Journal of Sustainability Science and Management
Special Issue Number 5, 2018 : 3-17
ISSN: 1823-8556
© Penerbit UMT

FORECASTING OF PARTICULATE MATTER (PM_{10}) CONCENTRATION BASED ON GASEOUS POLLUTANTS AND METEOROLOGICAL FACTORS FOR DIFFERENT MONSOONS OF URBAN COASTAL AREA IN TERENGGANU

SI YUEN FONG[1], SAMSURI ABDULLAH[2] AND MARZUKI ISMAIL[1, 3*]

[1]*School of Marine and Environmental Sciences, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia.*
[2]*School of Ocean Engineering, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia.*
[3]*Institute for Tropical Biodiversity and Sustainable Development, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia.*

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

Abstract: Particulate Matter (PM_{10}) is the most conspicuous pollutant in Peninsular Malaysia having the highest air pollutant index (API) value contrasted with the other criteria contaminations. Long-term exposure of PM_{10} may lead to a marked reduction in life expectancy due to increase in cardiopulmonary and lung disease mortality. Compelling forecasting models at local level to predict PM_{10} status is essential as the information generated permits the authority and local community to take prudent steps and lessen the effect of particulate contamination. The study aim to develop multiple linear regression (MLR) and principal component regression (PCR) model for PM_{10} concentration prediction in the East Coast of Peninsular Malaysia for the two main monsoon seasons. Daily observations of PM_{10} concentrations, meteorological variables and gaseous pollutants in Kuala Terengganu, Malaysia from January 2000 to December 2014 (14 years) were selected for predicting PM_{10} concentration level. Model comparison statistics through performance indicators using coefficient of determination (R^2), index of agreement (IA), normalized absolute error (NAE), mean absolute error (MAE) and root mean square error (RMSE) shows that MLR is a better model in forecasting next day PM_{10} concentration compared to PCR.

KEYWORDS: Air quality status, multiple linear regression, particulate matter (PM_{10}), forecasting, sustainable environment

Introduction

The air quality status in Malaysia is evaluated based on the concentration of the five criteria pollutants namely carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3) and particulate matter less than or equal 10 μm (PM_{10}). Out of the five criteria pollutants, PM_{10} record the highest air pollution index (API) value in the majority of the air quality monitoring stations in the country and the number of moderate air quality days per year were gradually increasing (Abdullah *et al.,* 2017). The major sources of PM_{10} in Malaysia listed by the Department of Environment (DOE) of Malaysia are motor vehicle emissions, power plants, industrial process emissions and open burning (Department of Environment, 2014). Prolonged exposure to PM_{10} has long been associated with health issues such as respiratory, cardiovascular diseases and increase mortality rate (Jahn *et al.,* 2011; Zhang *et al.,* 2014). Deposition of PM_{10} on the surface of vegetation instigated radiative heating, reduction of photosynthesis rate and alters cycles of decomposition which in turn, affecting animal communities (Grantz *et al.,* 2003). To address the elevating PM_{10} concentration issue, the Malaysian government have set two guidelines on controlling the PM_{10} concentration emission which states that the average daily PM10 concentration may not exceed 150 μg/m^3 and the annual PM_{10} concentration may not exceed 50 μg/m^3 (Department of Environment, 2014).

Monitoring is one way to evaluate air quality, and it is imperative due to the significant increase in air pollutants emissions. The air monitoring involves measurement and the measurement may be accurate, but it is costly and time-dependent. In Malaysia, the monitoring was done under the responsibility of DOE. The

high cost of installation, maintenance, and constant operation have led to the development of modeling techniques which facilitate the prediction of PM_{10} with sufficient accuracy (Pires *et al.,* 2008; Stadlober *et al.,* 2008). Statistical modeling can provide a good insight into the future air pollution level for short-term forecasting which allows government authorities to prepare suitable abatement actions in order to minimize casualties and monetary loss.

Regression models have widely been used for air pollution forecasting due to the simple computation method and straightforward implementation technique. Researchers have applied these models in PM_{10} concentration prediction in various study areas such as the industrial area in Haute-Normandie, France (Misiti *et al.,* 2015), Mediterranean City, Greece (Papanastasiou *et al.,* 2007) and urban industrial area, Penang, Malaysia (Ul-Saufie *et al.,* 2011). The most popular linear regression models were the Multiple Linear Regression (MLR) model and the Principle Component Regression (PCR) model which utilize the meteorological parameters and criteria pollutant data as input variables. Historical PM_{10} values in preceding time steps were also considered as an important predictor of future PM_{10} concentration as an explanatory variable which vastly improves the modeling results (Slini *et al.,* 2006). The study area, Kuala Terengganu is an urban coastal area that experienced two main monsoons annually and the weather characteristics for these two monsoons varies greatly whereby from November to March, northeast monsoon (NEM) will occur and during May to September, southwest monsoon (SWM) will take place. The different monsoon season was taken into consideration in the development of the forecasting model as the meteorological factors vastly affect the model outcome. Hence, the aim of this study is to develop the multiple linear regression (MLR) and principal component regression (PCR) model for PM_{10} concentration prediction for the two main monsoon seasons namely the NEM and the SWM in the east coast of Peninsular Malaysia. These models are very beneficial at the local level to provide prior notice which allows the local authority and environmental agencies to take preventive measures to avoid or limit their exposure to unhealthy levels of air quality and implement required actions focused on improving air quality on targeted areas.

Materials and Methods

Site Description

Terengganu is located on the east coast of Malaysia with the total area of 13,035 km^2. The total population of the state in year 2010 is 1,015,776 (Department of Statistics, 2011). There are three air quality monitoring stations in the state of Terengganu and the station located in Kuala Terengganu, the capital of the state were chosen (N 05°18.487", E 103°07.226"). The location of the study area was depicted in Figure 1.

