REGRESSION AND MULTIVARIATE MODELS FOR PREDICTING PARTICULATE MATTER CONCENTRATION LEVEL

AMINA NAZIF*, NURUL IZMA MOHAMMED, AMIRHOSSEIN MALAKAHMAD, AND MOTASEM S. ABUALQUMBOZ

Department of Civil and Environmental Engineering, Universiti Teknologi PETRONAS,
32610 Bandar Seri Iskandar, Perak Darul Ridzuan, Malaysia.
*aminanazif@yahoo.co.uk

Abstract

The devastating health effects of Particulate Matter (PM_{10}) exposure by susceptible populace has made it necessary to evaluate PM_{10} pollution. Meteorological parameters and seasonal variation increases PM_{10} concentration levels, especially in areas that have multiple anthropogenic activities. Hence, Stepwise Regression (SR), Multiple Linear Regression (MLR) and Principal Component Regression (PCR) analyses were used to analyse daily average PM_{10} concentration levels. The analyses were carried out using daily average PM_{10} concentration, temperature, humidity, wind speed and wind direction data from 2006 to 2010. The data was from an industrial air quality monitoring station in Malaysia. The SR analysis established that meteorological parameters had less influence on PM_{10} concentration levels having coefficient of determination (R^2) result from 23 % to 29 % based on seasoned and unseasoned analysis. While, the result of the prediction analysis showed that PCR models had better R^2 result than MLR methods. The results for the analyses based on both seasoned and unseasoned data established that MLR models had R^2 result from 0.50 to 0.60. While, PCR models had R^2 result from 0.66 to 0.89. In addition, the validation analysis using 2016 data also recognised that the PCR model outperformed the MLR model, with the PCR model for the seasoned analysis having the best result. These analyses will aid in achieving sustainable air quality management strategies.

Keyword: Air pollution; Particulate Matter; Prediction; Regression analysis

1. Introduction

Particulate Matter (PM) concentration has continuously remained an air pollutant of concern for the past decade. PM is an air pollutant mainly found in the atmosphere, it has the characteristics of being both a primary and a secondary pollutant (Abdullah et al, 2011). Sources of PM include industrial activities, combustion and non-combustion sources as well as land and forest fires which are usually in the form of dust, soot, and smoke (Juneng et al, 2011; Kassomenos et al, 2014). PM with aerodynamic measurement of ≤ 10 microns (PM_{10}) has been associated with a lot of devastating health effects including asthma, air way infections, heart attack and in severe cases death (Henderson et al, 2012; Sahani et al, 2014). Furthermore, the effect of PM_{10} is greatly influenced by seasonal variation and suitable atmospheric conditions

(Latif et al, 2014). Previous studies have emphasized that meteorological parameters can influence the behaviour of PM_{10} pollutants in the atmosphere (Gvozdić et al, 2011; Özbay 2012). Therefore understanding the effect of meteorological parameters on PM_{10} is vital to curtail the effects of PM_{10} concentration and have a sustainable PM_{10} management strategy (Wai et al, 2005).

One of the necessary air quality management strategy is to predict air pollutant concentration in advance, to be adequately prepared (Nejadkoorki et al, 2012). This alertness would reduce the impact of PM_{10} concentration and would help in curtailing future PM_{10} problems (Sfetsos et al, 2010). Previous studies have carried out tremendous studies on different statistical approaches to forecast PM_{10} concentration levels (Chen et al, 2013; Taşpınar 2015). However, there is the need to develop prediction models based on seasons and to validate these models using recent data to evaluate the performance of the models. Consequently, this study investigates the significance of two forecasting methods namely: Multiple Linear Regression (MLR) and Principal Component Regression (PCR) analysis in forecasting next day PM_{10} concentration of an industrial area based on seasoned and unseasoned data. Prior to the prediction analysis, the influential relationship between PM_{10} concentration and meteorological parameters was evaluated using Stepwise Regression (SR) analysis.

