International Journal of Design & Nature and Ecodynamics
Vol. 16, No. 1, February, 2021, pp. 53-59
Journal homepage: http://iieta.org/journals/ijdne

# Three-Hour-Ahead of Multiple Linear Regression (MLR) Models for Particulate Matter ($PM_{10}$) Forecasting

Amalina Abu Mansor[1], Samsuri Abdullah[2,3*], Nazri Che Dom[4], Nur Nazmi Liyana Mohd Napi[2], Ali Najah Ahmed[5], Marzuki Ismail[1,3], Mohammad Fakhratul Ridwan Zulkifli[2]

[1] Faculty of Science and Marine Environment, Universiti Malaysia Terengganu, Kuala Nerus 21030, Terengganu, Malaysia
[2] Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Kuala Nerus 21030, Terengganu, Malaysia
[3] Institute of Tropical Biodiversity and Sustainable Development, University Malaysia Terengganu, Kuala Nerus 21030, Terengganu, Malaysia
[4] Integrated Mosquito Research Group, Faculty of Health Sciences, Universiti Teknologi MARA, UiTM Cawangan Selangor, Puncak Alam 42300, Selangor, Malaysia
[5] Institute of Energy Infrastructure (IEI), Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), Kajang, Selangor Darul Ehsan 43000, Malaysia

Corresponding Author Email: samsuri@umt.edu.my

https://doi.org/10.18280/ijdne.160107

**Received:** 19 October 2020
**Accepted:** 22 December 2020

***Keywords:***
*air pollution, multiple linear regression, accuracy, forecasting, industrial*

**ABSTRACT**

The increase of air pollutants emission through anthropogenic activities and natural phenomena in the atmosphere can give an adverse impact on human health especially to some groups of people such as children, the elderly, and people that have cardiovascular problems. Multiple Linear Regression (MLR) model establishments for the particulate matter ($PM_{10}$) forecasting can be useful, as it provides early warning information to the local authorities and the communities. We aim to develop MLR models for $PM_{10}$ forecasting in Peninsular Malaysia, specifically in the southern part. In this study, the hourly data of $PM_{10}$, meteorological factors, and gaseous pollutants from the year 2009-2011 had been used. As a result, the next first hour of the MLR prediction model, $PM_{10,t+1}$ has been selected as the best-fitted model as compared to the second and third prediction hour models, $PM_{10,t+2,}$ and $PM_{10,t+3}$, respectively. The $PM_{10,t+1}$ model was explained 61.4% ($R^2$=0.614) variance in the data which is higher compared to model $PM_{10,t+2}$ and $PM_{10,t+3}$ with 42.3% ($R^2$=0.423) and 34.7% ($R^2$=0.347), respectively. Thus, the validation of $PM_{10,\ t+1}$ model also has a high accuracy value of $R^2$ (55.1%) as compared to the other two models. We conclude that the development of MLR models is adequate for $PM_{10}$ forecasting in the industrial area.

## 1. INTRODUCTION

Air pollution is considered the main issue while dealing with the environment. The high rate of urbanization, an increase in motor vehicles, population growth, and industrialization activities had caused an increase in the concentration of various air pollution [1]. One of the criteria pollutants are known as coarse particulate matter ($PM_{10}$) is the notorious air pollution towards human health [2, 3]. The dust has come in different shapes or sizes; hence it can be seen with the naked eye (0.01-100µm). It is also known as the molecular dimension of the dust. Dust or particulate matter can be divided into 3 categories which divided by its diameter ($PM_1$, $PM_{2.5}$, $PM_{10}$) [4]. PM is one of the major air pollutant contributors associated with the air quality status [5]. The emission of PM is influenced by several meteorological and climatology factors such as temperature, relative humidity, rain scavenging potential, radiation, dispersive conditions against re-circulation of air masses, and gaseous formation, dispersion, and transportation [6]. Besides, the rapid growth in the industrial sector causes the air pollutant to unstoppable emitted into the atmosphere layer and it becomes hard to monitor [7]. In 2015, $PM_{10}$ and $PM_{2.5}$ were top five leading towards mortality risk which contributes up to 7.6% of death cases worldwide [8]. It can decrease lung function and increase respiratory diseases such as asthma, sinusitis, shortness of breath, and the development of lung cancer [9]. As to protect public health, various statistical methods are used in PM studies in Malaysia for future prediction of PM [1, 10]. These prediction models are essential to determine the early air pollution information to prevent long-term and short-term health effects [1]. Simultaneous increasing rates in the population, transportation, industrial activities, and urbanization have given adverse indications of air pollutants and out of many gaseous pollutants which caused increasing health problems [11].

Modeling in the air quality field helps to determine the relative contribution between sources of air pollution and their relationship with meteorological factors in determination for a future scenario [11]. Big data in air quality studies can be associated with weather parameters in terms of dependent and independents response via a statistical tool of Multiple Linear Regression (MLR) [10, 11]. Algorithms of the association between dependent and independents variables are expected to be used for prediction purposes. This MLR technique had been widely used for its simple and direct computation by Fong et al. [12]. MLR is used as it can determine more than one predictor for a certain situation and at the same time can simultaneously predict using several independent variables. In air quality studies, the prediction of the dependent variable usually depends on several independent parameters, and these independent parameters are simultaneously inserted as input in the MLR model for real-world presentation. It has also the capability to determine the outliers of the dataset. Few studies applied the MLR or regression method in PM concentration prediction in rural and industrial areas [5, 10, 12]. MLR has several assumptions, including dealing with multicollinearity, and does not have any first-order auto-correlation problem [11]. The temporal prediction of $PM_{10}$ concentration is generally conducted for the next hour (t+1), there is less research conducted study in determining whether the next two (t+2) and three (t+3) hours can significantly predict $PM_{10}$ concentrations. The bias or error in prediction between the hours is the main aim of this study. Thus, in this study, we develop the MLR models for $PM_{10}$ forecasting in Peninsular Malaysia, specifically in the southern part.