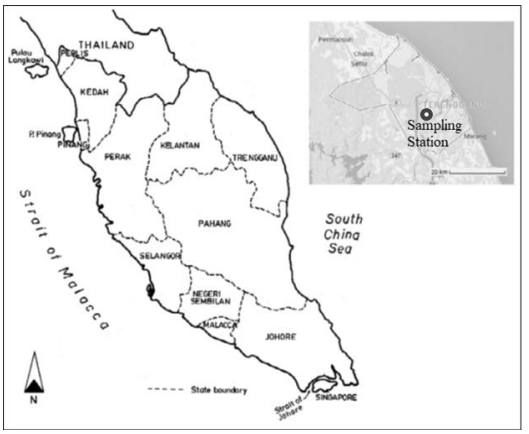


Figure 1: The location of the study area in Kuala Terengganu, Terengganu

Data Acquisition and Processing

The parameters used in this study are lagged PM_{10} (μg/m³), CO (ppm), NO_2 (ppm), SO_2 (ppm), ambient temperature (°C), relative humidity (%), wind speed (m/s), Mean Sea Level Pressure (atm) and rainfall amount (mm). Hourly air quality and meteorological data were monitored and recorded at the air quality monitoring station (AQMS) and local meteorological station administered by the Department of Environment Malaysia and the Malaysian Meteorological Department. The hourly criteria pollutants data (CO, NO_2, SO_2 and PM_{10}) and hourly meteorological data (relative humidity, temperature and wind speed) were obtained from the AQMS whereas the rainfall and mean sea level pressure data were acquired from the local meteorological station. The data from different air quality monitoring stations and meteorological stations were collectively used to develop models for PM_{10} forecasting (De Gennaro *et al.,* 2013). The PM_{10} concentration is predicted on daily basis for the predictions of the next day, next two-day and next three-day. The preciseness (accurate and error) of the models is evaluated through the evaluation of performance indicator.

The data used for this study were recorded during the period of year 2000 – 2014 and were screened and divided into the two main monsoons respectively prior to the analysis. The analysis of the data was performed using the statistical software SPSS® version 21.0 (Abdullah *et al.,* 2017) and the models developed for the forecasting of three days ahead. The missing data were found to be less than 20%. Thus, the data were replaced using the imputation technique by linear interpolation utilizing the closest available data points (Yu *et al.,* 2015).

Multiple Linear Regression

Multiple Linear Regression (MLR) is one of the statistical model techniques that allow us to predict the variability between the independent variable and the dependent variable by fitting the observed data into a linear equation. Generally, the equation of MLR is as follows:

$$Yi = b_0 + b_i x_{1i} + ... + b_k x_{ki} + \varepsilon_i \qquad \text{Equation 1}$$

Where bi is the regression coefficients, x_i is the independent variables, and ε is the stochastic error associated with the regression. It assumes that the data residuals to be normally distributed with zero mean and constant variance (Ul-Saufie *et al.,* 2012). Prior to the model development, the dataset was normalized as the unit for the dependent and independent variables were made up of parameters with different units. The data normalization process will produce a new dataset scaled within the value of 0 to 1 based on the following transformation formula (Abdul-Wahab *et al.,* 2005):

$$z_i = \frac{(x_i) - \min(x)}{\max(x) - \min(x)} \qquad \text{Equation 2}$$

Where $x = (x_1, \ldots, x_n)$ and z_1 is the normalized data.

The effect of multicollinearity on the variance of the regression output was verified by the Variance Inflation (VIF) where the average value of VIF should be below 10. The VIF is calculated with the following equation:

$$VIF_i = \frac{1}{1 - R_i^2} \qquad \text{Equation 3}$$

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. The autocorrelation issue was identified by the Durbin-Watson (DW) Test. The autocorrelation coefficient shows the capability of the lagged PM_{10} concentration in predicting future PM_{10} concentration whereby the values were range from 0 to 4, 2 signifies that the residuals are uncorrelated and values approaching 4 means negative autocorrelation. The DW is given by:

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

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

Principal Component Analysis (PCA) and Regression (PCR)

Principal component analysis (PCA) is a source identification method that enables us to certify that the variables were ideally correlated with one sole component only without correlating with other component and produce a new set of variables called principal component (PCs) (Baklouti *et al.,* 2015; Voukantsis *et al.,* 2011). The number of principal components will be less than or equals to the original variables. The most significant PC allow us to determine the relationship between the variables and the dataset, thus deducing the possible sources. PCR is a hybrid of PCA and MLR. PCR is a statistical technique, which can reveal the inherent linear structure of a dataset and thus reduce the dimension of predicting variables. In this study, PCA was used as a feature selection method for clustering the descriptors and choosing the best group of them as input for the MLR.

Performance Indicators

Performance indicators (PI) were used to determine the suitability of the model in predicting the future PM_{10} concentration and to select the best-fitted model for both monsoons. There are two categories of performance indicators (Table 1), one that measures the errors which are the normalized absolute error (NAE), mean absolute error (MAE) and root mean square error (RMSE), and the one that measures the accuracy of the prediction are the coefficient of determination (R^2) and index of agreement (IA).