2. Materials

2.1 Study area

The industrial air quality monitoring station is situated in Taiping industrial area located in the state of Perak in peninsular Malaysia. The monitoring station is at Latitude E100°40.782' and Longitude N04°53.940'. The area is surrounded by industries and have substantial air pollution treat.

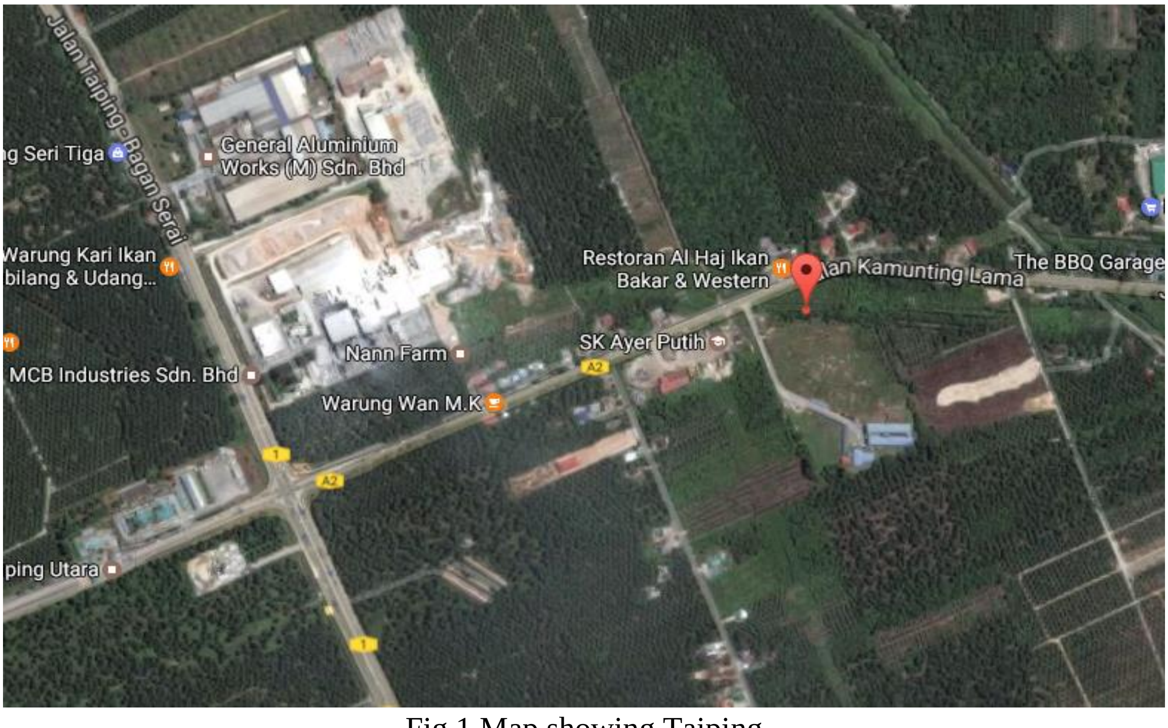

Fig.1 Map showing Taiping

Taiping industrial air quality monitoring station is located at Sekolah Kebangsaan Air Puteh, Taiping as shown in Figure 1. It is close to rubber plantations and industries. These industries include; Aluminium, Cement, and Mineral Industries.

2.2 Monitoring records

Two air quality monitoring equipment were used for this study. The Beta Attenuation Monitor (BAM) was used to obtain daily average data for five years (2006 to 2010), this data was used for the model development. Meanwhile, an E-sampler met one instrument was used to obtain data that was used for validating the prediction models developed. The daily average data was for four days (26-29th July 2016) and was used for the validation analysis. Both equipment sample real time PM_{10} concentration levels and meteorological parameters. The daily average data for prediction and validation analysis were: Particulate matter (PM_{10}), Temperature (T), Humidity (H), Wind speed (WS), and Wind direction (WD). The data was divided into two: seasoned and unseasoned. The seasoned data was divided into southwest monsoon (SWM) season from May to September while, northeast monsoon (NEM) season data was from November to March. Meanwhile, the unseasoned data was from January to December (Seasons were not acknowledged).

