric model that require several assumptions to be valid. MLR is a leverage. Some had used the model aiming for predicting the average or expected new data based on several independent variables and some had considered to use the model to investigate the relationship and the effect of the independent variables on the dependent variable O_3. MLR model is not robust to outliers. The model is also not robust to serial correlation. Something important that need to be highlighted is that, air quality data set often consist of outliers and the data are serial correlated as the data are time series. However, in the previous studies this important concern was seen neglected in the modelling procedure. Without a proper modelling procedure, high potential of default model and inaccurate sign of coefficients estimates would be obtained. The analysis results consequently direct to wrong interpretation and conclusion, thus very dangerous to the decision makers. Therefore, MLR modelling applied for air quality data set is supposed to be conducted with precaution.

This study mainly aims to investigate the relationship and the effect of several precursors ((NOx, NO, NO_2, SO_2, CO) towards O_3 using MLR model at an industrial location (Petaling Jaya) and an urban location (Shah Alam) in Selangor Malaysia. This study also aims to emphasize on the importance of a proper statistical procedure in MLR modelling when applied to time series observations or data.

2. Methodology

2.1 Data and the study locations.

Data used in this study consist of O_3 and the precursors such as NO_2, NO, SO_2, NO_x and CO in unit (ppm) which were obtained from Department of Environment (DOE), Putrajaya, Malaysia. The data are recorded by hourly basis from January 2015 until December 2017. The location of the study includes Petaling Jaya and Shah Alam. Petaling Jaya air quality monitoring sites which are categorized under an industrial background respectively. Figure 1 is the map location of the two sites. Petaling Jaya is a busy city center surrounded by Kuala Lumpur to the east, Sungai Buloh to the north, Shah Alam, the capital of Selangor, and Subang Jaya to the west and Bandar Kinrara (Puchong) to the south. The climate is warm with an average maximum of 30 °C. The area is approximately 97.2 square kilometres. Shah Alam is situated within the Petaling District and a small portion of the neighbouring Klang District with a total area of 290.3 square kilometres. The temperatures are consistent throughout the year with an average high temperature of 31.9 °C and an average low temperature of 23.2 °C. The location for air quality monitoring station of Shah Alam is at Section U2 (Taman TTDI Jaya Primary School, Shah Alam. The station is surrounded by residential area, near to major roads, industrial area and Subang airport with coordinate 3° 6’ 28” N, 101° 33’ 4” E. Meanwhile, Putrajaya is a mixed commercial-residential-industrial area which has the highest population in the Klang Valley region with the coordinate location of 3° 06' 26.14" N, 101° 36' 24.16" E. According to DOE, Petaling Jaya site is under the category of industrial while Shah Alam is urban background.

Figure 1. The location of Petaling Jaya and Shah Alam site in Selangor Malaysia.

The following Figure 2 shows the conceptual framework of the relationship between the independent variables (precursors) and the dependent variable (O_3) used to model using Multiple Linear Regression (MLR). In order to construct the MLR model, considering the effect of serial correlation the data since the study data ere multivariate time series, variable Lag (AR1) is decided to be included.

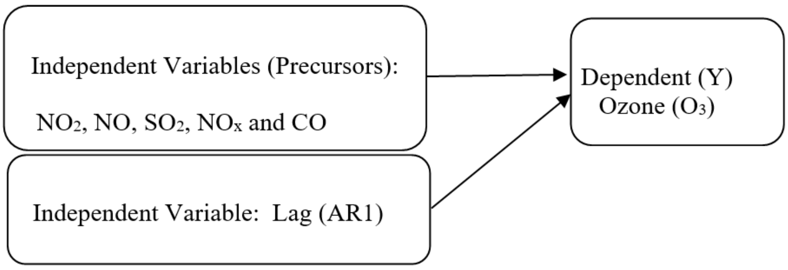

Figure 2. The theoretical framework between the precursors, lag and O_3 variables.

2.2 Multiple Linear Regression

This study had applied Multiple Linear Regression (MLR) analysis in order to achieve the main objective which is to study the effect of precursors and their relationship with O_3. At the same time, the study also aims to highlight the importance of air quality modelling using good quality data set. MLR is one of the most widely used of all statistical models to model the relationship between many independent variables and a single dependent variable where the scale level of the dependent variable must be at least interval. MLR model enables researcher to determine the variation of the model and the relative contribution of each independent variable in the total variance in the data set. Equation (1) depicts the MLR model equation for variables in the study theoretical framework shown in Figure 3.

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \varepsilon \quad (1)$$

where

y – the dependent or response variable
x – the independent variables (the precursors) and lag variable AR1
β_0 – the y-intercept
β_i – the i^{th} regression coefficient for $i = 1,2,\ldots,n$
ε – the error or residuals term

Least square method is used to estimate the parameter for the regression coefficient by minimizing the sum of squares of the residuals. The following are the assumptions of MLR. The validity of MLR model depends on several important assumptions that need to be considered prior to the model development and after the model has been developed. The fundamental MLR model assumption is that the error term $\varepsilon\varepsilon$ follows normal distribution with zero mean and constant variance.

2.3 Modelling and Design Framework

The following Figure 3 shows the considered systematic framework of how MLR model is constructed and established in this study.

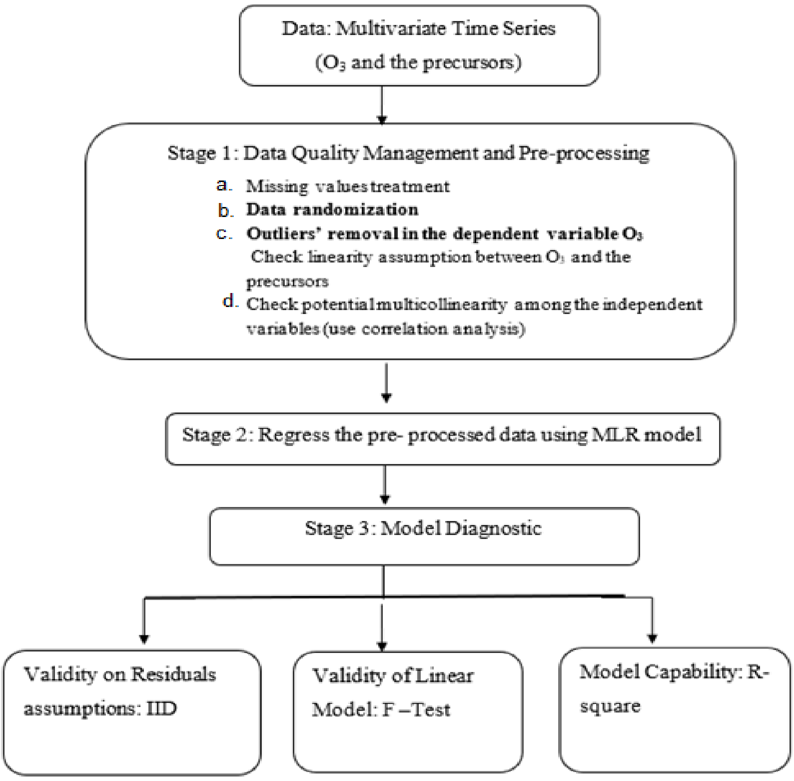

Figure 3. Methodology of MLR modelling framework for air quality data set.

3. Results of Analysis

3.1 Temporal Pattern of O3 and the precursors

Figure 4a and Figure 4b describe the distribution pattern of the precursors and O_3 in Petaling Jaya and Shah Alam.

a) Petaling Jaya

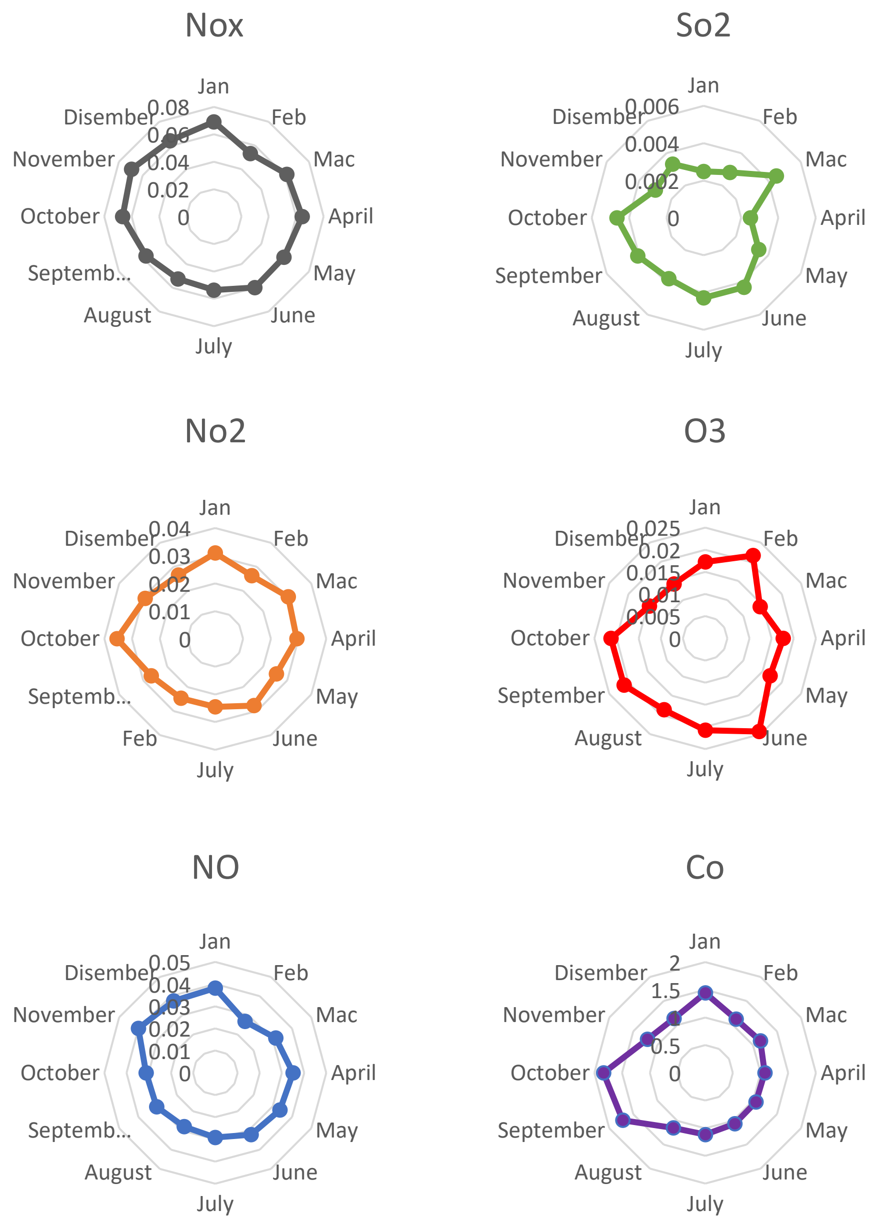

Figure 4a. Monthly Average Reading of O_3 and its Precursors in Petaling Jaya.

b) Shah Alam

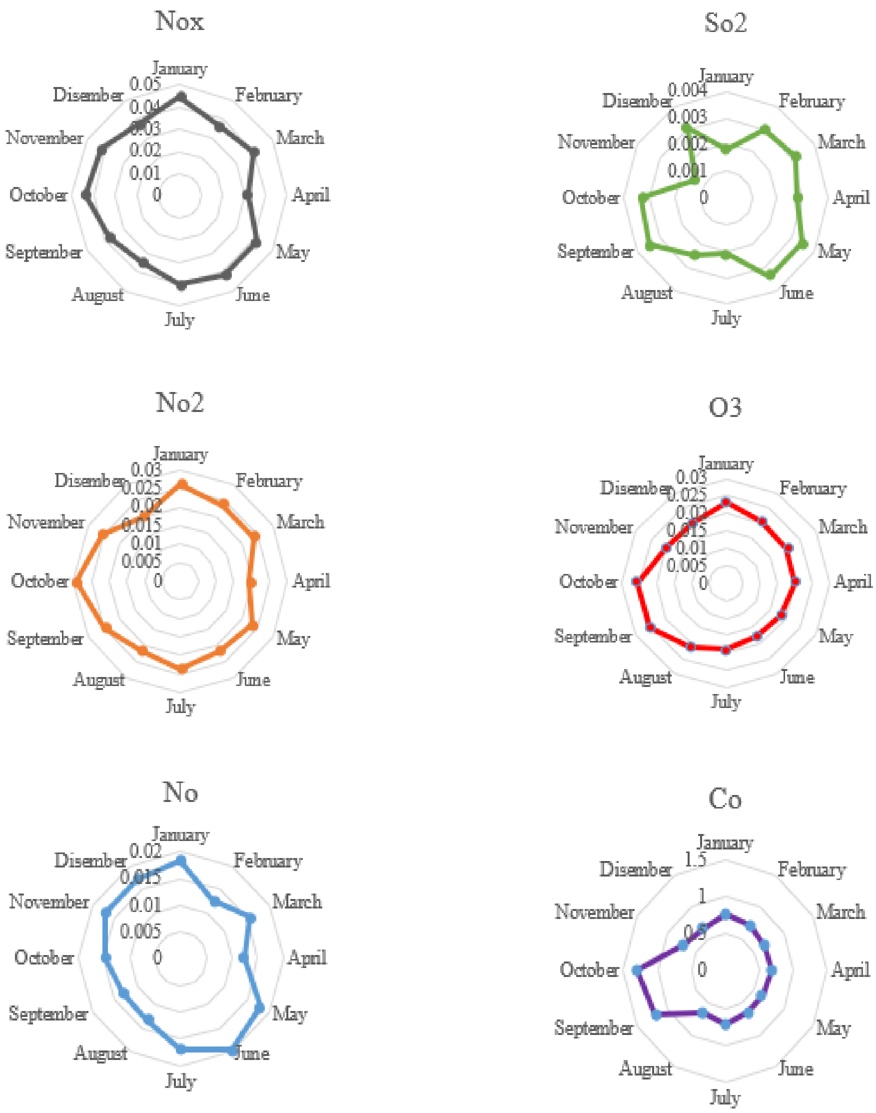

Figure 4b. Monthly average reading of O_3 and its precursors in Shah Alam.

The distribution pattern of NOx, NO_2, NO, and CO is about the same in Petaling Jaya and Shah Alam. However, in terms of the recorded level, higher variations were visualized in Shah Alam compared to Petaling Jaya. This is shown by a greater number of multiple layers of contour in the spider web plot. However, the magnitude levels of the precursors are higher in Petaling Jaya compared to Shah Alam. NOx has the peak in January, NO_2 has the peak in January and October and CO has the peak in September and October at both sites, Petaling Jaya and Shah Alam. However, NO has the peak in January, April and November in Petaling Jaya meanwhile in Shah Alam the peak is in January and June. SO_2 has the peak in Mac and October in Petaling Jaya, and in May, June, September, October and December in Shah Alam. The distribution pattern of O_3 shows a distinct pattern. The peak occurs in February and June in Petaling Jaya, meanwhile it occurred in January, September and October in Shah Alam. O_3 levels also higher than other months in September and October in Petaling Jaya.

3.2 MLR model and analysis

Towards the modelling process, the following are the results of several preliminary analyses to check for the basic requirements of MLR model. MLR model is a parametric and a leverage model to estimate O_3 which is sensitive to outliers. Thus, it is very important to check the normality distribution of O_3 (the dependent variable). The following histogram in Figure 5, shows that the O_3 data are having skewed distributions, thus contain outliers. In this study, outliers were identified using boxplot and was removed from the data set. The independency among the independent variables (precursors) also needs to be explored to avoid multicollinearity. Any pair of the variables that has high correlation reflects, the existence of multicollinearity indicating the two variables shared the same common information. One of them should be dropped. Since the air quality data set often recorded as time series observations, data serial correlation also needs to be considered. Thus, in this study, the observations were firstly randomized to ensure for autocorrelation free data set before fitting the model to the data set.

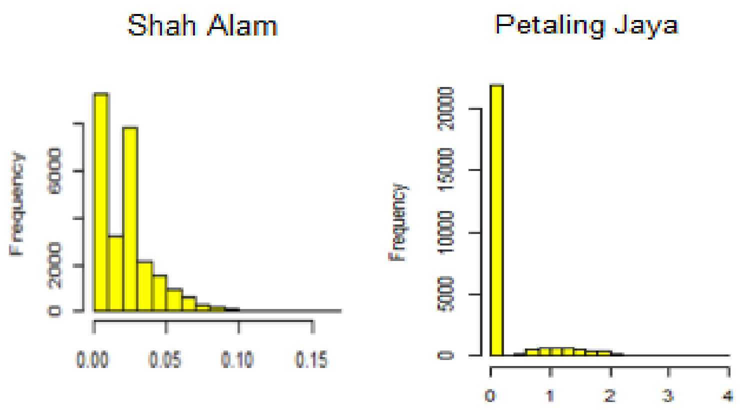

Figure 5. Histogram of O_3 in Shah Alam and Petaling Jaya.

Table 1 and Table 2 shows the matric correlation of the precursors variables from Petaling Jaya and Shah Alam. It is found that NO_X and NO having very high correlation with for coefficient correlation r=0.9405 (Petaling Jaya) and r=0.85357 (Shah Alam) exceeding 0.7, thus NO is dropped to be put in MLR model from Petaling Jaya while both NO and NO_2 were dropped from Shah Alam data set.

Table 1. Matrix correlation among the independent variables (precursors) in Petaling Jaya.

Precursors	**NO_X**	**NO**	**SO_2**	**NO_2**	**CO**	**Lag**
NO_X	1.0000	0.9405	-0.0267	0.6299	0.2118	-0.2073
NO	0.9405	1.0000	-0.0328	0.4419	0.1994	-0.2869
SO_2	-0.0267	-0.0328	1.0000	0.0709	0.3235	-0.2740
NO_2	0.6299	0.4419	0.0709	1.0000	0.2188	-0.0298
CO	0.2118	0.1994	0.3235	0.2188	1.0000	-0.5236
Lag	-0.2073	-0.2869	-0.2740	-0.0298	-0.5236	1.0000

Table 2. Matrix correlation among the independent variables (precursors) in Shah Alam.

Precursors	**NO_X**	**NO**	**SO_2**	**NO_2**	**CO**	**Lag**
NO_X	1.00000	0.85357	0.06958	0.73310	0.37574	-0.45917
NO	0.85357	1.00000	0.00713	0.42965	0.35104	-0.50355
SO_2	0.06958	0.00713	1.00000	0.15479	0.08629	0.04503
NO_2	0.73310	0.42965	0.15479	1.00000	0.35305	-0.30848
CO	0.37574	0.35104	0.08629	0.35305	1.00000	-0.29806
Lag	-0.45917	-0.50355	0.04503	-0.30848	-0.29806	1.00000

Using the remaining independent variables (precursors), two MLR model is fitted to the data; Model 1 is the model fitted to the data set with data quality treatment was applied (outlier removal and data randomization) was conducted and another Model 2 is the model without data quality has been considered. The results of the fitted Model 1 and Model 2 are shown in Table 3 and Table 4. The diagnostic checking on the MLR residual assumptions is depicted in Figure 6(a) and Figure 6(b).

Table 3. The Fitted Models and the performances.

Location	**Petaling Jaya**		**Shah alam**	
	Model 1 Modelling with quality data concern	Model 2 Modelling without data quality concern	Model 1 Modelling with data quality concern	Model 2 Modelling without data quality concern
ANOVA F-test	p-value < 0.05 (Significant)	p-value <0.05 (Significant)	p-value < 0.05 (Significant)	p-value <0.05 (Significant)
R-squared	0.70	0.72	0.54	0.75
Significant Variables and direction (p-value<0.05)	- NO_X +SO_2 +NO_2 + CO +Lag	-NO_X -SO_2 + NO_2 -CO +Lag	-NO_X -SO_2, - +CO + Lag	-NO_X +SO_2 - +CO +Lag

Results in Table 3 highlights on the impact of fitting MLR model using data set that is free from outliers and free of serial correlation (quality data) given by Model 1 in comparison to the results of Model 2, when model was fitted using data set without data quality consideration. It can be seen that for both models, the models F-test was not affected but the value of R-square, the sign or direction of the relationship among them can be affected. For the case of data set from Petaling Jaya, the R-square has slightly reduced for Model 2 compared to Model 1 but for Shah Alam, the R-squared has large reduction in the value. Using Model 1, all the precursors (NO_X, SO_2, NO_2, CO) including lag variable are significant to contribute to O3 levels with (-,+,+,+,+) relationship but with Model 1, the sign of CO is negative (-). For Shah Alam data set, the impact of data quality also changes the sign of SO_2 variable. Supposedly with good quality data, the sign of SO_2 is negative (-). But modelling using bad quality data, the results of sign become the opposite; that is (+). The results in table 3 provide evidence that modelling using bad quality data has led to false results.

The MLR model's diagnostic checking is shown in Figure 6(a) and Figure 6(b). The normality and independent, identically distributed (IID) assumptions on the Residuals of the model were examined which consist of plots of residuals versus fitted data and scale location plot to assess the homoscedasticity assumption as well as Normal q-q plot and residual vs leverage plot to assess for normality and no influential outlier's assumption.

(a) Petaling Jaya

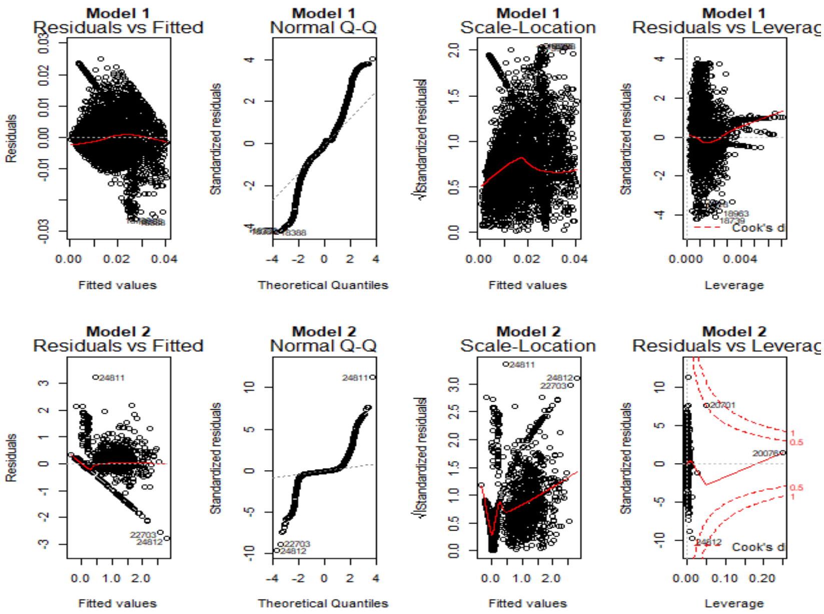

Figure 6 (a). Results of Diagnostic checking on Residual's assumption between Model 1 and Model 2 of Petaling Jaya.

(b) Shah Alam

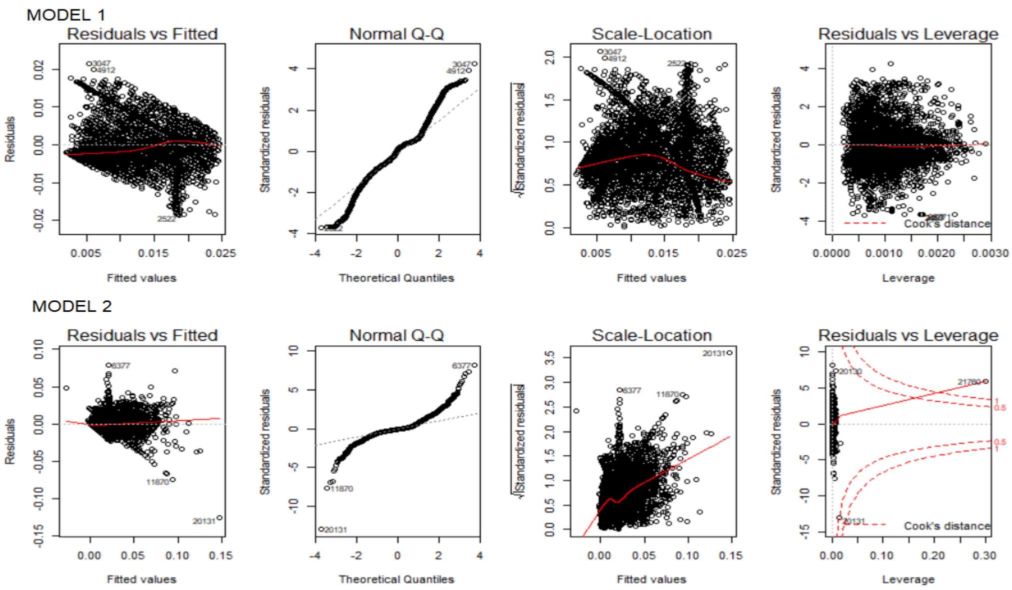

Figure 6 (b). Results of Diagnostic checking on Residual's assumption between Model 1 and Model 2 Shah Alam.

Based on Figure 6(a) and Figure 6(b), at both Petaling Jaya and Shah Alam sites, there is evidence that modelling with data quality provides better improvement on the diagnostic results. Models without data quality (Model 2), the IID assumptions were not satisfied but modelling with data quality (Model 1) all the IDD assumptions are satisfied. Both results from

Table 3 and Figure 6(a) and Figure 6(b), concluded that Model 1 should be used as the best model to explain the relationship and to study the effect of precursors towards O_3 in Petaling Jaya and Shah Alam sites.Results in Table 4 can be used to explain the relationship and the effect of the significant precursors towards O_3.

Table 4. The models coefficient and the relationship as well as effect of the precursors (independent variables) on O_3 based on Model 1.

	Petaling Jaya		**Shah Alam**	
	Estimates	**P-value**	**Estimates**	**P-value**
Intercept	0.0005793	0.153	0.00382	2e-16
NO_X	-0.0705566	2e-16	-0.05832	6.42e-16
SO_2	0.4755807	2.04e-08	-0.10838	0.209
NO_2	0.0870808	3.36e-05	-	-
CO	0.0029999	2e-16	0.00266	5.36e-11
Lag	0.8106473	2e-16	0.73103	2e-16

At 5% significance level, as all the p-values are very small (<0.05), the results in Table 4 provide evidence that the O_3 levels in Petaling Jaya has very strong significant negative association with NOx and significant positive association with SO_2, NO_2, CO and Lag variables. A one unit increase in NOx, resulted into 0.0705566 times reduction in O_3 meanwhile SO_2, NO_2 and CO affects the increment by 0.4755807, 0.0870808 and 0.0029999 times of O_3 respectively. Meanwhile in Shah Alam, both NOx and SO_2 contribute with negative relationship with O_3 whereby NOx has stronger influence, indicated by much smaller p-value compared to SO_2. A one unit increase in NOx resulted into 0.05832 reduction in O_3 in Shah Alam. While CO, affects O_3 levels with positive direction. A one unit increase in CO, there will be 0.00266-unit increment in O_3.

Based on the p-values ranking, the smallest value indicates stronger the significance of the variables, the results also shows that NOx is the most significant precursors, followed by CO, SO_2 and NO_2 in Petaling Jaya. NOx also is the most significant precursor in Shah Alam, but the second important precursors is CO then followed by SO_2. Lag variables are significant variable in the O_3 model.

4. Conclusion

This study had applied MLR model on the time series air quality data set to investigate the relationship between O_3 and the precursors including NOx, SO_2, NO_2 and CO. Lag variable is also included as an additional variable that is believed might affect O_3 levels.

The study has shown that MLR application must be used carefully on time series air quality data set to model the relationship between O_3 and the precursors. This is done by considering the negative impact of untreated data set towards the fitted model when data quality is ignored in the modelling process. Outliers and serial correlation of time series observation problem in the data set must be solved before MLR is fitted into the data. In this study, outlier's removal and data randomization were implemented as the approach to getting a good quality data set prior to regression analysis. The results have shown that MLR model with bad quality data has produced false fitted model. In specific, the existence of outliers resulted into failure diagnostic checking, where all the model assumptions were not satisfied. Meanwhile modelling with serial correlated data will affect the true sign of the relationship between the independent variables (precursors) and the dependent variable (O_3). Bad quality data also affects the value of coefficient determination R^2 reflecting false result of model performance.

In the context of application of MLR based on the best model, all the precursors (NOx, SO_2, NO_2, CO) and lag variable are found significant to affect changes in O_3 in Petaling Jaya site. All the precursors having positive relationship with O_3 accept NOx. Similarly, all the considered precursors (NOx, SO_2, CO) are also significant to affect O_3 in Shah Alam. CO has positive relationship meanwhile NOx and SO_2 have negative relationship. NOx is the most significant precursors, followed by CO, SO_2 and NO_2 in Petaling Jaya. NOx also is the most significant precursor in Shah Alam, but the second important precursors is CO then followed by SO_2. Lag variable is also proven significance in the O_3 model at both sites.

Acknowledgement

The authors would like to thank the Department of Environment, Putrajaya Malaysia for providing data to be used in this study.

Funding

The author(s) received no specific funding for this work.

Author Contribution

Norshahida Shaadan proposed the research topic, designed the research framework and write the manuscript. Emilya Shahira Azli, Nurazrin Syafiqah Mat Asri and Siti Nurliyani Ahmad Tajuddin prepared the literature review, wrote the research methodology, obtain the research data and conduct the analysis and interpret the results. Siti Afiqah Muhamad Jamil and Shamsiah Sapri oversaw the article writing and proofread the writing.

References

Abdullah, A. M., Ismail, M., Yuen, F. S., Abdullah, S., & 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. https://doi.org/10.15244/pjoes/65366

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.

Afroz, R., & Ibrahim, N.A. (2003). Review of air pollution and health impact in Malaysia. *Environmental Research*, 92(2), 71-77.

Ahmat, H., Musa, N.S., Nazamid, N., & Zaharin, N.A. (2019). Classification of High and Low level of PM10 Concentrations in Klang and Shah Alam, Malaysia. *Malaysian Journal of Computing*, 4 (2), 325-334.

Azmi, S. Z., Latif, M. T., Ismail, A. S., Juneng, L., & Jemain, A. A. (2010). Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. *Air Quality, Atmosphere and Health*, *3*(1), 53–64.

Baghini, N.S., Falabatkar, S., & Hassanvadh, M.S. (2022). Time series analysis and spatial distribution map of aggregate risk index due to tropospheric NO_2 and O_3 based on satellite observation. *Journal of Environmental Management*, 304, 114-202.

Carvalho, R.B., Marnett, B., Dorneles, G. et al. 2022. O3 concentration and duration of exposure are factors influencing the environmental health risk of exercising in Rio Grande, Brazil. *Environmental Geochemistry and Health* 44(2), 2733-2742.

Gordon, T. Stanek, L.W., & Brown J. (2014). *Pollution, Air in Encyclopaedia of Toxicology*: 3rd Edition, Elsevier, 995-1002.

Gao, M., Yin, L., & Ning, J. (2018). Artificial neural network model for ozone concentration estimation and Monte Carlo analysis. *Atmospheric Environment*, 184, 129-139

Hashim, N.I.M., Yusoff, N.A.I.M., Noor, N.M., & Ul-Saufie, A.Z. (2019). Assessment of Surface Ozone Concentration in Northern Peninsular Malaysia. In *International Conference on Green Engineering Technology & Applied Computing. IOP Conf. Series: Materials Science and Engineering 551 (2019) 012100*

Kumar, P., & Imam, B. (2013). Footprints of air pollution and changing environment on the sustainability of built infrastructure. *Science of The Total Environment*, *444*, 85–101.

Liu, S., Yi, Z., Ma, R., Liu, X., Liang, J., Lin., H., Shen, P., Zhang, J., Lu., P., Tang, X., and Gao, P. (2022). Long-term exposure to ozone and cardiovascular mortality in a large Chinese cohort. *Environmental International* 165, 1-16.

Mahidin, H., Latif, M.T., Hamdan, A. Salleh, J., Dominick, D., Ooi, M.C.G., Juneng, L., & Zawawi, S. *International Conference on Science and Technology Applications in Climate Change* 1 July, National University of Malaysia, Selangor, Malaysia

Marathe, S. A. (2018). Multiple Regression Analysis of Ground level Ozone and its Precursor Pollutants in Coastal Mega City of Mumbai, India. *MOJ Ecology & Environmental Sciences*, *2*(6). https://doi.org/10.15406/mojes.2017.02.00041

Mintz,R., Young, B., Svrcek, W.Y. (2005). Fuzzy logic modeling of surface ozone concentrations. *Computers & Chemical Engineering* 29(10), 2049-2059

Mohamed, N., Mohamad, H., & Yusof, S.Y. 2018. IOP Conf. Series: Materials Science and Engineering 374, 012097 doi:10.1088/1757-899X/374/1/012097

Monks, P. S., Archibald, A. T., Colette, A., Cooper, O., Coyle, M., Derwent, R., … Williams, M. L. (2015). Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmospheric Chemistry and Physics*, *15*(15), 8889–8973. https://doi.org/10.5194/acp-15-8889-2015

Noor, N.M., Mohamad, N.N., & Hashim, N.I.M. (2028). Variation of Ground-level Ozone in the West Coast of Peninsular Malaysia. *Environment Asia* 11(3), 235-250. DOI 10.14456/ea.2018.

Sharma, A., Sharma, S.K., Pathak, U., & Mandal, T.K. Study on surface ozone and its precursors at an urban site of Delhi, India. *Indian Journal of Radio and Space Physics* 45(3):95-100

Shaadan, N., Rosslan, R.R., Deni, S.M., & Ismail, A. 2018. The diurnal Ozone dynamics profile at several locations in Selangor, Malaysia: A functional data analysis approach. AIP Conference Proceedings 2013, 020011 (2018); https://doi.org/10.1063/1.5054210

Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M., & Pereira, M.C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software* 22 (1), 97-103

Verma, N., Satsangi, A., Lakhani, A., & Kumari, K. M. (2015). *Journal of Chemical , Biological and Physical Sciences Prediction of Ground level Ozone concentration in Ambient Air using Multiple Regression Analysis*. *5*(4), 3685–3696.

Zakaria, S.A., Amin, N.A.M., Radi, N.F.A., & Hamidin, N. (2021). Stationary and Non-Stationary Models of Extreme Ground-Level Ozone in Peninsular Malaysia. *Mathematics and Statistics*, 9(3), 357 - 370.

Zakaria, S.A., Amin, N.A.M., Radi, N.F.A., & Hamidin, N. (2022). AIP Conference Proceedings 2465, 030009; https://doi.org/10.1063/5.0078652