Table 1: Performance Indicators (Ul-Saufie *et al.,* 2012)

Performance Indicators	Equation	Description
Normalized absolute error (NAE)	$NAE = \frac{\sum_{i=1}^{n}\lvert P_i - O_i\rvert}{\sum_{i=1}^{n} O_i}$	NAE value closer to zero indicates better model
Mean absolute error (MAE)	$MAE = \frac{\sum_{i=1}^{n}\lvert P_i - O_i\rvert}{n}$	MAE value closer to zero indicates better model
Root mean square error (RMSE)	$RMSE = \frac{1}{n-1}\sum_{i=1}^{n}(P_i - O_i)^2$	RMSE value closer to zero indicates better model
Coefficient of determination (R^2)	$R^2 = \left(\frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{n.S_{pred}.S_{obs}}\right)$	R^2 value closer to one indicates better model
Index of agreement (IA)	$IA = 1 - \left[\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}\lvert P_i - \bar{O}\rvert + \lvert O_i - \bar{O}\rvert^2}\right]$	IA value closer to one indicates better model

Results and Discussion

Principal Component Analysis

To execute the PCA, the individual variable and the set of variables must obtain the Kaiser-Meyer-Olkin (KMO) of sampling adequacy value of greater than 0.50 and the probability associated with Bartlett's Test of Sphericity be less than the level of significance (Chua, 2009). The result of these test for both NEM and SWM were shown in Table 2. The KMO values were greater than 0.50 and the Barlett's test of sphericity ($p < 0.001$) for all prediction days of both monsoons shows a high degree of relationship between the input parameters and the dataset were suitable for the PCA.

Table 2: Kaiser-Meyer-Olkin (KMO) and Barlett's Test for Northeast Monsoon (NEM) and Southwest Monsoon (SWM)

		Next day	Next two-day	Next three-day
			NEM	
KMO Measure of Sampling Adequacy		0.652	0.618	0.608
Bartlett's Test of Sphericity	Approx. Chi-Square	2881.627	2139.342	2019.347
	Sig.	0.000	0.000	0.000
			SWM	
KMO Measure of Sampling Adequacy		0.585	0.567	0.571
Bartlett's Test of Sphericity	Approx. Chi-Square	1123.291	984.344	1090.854
	Sig.	0.000	0.000	0.000

For NEM, eight linear components were identified before extraction for the next day prediction dataset and after extraction, three factors with eigenvalues more than 1 were identified as principal component. The cumulative variance of the principal components is 70.142%, whereby 29.711% was explained by the first component, 21.067% was explained by the second component, and the remaining 19.363% was explained by the third component. The significance of the principal component was arranged by the first, followed by the second and third as determined by the percentage of the eigenvalues in explaining the amount of variance. The similar conclusion can be attained for the next two-day and next three-day prediction in NEM with the total percentage of variance of 68.728% and 67.713 respectively. As for SWM, six linear components were identified before extraction with three principal components established after extraction for the next day, next two-day and next three-day prediction dataset with the total percentage of variance of 69.610%, 68.456%, and 68.265% respectively.

The rotated component matrix utilizing Kaiser Normalization of the three components for all three days predictions of both NEM and SWM were portrayed in Table 3. The output is suppressed with values less than 0.4. The factor loadings correlate the factors and the variables, the higher the factor loading, the more the variable contributes to the variation of the PC. The factor scores obtained from the PCA for the NEM and SWM were displayed in Table 3. There are three principal components (PC) with similar loadings that take accounts for the concentration of next day, next two-day and next three-day PM_{10} prediction in both NEM and SWM.

Table 3: Rotated Component Matrix for Northeast Monsoon (NEM) and Southwest Monsoon (SWM)

Factors	Next day			Next two-day			Next three-day		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
NEM									
Lagged PM_{10}		0.731			0.737			0.733	
CO	0.800			0.796			0.804		
NO_2	0.856			0.857			0.859		
Wind speed	-0.748			-0.753			-0.748		
Temperature			0.827			0.825			0.822
Rainfall			-0.761			-0.769			-0.774
Pressure		0.826			0.819			0.817	
SWM									
Lagged PM_{10}			0.803			0.846			0.851
CO	0.796			0.809			0.813		
NO_2	0.750			0.769			0.775		
Wind speed	-0.680		0.513	-0.651		0.491	-0.637		0.489
Temperature		0.757			0.791			0.807	
Pressure		-0.823			-0.787			-0.763	

For NEM, PC1 have strong positive loadings of CO and NO_2 and negative loading of wind speed leading to the observation that PC1 is related to traffic emission parameters (Abdullah *et al.,* 2016). Kuala Terengganu was categorized as an urban background area and the AQMS were located in the city where traffic emission can be the main sources of CO and NO_2 (Chouykerd *et al.,* 2008; Hak *et al.,* 2010). Low wind speed was associated with high levels of CO and NO_2 as wind speed affects the dispersion of pollutants from the sources at street level (Grundström, 2015). The second factor (PC2) were loaded with positive influences from lagged PM_{10} and pressure. During high-pressure circumstances, the area above the source of PM_{10} gets limited mixing and dispersion due to subsidence inversion. Vertical air mixing was suppressed by the inversions and thus ground level emissions were trapped at the surface and cause the elevation of PM_{10} concentration (Grundström *et al.,* 2015; Li *et al.,* 2015). PC3 is composed of temperature and rainfall which represents seasonality changes. The changes in the main monsoon season were characterized by the increase in rainfall amount and declination in temperature whereby the higher amount of precipitation, the lower the temperature (Giri *et al.,* 2008).

During SWM, the PC1 was also loaded with positive loadings of CO and NO_2 and negative loading of wind speed, which is similar to NEM. PC2 were composed of positive loading of temperature and negative loading of pressure. During SWM, the weather was drier and less wet deposition recorded, thus leading to higher surface temperature and lower atmospheric pressure (Met Office, 2017). PC3 were associated with lagged PM_{10}. Historical PM_{10} concentration in preceding time step is an important predictor in affecting future PM_{10} concentration (Shahraiyni & Sodoudi, 2016).

Models Development

Table 4 shows the model summary for PM_{10} concentration predictions for both NEM and SWM based on the gaseous pollutants and meteorological factors as input for MLR as well as PCA results as input for PCR. The best MLR models for the next-day, next two-day and next three-day PM_{10} concentration prediction during NEM were obtained with R^2 of 0.715, 0.616

and 0.586 respectively whereas for SWM, the R^2 are 0.626, 0.505 and 0.448 subsequently. The results showed that the variance inflation factor (VIF) of the MLR models was below 10 for all three future PM_{10} concentration prediction in both monsoons which indicates that there is no multicollinearity problem between the independent variables. The Durbin-Watson statistic points out that the MLR models for next day NEM (DW = 1.668), next two-day NEM (DW = 0.997), next three-day NEM (DW = 0.807), next day SWM (DW = 1.808), next two-day SWM (DW = 1.050), and next three-day SWM (DW = 0.856) does not have any first order autocorrelation problem. The best PCR models developed achieved the R^2 of 0.669, 0.542 and 0.500 respectively during NEM and 0.593, 0.490 and 0.428 subsequently during SWM for the next-day, next two-day and next three-day PM10 concentration prediction. The summary of all three models for both NEM and SWM also implies that there are no multicollinearity (VIF = 1.000) nor autocorrelation problems (DW = 0.873 – 1.463). The residuals for all MLR and PCR models were normally distributed with zero mean and constant variances, and were distributed along the horizontal band which indicates the homogenity of the variances and uncorrelated. The graph of the residuals for the next day prediction by MLR and PCR during both monsoons were displayed in Figure 2 and Figure 3.

Table 4: Models summary of Multiple Liner Regression (MLR) and Principle Component Regression (PCR) for both Northeast Monsoon (NEM) and Southwest Monsoon (SWM)

Method	Models	R^2	Range of VIF	DW
	NEM			
	Next day			
MLR	$PM_{10, t+1}$ = -0.036 + 0.529 [lagged PM_{10}] + 0.398 [CO] – 0.187 [Relative Humidity] – 0.275 [Rainfall] + 0.191 [Wind speed] + 0.116 [Temperature] + 0.151 [Pressure] + 0.076 [SO_2] + 0.126 [NO_2]	0.715	1.040 - 1.909	1.668
PCR	$PM_{10, t+1}$ = 0.469 + 0.071 [PC1] + 0.126 [PC2] - 0.075 [PC3]	0.669	1.000	1.298
	Next two-day			
MLR	$PM_{10, t+2}$ = -0.032 + 0.341 [lagged PM_{10}] + 0.535 [CO] – 0.261 [Relative Humidity] – 0.320 [Rainfall] + 0.217 [Wind speed] + 0.149 [Temperature] + 0.232 [Pressure] + 0.106 [SO_2] + 0.165 [NO_2]	0.616	1.038 - 1.902	0.997
PCR	$PM_{10, t+2}$ = 0.469 + 0.059 [PC1] + 0.118 [PC2] + 0.064 [PC3]	0.542	1.000	1.081
	Next three-day			
MLR	$PM_{10, t+3}$ = -0.037 + 0.272 [lagged PM10] + 0.580 [CO] – 0.275 [Relative Humidity] – 0.351 [Rainfall] + 0.225 [Wind speed] + 0.151 [Temperature] + 0.266 [Pressure] + 0.121 [SO_2] + 0.190 [NO_2]	0.586	1.039 - 1.890	0.807
PCR	$PM_{10, t+2}$ = 0.469 + 0.059 [PC1] + 0.111 [PC2] + 0.064 [PC3]	0.500	1.000	0.906

Method	Models	R^2	Range of VIF	DW
	SWM			
	Next day			
MLR	$PM_{10, t+1}$ = -0.114 + 0.546 [lagged PM10] + 0.403 [CO] + 0.191 [Temperature] + 0.290 [NO2] -0.145 [Rainfall] + 0.066 [SO2] + 0.045 [Pressure]	0.626	1.034 - 1.344	1.808
PCR	$PM_{10, t+1}$ = 0.487 + 0.094 [PC1] + 0.021 [PC2] + 0.109 [PC3]	0.593	1.000	1.463
	Next two-day			
MLR	$PM_{10, t+2}$ = -0.136 + 0.383 [lagged PM_{10}] + 0.499 [CO] + 0.253 [Temperature] + 0.383 [NO_2] - 0.126 [Rainfall] + 0.081 [SO_2] + 0.066 [Pressure]	0.505	1.033 - 1.315	1.050
PCR	$PM_{10, t+2}$ = 0.487 + 0.094 [PC1] + 0.027 [PC2] + 0.089 [PC3]	0.490	1.000	1.034
	Next three-day			
MLR	$PM_{10, t+3}$ = -0.121 + 0.295 [lagged PM_{10}] + 0.536 [CO] + 0.270 [Temperature] + 0.396 [NO_2] - 0.141 [Rainfall] + 0.093 [SO_2] + 0.071 [Pressure]	0.448	1.036 - 1.304	0.856
PCR	$PM_{10, t+2}$ = 0.487 + 0.095 [PC1] + 0.030 [PC2] + 0.073 [PC3]	0.428	1.000	0.873

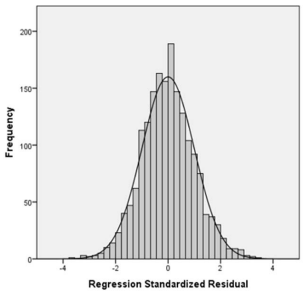
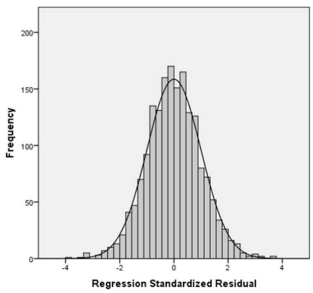
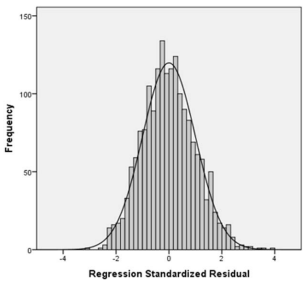
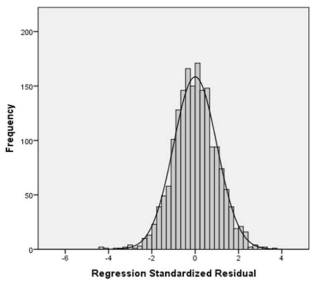
Models	NEM	SWM
Next day (MLR)		
Next day (PCR)		

Figure 2: Standardized residual analysis of next day PM_{10} concentration for multiple liner regression (MLR) and principle component regression (PCR) model.

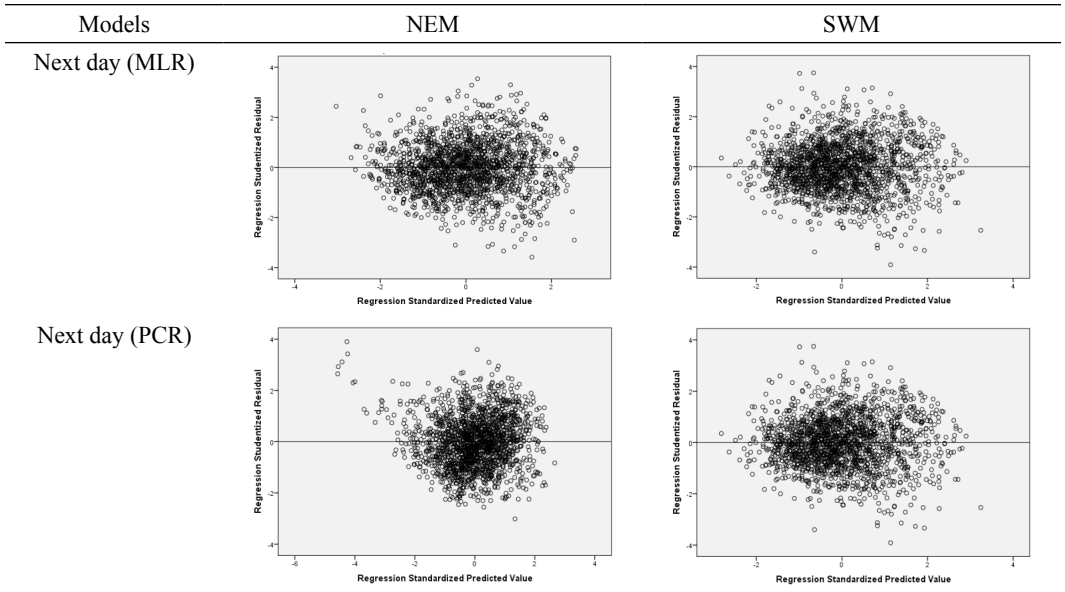


Figure 3: Constant and uncorrelated variance with mean equal to zero for multiple liner regression (MLR) and principle component regression (PCR) model.

According to the MLR models, during NEM, the variables such as lagged PM_{10}, CO, NO2, SO2, temperature, wind speed and atmospheric pressure have positive influences whereas relative humidity and rainfall have negative influences on the predicted PM_{10} concentration for the next day, next two-day and next three-day. As for the SWM, the predicted PM_{10} concentration for the next day, next two-day and next three day increase with lagged PM_{10}, CO, NO_2, SO_2, temperature and atmospheric pressure while decrease with rainfall. All inputs for PCR were selected in the regression model as siginificant predictors for all the three PCs, for both the NEM and SWM.

The predicted daily PM_{10} concentrations for the model derived for both NEM and SWM were plotted in figures 4 and 5 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 R^2 for both MLR and PCR models decreases linearly from the next day prediction to the next three-day prediction. The R^2 is between 0.360 to 0.604 and 0.440 to 0.677 for MLR during NEM and SWM respectively whereas for PCR, the R^2 falls in the range of 0.383 to 0.533 and 0.432 to 0.685 during NEM and SWM, respectively.

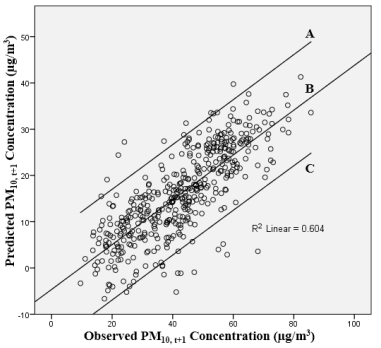
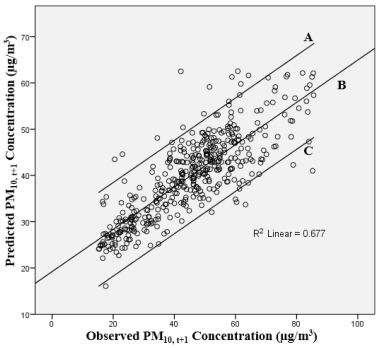
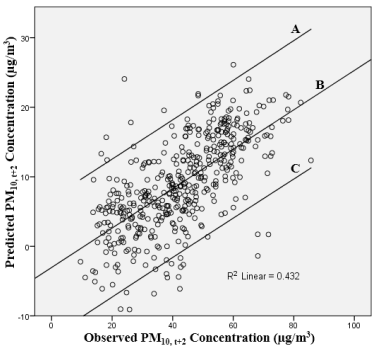
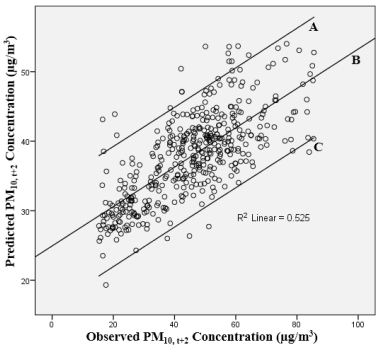
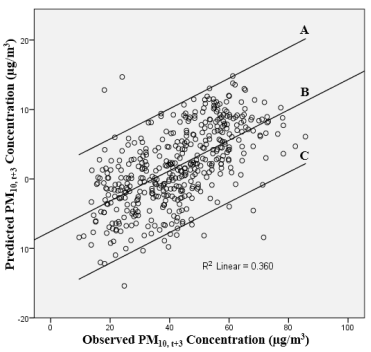
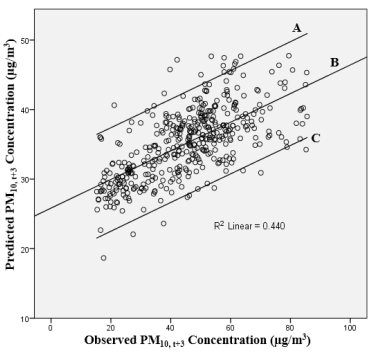
Models	NEM	SWM
Next day		
Next two-day		
Next three-day		

Figure 4: Scatter plot of predicted PM_{10} concentration (µg/m^3) against observed PM_{10} concentration (µg/m^3) for next day, next two-day and next three-day multiple liner regression models.

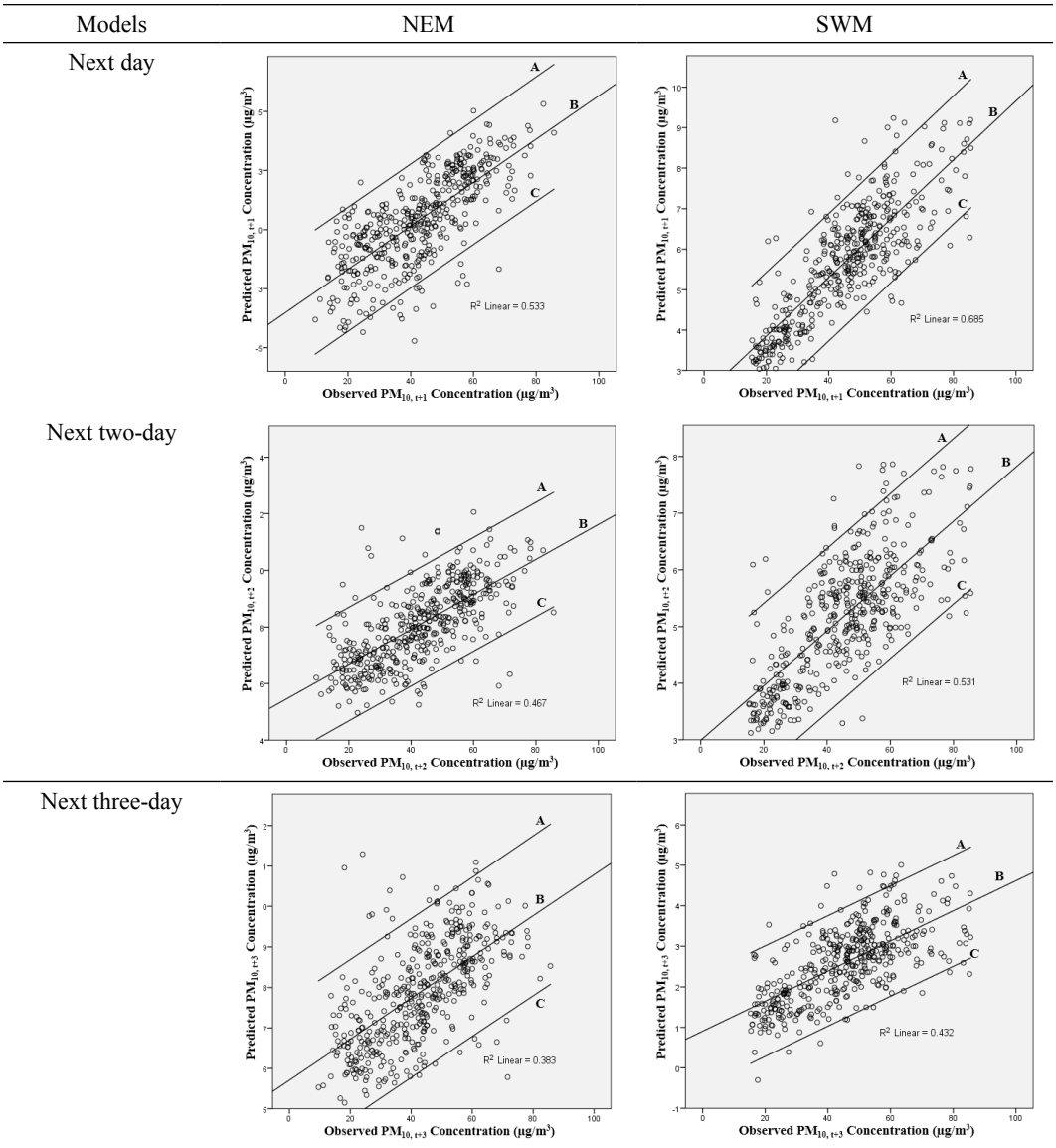


Figure 5: Scatter plot of predicted PM_{10} concentration (µg/m^3) against observed PM_{10} concentration (µg/m^3) for next day, next two-day and next three-day principle component regression models.

Model Performance Comparison

In order to evaluate the performance of the models and to test the best-fitted model, five performance indicators were used to measure the error and accuracy of the models. Table 5 displays the performance indicators between MLR and PCR models for NEM and SWM PM_{10} prediction models.

Normalized Absolute Error (NAE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were utilized to compute the error of the model where the value closer to zero indicates the better model. The coefficient of determination (R^2) and index of agreement (IA) were used to calculate the accuracy of the models' result where the value closer to 1 signifies higher accuracy. The best prediction

model for next-day, next two-day and next three-day PM10 concentration during the NEM is MLR. The error ranges from 25.743 to 32.752 (MAE), 0.623 to 0.792 (NAE) and 27.997 to 35.710 (RMSE) whereas the accuracy falls between 0.390 to 0.445 (IA) and 0.584 to 0.715 (R^2). As for the SWM, the best prediction model for all five model from next-day, next two-day and next three-day is also MLR. The error ranges from 8.675 to 13.213 (MAE), 0.195 to 0.298 (NAE) and 11.389 to 16.641 (RMSE) whereas the accuracy falls between 0.455 to 0.596 (IA) and 0.448 to 0.626 (R^2).

The errors and accuracy computed from these MLR models were in agreement with the values obtained from Grivas and Chaloulakou (2006) where their results reported the RMSE and R^2 were in the range of 17.06 to 35.14 µg/m^3 and 0.280 to 0.348 respectively (Grivas and Chaloulakou, 2006). MLR reduces the error of the models during NEM by as much as 28.4%, 3.7% and 3.4% for the next-day, next two-day and next three-day predictions and increases the accuracy up to 13.9%, 8.7%, and 9.9%, respectively. During SWM, the MLR lessens the error of the predictions of the next day, next two-day and next three-day models by 76.1%, 69.7% and 67.5%, respectively whereas the accuracy of the models were increased as much as 29.0%, 16.6% and 18.3% subsequently. The MLR model has shown to be a better model for short-term PM_{10} concentration forecasting as the error is lower and have higher accuracy for both monsoons. The predictions for SWM show better results compared to NEM as the seasonal characteristic of the monsoon is more stable and less extreme weather events occurred.

Table 5: Performance indicators between multiple linear regression (MLR) and principle component regression (PCR) models for northeast monsoon (NEM) and southwest monsoon (SWM).

PI	NEM		SWM	
	PCR	MLR	PCR	MLR
Next day				
MAE (µg/m^3)	35.958	25.743	38.802	8.675
NAE	0.869	0.623	0.874	0.195
RMSE (µg/m^3)	39.214	27.997	41.939	11.389
IA	0.368	0.445	0.391	0.596
R^2	0.669	0.715	0.593	0.626
Next two-day				
MAE (µg/m^3)	33.541	32.490	39.299	11.238
NAE	0.811	0.786	0.886	0.253
RMSE (µg/m^3)	37.035	35.212	42.545	14.369
IA	0.380	0.394	0.361	0.470
R^2	0.542	0.616	0.490	0.505
Next three-day				
MAE (µg/m^3)	33.744	32.752	43.104	13.213
NAE	0.816	0.792	0.971	0.298
RMSE (µg/m^3)	37.290	35.710	46.187	16.641
IA	0.379	0.390	0.345	0.455
R^2	0.500	0.584	0.428	0.448

Conclusion

The aim of this study was to compare a multiple linear regression model and principal component regression model for predicting PM_{10} concentration in the coastal city of Kuala Terengganu. The performance and reliability of the developed models were evaluated by the performance indicators (MAE, NAE, RMSE, IA and R^2). Through the PI assessment, MLR was found to be a better model in predicting the next-day PM10 concentration for both NEM and SWM with the R^2 of 0.715 and 0.626 respectively. In conclusion, particulate matter concentration is strongly influence by gaseous pollutant (NO_2, SO_2, CO and lagged PM_{10}) and meteorological parameters (temperature, wind speed, relative humidity, atmospheric pressure and rainfall). Consequently, management of particulate matter is very vital because unpredictability of weather, climate change and air pollution are all intrinsically connected (Jacob and Winner, 2009; Noyes *et al.,* 2009). Thus, the developed MLR models in this study is pertinent for PM_{10} forecasting as it allows the local authorities and the public to implement strategies for better management of air quality status in 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 and the Malaysian Meteorological Department for the meteorological data.

References

Abdullah, S., Ismail, M., Fong, S. Y., & Ahmed, A. L. (2016). Evaluation for long term PM_{10} concentration forecasting using multi linear regression (MLR) and principal component regression (PCR) models. *EnvironmentAsia*, 9(2): 101-109.

Abdullah, S., Ismail, M., Fong, S. Y., & Ahmed, A. L. (2016). Evaluation for long term PM_{10} concentration forecasting using multi linear regression (MLR) and principal component regression (PCR) models. *Environment Asia*, 9(2): 101-109.

Abdul-Wahab, S. A., Bakheit, C. S., & Al-Alawi, S. M. (2005). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling and Software,* 20(10): 1263–1271.

Baklouti, R., Mansouri, M., Nounou, M., Nounou, H., & Hamida, A. Ben. (2015). Iterated robust kernel fuzzy principal component analysis and application to fault detection. *Journal of Computational Science,* 15: 34-49.

Chouykerd, P., Laosiripojana, N., & Sorapipatana, C. (2008). Economic assessment of compressed natural gas for diesel vehicle in Thailand. *Songklanakarin Journal of Science and Technology,* 30(6): 747–754.

Chua, Y.P. (2009). *Advanced Research Statistics: Regression test, factor analysis and SEM analysis.* McGraw-Hill Education. ISBN 978-983-3850-51-8

De Gennaro, G., Trizio, L., Di Gilio, A., Pey, J., Perez, N., Cusack, M., Alastuey, A., & Querol, X. (2013). Neural network model for the prediction of PM10 daily concentrations in two sites in the Western Mediterranean. *Science of the Total Environment,* 463–464: 875–883.

Department of Environment. (2014). *Environmental Quality Report 2014.* Department of Environment, Ministry of Sciences, Technology, and the Environment, Malaysia. Kuala Lumpur

Department of Statistics (2011). Population Distribution by Local Authority Areas ad Mukims, Putrajaya: Department of Statistics Malaysia.

Giri, D., Murthy, K., & Adhikary, P.R. (2008). The influence of meteorological conditions on PM10 concentrations in Kathmandu Valley. *International Journal of Environmental Research,* 2(1): 49–60.

Grantz, D. A., Garner, J. H. B., & Johnson, D. W. (2003). Ecological effects of particulate matter. *Environment International,* 29(2–3): 213–239.

Grivas, G., & Chaloulakou, A. Ã. (2006). Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. *Atmospheric Environment,* 40: 1216–1229.

Grundström, M. (2015). The influence of atmospheric circulation and meteorology on urban air pollution and pollen exposure (Doctoral thesis). Faculty of Science, University of Gothenburg.

Grundström, M., Hak, C., Chen, D., Hallquist, M., & Pleijel, H. (2015). Variation and co-variation of PM10, particle number concentration, NO_x and NO_2 in the urban air – Relationships with wind speed, vertical temperature gradient and weather type. *Atmospheric Environment,* 120: 317–327.

Hak, C., Larssen, S., Randall, S., Guerreiro, C., Denby, B., & Horalek, J. (2010). *Traffic and Air Quality - Contribution of Traffic to Urban Air Quality in European Cities*. Retrieved from http://acm.eionet.europa.eu/docs/ETCACC TP 2009 12 transport and air quality in cities.pdf.

Jacob, D.J., & Winner, D.A. (2009). Effect of climate change on air quality. *Atmospheric Environment,* 43: 51–63.

Jahn, H. J., Schneider, A., Breitner, S., Eißner, R., Wendisch, M., & Krämer, A. (2011). Particulate matter pollution in the megacities of the Pearl River Delta, China – A systematic literature review and health risk assessment. *International of Hygience and Environmental Health,* 214(4): 281–295.

Li, H., Guo, B., Han, M., Tian, M., & Zhang, J. (2015). Particulate matters pollution characteristic and the correlation between PM (PM2.5, PM10) and meteorological factors during the summer in Shijiazhuang. *Journal of Environmental Protection,* 6: 457–463.

Maté, T., Guaita, R., Pichiule, M., Linares, C., & Díaz, J. (2010). Short-term effect of fine particulate matter (PM2.5) on daily mortality due to diseases of the circulatory system in Madrid (Spain). *Science of the Total Environment,* 408(23): 5750–5757.

Met Office. (2017). High and low pressure. http://www.metoffice.gov.uk/learning/learn-about-the-weather/how-weather-works/ highs-and-lows/pressure, [accesed 17 August 2017]

Misiti, M., Misiti, Y., Poggi, J., & Portier, B. (2015). Mixture of linear regression models for short term PM10 forecasting in Haute Normandie (France). *Case Studies in Business, Industry and Government Statistics.* 6(1): 47–60.

Noyes, P.D., McElwee, M.K., Miller, H.D., Clark, W.B., Van Tiem, L.A., Walcott, K.C., Erwin, K.N., & Levin, E.D. (2009). The toxicology of climate change: Environmental contaminants in a warming world. *Environment International,* 35: 971–986.

Papanastasiou, D. K., Melas, D., & Kioutsioukis, I. (2007). Development and assessment of neural network and multiple regression models in order to predict PM10 levels in a medium-sized Mediterranean city. *Water, Air, and Soil Pollution,* 182(1–4): 325–334.

Pires, J. C. M., Martins, F. G., Sousa, S. I. V., Ferraz, M. C. M. A., & Pereira, M. C. (2008). Prediction of the daily mean PM10 concentrations using linear models. *American Journal of Environmental Sciences,* 4(5): 445-453.

Shahraiyni, H. T., & Sodoudi, S. (2016).

Statistical modeling approaches for pm10 prediction in urban areas; A review of 21st-century studies. *Atmosphere,* 7(2): 10–13.

Slini, T., Kaprara, A., Karatzas, K., & Moussiopoulos, N. (2006). PM10 forecasting for Thessaloniki, Greece. *Environmental Modelling & Software,* 21(4): 559–565.

Stadlober, E., Hörmann, S., & Pfeiler, B. (2008). Quality and performance of a PM10 daily forecasting model. *Atmospheric Environment,* 42(6): 1098-1109.

Ul-Saufie, A. Z., Yahaya, A. S., Ramli, N. A., & Abdul Hamid, H. (2011). Comparison between multiple linear regression and feed forward back propagation neural network models for predicting PM10 concentration level based on gaseous and meteorological parameters. *International Journal of Applied Science and Technology,* 1(4): 42–49.

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: 1488-1494.

Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., & Kolehmainen, M. (2011). Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki. *Science of the Total Environment,* 409(7): 1266–76.

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: Transport and Environment,* 34: 104–115.

Zhang, L. W., Chen, X., Xue, X. D., Sun, M., Han, B., Li, C. P., Ma, J., Yu.H., Sun, Z. R., Zhao, L. J., Zhao, B. X., Liu, Y. M., Chen, J., Wang, P. P., Bai, Z. P., & Tang, N. J. (2014). Long-term exposure to high particulate matter pollution and cardiovascular mortality: a 12-year cohort study in four cities in northern China. *Environment International,* 62: 41–7.