In this study, imputation procedures were not used for missing data analysis. Alternatively, missing data was omitted from the analysis, this was done to have a true representation of the data. Similar procedure was used previously (Juneng et al, 2011). Excel 2010 and Minitab 16 were used for the analyses.

3. Methods

3.1 Descriptive Statistics

Descriptive statistics was applied on the daily average PM_{10} data to understand the PM_{10} concentration of the area. These included maximum, minimum, and mean of the PM_{10} data. The descriptive statistics result for each year was displayed using a Box plot.

3.2 Multiple Linear Regressions (MLR) Analysis

MLR is widely used in assessing the relationship between a dependent variable and independent variables. The MLR equation is displayed in Eq.1.

$$y = b_o + \sum_{i=1}^{n} b_i X_i + \varepsilon \qquad \text{Eq.1}$$

where, b_o is the intercept, b_i is the regression coefficient, X_i represent the independent variables and ε is the error linked to the regression analysis.

3.3 Stepwise Regression (SR) Analysis

Stepwise regression (SR) analysis is a step by step approach, were inconsequential variables are removed from the regression analysis, allowing only important variables to be present. SR analysis has the ability to transform from a linear to a multiple linear equation (Thomas et al, 2007). The forward selection was used for this study. The analysis starts by choosing the important variables that contribute substantially to the analysis and subsequently add the variable that would improve the data most. This selection process is repeated on all the influential variables until no further improvement is achieved. SR equation is as shown in Eq.2.

$$y=\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n+\mathcal{E} \qquad \text{Eq.2}$$

Where, y is the dependent variable, β_o is the intercept, β_1, β_2,….β_n are the regression coefficients of the independent variables X_1, X_2,….. X_n and Ɛ is the residual error.

3.4 Hybrid Models

Hybrid models are a combination of more than one analytical method to enhance a desired output (Ul-Saufie et al, 2013). Hybrid model in this study combined multivariate and MLR analysis to reduce the complexity of the model, decrease multicollinearity and yield better result. The architecture of the hybrid model is displayed in Figure 1. Principal Component Analysis (PCA) was carried out using daily average PM_{10}, Temperature, Humidity, Wind speed, and Wind direction data. The principal component's (PC) with Eigenvalue ≥0.9 were used, while the PC's with Eigenvalues <0.9 were discarded from the analysis.

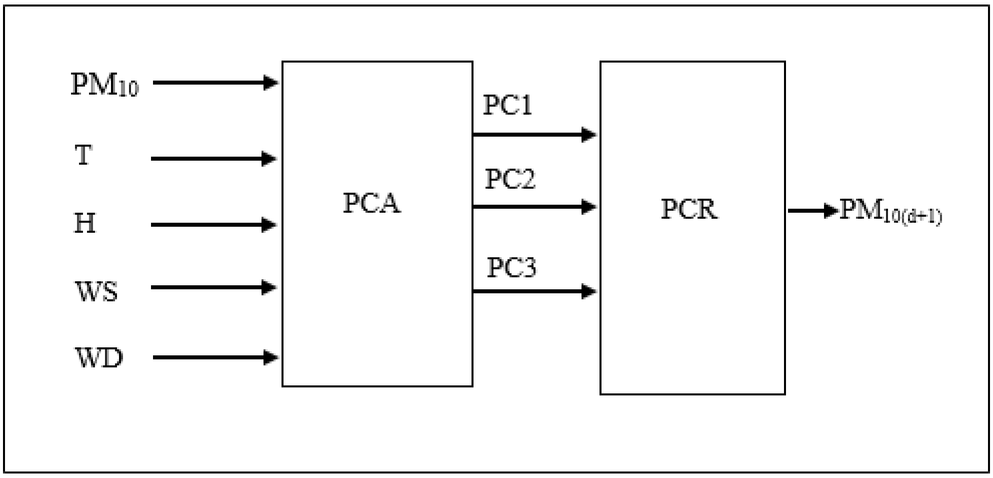

Figure 1: Architecture of Hybrid Model

The significant PC's coefficient were then analysed with the daily average data to develop the Principal Component Regression (PCR) models. The equations are regressed in multiple linear forms to develop the PCR prediction models (Pires et al, 2008).