## 2. MATERIALS AND METHODS

Economic enhancement in a particular area is in line with industrial development, which might cause environmental impacts. Pasir Gudang was selected because several industries operated in that area which possibly contributes to pollutant emissions such as oleochemical, oil, plastic products, gas, and petrochemical. Air pollution becomes a more significant problem due to rapid growth in the industrial sector, the higher density of vehicles in the traffic, and urbanization activities [3]. The site is located at Pasir Gudang 2, Secondary School (1.4707ºN, 103.895ºE) which is one of the Air Quality Monitoring Stations (AQMSs) in Malaysia (Figure 1).

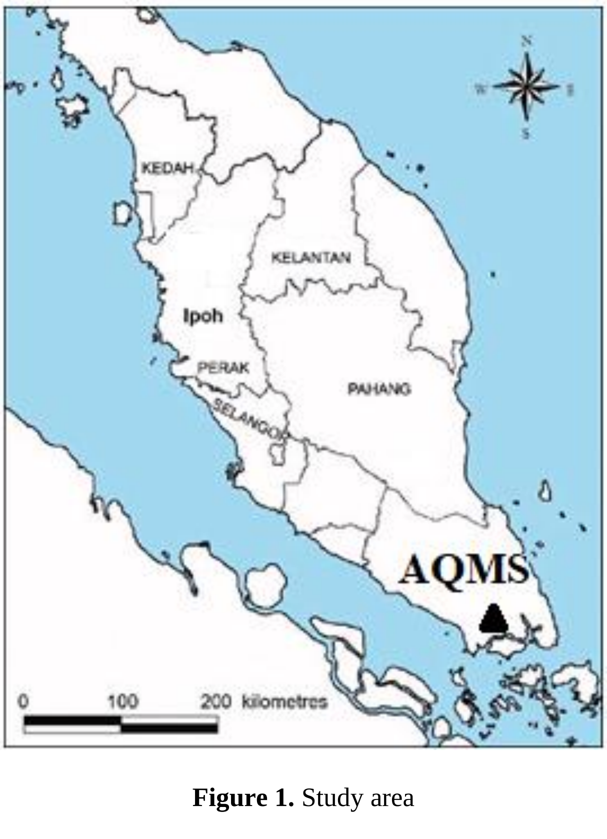


**Figure 1.** Study area

Acquired hourly data from the Malaysian Department of Environment (DOE) for three years (the year 2009-2011) were used. The parameters include carbon monoxide (CO, ppm), ozone ($O_3$, ppm), coarse particulate matter ($PM_{10}$, µg/m$^3$), sulphur dioxide ($SO_2$, ppm), nitrogen oxide ($NO_X$, ppm), nitrogen dioxide ($NO_2$, ppm), wind speed (WS, km/hr), and temperature (T, °C), relative humidity (RH, %).

Incomplete data due to several factors may affect the reliability of the developed model. In this study, the deletion method was used in handling the missing data. Removing the missing value is the common approach in handling data, as long as the minimum data capture is more than 90% completeness or maximum missing data is 10% per year in the study period [10, 11]. All data captured for each year in this study fulfill this standard and the incomplete data row as shown in Table 1. We used Microsoft Excel Spreadsheet® 2018 and Statistical Package for Social Sciences (SPSS®) Software version 23 for data analysis.

**Table 1.** The percent of incomplete data rows

| Year | % of incomplete data rows |
|---|---|
| 2009 | 9.84% |
| 2010 | 3.51% |
| 2011 | 4.25% |

The Multiple Linear Regression (MLR) model establishes the relationship between the dependent and independent variables. Mathematically, the MLR is as in Eq. (1).

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

where, $b_i$ are regression coefficients, $X_i$ are independent variables such as carbon monoxide (CO), ozone ($O_3$), coarse particulate matter ($PM_{10}$), sulphur dioxide ($SO_2$), nitrogen oxide ($NO_X$), nitrogen dioxide ($NO_2$), temperature (T), relative humidity (RH) and wind speed (WS) and $\varepsilon$ is a stochastic error related with the regression. There are two assumptions of the MLR model, there are no multicollinearity problems and the models do not have any first auto-correlation problem. The stepwise MLR model was developed based on a 95% confidence interval. 70% of the dataset is used for model development and the rest for model validation. The residuals are assumed to have a normal distribution and constant variance for all models.

Normalization is required as each parameter contains different types of units. The normalization is ranged all the parameters value from 0 to 1 [0 1] to avoid the biased [6]. The normalization data is obtained by applying Eq. (2).

$$Z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{2}$$

where, $x = (x_1, \ldots, x_n)$ and $Z_i$ is the *ith* normalized data.

The multicollinearity problem happens when the independent variables are correlated with each other. In this study, the model assumed no multicollinearity problem has occurred. This multicollinearity issue is proved by Variable Inflation Factor (VIF) (VIF<10). The VIF Eq. as in Eq. (3).

$$VIF_i = \frac{1}{1 - R_i^2} \tag{3}$$

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

Autocorrelation is recognized via the Durbin-Watson (D-W) Test as shown in Eq. (4). It confirms the capability of the dependent parameter in the current state to predict for the next state. Durbin-Watson (D-W) test value must range between 0-4 and if the value is 2, it shows that the residuals are uncorrelated.

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

where, $n$= total number measurement at a particular site, $e_i = y_i - \bar{y}_i$, as $y_i$ is observed value and $\bar{y}_i$ is predicted value.

The Coefficient of Determination ($R^2$) is an indicator to identify the relationship and strength of each variable whether the prediction Eq. is fitted with data. It is also can be used to prove that the model was able to convey adequate information for the forecasting of $PM_{10}$ concentrations. The Coefficient of Determination ($R^2$) Eq. is stated in Eq. (5):

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

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

## 3. RESULTS AND DISCUSSION

$PM_{10}$ concentration for the year 2009-2011 was between 7 µg/m$^3$ to 488 µg/m$^3$. The maximum value of $PM_{10}$ was exceeded by New Malaysian Ambient Air Quality Standards (NMAAQS) which was caused by transboundary haze, industrial emission, and also motor vehicles [13]. The maximum concentration was caused by haze in October 2010 that affected the southern region of Peninsular Malaysia which the main source caused by biomass burning [14, 15]. Table 2 shows descriptive statistics for all parameters used in this study. All parameters were found within the NMAAQS except $PM_{10}$.

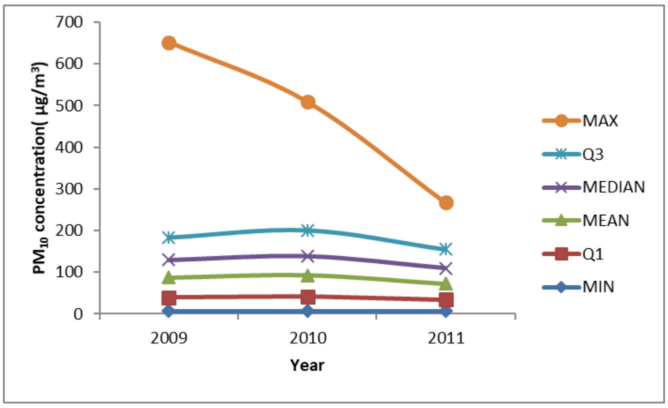


**Figure 2.** Daily $PM_{10}$ concentration

The descriptive statistics for $PM_{10}$, gaseous pollutants, and meteorological factors are summarized in Table 2. Figure 2 shows that the higher maximum value for $PM_{10}$ was in 2009 which is 468 µg/m$^3$. The data set shows that averaged 1 year exceeding the NMAAQS which is 40 µg/m$^3$ and the average of observed years were 47 µg/m$^3$ (2010) and 51 µg/m$^3$ for the year 2011 due to local sources such as transportation, industrial activities and open burning effect [11, 13]. The mean concentrations for other gaseous pollutants complied with NMAAQS and the temperature (T), relative humidity (RH), and wind speed (WS) was in between the normal mean range in Malaysia which is in between 26.45-27.89°C for temperature, 79.36%-83.9% for relative humidity and 1.01-1.82 km/hr for wind speed [6]. The transboundary haze hit Malaysia as a result of a forest fire from May to September 2011 in Indonesia. In turn, this episode deteriorates the air quality in Malaysia [16]. Generally, transboundary haze hit Malaysia every year, as a part of air pollution problems besides two main sources of emission in the country which are stationary and mobile. The standard deviation for $PM_{10}$ was 26.48 µg/m$^3$ and the variance was 701.38 which shows that the data were dispersed randomly with a narrower spread of measurement and proved that the data have comparatively fewer higher or lower value [17].

Correlation analysis in this study was conducted using Spearman correlation as the data are not normally distributed [18]. The Spearman correlation analysis is shown in Table 3, signifying the strong relationship of $PM_{10}$ with other gaseous pollutants.

$PM_{10}$ showed that there is a positive correlation with wind speed (r=0.248, p<0.01) and the other meteorological factors show an inverse correlation with relative humidity (r=-0.124) and temperature (r=-0.043, p<0.01). This is due to $PM_{10}$ concentrations are high caused by the lower temperature ranges. Particulate matter is favorable to be transported at higher temperature [19]. Wind speed determines the air pollutant transportation and dispersion which positively correlated in this study. High wind speed will inhibit the high diffusion of $PM_{10}$ and make a high concentration of $PM_{10}$ concentration in the atmosphere. Relative humidity (RH) and particle deposition are directly proportionate. Particle deposition due to gas-to-particle conversion is favorable at higher RH and low temperature by the mechanisms of evapotranspiration [3, 13, 19]. This theory proved there is a negative correlation between $PM_{10}$ and relative humidity in this study.

The other pollutants such as $SO_2$, $NO_X$, and $NO_2$ showed a moderate and strong positive correlation with $PM_{10}$. NO and $NO_X$ showed highly positive correlation with $PM_{10}$ and $NO_x$ (r=0.493, p<0.01) and $NO_2$ (R=0.501, p<0.01). Transportation emission released the NO into the atmosphere, which then transformed into $NO_2$ as a secondary pollutant. The $NO_x$ is important for the production of ozone. The transformation depending on the local combustion and meteorology [20]. $SO_2$ is released as a by-product when the combustion of coal and petroleum is initiated [8]. Weather parameters such as wind speed, relative humidity, and temperature are important factors in influencing the dispersion and deposition of air pollutions at different locations [17]. Table 3 showed that most of the pollutants, especially $NO_2$, NO, and $NO_X$ have a weak negative correlation with temperature; NO (r=-0.228, p<0.05), $NO_2$ (r=-0.151, p<0.01) and $NO_X$= (r=-0.262, p<0.01) and same go with wind speed.

The model summary of MLR is shown in Table 4. Before the inputs are feed in the MLR model, the inputs are normalized accordingly with Eq. (2) as they have different units of measurement. This to ensure the inputs have similar contributions towards the dependent variable. The models in this study included the forecasting for next one hour ($PM_{10,t+1}$), 2 hours ($PM_{10,t+2}$) and 3 hours ($PM_{10,t+3}$). The highest $R^2$ ($R^2$ = 0.614) was obtained for model $PM_{10,t+1}$, explaining 61.4% variance in data, 42.3% for model $PM_{10,t+2,}$ and 34.7% for $PM_{10,t+3}$. The VIF values are less than 10 which signifies there is no multicollinearity problem in the MLR models. The range of this study was in between the previous forecasting study for $PM_{10}$ in the industrial area which is 1.045-4.203 [2, 6]. The models did not face any autocorrelation problems as the Durbin-Watson (D-W) Test was less than 4 which ranged from 0.017-3.199 for all models. The Durbin-Watson (D-W) statistics point out that the MLR models for the next one hour (D-W=2.043), next two hours (D-W=1.211), and next three day (D-W= 0.995) did not face any first-order autocorrelation problem. The D-W statistics are calculated accordingly with Eq. (4).

Overall, models display the temperature, $PM_{10}$, $NO_2$ and CO have positive influenced towards dependent variables which are $PM_{10,\ t+1}$, $PM_{10,t+2,}$ $PM_{10,t+3}$ concentration. In a tropical country of Malaysia, the parameter of temperature played an important role in influencing the particulate matter, simultaneously, it causes variation in the wind, and dilution of pollutants might occur [15]. From the Eq.s in Table 4, we can determine air pollution mostly composed of $PM_{10}$, $NO_{2,}$ and CO which is usually caused by local traffic emissions. It was found that the model for $PM_{10,\ t+1}$ has significant predictors of $PM_{10}$, $NO_2$, WS, RH, CO, and T. $PM_{10,t+1}$ concentration increased by 0.551 unit when $PM_{10}$ variable increases by one unit, 0.035 unit when $NO_2$ increased, caused increased of a dependent variable ($PM_{10,t+1}$) up to 0.027 unit and 0.014 when increasing 1 unit of wind speed and decreasing relative humidity. Increasing one unit of carbon monoxide and temperature can increase 0.025 and 0.019 units of 1-hour $PM_{10}$ concentration for the developed model. In deciding the adequacy of the statistical models, residual or error played an important role to shows any systematic information by considered the pattern of the model. Figure 3 shows the graph residuals for the next hours' prediction. Normal distributions were depicted via the symmetrical graph. The residuals at all stations also show constant variances plots in Figure 4.

**Table 2.** Summary of descriptive statistics for the study area

| Descriptive statistics | Mean | Median | Std. Deviation | Variance | Min | Max |
|---|---|---|---|---|---|---|
| CO (ppm) | 0.60 | 0.46 | 0.48 | 0.23 | 0.00 | 4.85 |
| $O_3$ (ppm) | 0.13 | 0.01 | 1.01 | 0.00 | 0.00 | 0.12 |
| $PM_{10}$ (µg/m$^3$) | 48.06 | 43.00 | 26.48 | 701.38 | 7.00 | 468.00 |
| $SO_2$ (ppm) | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.12 |
| $NO_X$ (ppm) | 0.03 | 0.02 | 0.02 | 0.00 | 0.00 | 0.24 |
| NO (ppm) | 0.01 | 0.00 | 0.00 | 0.02 | 0.00 | 0.21 |
| $NO_2$ (ppm) | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.06 |
| T (ºc) | 27.49 | 26.70 | 3.05 | 9.29 | 21.40 | 37.80 |
| RH (%) | 82.48 | 86.00 | 11.37 | 129.26 | 41.00 | 98.00 |
| WS (km/hr) | 1.05 | 1.85 | 2.99 | 8.97 | 0.8 | 20.90 |

**Table 3.** Spearman correlation of gaseous pollutants and meteorological factors

| Parameter | CO | $O_3$ | $PM_{10}$ | $SO_2$ | $NO_X$ | NO | $NO_2$ | TEMP | RH | WS |
|---|---|---|---|---|---|---|---|---|---|---|
| CO | 1.000 | -.626** | .413** | .010 | .620** | .569** | .479** | -.462** | .535** | -.511** |
| $O_3$ | | 1.000 | -.248** | -.074** | -.662** | -.641** | -.501** | .663** | -.757** | .686** |
| $PM_{10}$ | | | 1.000 | .389** | .493** | .332** | .501** | -.043** | -.124** | .248** |
| $SO_2$ | | | | 1.000 | .562** | .398** | .631** | .265** | -.164** | -.058** |
| $NO_X$ | | | | | 1.000 | .883** | .895** | -.262** | .388** | -.474** |
| NO | | | | | | 1.000 | .650** | -.228** | .365** | -.379** |
| $NO_2$ | | | | | | | 1.000 | -.151** | .237** | -.398** |
| TEMP | | | | | | | | 1.000 | -.902** | .593** |
| RH | | | | | | | | | 1.000 | -.661** |
| WS | | | | | | | | | | 1.000 |

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

**Table 4.** Summary of the MLR model for $PM_{10}$

| MLR Algorithms | $R^2$ | Range of VIF | Durbin Watson Statistics |
|---|---|---|---|
| $PM_{10,t+1}$ = 0.044+0.551$PM_{10}$+ 0.035$NO_2$ -0.027WS-0.014RH + 0.025 CO + 0.019T | 0.614 | 1.240-3.199 | 2.043 |
| $PM_{10,t+2}$=0.061+0.340$PM_{10}$+0.041$NO_2$ + 0.030T -0.31WS -0.016RH +0.035 $SO_2$+0.031CO -0.02 NO | 0.423 | 1.240-4.113 | 1.211 |
| $PM_{10,t+3}$ =0.069+ 0.252$PM_{10}$ + 0.050$NO_2$+ 0.038T +0.043 CO- 0.027WS -0.017 RH | 0.347 | 0.017-0.078 | 0.995 |

Notes: $PM_{10}$ = particulate matter, $NO_2$ = nitrogen dioxide, WS = wind speed, RH = relative humidity, CO = carbon monoxide, T = temperature, $SO_2$ = sulphur dioxide, NO = nitrogen oxide

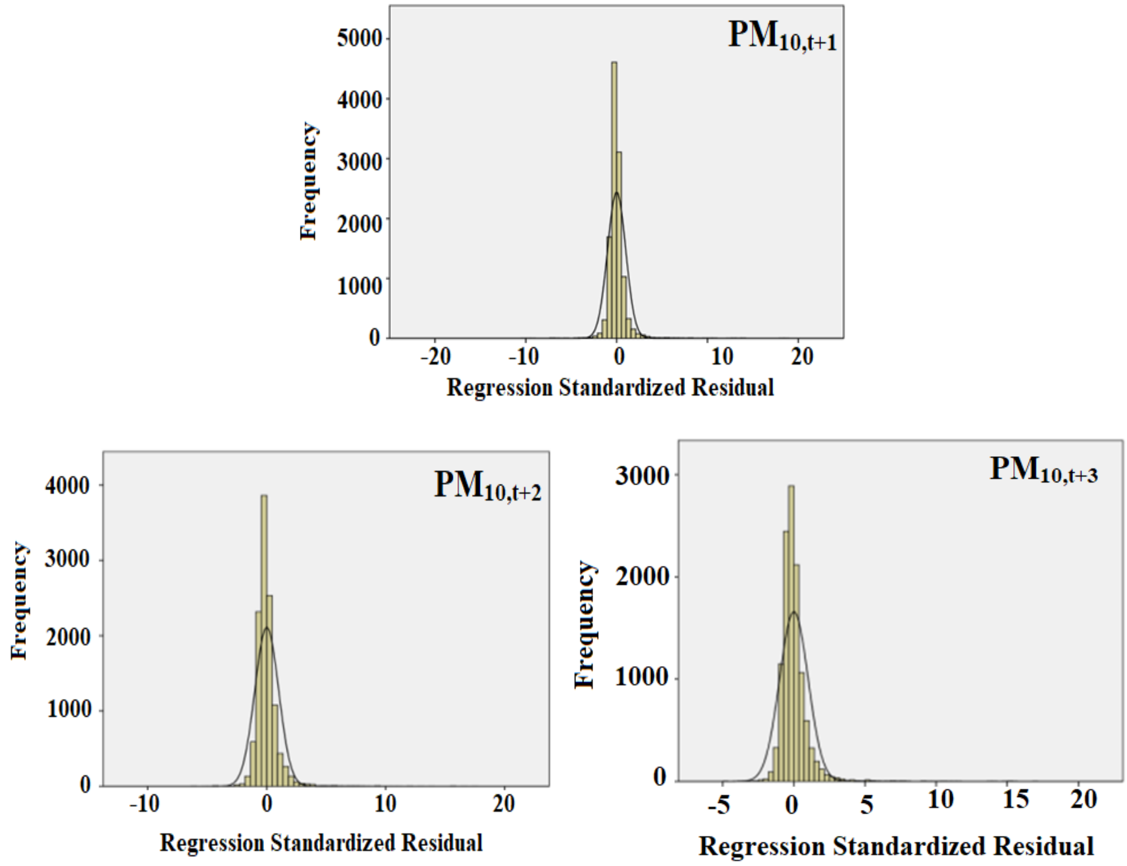


**Figure 3.** Standardized residual analysis of $PM_{10}$ for $PM_{10,t+1}$, $PM_{10,t+2}$, $PM_{10,t+3}$

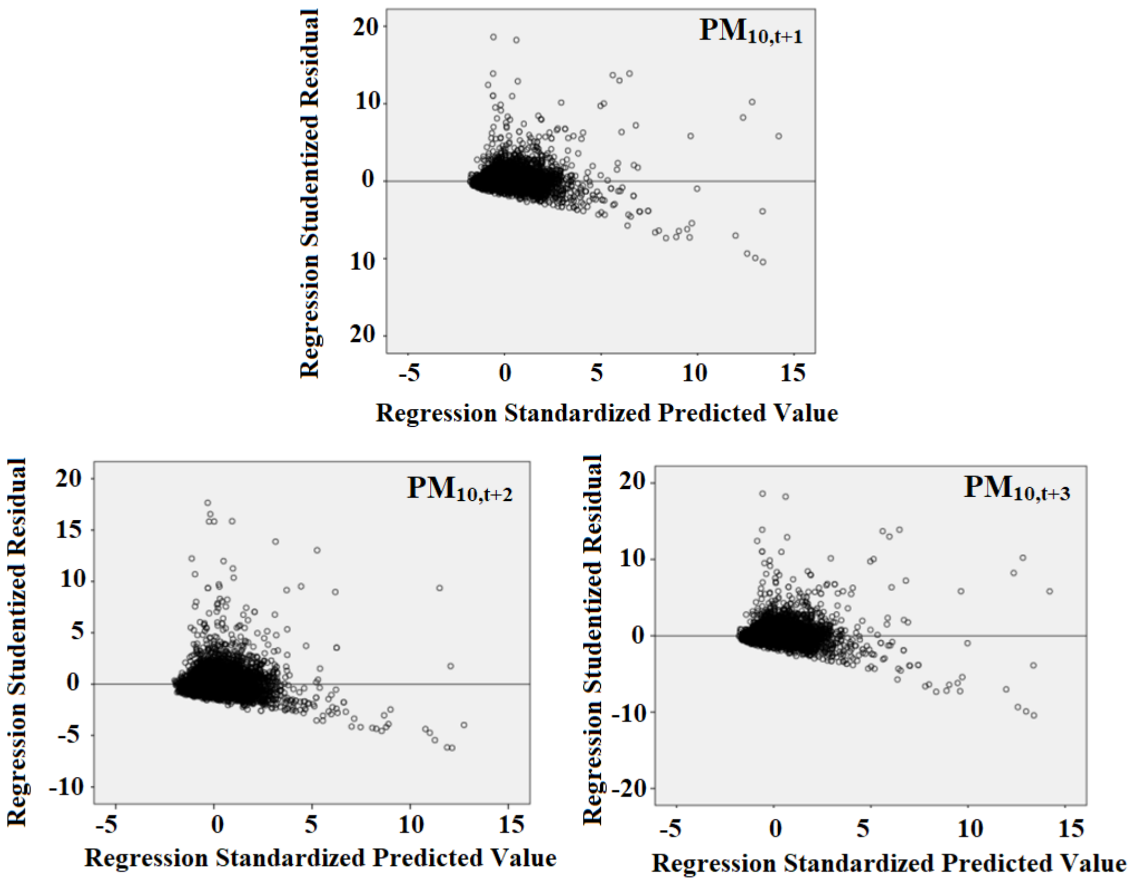


**Figure 4.** Testing assumption of variance and uncorrelated with mean equal zero for $PM_{10}$ for $PM_{10,t+1}$, $PM_{10,t+2}$, $PM_{10,t+3}$

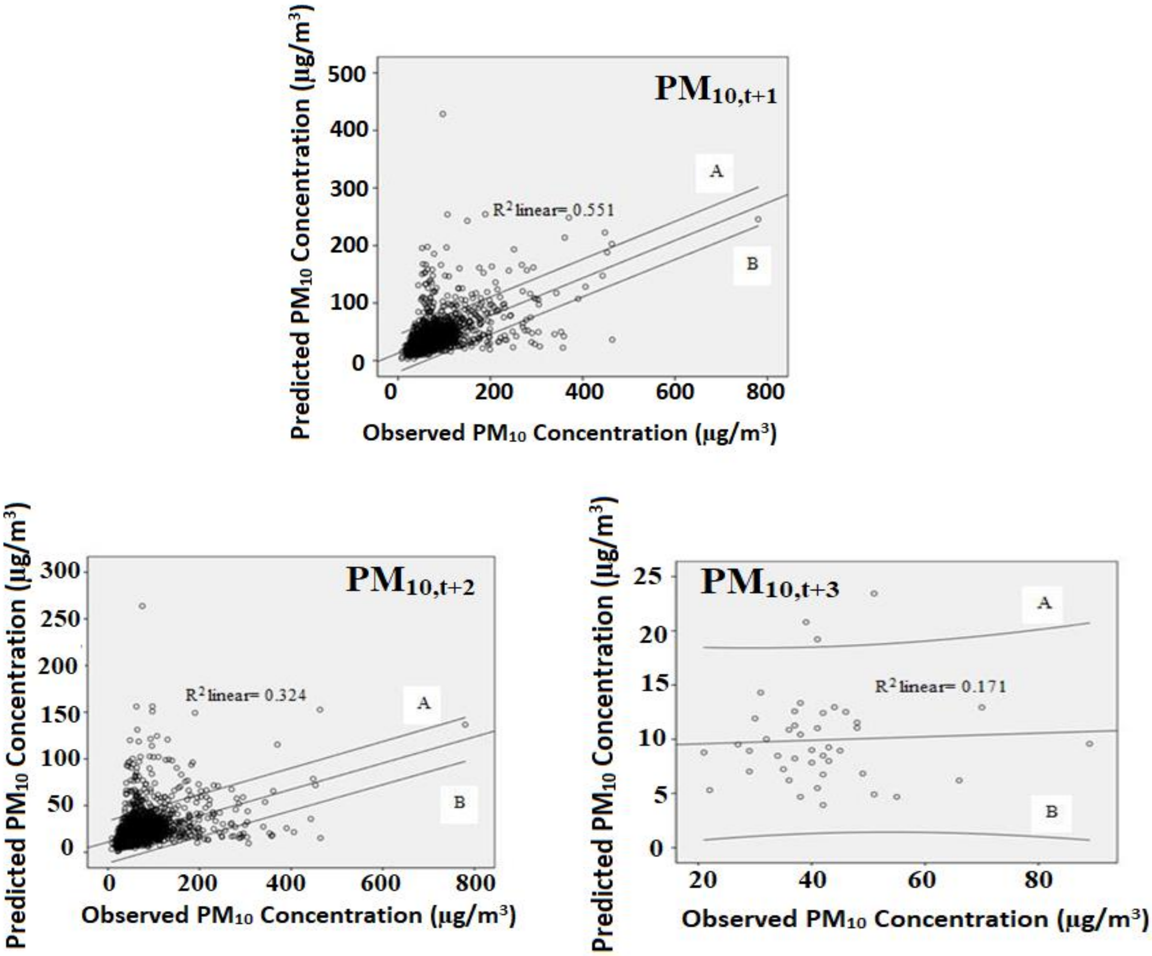


**Figure 5.** Scatter plot for predicted $PM_{10}$ concentration (µg/m$^3$) of $PM_{10}$ of $PM_{10,t+1}$, $PM_{10,t+2}$, $PM_{10,t+3}$

Model verification was conducted to predict the next hours' $PM_{10}$ concentration. The goodness-of-fit for the MLR model is shown in Figure 5. The confidence interval is set at 95%. The upper and lower limit is notified by Point A and point B, respectively. The point that exceeds the upper and lower limit shows the extreme value that appeared from haze episodes that happened in June, July 2009, and October 2010 [15]. $PM_{10}$ was the dominant fraction during the haze and triggered an increased concentration of coarse particulate matter $PM_{10}$, 4-5 times higher than a non-haze day, and 2 times higher for fine particulate ($PM_{2.5}$) besides surrounding activities in that area itself [16]. $R^2$ ranges between 0.171-0.551. The first model which is the prediction of the next one hour shows 55.1%, the second model which is next to two hours shows 32.4% of $R^2$ and the third model which is the prediction next three hours show 17.1%.

## 4. CONCLUSION

Modeling in the air quality field helps to determine the relative contribution between sources of air pollution and their relationship with meteorological factors in determination for a future scenario. In the development of MLR models, it was found that the model for prediction of the next one-hour ($R^2$ = 0.614) is better as compared to the next two ($R^2$ = 0.423), and three ($R^2$ = 0.347) hours for $PM_{10}$ concentrations. Particulate matter ($PM_{10}$) concentration is strongly influenced by meteorological parameters (wind speed (r = 0.248), temperature (r = -0.043), and relative humidity (r = -0.124)) and gaseous pollutants ($NO_2$ (r = 0.501), $SO_2$ (r = 0.389), CO (r = 0.413), $O_3$ (r = -0.248), and NO (r = 0.332)). The developed MLR models in this study are relevant for $PM_{10}$ forecasting as it allows the local authorities to implement strategies for better management on air quality as the management on particulate matter pollution is very dynamic caused by the unpredictability of climate change, and weather all intrinsically connected.

## ACKNOWLEDGMENT

This study is funded by the Fundamental Research Grant Scheme (FRGS) by the Malaysian Ministry of Education (FRGS/1/2019/TK10/UMT/02/4) (VOT 59606). We also would like to thanks the Air Quality Division, Malaysian Department of Environment (DOE) for air quality data.

## REFERENCES

[1] Mohamed, R.M.S.R., Rahim, A.F.H., Kassim, A.H.M. (2016). A monitoring of air pollutants (CO, $SO_2$ and NO) in ambient air near an industrial area. MATEC Web of Conferences, 47(05022): 1-6. https://doi.org/10.1051/matecconf/20164705022

[2] Ng, K.Y, Awang, N. (2018). Multiple linear regression and regression with time series error models in

forecasting $PM_{10}$ concentrations in Peninsular Malaysia. Environment Monitoring Assessment, 190(63): 5-11. https://doi.org/10.1007/s10661-017-6419-z

[3] Yusuf, K.M.K.K., Azid, A., Sani, M.S.A., Samsudin, M.S., Amin, S.N.S.M., Rani, N.L.A., Jamalani, M.A. (2019). The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models over particulate matter ($PM_{10}$) variability during haze and non-haze episodes: A decade case study. Malaysian Journal of Fundamental and Applied Science, 15(2): 164-172. https://doi.org/10.11113/mjfas.v15n2.1004

[4] Chamseddine, A., Alameddine, I., Hatzopoulou, M., El-Fadel, M. (2019). Seasonal variation of air quality in hospitals with indoor–outdoor correlations. Building and Environment, 148: 689-700. https://doi.org/10.1016/j.buildenv.2018.11.034

[5] Ahmad, M., Alam, K., Tariq, S., Anwar, S., Nasir, J., Mansha, M. (2019). Estimating fine particulate concentration using a combined approach of linear regression and artificial neural network. Atmospheric Environment, 219: 117050. https://doi.org/10.1016/j.atmosenv.2019.117050

[6] Abdullah, S., Ismail, M., Samat, N.N.A., Ahmed, A.N. (2018). Modelling Particulate Matter ($PM_{10}$) concentration in industrialized area: A comparative study of linear and nonlinear algorithms. ARPN Journal of Engineering and Applied Sciences, 13(20): 8227-8235.

[7] Malashock, D., Khwaja, H.A., Fatmi, Z., Siddique, A., Lu, Y., Lin, S., Carpenter, D. (2018). Short-Term association between black carbon exposure and cardiovascular diseases in Pakistan's largest megacity. Atmosphere, 9(11): 420. https://doi.org/10.3390/atmos9110420

[8] Zielinska, M.A., Hamulka, J. (2019). Protective effect of breastfeeding on adverse health effects induced by air pollution: Current evidence and possible mechanisms. Environmental Research and Public Health, 16(21): 4181-4211. https://doi.org/10.3390/ijerph16214181

[9] Oliveira, M., Slezakova, K., Delerue-Matos, C., Pereira, M.C., Morais, S. (2019). Children environmental exposure to particulate matter and polycyclic aromatic hydrocarbons and biomonitoring in school environments: A review on indoor and outdoor exposure levels, major sources and health impacts. Environment International, 124: 180-204. https://doi.org/10.1016/j.envint.2018.12.052

[10] Ul-Saufie, A.Z., Yahaya, A.S., Ramli, N.A., Rosaida, N., Hamid, H.A. (2013). Future daily $PM_{10}$ concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA). Atmospheric Environment, 77: 621-630. https://doi.org/10.1016/j.atmosenv.2013.05.017

[11] Elbayoumi, M., Ramli, N.A., Yusuf, N.F.F.M. (2015). Development and comparison of regression models and feedforward backpropagation neural network models to predict seasonal indoor $PM_{2.5-10}$ and $PM_{2.5}$ concentrations in naturally ventilated schools. Atmospheric Pollution Research, 6(6): 1013-1023. https://doi.org/10.1016/j.apr.2015.09.001

[12] Fong, S.Y., Abdullah, S., Ismail, M. (2018). Forecasting of particulate matter ($PM_{10}$) concentration based on gaseous pollutants and meteorological factors for different monsoons of urban coastal area in Terengganu. Journal of Sustainability Science and Management, 13(5): 3-17.

[13] Abdullah, S., Napi, N.N.L.M., Ahmed, A.N., Mansor, W.N.W., Mansor, A.A., Ismail, M., Abdullah, A.M., Ramly, Z.T.A. (2020). Development of multiple linear regression for particulate matter ($PM_{10}$) forecasting during episodic transboundary haze event in Malaysia. Atmosphere, 11(3): 289. https://doi.org/10.3390/atmos11030289

[14] George, S., Chua, M.L., Wei, D.Z.Z., Das, R., Bijin, V.A., Connolly, J.E., Lee, K.P., Yung, C.F., Teoh, O.H., Thomas, B. (2020). Personal level exposure and hazard potential of particulate matter during haze and non-haze periods in Singapore. Chemosphere, 243: 125401. https://doi.org/10.1016/j.chemosphere.2019.125401

[15] Othman, J., Sahani, M., Mahmud, M., Ahmad, M.K.S. (2014). Transboundary smoke haze pollution in Malaysia: Inpatient health impacts and economic valuation. Environmental Pollution, 189: 191-201. https://doi.org/10.1016/j.envpol.2014.03.010

[16] Manan, N.A., Manaf, M.R.A., Hod, R. (2018). The Malaysia haze and its health economic impact: A literature review. Malaysian Journal of Public Health Medicine, 18(1): 38-45.

[17] Yotova, G.I., Tsitouridou, R., Tsakovski, S.L., Simeonov, V.D. (2016). Urban air quality assessment using monitoring data of fractionized aerosol samples, chemometrics and meteorological conditions. Journal of Environmental Science and Health - Part A Toxic/Hazardous Substances and Environmental Engineering, 51(7): 544-552. https://doi.org/10.1080/10934529.2016.1141620

[18] Ismail, M., Suroto, A., Abdullah, S. (2015). Response of Malaysian local rice cultivars induced by elevated ozone stress. Environment Asia, 8(1): 86-93. https://doi.org/10.14456/ea.2015.11

[19] Li, X., Feng, Y.J., Liang, H.Y. (2017). The impact of meteorological factors on $PM_{2.5}$ variations in Hong Kong. IOP Conference Series: Earth and Environmental Science, 78: 012003. https://doi.org/10.1088/1755-1315/78/1/012003

[20] Dandotiya, B., Jadon, N., Sharma, H.K. (2018). Effects of meteorological parameters on gaseous air pollutant concentrations in urban area of Gwalior city, India. Environmental Claims Journal, 31(1): 32-43. https://doi.org/10.1080/10406026.2018.1507508