Journal of Sustainability Science and Management
Volume 12 *Number* 1, June 2017: 60-69
ISSN: 1823-8556

MULTIPLE LINEAR REGRESSION (MLR) MODELS FOR LONG TERM PM_{10} CONCENTRATION FORECASTING DURING DIFFERENT MONSOON SEASONS

SAMSURI ABDULLAH, MARZUKI ISMAIL* AND SI YUEN FONG

School of Marine Science and Environment, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia.

**Corresponding author: marzuki@umt.edu.my*

Abstract: Particulate matter is the most prevailing pollutant in Peninsular Malaysia having the highest API value compared to the other criteria pollutants. Long-term exposure to small particles less than 10 micrometres may lead to a marked reduction in life expectancy due to increase cardio-pulmonary and lung cancer mortality. Effective forecasting models at the local level predict the concentrations of particulate matter is crucial as the information generated allows the authority and people within a community to take precautionary measures to avoid exposure to unhealthy levels of air quality and implement strategic measures that improve air quality status. The aim of this study is to establish MLR models for different monsoon seasons with meteorological factors as predictors. Daily observations of PM_{10} concentrations in Kuala Terengganu, Malaysia from January 2005 to December 2011 were selected for predicting PM_{10} concentration level. The MLR models for NEM, Inter Monsoon 1, SWM and Inter Monsoon 2 disclose R^2 of 0.68, 0.58, 0.57, and 0.63, respectively. Wind speed, relative humidity and rainfall exhibit negative relationship whilst temperature and atmospheric pressure are directly correlated with PM_{10} concentrations. In conclusion, the developed MLR models are appropriate for forecasting PM_{10} concentrations at local level for each monsoon.

Keywords: Air quality, PM_{10}, Forecasting, Malaysia, Meteorological factors, sustainability.

Introduction

Air quality in developing country such as Malaysia has decreased gradually because of rapid urbanization, industrialization and population growth (Latif *et al.,* 2010). In Malaysia, mobile sources such as the emission from motor vehicles, stationary sources such as the emission from power plants and factories and open burning are the three listed major sources of air pollution (Afroz *et al*., 2003). Air pollution status in Peninsular Malaysia is dominated by particulate matter, proven always having the highest API value compared to the other pollutants at most part of the country. Particulate matter (PM) is a mixture of solid and liquid particles that suspended in air. Meteorological factors such as ambient temperature, relative humidity, wind speed, atmospheric pressure and rainfall amount have greatly influence the PM_{10} concentration in atmosphere (Querol *et al*., 2004). Therefore, the changing of meteorological factors either especially in different monsoon seasons affect the concentration of PM_{10} in atmosphere. Long-term exposure to small particles less than 10 micrometres may lead to a marked reduction in life expectancy. The reduction in life expectancy is primarily due to increase cardio-pulmonary and lung cancer mortality. Increases are likely in lower respiratory symptoms and reduced lung function in children while chronic obstructive pulmonary disease and reduced lung function in adults. Recently, the association between air borne particulate matter and wide range of health effects has been found. It is estimated that approximately 3% of cardiopulmonary and 5% of lung cancer deaths are attributable to particulate matter globally (Shin, 2007).

Regression techniques had been used for a long time ago as forecasting tools in many fields, especially in air pollution forecasting. It is because, regression has two main advantages; simple computation and ease of implementation. There are many studies on air pollution forecasting using multi linear regression in Malaysia. However, there is a need to consider the models development on different monsoon seasons as these monsoons will be faced by Terengganu. According to Malaysian Meteorological Department (2012), Malaysia has four seasons; Northeast monsoon (November to March), Inter Monsoon 1 (April), Southwest Monsoon (June to September), and Inter Monsoon 2 (October). The establishment of such models by taking into consideration of meteorological factors are very important in forecasting PM_{10} concentrations. Therefore, the aim of this study is the establishment of MLR models for the four different monsoons in Terengganu. These models are very useful at the local level to provide information which allows the authority and people within a community to take precautionary measures to avoid or limit their exposure to unhealthy levels of air quality and implement significant actions oriented to improve air quality on specific locations.

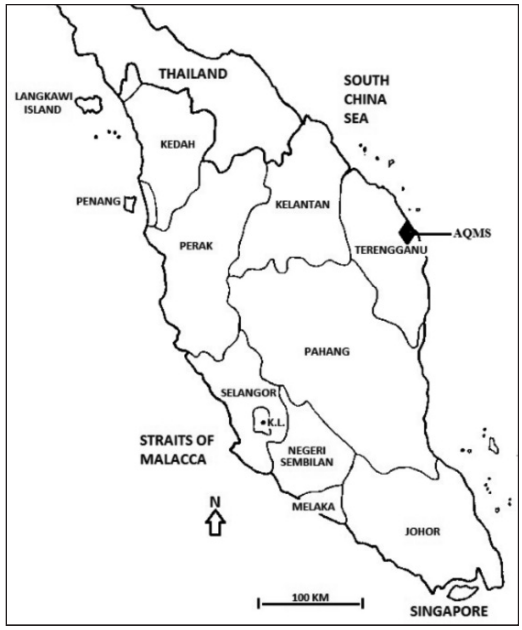

Figure 1: The location of the continuous monitoring station in Kuala Terengganu, Terengganu

Materials and Methods

Site Description

Terengganu is located along the east coast of Peninsular Malaysia facing South China Sea. Kuala Terengganu air monitoring station (N05˚18.455'; E103˚07.213') is situated at SK Pusat Chabang Tiga, located near to the Kuala Terengganu city center (Figure 1). This monitoring station is affected by busy traffic, especially during the rush hour in the morning and late afternoon and the meteorological condition in this region is influence by the South West monsoon, North East monsoon and the inter monsoon seasons. The factors influencing air pollution in this area were associated with local traffic, seasonality and open burning (Abdullah *et al*., 2015).

Monitoring Records

The study was based on the data measured for the 7 years period, from January 2005 to December 2011. The PM_{10} concentration data used in this study was recorded as part of a Malaysian Continuous Air Quality Monitoring (CAQM) program, using the β-ray attenuation mass monitor (BAM-1020) as manufactured by Met One Instruments Inc. The monitoring network was installed, operated and maintained by Alam Sekitar Malaysia Sdn. Bhd (ASMA) on behalf of the Malaysian Department of Environment (Afroz *et al.,* 2003). Five daily averaged parameters were used in order to gain a better understanding of PM_{10} variability during different monsoon seasons. The variation of PM_{10} concentration is controlled by other meteorological parameters and these parameters might be different in influencing PM_{10} concentration during different monsoon seasons. The parameters that were used in this study are particulate matter with aerodynamic diameter less than 10 µm of previous day ($PM_{10, t-1}$, µg/m^3), ambient temperature (°C), relative humidity (%), wind speed (m/s), Mean Sea Level Pressure (hPa) and rainfall amount (mm). The monitoring records were obtained from the Air Quality Division, Department of Environment (DOE), Ministry of Natural Resources and Environment of Malaysia for $PM_{10,t-1}$, ambient temperature, relative humidity, and wind speed. Mean Sea Level Pressure and rainfall amount were two parameters of meteorological factors that acquired from MMD due to the limited meteorological data from the DOE. The MMD station selected is from the nearest AQMS of DOE. Several studies performed showed that they are also using the meteorological data from the nearby or nearest meteorological stations. The data from different air quality monitoring stations and meteorological stations are then combined together to develop models for air quality forecasting (Gennaro *et al.,* 2013; Singh *et al*., 2012). In this study, the nearest meteorological station for the Chabang Tiga is at Kuala Terengganu Meteorological Station, which is about 8.715 km North-North-West (NNW) of the Chabang Tiga (Figure 2). This station is located at height of 5.0 m above MSL. In literature, it can be up to 20km distance (Niska *et al.,* 2004) and therefore the data from these two stations can be combined to forecast PM_{10} concentration.

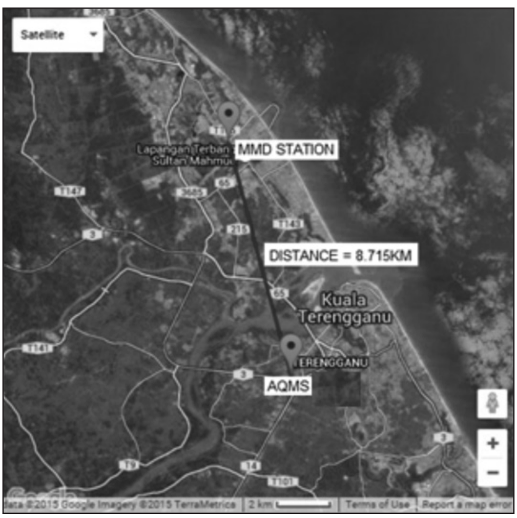

Figure 2: The measured distance from each AQMS with relative to MMD Stations

The data was tabulated using Microsoft Excel Spreadsheet® and analysis of the data was carried out using statistical software, SPSS® version 22.0. In air pollution studies, missing values may occur because of equipment malfunctioned or of errors in measurements (Noor & Zainudin, 2008). Incomplete datasets may lead to results that are different from those that would have been obtained from a complete dataset (Hawthrone & Elliot, 2005). In this study, the incomplete data was treated by imputation of missing values of linear interpolation. This imputation technique also has been used by Yu *et al*. (2015). Interpolation is a method of finding new values for any function using the given set of values. In SPPS®, replacing missing values using a linear interpolation meaning that the last valid value before the missing value and the first valid value after the missing value are used for the interpolation. The unknown value at a particular point can be found using many interpolation formula. The Linear Interpolation Formula (Chapra & Canale, 1998) is given as;

$$f_1(x) = b_0 + b_1 (x - x_0) \qquad \text{Equation 1}$$

where x is independent variable, x_0 is a known value of the independent variable and $f_1(x)$ is the value of the dependent variable for a value x of the independent variable. Then from Equation (1);

$$b_0 = f(x_0) \qquad \text{Equation 2}$$

and

$$b_1 = \frac{f(x_1) - f(x_0)}{x - x_0} \qquad \text{Equation 3}$$

Multiple Linear Regression

Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. This relationship is expressed in mathematical equation. Generally, the equation of MLR is as follows:

$$y = b_0 + \sum_{i=1}^{n} b_i X_i + \varepsilon \qquad \text{Equation 4}$$

Where b_i are the regression coefficients, are the independent variables and ε is stochastic error associated with the regression. MLR assumes that the residuals have normal distribution with zero mean, uncorrelated and constant variance. The method used in obtaining the model was stepwise multiple linear regression. This method involves entering independent variables into the regression model one at a time, which each independent variable entered results in increase of R^2 value. Over the years, the MLR has been used in PM_{10} concentration forecasting as well as forecasting the ozone concentration in Malaysia. In this study, the data of dependent and independent variables consist of different units and therefore normalization of data is required. The normalization produce the data are scaled within the range of 0 to 1 [0 1]. This scaling is suitable for improving the accuracy of numeric computation carry out by the MLR models for the better outputs utilizing the min-max technique. The advantage is preserving exactly all relationship in the data and it does not introduce bias. The normalization of data is obtained by the following transformation (Abdul-Wahab *et al.,* 2005):

$$z_i = \frac{(x_i) - min(x)}{max(x) - min(x)} \quad \text{Equation 5}$$

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

Variable of Inflation (VIF)

Multi-collinearity assumption will be verified by Variable of Inflation (VIF) accompanied with the regression output, where as long as the average VIF is under 10 the conducted regression should be fine, where there is no multicollinearity between the independent variables (Ul-Saufie *et al.,* 2011). The VIF is given by:

$$VIF_i = \frac{1}{1 - R_i^2} \quad \text{Equation 6}$$

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

Durbin-Watson (D-W) Test

By performing Durbin-Watson (D-W) Test, autocorrelation can be identified. Autocorrelation essentially reveals the ability of PM_{10} concentration from previous day to predict PM_{10} concentration in the current day. The test statistic can vary between 0 and 4 with a value of 2 meaning that the residual are uncorrelated (Ul-Saufie *et al.,* 2012). The DW is given by:

$$d = \frac{\sum_{i=1}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \quad \text{Equation 7}$$

Where n is number of observations, $\varepsilon_i = y_i - \bar{y}_i$ (y_i = observed values and $\bar{y}_i$ is the predicted values).

Correlation Coefficient (R^2)

Coefficient of determination (R^2) is an indicator of how well the prediction equation fits the data. It can also use to determine whether the data provide sufficient evidence to indicate that the overall models contribute information for the prediction of PM_{10} concentrations. It is given as (Ul-Saufie *et al.,* 2012):

$$R^2 = \left(\frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{n . S_{pred} . S_{obs}}\right)^2 \quad \text{Equation 8}$$

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.

Results and Discussion

Statistical Characteristics of PM_{10} Concentrations

Figure 3 shows the box plot of daily PM_{10} concentrations in Kuala Terengganu, Terengganu during different monsoon seasons. The box plot is a simple graphical display that is ideal for making comparisons (Ramli *et al.,* 2010). The highest daily average of PM_{10} concentration was observed during SWM that being 145.8 μg/m^3, while the lowest was observed during Inter Monsoon 2 that being 17.1 μg/m^3. The descriptive statistics for PM_{10} concentrations during the study period (2005-2011) are summarized in Table 1. The highest mean concentration of PM_{10} was recorded as 53.2 μg/m^3 (26.3- 93.1 μg/m^3) during Inter Monsoon 1. The lowest mean value of PM_{10} concentration was during Inter Monsoon 2 with 43.0 μg/m^3 (17.1 – 127.7 μg/m^3). Department

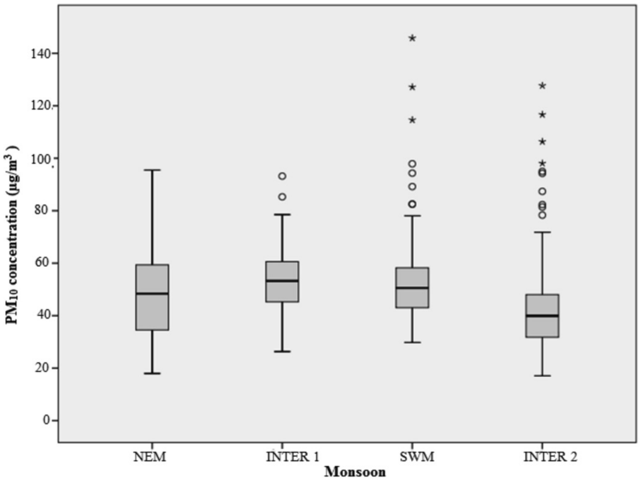

Figure 3: Box plots for daily PM_{10} concentrations in Kuala Terengganu

of Environment Malaysia (DOE) has set the national ambient air quality standards for PM_{10} concentration in ambient air which the maximum daily 24 h average has been determined as 150 µg/m³ (DOE, 2010; Afroz *et al.,* 2003). All PM_{10} concentrations from this study were found to be within the Recommended Malaysian Air Quality Guidelines (RMAQG).

The increasing number of motor vehicles, industries and street dust level are likely to contribute to the total of suspended particulate matter in the atmosphere at this study area, with urban background (Afroz *et al.,* 2003). The high value of skewness during SWM (1.3) and Inter Monsoon 2 (1.9), number of outliers (q_3 + 1.5s.d) and extreme values (q_3 + 3s.d) indicated that Kuala Terengganu experienced high particulates events as well as extreme events which promote increasing of PM_{10} concentrations. The SWM usually brings the high amount of particulate matter to Malaysia due to biomass burning in Sumatera and Kalimantan, Indonesia (Juneng *et al.,* 2009; Ismail *et al.,* 2015)

Models Establishment

The analysis of the air quality data and meteorological data sets are continued by applying Multiple Linear Regression (MLR)

Table 1: Summary of Descriptive Statistic for different monsoon seasons

Descriptive Statistics	NEM (N=1058)	INTER 1 (N=210)	SWM (N=1071)	INTER2 (N=219)
Mean	51.09	53.15	50.93	42.96
Median	50.48	53.25	49.19	39.93
Std Dev	15.20	11.20	13.79	16.72
Skewness	0.332	0.119	1.289	1.937
Min	17.63	26.35	20.09	17.09
Max	105.46	93.17	145.79	127.67

Table 2: Summary model for PM_{10} forecasting on different monsoons

Monsoon	Model	R^2	Range of VIF
NE	$PM_{10} = 0.165 + 0.612PM_{10,\ t-1} - 0.221RH - 0.187R + 0.172AP + 0.102T$	0.681	1.201-1.411
Inter 1	$PM_{10} = 0.153 + 0.670PM_{10,\ t-1} - 0.198R$	0.578	1.035
SWM	$PM_{10} = 0.189 + 0.617PM_{10,\ t-1} - 0.220WS + 0.120T - 0.084RH - 0.067R$	0.570	1.045-1.115
Inter 2	$PM_{10} = 0.124 + 0.688PM_{10,\ t-1} + 0.174T - 0.196WS - 0.163RH + 0.120AP$	0.628	1.209-1.388

with the additional aim of developing effective operational forecasting models for PM_{10} concentrations. The multiple linear regression models were developed and the models summary is depicted in Table 2. The best MLR models for Northeast monsoon, Inter Monsoon 1, Southwest Monsoon, and Inter Monsoon 2 were obtained with R^2 of 0.681, 0.578, 0.570, and 0.628 respectively. The range of Variance Inflation Factor (VIF) for the independent variables for MLR models during all monsoons was 1.035-1.411. The VIF values are lower than 10, which indicate there is no multi-collinearity between the independent variables. Durbin Watson statistic shows that the models do not have any first order autocorrelation problem as the values were in range of 1.704-2.074. It was found that, during the Northeast monsoon, the significant predictors were $PM_{10,\ t-1,}$ relative humidity, rainfall, atmospheric pressure and temperature. PM_{10} concentrations increased by 0.612 unit when $PM_{10,\ t-1}$ variable goes up by one unit, 0.221 unit in decreasing one unit of relative humidity, 0.187 unit for the decreased in one unit of rainfall, 0.172 unit when atmospheric pressure increased by one unit and 0.102 change in PM_{10} concentrations when there is an increase in one unit of temperature. The $PM_{10,\ t-1}$ and rainfall parameters were found to be two significant predictors during the inter monsoon 1. The PM_{10} concentrations increased by 0.670 units and 0.198 unit for the increased of one unit of $PM_{10,\ t-1}$ and the decreased in one unit of rainfall. $PM_{10,\ t-1}$, wind speed, temperature, relative humidity and rainfall were found be the significant predictors during southwest monsoon. The increased in one unit of $PM_{10,\ t-1}$, and temperature, and decreased in one unit of wind speed, relative humidity, and rainfall were known to increase 0.617, 0.120, 0.220, 0.084, and 0.067 unit of PM_{10} concentrations, respectively. PM_{10} concentrations increased by 0.668 unit when there is an increase of one unit of $PM_{10,\ t-1}$, 0.174 unit for an increasing one unit of temperature, 0.120 unit increased when one unit increased in atmospheric pressure, during inter monsoon 2. The increasing of one unit of wind speed and relative humidity will decrease 0.196 and 0.163 units of PM_{10} concentrations, during inter monsoon 2. In general, it was found that different meteorological factors influence on PM_{10} concentrations during different monsoon seasons. During all monsoon seasons, $PM_{10,\ t-1}$, temperature and atmospheric pressure have positive influence on PM_{10} concentrations, while wind speed, relative humidity and rainfall have negative influence on PM_{10} concentrations.

Overall, the average temperature has positive influence on PM_{10} concentrations. Malaysia, as a tropical country has a high temperature which subsequently increases the amount of biomass burning and high temperature also increase the evaporation of materials from the earth's surfaces (Ramli *et al*., 2010). Temperature also affects pollutant concentrations by causing variations in wind circulation and at the same time dilute the concentrations of air pollutant (Banerjee *et al.,* 2011). Atmospheric pressure and PM_{10} concentrations have positive relationship. When the ground is controlled by a low pressure, high pressure air mass is counter-clockwise around the centre of the flow. Subsequently, the updraft is formed in the centre and the wind increases which will help pollutants evacuate upwards. Then PM_{10} concentrations are getting lowly (Zhao *et al.,* 2014). Wind speed showed a negative influence on PM_{10} concentrations, which means the concentrations of PM_{10} tends to be higher in low wind speed areas. When wind speed is high, pollutants are diluted by dispersion (Turaliolu *et al.,* 2005). Rainfall and relative humidity also have negative relationship with PM_{10} concentration. It was believed that rainfall washes out the atmospheric pollutant including particulates in the ambient air (Abas *et al.,* 2004; Bhaskar & Mehta*,* 2004). Wet deposition by precipitation or wet removal is one of the main mechanisms for removal of aerosols in atmosphere (Jaenicke, 1993). The increasing of rainfall amount will increase the amount of water vapour in atmosphere, and therefore the relative humidity also increases. The increasing of relative humidity will subsequently decreased the PM_{10} concentrations in atmosphere.

The residuals (error) are important in deciding the adequacy of the statistical model.

If the error shows any kind of pattern, then it is considered that the model is not taking care of all the systematic information.

Figure 4 indicates histograms of the residuals of PM_{10} models. The residual analysis shows that the residuals are normally distributed. The plots of fitted values with residuals for PM_{10} model are shown in Figure 5, indicating that the residuals are uncorrelated because the residuals are contained in a horizontal band and hence obviously that variance are constant.

Models Verification

The predicted daily PM_{10} concentrations for the model derived for Kula Terengganu during different monsoon seasons were plotted in Figure 6 against the observed values to determine a goodness-of-fit of the models. The regression lines showing 95% confidence interval were also drawn. Most of the points fall in the range of 95% confidence interval. Lines A and C are the upper and lower 95% confidence limit for regression model. The accuracy of the predicted models is 95%. R^2 is between 0.386-0.608.

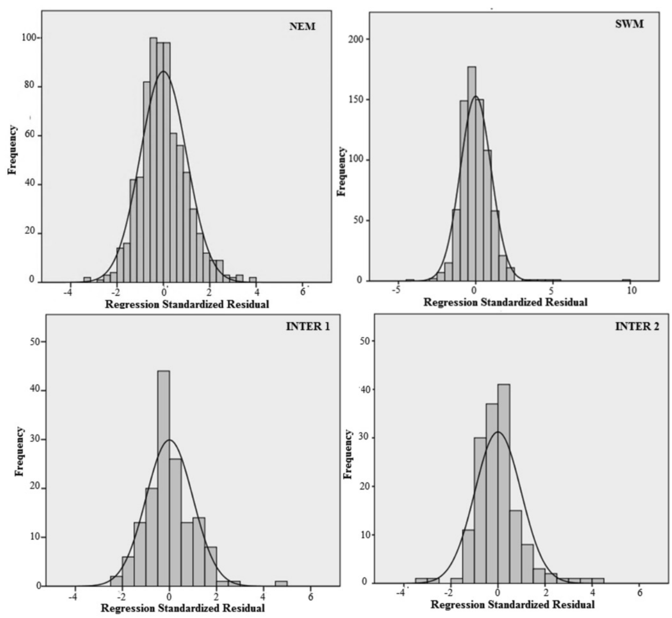

Figure 4: Standardized residual analysis of PM_{10}

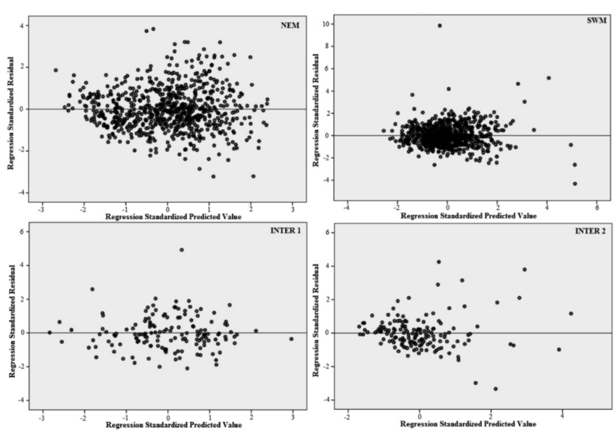

Figure 5: Testing assumption of variance and uncorrelated with mean equal to zero

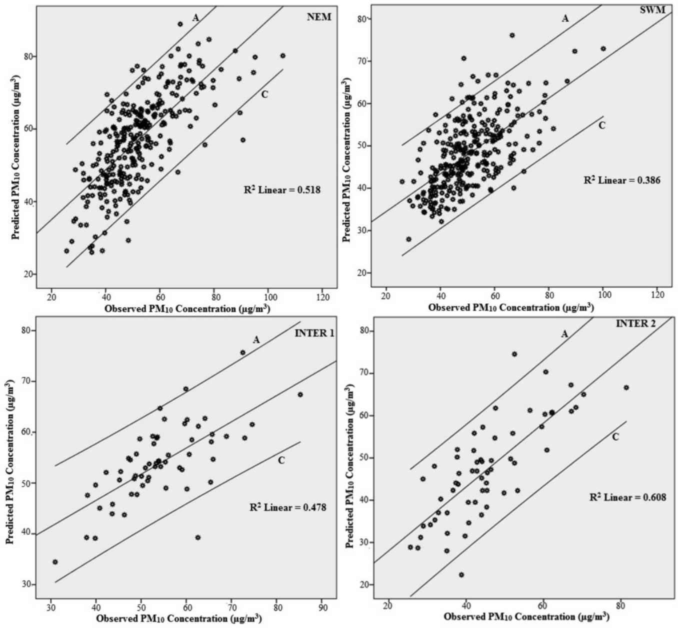

Figure 6: Scatter plot of predicted PM_{10} concentration (µg/m^3) against observed PM_{10} concentration (µg/m^3) for Kuala Terengganu

Conclusion

The daily data of PM_{10} concentrations and meteorological factors from year 2005 to 2011 were used to develop MLR models. This study showed PM_{10} concentrations were below the limit set by Recommended Malaysian Air Quality Guidelines (RMAQG) of 150 µg/m^3 during the study period. The MLR analysis indicates that the influence of meteorological factors on the variability of PM_{10} concentrations was found to be highest during NEM, followed by Inter Monsoon 1, Inter Monsoon 2 and lastly SWM. The significant meteorological factors that influence on PM_{10} concentrations during; (1) NEM; relative humidity, temperature, rainfall and atmospheric pressure (2) Inter Monsoon 1; rainfall (3) Inter Monsoon 2; wind speed, relative humidity, temperature, and atmospheric pressure (4) SWM; wind speed, relative humidity, temperature and rainfall. Meanwhile, wind speed, relative humidity and rainfall have negative relationship on PM_{10} concentration. The validated MLR models shown that the R^2 values of 0.518 (NEM), 0.478 (Inter Monsoon 1), 0.386 (SWM) and 0.608 (Inter Monsoon 2), respectively. The developed MLR models are appropriate for forecasting PM_{10} concentrations intended for early warnings system for public health as well as for local authorities to formulate strategies in improving the air quality at Kuala Terengganu.

Acknowledgements

This study was funded by Fundamental Research Grant Scheme (FRGS) FRGS/2/2013/STWN01/UMT/02/1 (VOT 59312) and Universiti Malaysia Terengganu Scholarship Scheme (BUMT). The authors also would like to thank the Air Quality Division, Malaysian Department of Environment (DOE) for the air quality data.

References

Abas, M. R. B., Oros, D. R., & Semoneit, B. R. T. (2004). Biomass Burning as the Main Source of Organic Aerosol Particulate Matter in Malaysia during Haze Episode. *Chemosphere,* 55(8): 1089-1095.

Abdullah, S., Ismail, M., Yuen, F. S., & Ahmed, A. N. (2015). Principal Component Regression (PCR) for PM_{10} Forecasting in Kuala Terengganu, Terengganu. *National Conference on Wood based Technology, Engineering and Innovation*. Sabah, Malaysia, 82-88.

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. *Environ. Model. Model. Softw*., 20: 1263-1271.

Afroz, R., Hassan, M. N., & Ibrahim, N. A. (2003). Review of Air Pollution and Health in Malaysia. *Environmental Research*, 92: 71-77.

Banerjee, T., Singh, S. B., & Srivastava, R. K. (2011). Development and Performance Evaluation of Statistical Models Correlating Air Pollutants and Meteorological Variables at Pantnagar, India. *Atmospheric Research*, 99(3-4): 505-517.

Bhaskar, B. V., & Mehta, V. M. (2004). Atmospheric Particulate Pollutants and Their Relationship with Meteorology in Ahmedabad. *Aerosol and Air Quality Research,* 10(4): 301-315.

Chapra, S. C., & Canale, R. P. (1998). *Numerical Methods for Engineers.* Singapore: McGraw-Hill.

Department of Environment, Malaysia. (2010). Malaysia Environmental Quality Report 2010. Department of Environment, Ministry of Sciences, Technology, and the Environment, Malaysia. Kuala Lumpur.

Gennaro, G. D., Trizio, L., Gilio, A. D., Pey, J., Perez, N., Cusack, M., Alastuey, A., & Querol, X. (2013). Neural Network Model for the Prediction of PM_{10} Daily Concentrations in Two Sites in the Western Mediterranean. *Science of the Total Environment,* 463-464: 875-883.

Hawthrone, G., & Elliot, P. (2005). Imputing Cross-sectional Missing Data: Comparison of Common Techniques. *Australian and New Zealand Journal of Psychiatry,* 39: 583-590.

Ismail, M., Yuen, F. S., & Abdullah, S. (2015). Trend and Status of Particulate Matter (PM_{10}) Concentration at Three Major Cities in East Coast of Peninsular Malaysia. *Research Journal of Chemical and Environmental Sciences,* 3(5): 25-31.

Jaenicke, R. (1993). *Aerosol-cloud Climate Interactions*. San Diego CA: Academic Press.

Juneng, L., Latif, M. T., Tangang, F. T., & Mansor, H. (2009). Spatio-temporal Characteristics of PM_{10} Concentrations across Malaysia. *Atmospheric Environment,* 43(30): 4584-4594.

Latif, M. T., Azmi, S. Z., Noor, A. D. M., Ismail, A. S., Johny, Z., Idrus, S., Mohamed, A. F., & Mohktar, M. (2010). The Impact of Urban Growth on Regional Air Quality Surrounding the Langat River Basin. *Malaysia Environmentalist,* 31: 315-324.

Malaysian Meteorological Department. [homepage on the Internet]. (2012) [cited 2016 Jan 15]. Available from: http://www.met.gov.my.

Niska, H., Hiltunen, T., Karppinen, A., Ruuskanen, J., & Kolehmainen, M. (2004). Evolving the Neural Network Model for Forecasting Air Pollution Time Series. *Engineering Applications of Artificial Intelligence,* 17: 159-167.

Noor, N. M., & Zainudin, M. L. (2008). A Review: Missing Values in Environmental Data Sets. *Proceeding of International Conference on Environment, Malaysia.*

Querol, X., Alastuey, A., Ruiz, C. R., Artinano, B., Hansson, H. C., Harisson, R. M., Buringh, E., ten Brink, H. M., Lutz, M., Bruckmann, P., Straehl, P., & Schneider, J. (2004). Speciation and Origin of PM_{10} and $PM_{2.5}$ in Selected European Cities. *Atmospheric Environment,* 38: 6547-6555.

Ramli, N. A., Ghazali, N. A., & Yahaya A. S. (2010). Diurnal Fluctuations of Ozone Concentrations and Its Precursors and Prediction of Ozone Using Multiple Linear Regressions. *Malaysian Journal of Environmental Management*, 11(2): 57-69.

Shin, D. C. (2007). Health Effects of Ambient Particulate Matter. *Journal of the Korean Medical Association,* 50(2): 175-182.

Singh, K. P., Gupta, S., Kumar. A., & Shukla. S. P. (2012). Linear and Nonlinear Modeling Approaches for Urban Air Quality Prediction. *Science of the Total Environment,* 426: 244-55.

Turaliolu, F. S., Nuhoglu, A., & Bayraktar, H. (2005). Impacts of Some Meteorological Parameters on SO_2 and TSP Concentrations in Erzurum, Turkey. *Chemosphere,* 59(11): 1633-1642.

Ul-Saufie, A. Z., Yahaya, A. S., Ramli, N. A., & Hamid, H. A. (2012). Performance of Multiple Linear Regression Model for Long-term PM10 Concentration Prediction Based on Gaseous and Meteorological Parameters. *Journal of Applied Sciences* 12(14): 1488-1494.

Ul-Saufie, A. Z., Yahya, A. S., Ramli, N. A., & Hamid, H. A. (2011). Comparison between Multiple Linear Regression and Feed Forward Back Propagation Neural Network Models for Predicting PM_{10} Concentration Level Based on Gaseous and Meteorological Parameters. *International Journal of Applied Science and Technology,* 1(4): 42-49.

Yu, R., Liu, X. C., Larson, T., & Wang, Y. (2015). Coherent Approach for Modeling and Nowcasting Hourly Near-road Black Carbon Concentrations in Seattle, Washington. *Transportation Research Part D,* 34: 104-115.

Zhao, C. X., Wang, Y. Q., Wang, Y. J., Zhang, H. L., & Zhao, B. Q. (2014). Temporal and Spatial Distribution of $PM_{2.5}$ and PM_{10} Pollution Status and the Correlation of Particulate Matters and Meteorological Factors during Winter and Spring in Beijing. *Environmental Science,* 35: 418-427.