3.5 Performance Indicators

Performance indicators as shown in Table 1 were used to assess the performance of the prediction models using; Coefficient of determination (R^2), Variance Inflation Factor (VIF), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Prediction Accuracy (PA), and Index of Agreement (IA).

Table 1: Performance Indicators

Performance Indicators	Equation
Coefficient of determination (R^2)	$R^2 = \left[\frac{\sum_{i=1}^{n}(O_i-\bar{O})^2.(P_i-\bar{P})}{n.\sigma_0.\sigma_p}\right]^2$
Variance Inflation Factor (VIF)	$VIF_i = \frac{1}{1-R_i^2}$
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i-O_i)^2}{n}}$
Mean Absolute Error (MAE)	$MAE = \frac{\sum_{i=1}^{n}\lvert P_i-O_i\rvert}{n}$
Prediction Accuracy (PA)	$PA = \frac{\sum_{i=1}^{n}(P_i-\bar{O})^2}{\sum_{i=1}^{n}(O_i-\bar{O})^2}$
Index of Agreement (IA)	$IA = 1 - \left[\frac{\sum_{i=1}^{n}(P_i-O_i)^2}{\sum_{i=1}^{n}(\lvert P_i-\bar{O}\rvert - \lvert O_i-\bar{O}\rvert)^2}\right]$

Where, n is the total number of measured data, Pi = predicted values, Oi = observed values, $\bar{O}$ = the mean observed values, $\bar{P}$ = mean of the predicted values, σ_p = standard deviation of the predicted values, and σ_o = standard deviation of the observed values. For the VIF analysis, VIF_i is the variance inflation factor while R_i^2 is the coefficient of determination of the predictor variables.

The models ability to predict daily average PM_{10} concentration levels >100 µg/m^3 (newly proposed limit for Malaysia) was evaluated using statistical evaluation methods: Probability of Detection (POD), False Alarm Rate (FAR), and Critical Success Index (CSI) as shown in Table 2 . These are statistical evaluation methods used for assessing a prediction model's ability to predict a particular benchmarked level (Chaloulakou et al, 2003; Papanastasiou et al, 2007).

Table 2: Statistical Evaluation for predicting high PM_{10} levels (100mg/m^3)

Index	**Equation**
Probability of Detection (POD)	A/(A+B)
False Alarm Rate (FAR)	C/(C+A)
Critical Success Index (CSI)	A/(A+B+C)

A= observed and predicted exceedances, B= observed but not predicted, and C= Predicted but not observed.

4. Results and discussion

The descriptive statistical analysis for Taiping as shown in Figure 2 illustrates that fluctuating PM_{10} concentrations levels were recorded from 2006 to 2010 in the study area. The difference in PM_{10} concentration levels from 2006 to 2010 could be attributed to dry and hot weather conditions observed in a particular year. This atmospheric condition is favourable for the formation of secondary PM_{10} pollutants that aid in increasing PM_{10} concentration levels especially in 2007 (D.o.E 2007). The whiskers represent the maximum and minimum, while the middle line represent the average PM_{10} concentration levels for a particular year. The analysis showed that all years had maximum daily average PM_{10} concentration above the World Health Organization (WHO) guideline of 50 µg/m^3. Although it was lower than the Malaysian Ambient Air Quality Guidelines (MAAQG) of 150 µg/m^3. The maximum level of 140 µg/m^3 in 2007 was the peak of PM_{10} concentration level in the area.

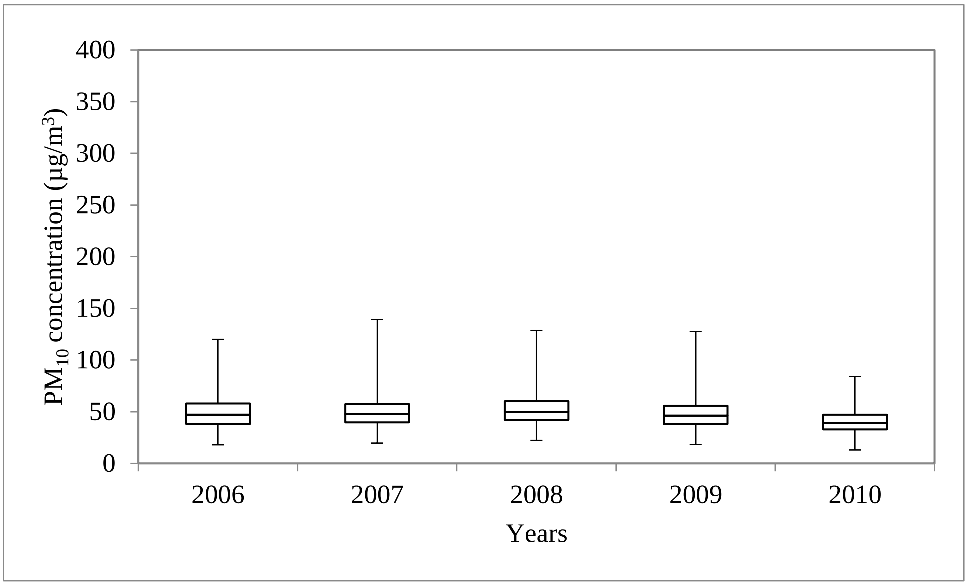

Figure 2: Box Plot of Taiping (2006 to 2010)

Table 3 showed that a total of 1245 observations were used for the five years analysis, 68 % of data was used while the missing data was 31 %. The skewness result was 1.44 showing substantial PM_{10} concentration levels for the five years data combined. A mean of 51 µg/m^3 was ascertained for the five years analysis. This signifies that the area had annual PM_{10} concentration level above the annual WHO guideline limit (annual mean 20 µg/m^3) and slightly exceeding the annual MAAQG limit (50 µg/m^3).

Table 3: Descriptive Statistics of Taiping from 2006 to 2010

Mean	Median	Mode	Maximum	Minimum	Count	Standard Deviation	Standard Error
51	47	45	140	21	1245	16.78	0.48

The SR analysis result in Table 4 showed that the unseasoned analysis had four significant variables: $PM_{10(d-1)}$, T, WS and H. In the analysis, WD was not chosen by the SR model as an important input in the area as shown in the model. Thus, establishing that WD had less significance in influencing PM_{10} concentration levels. Meteorological parameters and previous day PM_{10} ($PM_{10(d-1)}$) concentration explained 23 % and 47 % variability of daily average PM_{10} concentration, respectively. For the seasoned analysis, meteorological parameters explained more than 20 % PM_{10} variability for both seasons. Previous day PM_{10} concentration levels accounted for 40 % and 50 % PM_{10} variability for NEM and SWM seasons, respectively.

Table 4: Stepwise Regression Models

PM_{10} Models		
Unseasoned Analysis	R^2 (%)	AdjR (%)
$Y=15.77+0.69PM_{10(d-1)}$	47	47
$Y= -102.27+0.61PM_{10(d-1)}+4.46T$	56	56
$Y=-100.23+0.61PM_{10(d-1)}+4.88T-3.00WS$	57	57
$Y=-110.97+0.61PM_{10(d-1)}+5.09T-3.00WS+0.06H$	57	57
$Y=-160.89+7.73T+0.14H-2.46WS$	23	23
Seasoned Analysis		
Southwest Monsoon		
$Y=16.21+0.703\ PM_{10(d-1)}$	50	49
$Y=-130.68+0.633\ PM_{10(d-1)}+5.44T$	59	59
$Y=-120.04+0.643\ PM_{10(d-1)}+5.72T-4.04WS$	61	61
$Y=-159.98+0.634\ PM_{10(d-1)}+6.50T-3.91WS+0.24H$	61	61
$Y=-189.06+0.616\ PM_{10(d-1)}+7.22T-3.68WS+0.40H-0.022WD$	62	61
$Y=-344.42+12.05T+1.11H-1.608WS-0.0761WD$	29	28
Northeast Monsoon		
$Y=17.94+0.64\ PM_{10(d-1)}$	40	40
$Y=-81.00+0.541\ PM_{10(d-1)}+3.82T$	49	48
$Y=-83.20+0.539\ PM_{10(d-1)}+4.13T-1.38WS$	49	49
$Y=-80.85+0.530\ PM_{10(d-1)}+4.00T-1.68WS+0.0136WD$	50	49
$Y=-82.327+5.00T-3.379WS+0.048WD$	24	24

It was ascertained that increasing Temperature and Humidity as well as decreasing Wind speed and Wind direction can support the accumulation of PM_{10} concentration levels (Kassomenos et al, 2014). Increasing humidity results in the decrease in aerosol particles that can reduce the concentration of PM_{10} pollution especially in the SWM season. However, Humidity was not assigned any significance in influencing PM_{10} concentration in the NEM season. This suggests that increasing Temperature and Wind direction as well as decreasing wind speed is attributed to the increase in PM_{10} particles (Kassomenos et al, 2012). This increase is through the chemical formation of secondary PM_{10} particles (Majumder et al, 2012), as well as less dilution of PM_{10} which increases the PM_{10} pollution concentration in the area (Huang et al, 2016). This combined condition is attributed to temperature inversion that leads to the increase in PM_{10} pollution concentration (Kingham et al, 2006). Also, it can be stated that various meteorological parameters have peculiar influence on PM_{10} concentration levels. These effects are distinct due to discrete site location and divergent influential capacity of each meteorological parameter (Pant et al, 2012). From the SR analysis, it can be acknowledged that previous day PM_{10} concentration have influence on PM_{10} concentration levels and establishes the particles retentive capacity. This was also established in a previous study (Kassomenos et al, 2014).

Next was the PCA that was used for the hybridization process, the PCA result was used to develop the PCR models. The PCA varimax rotation result for both seasoned and unseasoned analysis is shown in Figure 3 and 4. The Unseasoned analysis result showed that three PCs had Eigenvalue of 1.78, 1.26, and 0.96. PC1, PC2, and PC3 had a cumulative percentage of 80 % and the variance percentage of 20 % each.

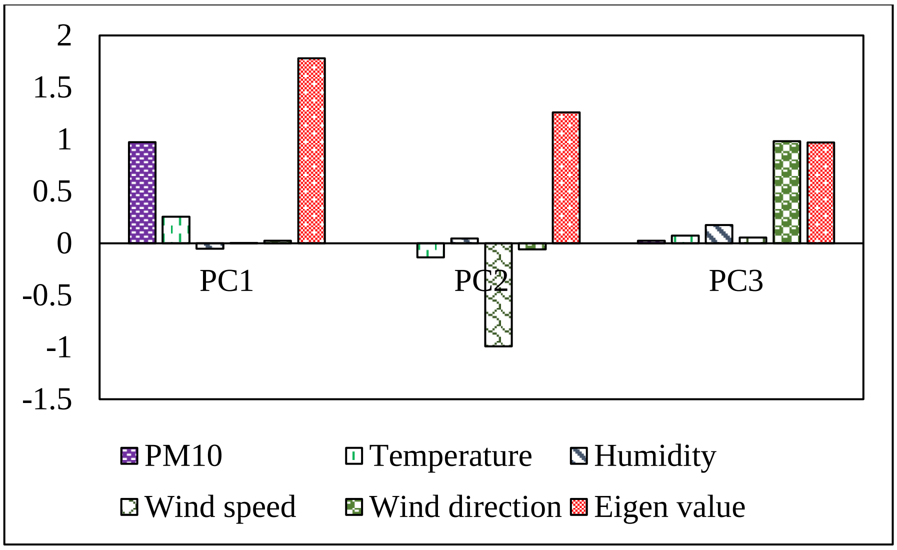

Figure 3: Varimax Rotation Result for Unseasoned Analysis

The heavily loaded variable for PC1 was previous day PM_{10} having a positive relation. While, wind speed had a negative correlation in PC2 and wind direction had a positive correlation in PC3.

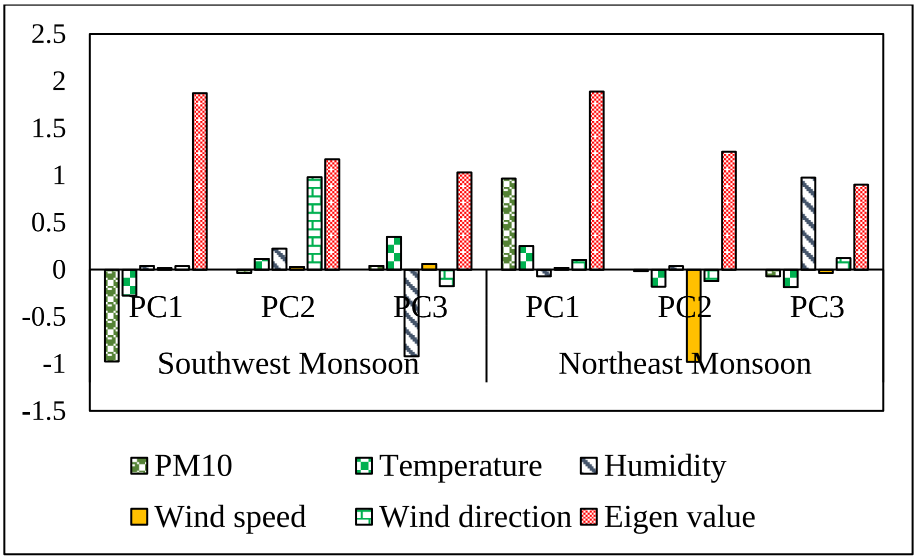

Figure 4: Varimax Rotation Result for Seasoned Analysis

The Varimax rotation result for seasoned analysis showed that SWM and NEM PC 1 to 3 had Eigenvalues >0.9 and a variance of 20 % for each PC.

Subsequently, the prediction models for both seasoned and unseasoned analysis were shown in Table 4 and Table 5.

Table 4: Unseasoned Prediction Models

Methods	R^2	RMSE	MAE	NAE	IA	PA
MLR	0.57	11.02	8.06	0.16	0.85	0.57
PCR	0.84	6.82	4.93	0.16	0.95	0.84

The analysis established that the PCR models outperformed the MLR models having higher R^2 result, lower error levels and higher predictability for both the seasoned and unseasoned analysis. Similar performance was achieved by previous studies although the seasonal analysis was not regarded (Ul-Saufie et al, 2013).

Table 5: Seasoned Prediction Models

Seasons	Methods	R^2	RMSE	MAE	NAE	PA	IA
SWM	MLR	0.62	12.83	9.68	0.18	0.62	0.87
	PCR	**0.89**	**6.73**	**5.12**	**0.09**	**0.89**	**0.97**
NEM	MLR	0.50	8.69	6.67	0.14	0.50	0.81
	PCR	**0.66**	**7.08**	**4.94**	**0.10**	**0.66**	**0.89**

However, the statistical evaluation result established that the PCR model for the SWM season performed better in predicting daily average PM_{10} concentration >100 µg/m^3. Having high POD and CSI result as shown in Table 6, the FAR result also established that low false alarm would be achieved for the SWM PCR model. No results were established for the NEM season as the daily average PM_{10} concentration recorded for the season was < 100 µg/m^3.

Table 6: Statistical Evaluation of Prediction Models

Limit	Unseasoned Analysis				Seasoned Analysis					
					Southwest Monsoon			Northeast Monsoon		
µg/m^3	Methods	POD	FAR	CSI	POD	FAR	CSI	POD	FAR	CSI
100	MLR	0.09	0.33	0.09	0.17	0.50	0.16	-	-	-
	PCR	0.32	0.13	0.30	0.72	0.13	0.72	-	-	-

The validation analysis result as shown in Table 7 established that the PCR models outperformed the MLR models in terms of predictability and low error levels. Also, the seasonal PCR model was the best prediction model having the highest predictability and lowest error levels as compared with the other models.

Table 7: Validation of Prediction Models

	Model	RMSE	MAE	NAE	IA
Unseasoned	MLR	41.50	37.75	2.36	0.90
	PCR	12.90	12.50	0.78	0.99
Seasoned	MLR	50.00	47.75	2.99	0.86
	PCR	**10.26**	**9.75**	**0.61**	**0.99**

Conclusion

The study ascertained that the influence of previous day PM_{10} concentration levels were higher than meteorological parameters for both seasoned and unseasoned analysis. Also, the influence of meteorological parameters on PM_{10} concentration level was distinct depending on the season. Based on the prediction analysis, the PCR model performed better than the MLR model in predicting next day average PM_{10} concentration levels. However, the seasonal PCR model was a better predictor of next day PM_{10} concentration levels. Also, the statistical evaluation method established that the PCR model had better ability to detect daily average PM_{10} concentration levels above 100 $\mu g/m^3$.

Acknowledgement

Appreciation goes to Universiti Teknologi PETRONAS for making this study possible, the Department of Environment (DOE) Malaysia for providing the data used for this study, and Clean Air Research group Universiti Sains Malaysia for providing the E-sampler met one equipment used for this study.

References

Abdullah, N., Shuhaimi, S., Toh, Y., Shafee, A., & Maznorizan, M. (2011). The Study of Seasonal Variation of PM10 Concentration in Peninsula, Sabah and Sarawak. *Malaysian Meteorological Department*(9).

Chaloulakou, A., Grivas, G., & Spyrellis, N. (2003). Neural network and multiple regression models for PM10 prediction in Athens: a comparative assessment. *Journal of the Air & Waste Management Association, 53*(10), 1183-1190.

Chen, Y., Shi, R., Shu, S., & Gao, W. (2013). Ensemble and enhanced PM 10 concentration forecast model based on stepwise regression and wavelet analysis. *Atmospheric Environment, 74,* 346-359.

D.o.E. (2007). Malaysia Environmental Quality Report. Malaysia: Ministry of Natural Resources.

Gvozdić, V., Kovač-Andrić, E., & Brana, J. (2011). Influence of meteorological factors NO2, SO2, CO and PM10 on the concentration of O3 in the urban atmosphere of Eastern Croatia. *Environmental modeling & assessment, 16*(5), 491-501.

Henderson, S. B., & Johnston, F. H. (2012). Measures of forest fire smoke exposure and their associations with respiratory health outcomes. *Current opinion in allergy and clinical immunology, 12*(3), 221-227.

Huang, L., Chen, M., & Hu, J. (2016). Twelve-Year Trends of PM10 and Visibility in the Hefei Metropolitan Area of China. *Advances in Meteorology, 2016*.

Juneng, L., Latif, M. T., & Tangang, F. (2011). Factors influencing the variations of PM10 aerosol dust in Klang Valley, Malaysia during the summer. *Atmospheric environment, 45*(26), 4370-4378.

Kassomenos, P., Vardoulakis, S., Chaloulakou, A., Grivas, G., Borge, R., & Lumbreras, J. (2012). Levels, sources and seasonality of coarse particles (PM 10–PM 2.5) in three European capitals–Implications for particulate pollution control. *Atmospheric environment, 54,* 337-347.

Kassomenos, P., Vardoulakis, S., Chaloulakou, A., Paschalidou, A., Grivas, G., Borge, R., & Lumbreras, J. (2014). Study of PM 10 and PM 2.5 levels in three European cities: analysis of intra and inter urban variations. *Atmospheric environment, 87*, 153-163.

Kingham, S., Durand, M., Aberkane, T., Harrison, J., