

Article

Development of Multiple Linear Regression for Particulate Matter (PM₁₀) Forecasting during Episodic Transboundary Haze Event in Malaysia

Samsuri Abdullah ^{1,*}, Nur Nazmi Liyana Mohd Napi ¹, Ali Najah Ahmed ^{2,3},
Wan Nurdiyana Wan Mansor ¹, Amalina Abu Mansor ⁴, Marzuki Ismail ^{4,5},
Ahmad Makmom Abdullah ⁶ and Zamzam Tuah Ahmad Ramly ^{6,7}

- ¹ Air Quality and Environment Research Group, Faculty of Ocean Engineering Technology and Informatics, University Malaysia Terengganu, Kuala Nerus 21030, Malaysia; p4069@pps.umt.edu.my (N.N.L.M.N.); nurdiyana@umt.edu.my (W.N.W.M.)
- ² Faculty of Engineering, University Tenaga Nasional, Bangi 43650, Malaysia; Mahfoodh@uniten.edu.my
- ³ Institute of Engineering Infrastructures, University Tenaga Nasional, Bangi 43650, Malaysia
- ⁴ Faculty of Science and Marine Environment, University Malaysia Terengganu, Kuala Nerus 21030, Malaysia; linadila@gmail.com (A.A.M.); marzuki@umt.edu.my (M.I.)
- ⁵ Institute of Tropical Biodiversity and Sustainable Development, University Malaysia Terengganu, Kuala Nerus 21030, Malaysia
- ⁶ Faculty of Environmental Studies, University Putra Malaysia, Serdang 43400, Malaysia; amakmom@upm.edu.my (A.M.A.); zamzam@eetsb.com (Z.T.A.R.)
- ⁷ Enviro Excel Tech Sdn Bhd., A-G-09, Univ 360 Places, Seri Kembangan 43300, Malaysia
- * Correspondence: samsuri@umt.edu.my; Tel.: +60-9668-3491

Received: 5 February 2020; Accepted: 13 March 2020; Published: 16 March 2020



Abstract: Malaysia has been facing transboundary haze events every year in which the air contains particulate matter, particularly PM₁₀, which affects human health and the environment. Therefore, it is crucial to develop a PM₁₀ forecasting model for early information and warning alerts to the responsible parties in order for them to mitigate and plan precautionary measures during such events. Therefore, this study aimed to develop and compare the best-fitted model for PM₁₀ prediction from the first hour until the next three hours during transboundary haze events. The air pollution data acquired from the Malaysian Department of Environment spanned from the years 2005 until 2014 (excluding years 2007–2009), which included particulate matter (PM₁₀), ozone (O₃), nitrogen oxide (NO), nitrogen dioxide (NO₂), carbon monoxide (CO), sulfur dioxide (SO₂), wind speed (WS), ambient temperature (T), and relative humidity (RH) on an hourly basis. Three different stepwise Multiple Linear Regression (MLR) models for predicting the PM₁₀ concentration were then developed based on three different prediction hours, namely t+1, t+2, and t+3. The PM_{10, t+1} model was the best MLR model to predict PM₁₀ during transboundary haze events compared to PM_{10, t+2} and PM_{10, t+3} models, having the lowest percentage of total error (28%) and the highest accuracy of 46%. A better prediction and explanation of PM₁₀ concentration will help the authorities in getting early information for preserving the air quality, especially during transboundary haze episodes.

Keywords: transboundary haze; prediction; multiple linear regression; accuracy; error; Malaysia

1. Introduction

Haze pollution has become an international issue as it affects the local, regional, and global air quality, especially in Southeast Asian (SEA) countries, including Malaysia, Singapore, Indonesia, Brunei Darussalam, and Southern Thailand [1–3]. Haze occurs annually due to periodic biomass

burning, such as forest fires, peatland combustion, and agricultural burning in Sumatra and Kalimantan, Indonesia. The influence of unpredictable monsoon rain and El-Nino phenomenon both worsens the transboundary haze due to the prolonged and drier weather conditions, especially during the southwest monsoon season [3–8]. These conditions increase the severity of air quality due to it containing a noxious mix of air pollutants, such as particulate matter and noxious gases, which are collectively influenced by meteorological factors like ambient temperature, relative humidity, and wind speed [9,10]. Malaysia has been facing this awful phenomenon for almost over a decade, which started in the 1980s up until recent years [5,11,12]. Haze is detected when the atmosphere contains a higher concentration of particulate matter suspended in the air, which either can or cannot be observed by the naked eye when the relative humidity is lower than 80% and the visibility is reduced to 10 km [5,13–15]. Daily particulate matter with 10 micro aerodynamic diameters (PM_{10}) exceeding $150 \mu\text{g}/\text{m}^3$ for 24 h of permissible limit is thus notified as haze, which is set under the new Malaysian Ambient Air Quality Standard (NAAQS) by the Department of Environment, Malaysia.

Biomass peat soil combustion generally emits the particulate matter and gaseous pollutants, as well as their components such as anions, cations, heavy metals, and levoglucosan [16–19]. Carbon monoxide (CO) is higher during haze events compared to other gases due to the characteristics of peatland areas, which are rich in carbon materials. It is emitted through the incomplete combustion process of carbonic materials in the presence of a minimum amount of oxygen. Meanwhile, PM_{10} contains a high concentration of NO_3^- anion produced through bacterial nitrification and oxidation process during combustion, while SO_4^{2-} cation is also found in a high amount due to pesticide application in the peat soil, thereby slowing down the accumulation of sulfur [18]. Those ions originate from sulfurous and nitrogenous materials as they will be converted to such particles via gas-to-particle conversion. Similarly, heavy metals from the peatland soil are converted to particulate matter by the combustion process in which the characteristics of heavy metals will be contained in the ash suspended in the air [17,18]. The isomeric anhydrous sugar levoglucosan is found in the fine particulate matter during transboundary haze as it is known as a specific and general indicator of biomass emission composition directly emitted into the atmosphere. It is yielded by some anthropogenic activities such as wood burning and biogenic activities, as well as the by-product of secondary photochemical oxidation of organic precursors [16].

Every time the haze phenomenon hits the country, it will inhibit the economic and social developments; the country will need to bear higher medical costs due to drastically increased outpatient and inpatient numbers. This was estimated to be approximately MYR273,000 (USD91,000) based on the data from 2005 until 2009 for 14 haze-related illnesses suffered by the public [5,20]. Moreover, when the haze event worsens and it is declared unsafe, hundreds of schools in the affected areas are closed, communication becomes limited, and people will need to reduce outdoor activities in order to minimize their exposure to it [21,22]. Therefore, the tourism sector is also affected during these events in terms of the economy, human perception, and the environment. Similarly, huge losses occur when the atmospheric visibility decreases; many flights are cancelled due to unsafe conditions to fly, which decreases the number of tourism activities. Furthermore, the number of extinct plants, animals, and biodiversity will increase due to extreme environmental events, while the circumstances can also change tourism moods due to unsatisfactory environmental conditions and lead to a refusal to revisit [5,10,21,22].

Toxic particles in haze harm human health either in cumulative or acute effects like respiratory mortality, cardiovascular illness, and lung cancer to all groups of people, such as adults and children [9,20]. The small particulate size contained in haze can easily penetrate the lung; at worst, it also goes into the bloodstream [22]. Based on the previous study by [4], long-term exposure to air pollution causes an increase in neurological diseases such as cerebrovascular disease and stroke, along with headache and migraine symptoms. Skin is one of the organs that are directly in contact with air pollution, especially during transboundary haze events. The small particulate matter enters through the skin via the pores, causing allergic symptoms and skin damage, such as skin irritation

and inflammation, damaged skin barrier function, increased rate of ageing, and influences on the composition of skin lipid [23–25]. In the worst-case scenario, long-term exposure to air pollution causes an increase in the degree of skin sensitivity and causes skin defect, as well as atopic dermatitis [23].

During transboundary haze events, particulate matter, especially the harmful PM₁₀ pollutant, needs to be predicted for determining and understanding its dispersion behavior in the atmosphere. This can provide information to the concerned people and their awareness to reduce outdoor activities at the affected areas [26]. The Multiple Linear Regression (MLR) model is globally and widely used over many years as a method for air pollution forecasting, which can help to attempt the uncertainty of the future simply by relying on past and current data for decision-making [27–29]. The fundamental basis of this model represents the relationship between the dependent variable and several independent variables, such as meteorological factors and gaseous pollutants, by uncomplicated computation and easy implementation [30,31]. Previous studies in Malaysia have developed the MLR model to predict PM₁₀ concentration, specifically in the East Coast Peninsular Malaysia, based on different site classifications of rural, suburban, urban, and industrial areas and during different types of monsoon to determine its variation during non-haze periods [27,32,33]. Moreover, [34] used the MLR model for PM₁₀ forecasting during haze events that hit Malaysia based on the data from the affected Air Quality Monitoring Station (AQMS). Therefore, this study aims to determine the annual variation of PM₁₀ concentration during transboundary haze events and develop the models to investigate the relationship linking PM₁₀ with meteorological factors and gaseous pollutants, which are contributing to the high PM₁₀ concentration. PM₁₀ is used in this study as the dependent parameter for the prediction purposes as the Malaysian Department of Environment had started to monitor the PM_{2.5} concentration in 2018. However, the episodic haze events need a long period of dataset, which is the main reason to use PM₁₀ in this study. These research findings will help the responsible parties in providing early warning information to come out alongside mitigation and precautionary plans to improve the air quality during haze events, as well as protect the public health.

2. Materials and Methods

2.1. Study Area and Data Collection

Malaysia faces a periodic transboundary haze event due to the large-scale biomass combustion in Sumatra, Indonesia, which happens almost throughout the year and affects the Peninsular Malaysia, Sabah, and Sarawak regions [6]. The haze event usually occurs during the months of January to February and June to August every year [22]. Figure 1 and Table 1 show the location of the affected Air Quality Monitoring Station (AQMS) where the air pollution data are obtained. Data were obtained from the years 2005 until 2014 with the exclusion of years 2007 to 2009 from the Air Quality Division, Department of Environment (DOE), Ministry of Natural Resources and Environment of Malaysia. They were tabulated and arranged based on the chronological timeline of haze episodes in Malaysia [14] in Microsoft Excel Spreadsheet® 2016 and analyzed using Statistical Packages for Social Sciences (SPSS®) version 25. The parameters used in this study were particulate matter with an aerodynamic diameter less than 10 µm (PM₁₀, µg/m³), ambient temperature (T, °C), relative humidity (RH, %), wind speed (WS, km/hr), ground-level ozone (O₃, ppm), nitrogen oxide (NO, ppm), nitrogen dioxide (NO₂, ppm), carbon monoxide (CO, ppm), and sulphur dioxide (SO₂, ppm) to gain a better picture of PM₁₀ variability during transboundary haze. The relationship between these parameters was examined using bivariate correlation analysis in which the PM₁₀ concentration was measured using the β-ray attenuation mass monitor (BAM-1020) as manufactured by Met One Instruments Inc. [27]. Meanwhile, the meteorological parameters were measured using Met One 062 sensor for ambient temperature, Met One 083D sensor for relative humidity, and Met One 010C sensor for wind speed [35]. The Teledyne API Model 400/400E instrument hourly through UV absorption (Beer-Lambert) method with a 0.4ppb detection limit and 0.5% of precision level and Model 200A NO/NO₂/NO_x analyzer, which applies chemiluminescence detection principles was used to detect O₃, NO₂, and NO concentrations

in ambient air, respectively [36,37]. The SO₂ and CO concentrations were monitored and measured by the Teledyne API Model 100A/100E and Teledyne API Model 300/300E, with the lowest detection level of SO₄ at 0.04 ppb by the UV fluorescence method. Meanwhile, the CO level was measured with 0.5% precision and the lowest detection was found at 0.04 ppm by the non-dispersive and infrared absorption (Beer-Lambert) method [37]. Therefore, all instruments were set to have daily calibration using zero air and standard gas concentration to guarantee the quality control and assurance of data monitoring and validation processes, which were revised before they were transferred to the DOE [38]. Missing data might occur due to equipment malfunction and the calibration program, which were thus removed to reduce the bias in the prediction and conservative results [39].

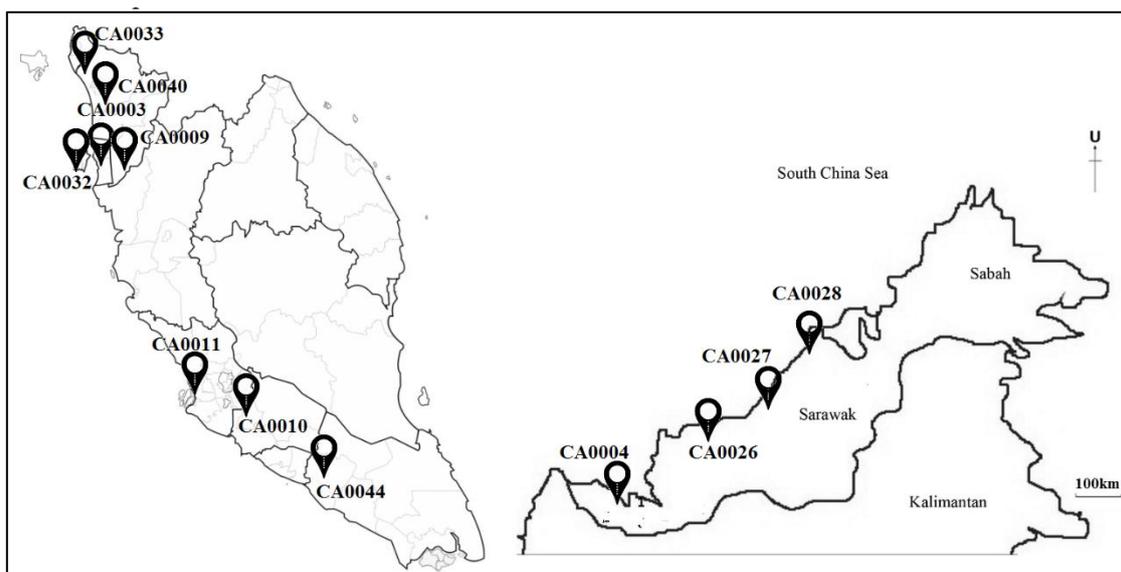


Figure 1. Study Areas.

Table 1. Selected air monitoring stations in Malaysia affected during haze events.

Classification	Station ID	Location	Latitude	Longitude
Industrial	CA0003	Cenderawasih Primary School, Perai	N05° 23.470	E100° 23.213
Industrial	CA0004	Medical Store, Kuching, Sarawak	N01° 33.734	E110° 23.329
Sub-Urban	CA0009	Seberang Jaya II Primary School, Perai, Penang	N05° 23.890	E100° 24.194
Industrial	CA0010	Taman Semarak (Phase II), Nilai, Negeri Sembilan	N02° 49.246	E101° 48.877
Urban	CA0011	Raja Zarina Secondary School, Klang, Selangor	N03° 00.620	E101° 24.484
Sub-Urban	CA0026	Sibu Police Main Office, Sarawak	N02° 18.856	E111° 49.906
Sub-Urban	CA0027	Bintulu Police Station, Sarawak	N03° 10.587	E113° 02.433
Sub-Urban	CA0028	Dato Permaisuri Secondary School, Miri, Sarawak	N04° 25.456	E114° 00.731
Sub-Urban	CA0032	Langkawi Sport Complex, Langkawi, Kedah	N06° 19.903	E099° 51.517
Sub-Urban	CA0033	MADA, Behor Temak, Kangar, Perlis	N06° 25.424	E100° 11.046
Urban	CA0040	Islamic Religious Secondary School, Mergong, Alor Setar, Kedah	N06° 08.218	E100° 20.880
Sub-Urban	CA0044	Vocational Secondary School, Muar, Johor	N02° 03.715	E102° 35.587

2.2. Multiple Linear Regression

Stepwise Multiple Linear Regression (MLR) statistical models were developed in this study with a 95% confidence interval and the dataset was separated according to 7:3 ratios for model development and validation, respectively [40,41]. Based on the data tabulated in the Microsoft Excel Spreadsheet®

2016, 70% of the earlier data in this spreadsheet were used for model development, while the remaining 30% were used for model validation. This statistical model can relate the single dependent variable with two or more independent variables and the equation of the MLR model is stated in Equation (1) [27].

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

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

In the air pollution modelling, the dataset consisted of different types of units and it required normalizing before the development of models in order to improve the accuracy of the numeric computation, which was scaled within the range of 0 to 1. Hence, these normalization steps can interpret all relationships in the data precisely and reduce bias [27,31]. Equation (2) shows the normalization equation used in this study.

$$z_i = (x_i - \min(x))/\max(x) - \min(x) \quad (2)$$

where $x = (x_1, \dots, x_n)$ and z_i is normalised data.

Variance Inflation Factor (VIF) is the multicollinearity assumption together with regression output, whereby the average of VIF values must be below ten to indicate no multicollinearity problem between the independent variables [27,42]. The VIF equation is shown in Equation (3).

$$VIF_i = \frac{1}{1 - R_i^2} \quad (3)$$

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

The Durbin–Watson (DW) test was used to determine the autocorrelation ability of PM_{10} concentration during the current hour to predict PM_{10} in the subsequent hour. The range values of the test must be between 0 and 4 with a value of 2, showing that the residuals are uncorrelated [27,42]. The DW equation is shown per Equation (4).

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

where, n is the number of observations, $e_i = y_i - \hat{y}_i$; (y_i = observed values and \hat{y}_i is the predicted values).

Coefficient of Determination (R^2) was one of the indicators used to determine whether the data provided enough evidence to indicate that the overall models contributed sufficient information for the prediction of PM_{10} concentrations. It also acts as an indicator to measure the extent to which the prediction models fit the data [27,42]. The R^2 equation is shown per Equation (5).

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

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.

2.3. Performance Indicator

The models were evaluated based on the model's error and accuracy by using several performance indicators, namely Root Mean Square Error (RMSE), Normalized Absolute Error (NAE), and Prediction of Accuracy (PA). The best-fitted model is chosen when it has high accuracy in which the PA is closer

to 1 while the minimal error (i.e., RMSE and NAE) is close to 0 [32,35,40]. Equations (6)–(8) show the performance indicators’ formula used in this study.

(a) Root Mean Square Error

$$RMSE = \left(\frac{1}{n} \sum_{i=1}^n [P_i - O_i]^2 \right)^{1/2} \tag{6}$$

(b) Normalized Absolute Error

$$NAE = \frac{\sum_{i=1}^n |P_i - O_i|}{\sum_{i=1}^n O_i} \tag{7}$$

(c) Prediction of Accuracy

$$PA = \sum_{i=1}^n \frac{(P_i - \bar{P})}{(n-1)S_{pred} S_{Obs}} \tag{8}$$

where, n = total number of data; P_i = predicted values; O_i = observed values; \bar{P} = mean of predicted values; S_{pred} = standard deviation of predicted values; S_{obs} = standard deviation of observed values

3. Results and Discussion

The maximum daily average of PM₁₀ concentration during a transboundary haze event in Malaysia in the years of 2014 and 2013 was 995 µg/m³, while the minimum value of PM₁₀ concentration was 150 µg/m³ in all years as this was the minimum concentration for notification as haze [5,40,43]. Table 2 summarizes the descriptive statistics during the chronology of haze events that happened in Malaysia from the years 2005 until 2014. The highest mean concentration of PM₁₀ was recorded in the year 2005 at 274.860 µg/m³ (150.000–994.000 µg/m³), while the lowest mean concentration of PM₁₀ was 178.930 µg/m³ in the year 2011 (150.000–497.000 µg/m³). The Department of Environment Malaysia sets a guideline to justify the status of air quality in the country under the New Ambient Air Quality Standard (NAAQS), whereby the daily 24 h PM₁₀ concentration is 150 µg/m³ [33,44]. Combustion activities due to the agriculture sector such as biomass and peatland combustion in Indonesia for land clearing thus creates haze pollution that has further affected the neighboring countries, especially Malaysia. Altogether, the higher concentration of hazardous PM₁₀ are transported thousands of kilometers by the wind and monsoon [13,17]. The transboundary haze events frequently happen during the dry season in July to October and are extended to southwest monsoon in February to March, which prolongs the combustion activities due to less amount of rainfall and drier condition of the land [5,13].

Table 2. Summary of descriptive statistics during periodic haze events in Malaysia.

Descriptive Statistical	2005 (N = 369)	2006 (N = 918)	2010 (N = 61)	2011 (N = 260)	2012 (N = 402)	2013 (N = 236)	2014 (N = 613)
Mean (µg/m ³)	274.860	199.100	275.100	178.930	204.270	260.650	233.520
Median (µg/m ³)	237.000	180.000	196.000	168.000	181.000	210.000	187.000
Std. Deviation (µg/m ³)	136.103	56.435	170.717	37.299	63.071	141.508	131.163
Variance	18524.096	3184.955	29144.323	1392.180	3977.931	20024.500	17203.855
Minimum (µg/m ³)	150.000	150.000	150.000	150.000	150.000	150.000	150.000
Maximum (µg/m ³)	994.000	573.000	866.000	497.000	533.000	995.000	995.000

The Spearman bivariate analysis was applied as the data were not normally distributed and non-parametric in order to see the relationship linking PM₁₀ with meteorological factors and other gaseous pollutants during the periodic haze events in Malaysia as tabulated in Table 3. The CO ($r = 0.512, p < 0.01$) showed a strong and positive correlation by increasing the concentration of PM₁₀

at the atmosphere. Meanwhile, T ($r = 0.055, p < 0.05$), SO_2 ($r = 0.131, p < 0.01$), NO_2 ($r = 0.059, p < 0.01$), and O_3 ($r = 0.046, p < 0.05$) were weakly and positively correlated to PM_{10} concentration. Similarly, RH was weakly and negatively correlated with $r = -0.076, p < 0.01$. Open burning from land clearing activities at Sumatra (Indonesia) has been contributing to a high concentration of carbon dioxide in the atmosphere during such periodic haze events [3]. The incomplete combustion of carbon thus generates CO_2 concentration via some chemical reactions with another atmospheric constituent [3,5]. Hence, high emission of CO also causes a high emission of particulate matter and toxic organic air contaminants, which have an adverse impact on human health [45,46].

Table 3. Spearman correlation of PM_{10} between meteorological factors and other gaseous pollutants during periodic haze events.

Parameter	PM_{10}	WS	T	RH	NO	SO_2	NO_2	O_3	CO
PM_{10}	1.00	0.043	0.055 *	-0.076 **	-0.025	0.131 **	0.059 **	0.046 *	0.512 **
WS		1.00	0.517 **	-0.537 **	-0.380 **	-0.029	-0.372 **	0.615 **	-0.206 **
T			1.00	-0.820 **	-0.259 **	0.050 *	-0.088 **	0.621 **	0.006
RH				1.00	0.230 **	-0.196 **	0.003	-0.590 **	-0.038
NO					1.00	0.259 **	0.617 **	-0.671 **	0.325 **
SO_2						1.00	0.458 **	-0.141 **	0.356 **
NO_2							1.00	-0.443 **	0.437 **
O_3								1.00	-0.243 **
CO									1.00

Note ** Correlation is significant at the 0.01 level (2-tailed), * Correlation is significant at the 0.05 level (2-tailed).

A statistical model to precisely forecast the PM_{10} concentrations during haze events using MLR was founded using 70% of the dataset, while the remaining 30% were used for model validation. The MLR models were developed based on three different prediction hours, which were to predict the first hour of PM_{10} concentration up until the next three hours of PM_{10} concentration. A summary of the models is tabulated in Table 4. All models were developed from the PM_{10} concentration during the transboundary haze event occurring in Malaysia. Based on the models, the first prediction hour of $PM_{10,t+1}$ had a higher coefficient of determination (R^2) at 0.638 compared to the second ($PM_{10,t+2}$) and third ($PM_{10,t+3}$) prediction hour at 0.452 and 0.353, respectively. Furthermore, the VIF values for all three models (i.e., range between 1.006 and 2.149) were lower than 10, which indicated no multicollinearity problem present between the independent variables. Moreover, the DW test showed that the range values for all models were 2.125, 1.201, and 0.932 for $PM_{10,t+1}$, $PM_{10,t+2}$, and $PM_{10,t+3}$, respectively. Hence, it indicated that all of the models did not have any first-order autocorrelation problem as the range values were still within 0–4 [27,42]. Thus, the $PM_{10,t+1}$ forecasting model, which was considered as the best-fit model in this study, revealed acceptable values of R^2 as discovered in the previous study of PM_{10} forecasting models by [27,40,47] and higher R^2 values compared to the other two models.

Table 4. Model summary for PM_{10} concentration forecasting during transboundary haze events in Malaysia.

Prediction Hour	Model	R^2	VIF	Durbin-Watson
Hour 1, t+1	$PM_{10,t+1} = 0.004 + 0.673 PM_{10} + 0.230 CO - 0.057 NO$	0.638	1.275–2.149	2.125
Hour 2, t+2	$PM_{10,t+2} = 0.019 + 0.597 PM_{10} + 0.141 CO - 0.053 NO$	0.452	1.275–2.149	1.201
Hour 3, t+3	$PM_{10,t+3} = 0.069 + 0.586 PM_{10} - 0.047 RH - 0.046 O_3$	0.353	1.006–1.629	0.932

For the first prediction hour of model forecasting, the significant predictors for $PM_{10,t+1}$ was CO and NO. The predicted $PM_{10,t+1}$ concentration increased by 0.673 unit when the PM_{10} variables increased by one unit, an increase of 0.230 unit by one unit of CO, and a decrease of 0.057 unit by

one unit of NO. Meanwhile, this was also applicable for $PM_{10,t+2}$: the same variable influenced it with $PM_{10,t+1}$ in which one unit of PM_{10} increased one unit of $PM_{10,t+2}$. There was an increase of 0.141 and decrease of 0.053 $PM_{10,t+2}$ by one unit of CO and NO, respectively. Moreover, $PM_{10,t+3}$ was influenced by three different variables, namely PM_{10} , RH, and O_3 . The increase in unit for PM_{10} and one unit decrease of RH and O_3 led to the increase of 0.568 unit and decrease of 0.047 and 0.046 units of $PM_{10,t+3}$, respectively. Generally, the biomass combustion activities emit several compounds such as aerosols in the form of organic carbon, black carbon, and inorganic carbon; greenhouse gases such as carbon dioxide (CO_2); and photochemical reactive gases like CO and nitrogen oxide (NO_x) [48]. The O concentration represents the dominant species after carbon dioxide and it is used as a tracer of biomass burning plumes in the remote troposphere in the form of smoke particles [19,48].

The normal distribution of residuals is illustrated with zero mean and constant variances for all three models in Figure 2a–c. Meanwhile, the fitted values plot with PM_{10} prediction model’s residuals show that the residuals are uncorrelated due to the residuals accumulating around the horizontal band, further indicating that the variance is constant and uncorrelated as depicted in Figure 2d–f. Unfortunately, more residuals dispersed away from the horizontal band as the prediction hour increased to the second to third hour of PM_{10} prediction time. Hence, this condition would hinder the best fit of the models, which reduced the precision of PM_{10} concentration forecasting.

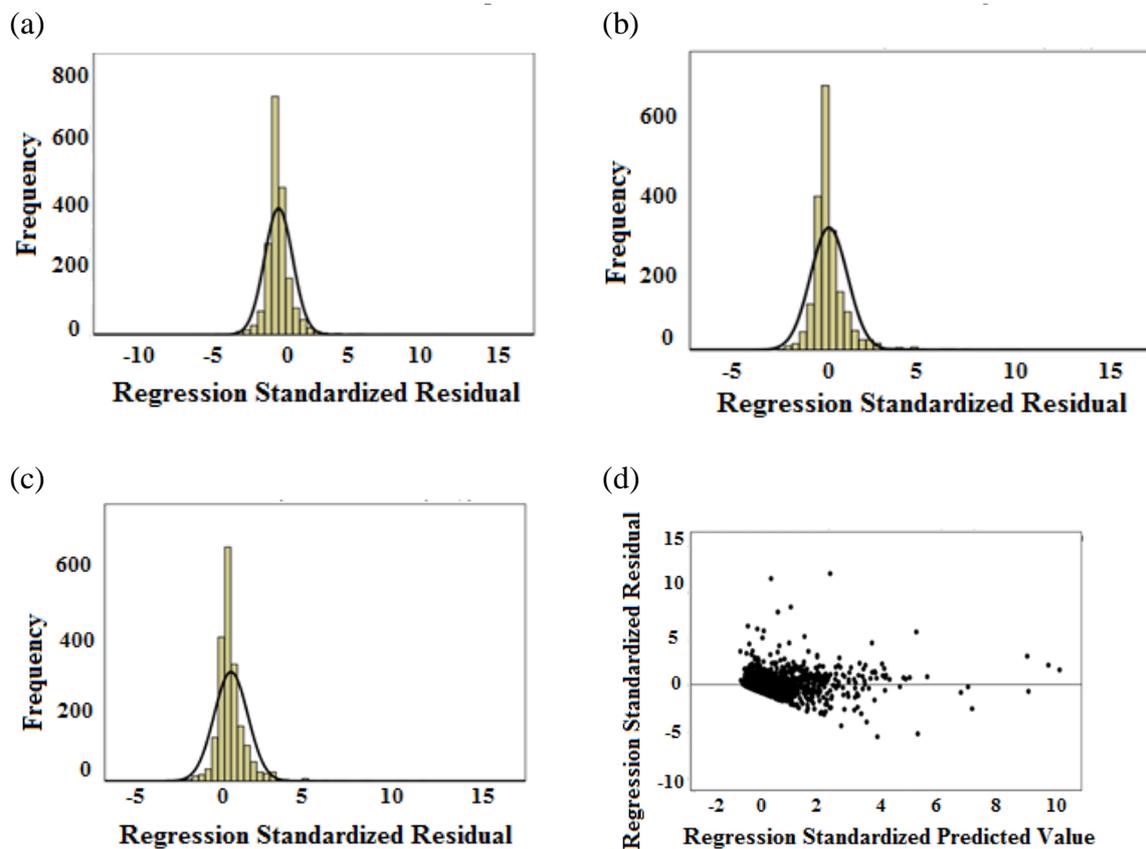


Figure 2. Cont

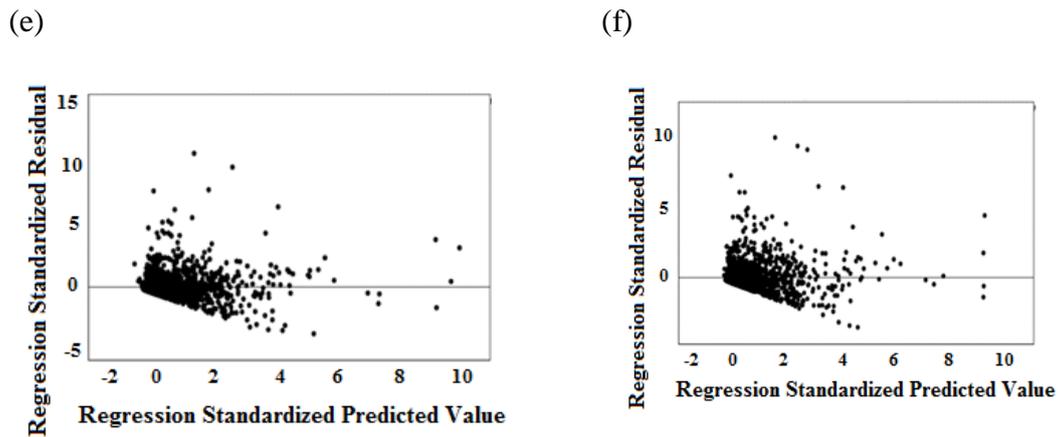


Figure 2. (a) Standardized residual analysis of $PM_{10,t+1}$, (b) standardized residual analysis of $PM_{10,t+2}$, (c) standardized residual analysis of $PM_{10,t+3}$, (d) testing assumption of variance and uncorrelated with mean equal to zero for model $PM_{10,t+1}$, (e) testing assumption of variance and uncorrelated with mean equal to zero for model $PM_{10,t+2}$, (f) testing assumption of variance and uncorrelated with mean equal to zero for model $PM_{10,t+3}$.

The predicted daily PM_{10} concentration was plotted against the observed PM_{10} concentration by prediction hour based on the 30% of haze chronology event data. Figure 3a–c illustrates the goodness-of-fit of the PM_{10} forecasting models in Malaysia. The regression lines drawn showing 95% confidence interval and most of the points were accumulated in between lines A and B, which were the upper and lower ranges of 95% confidence interval for the regression model. The accuracy of the predicted model for the next hour ($PM_{10,t+1}$) was thus proven by having the highest R^2 values of 0.447 compared to models $PM_{10,t+2}$ (0.186) and $PM_{10,t+3}$ (0.129). Hence, the increase of prediction hours of the models reduced their respective performance by increasing some of the errors of the prediction models, rendering it less precise [29,49].

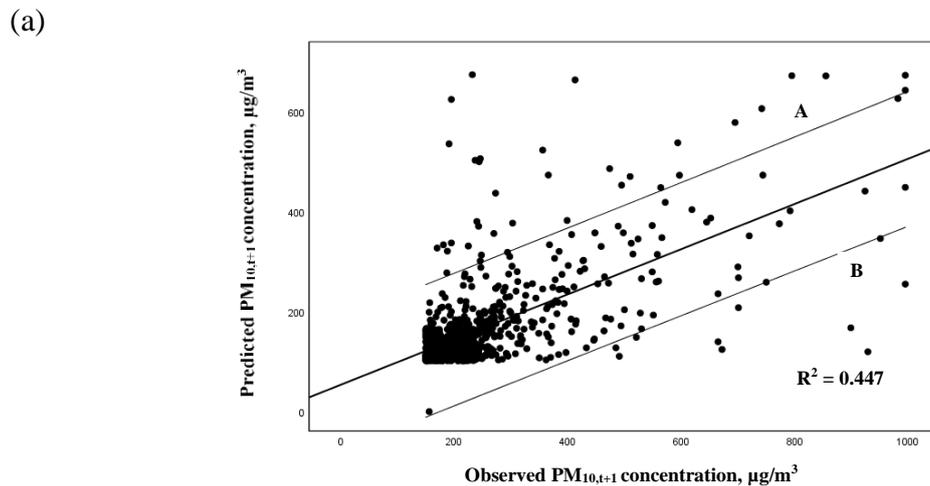
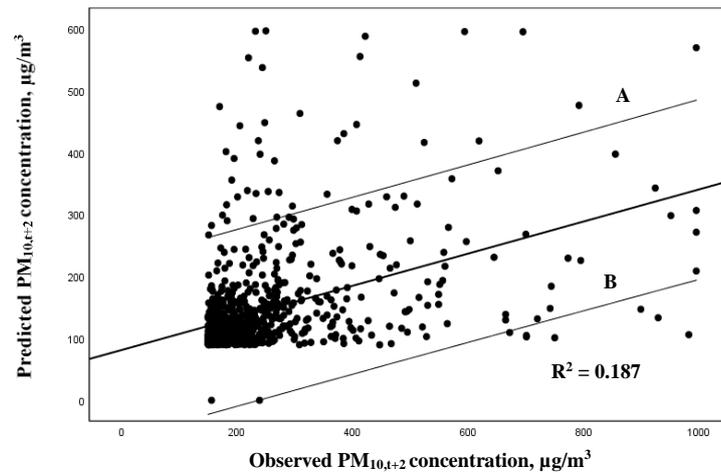


Figure 3. Cont.

(b)



(c)

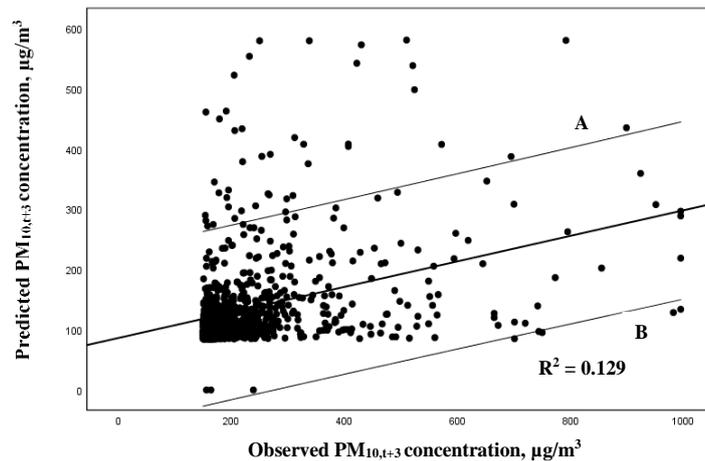


Figure 3. Scatter plot of predicted PM_{10} concentration ($\mu g/m^3$) against observed PM_{10} concentration ($\mu g/m^3$) for (a) $PM_{10,t+1}$ model, (b) $PM_{10,t+2}$ model, and (c) $PM_{10,t+3}$ model.

The RMSE, NAE, and PA components were chosen in this study to measure the performance of the models. RMSE and NAE are considered as suitable indexes for determining model accuracy in which the model is noted as having a high accuracy when their values are close to zero, while the PA value is nearest to 1 [32,50,51]. According to a previous study [32], a comparison of the best statistical PM_{10} forecasting methods with the lowest values of RMSE and NAE and the highest PA value has been conducted to select the best-fit prediction model. Table 5 depicts the performance indicator values for all PM_{10} forecasting models in this study. It showed that the $PM_{10,t+1}$ model yielded the lowest value of RMSE (126.728) and NAE (0.325), while it had the highest PA value (0.668) compared to the $PM_{10,t+2}$, and $PM_{10,t+3}$ models. Based on the results from this performance error and accuracy measures, the $PM_{10,t+1}$ model provided better PM_{10} concentration forecasting capacity compared to model $PM_{10,t+2}$, and model $PM_{10,t+3}$ in Malaysia during transboundary haze events.

Table 5. Summary of performance indicator for all PM₁₀ forecasting models during the transboundary haze events.

Prediction Hour	RMSE	NAE	PA
Hour 1, t+1	126.728	0.325	0.668
Hour 2, t+2	156.506	0.403	0.431
Hour 3, t+3	164.978	0.429	0.359

4. Conclusions

The maximum PM₁₀ concentration was higher during transboundary haze events in the years 2013 and 2014 at 995 µg/m³ and the higher mean values of PM₁₀ concentration was 274.860 µg/m³ (150–994 µg/m³) in the year 2005. Meanwhile, the lowest mean value obtained was 178.930 µg/m³ (150–497 µg/m³) in the year 2011. Next, the Spearman correlation analysis showed that CO was strongly and positively correlated with $r = 0.512$, $p < 0.01$. This was due to its higher emission through large-scale biomass combustion from neighboring countries. Furthermore, the best-fitted MLR model for PM₁₀ prediction during transboundary haze events was PM_{10,t+1}, which had a higher R² of 0.447 compared to PM_{10,t+2} (0.186) and PM_{10,t+3} (0.129). On the other hand, PM_{10,t+1} was the best-fitted model with a higher accuracy of 0.668 and lower values of RMSE (126.728) and NAE (0.325) compared to models PM_{10,t+2} and PM_{10,t+3}. The development of this model may thus aid the responsible parties in obtaining early information that can improve or initiate strategies in mitigating for better air quality.

Author Contributions: Conceptualization, S.A. and M.I.; methodology, S.A. and N.N.L.M.N.; software, W.N.W.M.; A.N.A.; validation, S.A. and N.N.L.M.N.; formal analysis, N.N.L.M.N.; writing—original draft preparation, N.N.L.M.N. and A.A.M.; writing—review and editing, S.A.; visualization, Z.T.A.R.; supervision, S.A., M.I. and W.N.W.M.; project administration, A.M.A.; funding acquisition, S.A., M.I. and A.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Grant Scheme for Research Acculturation of Early Career Researchers by the Malaysian Ministry of Education (RACER/1/2019/TK10/UMT//1) and the APC was funded by Research and Management Center, University Malaysia Terengganu.

Acknowledgments: The authors gratefully acknowledge the Malaysian Department of Environment for the air quality data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Thepnuan, D.; Chantara, S.; Lee, C.T.; Lin, N.H.; Tsai, Y.I. Molecular Markers for Biomass Burning Associated with the Characterization of PM_{2.5} and Component Sources during Dry Season Haze Episodes in Upper South East Asia. *Sci. Total Environ.* **2019**, *658*, 708–722. [[CrossRef](#)] [[PubMed](#)]
2. Dotse, S.Q.; Dagar, L.; Petra, M.I.; Silva, L.C.D. Influence of Southeast Asian Haze Episodes on High PM₁₀ Concentrations across Brunei Darussalam. *Environ. Pollut.* **2016**, *219*, 337–352. [[CrossRef](#)] [[PubMed](#)]
3. Wen, Y.S.; Nor, A.F.M.; Fazilan, N.N.; Sulaiman, Z. Transboundary Air Pollution in Malaysia: Impact and Perspective on Haze. *Nova J. Eng. Appl. Sci.* **2016**, *5*, 1–11.
4. Cheong, K.H.; Ngiam, N.J.; Morgan, G.G.; Pek, P.P.; Tan, B.Y.; Lai, J.W.; Koh, J.M.; Ong, M.E.H.; Ho, A.F.W. Acute Health Impacts of the Southeast Asian Transboundary Haze Problem—A Review. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1–18. [[CrossRef](#)]
5. Latif, M.T.; Othman, M.; Idris, N.; Juneng, L.; Abdullah, A.M.; Hamzah, W.P.; Khan, M.F.; Sulaiman, N.M.N.; Jewaratnam, J.; Aghamohammadi, N.; et al. Impact of Regional Haze towards Air Quality in Malaysia: A Review. *Atmos. Environ.* **2018**, *177*, 28–44. [[CrossRef](#)]
6. Sahani, M.; Zainon, N.A.; Mahiyuddin, W.R.W.; Latif, M.T.; Hod, R.; Khan, M.F.; Tahir, N.M.; Chan, C.C. A Case-Crossover Analysis of Forest Fire Haze Events and Mortality in Malaysia. *Atmos. Environ.* **2014**, *96*, 257–265. [[CrossRef](#)]
7. Forsyth, T. Public Concerns about Transboundary Haze: A Comparison of Indonesia, Singapore, and Malaysia. *Glob. Environ. Chang.* **2014**, *25*, 76–86. [[CrossRef](#)]

8. Oanh, N.T.K.; Ly, B.T.; Tipayarom, D.; Manandhar, B.R.; Prapat, P.; Simpson, C.D.; Liu, L.J.S. Characterization of Particulate Matter Emission from Open Burning of Rice Straw. *Atmos. Environ.* **2011**, *45*, 493–502. [[CrossRef](#)]
9. Li, Y.; Zheng, C.; Ma, Z.; Quan, W. Acute and Cumulative Effects of Haze Fine Particles on Mortality and the Seasonal Characteristics in Beijing, China, 2005–2013: A Time-Stratified Case-Crossover Study. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1–11. [[CrossRef](#)]
10. Sun, J.; Zhang, J.E.; Wang, C.; Duan, X.; Wang, Y. Escape or Stay? Effects of Haze Pollution on Domestic Travel: Comparative Analysis of Different Regions in China. *Sci. Total Environ.* **2019**, *690*, 151–157. [[CrossRef](#)]
11. How, C.Y.; Ling, Y.E. The Influence of PM_{2.5} and PM₁₀ on Air Pollution Index (API). Environmental Engineering, Hydraulics and Hydrology. In Proceedings of the Civil Engineering, Johor, Malaysia, 7–8 June 2016.
12. Afroz, R.; Hassan, M.N.; Ibrahim, N.A. Review of Air Pollution and Health Impacts in Malaysia. *Environ. Res.* **2003**, *92*, 71–77. [[CrossRef](#)]
13. Kusumaningtyas, S.D.A.; Aldrian, E. Impact of the June 2013 Riau Province Sumatera Smoke Haze Event on Regional Air Pollution. *Environ. Res. Lett.* **2016**, *11*, 075007. [[CrossRef](#)]
14. Department of Environment, Malaysia. Malaysia Environmental Quality Report 2015. Available online: <https://www.doe.gov.my/portalv1/en/> (accessed on 15 December 2019).
15. Sun, J.; Wu, F.; Hu, B.; Tang, G.; Zhang, J.; Wang, Y. VOC Characteristics, Emissions and Contributions to SOA Formation during Hazy Episodes. *Atmos. Environ.* **2016**, *141*, 560–570. [[CrossRef](#)]
16. Hoyle, C.R.; Boy, M.; Donahue, N.M.; Fry, J.L.; Glasius, M.; Guenther, A.; Hallar, A.G.; Huff Hartz, K.; Petters, M.D.; Petäjä, T.; et al. A Review of the Anthropogenic Influence on Biogenic Secondary Organic Aerosol. *Atmos. Chem. Phys.* **2011**, *11*, 321–343. [[CrossRef](#)]
17. Jaafar, S.A.; Latif, M.T.; Razak, I.S.; Wahid, N.B.A.; Khan, M.F.; Srithawirat, T. Composition of Carbohydrates, Surfactants, Major Elements and Anions in PM_{2.5} during the 2013 Southeast Asia High Pollution Episode in Malaysia. *Particuology* **2018**, *37*, 119–126. [[CrossRef](#)]
18. Othman, M.; Latif, M.T. Dust and Gas Emissions from Small-Scale Peat Combustion. *Aerosol Air Qual. Res.* **2013**, *13*, 1045–1059. [[CrossRef](#)]
19. Koppmann, R.; Czapiewski, K.V.; Reid, J.S. A Review of Biomass Burning Emissions, Part I: Gaseous Emissions of Carbon Monoxide, Methane, Volatile Organic Compounds, and Nitrogen Containing Compounds. *Atmos. Chem. Phys. Discuss.* **2005**, *5*, 10455–10516. [[CrossRef](#)]
20. Othman, J.; Sahani, M.; Mahmud, M.; Ahmad, M.K.S. Transboundary Smoke Haze Pollution in Malaysia: Inpatient Health Impacts and Economic Valuation. *Environ. Pollut.* **2014**, *189*, 194–201. [[CrossRef](#)]
21. Hou, J.; An, Y.; Song, H.; Chen, J. The Impact of Haze Pollution on Regional Eco-Economic Treatment Efficiency in China: An Environmental Regulation Perspective. *Int. J. Environ. Res. Public Health* **2019**, *16*, 4059. [[CrossRef](#)]
22. Rahman, H.A. Haze Phenomenon in Malaysia: Domestic or Transboundary Factor? *Int. J. Chem. Environ. Biol. Sci.* **2013**, *1*, 597–599.
23. Song, Y.; Guo, S.; Zhang, M. Assessing Customers' Perceived Value of the Anti-Haze Cosmetics under Haze Pollution. *Sci. Total Environ.* **2019**, *685*, 753–762. [[CrossRef](#)] [[PubMed](#)]
24. Rembiesa, J.; Ruzgas, T.; Engblom, J.; Holfors, A. The Impact of Pollution on Skin and Proper Efficacy Testing for Anti-Pollution Claims. *Cosmetics* **2018**, *5*, 1–9. [[CrossRef](#)]
25. Vierkötter, A.; Schikowski, T.; Ranft, U.; Sugiri, D.; Matsui, M.; Krämer, U.; Krutmann, J. Airborne Particle Exposure and Extrinsic Skin Aging. *J. Investig. Dermatol.* **2010**, *130*, 2719–2726. [[CrossRef](#)] [[PubMed](#)]
26. Delavar, M.R.; Gholami, A.; Shiran, G.R.; Rashidi, Y.; Nakhaeizadeh, G.R.; Fedra, K.; Afshar, S.H. A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran. *Int. J. Geo Inf.* **2019**, *8*, 99. [[CrossRef](#)]
27. Abdullah, S.; Ismail, M.; Fong, S.Y. Multiple Linear Regression (MLR) Models for Long Term PM₁₀ Concentration Forecasting during Different Monsoon Seasons. *J. Sustain. Sci. Manag.* **2017**, *12*, 60–69.
28. Mishra, D.; Goyal, P.; Upadhyay, A. Artificial Intelligence Based Approach to Forecast PM_{2.5} during Haze Episodes: A Case Study of Delhi, India. *Atmos. Environ.* **2015**, *102*, 239–248.
29. Zounemat-Kermani, M. Hourly Predictive Levenberg-Marquardt ANN and Multi Linear Regression Models for Predicting of Dew Point Temperature. *Meteorol. Atmos. Phys.* **2012**, *117*, 181–192. [[CrossRef](#)]

30. Yuen, F.S.; Abdullah, S.; Ismail, M. Forecasting of Particulate Matter (PM₁₀) Concentration based on Gaseous Pollutants and Meteorological Factors for Different Monsoons of Urban Coastal Area in Terengganu. *J. Sustain. Sci. Manag.* **2018**, *13*, 3–17.
31. Abdul-Wahab, S.A.; Bakheit, C.S.; Al-Alawi, S.M. Principal Component and Multiple Regression Analysis in Modelling of Ground-Level Ozone and Factors Affecting Its Concentrations. *Environ. Model. Softw.* **2005**, *20*, 1263–1271. [[CrossRef](#)]
32. Abdullah, S.; Ismail, M.; Ahmed, A.N.; Abdullah, A.M. Forecasting Particulate Matter Concentration Using Linear and Non-Linear Approaches for Air Quality Decision Support. *Atmosphere* **2019**, *10*, 667. [[CrossRef](#)]
33. Abdullah, S.; Ismail, M.; Fong, S.Y.; Ahmed, A.N. Evaluation for Long Term PM₁₀ Concentration Forecasting Using Multi Linear Regression (MLR) and Principal Component Regression (PCR) Models. *EnvironmentAsia* **2016**, *9*, 101–110.
34. Hashim, N.I.M.; Noor, N.M. Variations of Particulate Matter (PM₁₀) Concentration during Haze Episodes in Malaysia. In Proceedings of the 2017 Bangkok International Intellectual Property, Invention, Innovation and Technology Exposition, Bangkok, Thailand, 4 January 2017.
35. Awang, N.R.; Ramli, N.A.; Yahaya, A.S.; Elbayoumi, M. Multivariate Methods to Predict Ground Level Ozone during Daytime, Nighttime, and Critical Conversion Time in Urban Areas. *Atmos. Pollut. Res.* **2015**, *6*, 726–734. [[CrossRef](#)]
36. Awang, N.R.; Elbayoumi, M.; Ramli, N.A.; Yahaya, A.S. Diurnal Variations of Ground-Level Ozone in Three Port Cities in Malaysia. *Air Qual. Atmos. Health* **2016**, *9*, 25–39. [[CrossRef](#)]
37. Latif, M.T.; Dominick, D.; Ahamad, F.; Khan, M.F.; Juneng, L.; Hamzah, F.M.; Nadzir, M.S.M. Long Term Assessment of Air Quality from a Background Station on the Malaysian Peninsula. *Sci. Total Environ.* **2014**, *482–483*, 336–348. [[CrossRef](#)] [[PubMed](#)]
38. Banan, N.; Latif, M.T.; Juneng, L.; Ahamad, F. Characteristics of Surface Ozone Concentrations at Stations with Different Backgrounds in the Malaysian Peninsula. *Aerosol Air Qual. Res.* **2013**, *13*, 1090–1106. [[CrossRef](#)]
39. Kang, H. The Prevention and Handling of the Missing Data. *Korean J. Anesthesiol.* **2013**, *64*, 402–406. [[CrossRef](#)] [[PubMed](#)]
40. Abdullah, S.; Ismail, M.; Samat, N.N.A.; Ahmed, A.N. Modelling Particulate Matter (PM₁₀) Concentration in Industrialized Area: A Comparative Study of Linear and Nonlinear Algorithms. *ARPN J. Eng. Appl. Sci.* **2018**, *13*, 8227–8235.
41. Roy, K.; Ambure, P. The Double Cross-Validation Software Tool for MLR QSAR Model Development. *Chemom. Intell. Lab. Syst.* **2016**, *159*, 108–126. [[CrossRef](#)]
42. Ul-Saufie, A.Z.; Yahya, A.S.; Ramli, N.A.; Hamid, N.A. Comparison Between Multiple Linear Regression and Feed Forward Back Propagation Neural Network Models for Predicting PM₁₀ Concentration Level Based On Gaseous and Meteorological Parameters. *Int. J. Appl. Sci. Technol.* **2011**, *1*, 42–49.
43. Department of Environment, Malaysia. Malaysia Environmental Quality Report 2016. Available online: <https://www.doe.gov.my/portalv1/en/> (accessed on 15 December 2019).
44. Latif, M.T.; Abidin, E.Z.; Praveena, S.M. The Assessment of Ambient Air Pollution Trend in Klang Valley. *World Environ.* **2015**, *5*, 1–11.
45. Rajab, J.M.; Tan, K.C.; Lim, H.S.; Matjafri, M.Z. Investigation on the Carbon Monoxide Pollution over Peninsular Malaysia Caused by Indonesia Forest Fires from AIRS Daily Measurement. *Adv. Air Pollut.* **2011**, *8*, 115–136.
46. Bartington, S.E.; Bakolis, I.; Devakumar, D.; Kurmi, O.P.; Gulliver, J.; Chaube, G.; Manandhar, D.S.; Saville, N.M.; Costello, A.; Osrin, D.; et al. Patterns of Domestic Exposure to Carbon Monoxide and Particulate Matter in Households Using Biomass Fuel in Janakpur, Nepal. *Environ. Pollut.* **2017**, *220*, 38–45. [[CrossRef](#)] [[PubMed](#)]
47. Yusof, K.M.K.K.; Azid, A.; Sani, M.S.A.; Samsudin, M.S.; Amin, S.N.M.; Rani, N.L.A.; Jamalani, M.A. The Evaluation on Artificial Neural Networks (ANN) and Multiple Linear Regressions (MLR) Models over Particulate Matter (PM₁₀) Variability during Haze and Non-Haze Episodes: A Decade Case Study. *Malays. J. Fundam. Appl. Sci.* **2019**, *15*, 164–172. [[CrossRef](#)]
48. Petetin, H.; Sauvage, B.; Parrington, M.; Clark, H.; Fontaine, A.; Athier, G.; Blot, R.; Boulanger, D.; Cousin, J.M.; Nédélec, P.; et al. The Role of Biomass Burning as Derived from the Tropospheric CO Vertical Profiles Measured by IAGOS Aircraft in 2002–2017. *Atmos. Chem. Phys.* **2018**, *18*, 17277–17306. [[CrossRef](#)]

49. Wang, Z.; Wang, Y.; Zeng, R.; Srinivasan, R.S.; Ahrentzen, S. Random Forest Based Hourly Building Energy Prediction. *Energy Build.* **2018**, *171*, 11–25. [[CrossRef](#)]
50. Vakili, M.; Sabbagh-Yazdi, S.R.; Khosrojerdi, S.; Kalhor, K. Evaluating the Effect of Particulate Matter Pollution on Estimation of Daily Global Solar Radiation Using Artificial Neural Network Modeling Based on Meteorological Data. *J. Clean. Prod.* **2017**, *141*, 1275–1285. [[CrossRef](#)]
51. Domańska, D.; Wojtylak, M. Explorative Forecasting of Air Pollution. *Atmos. Environ.* **2014**, *92*, 19–30. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).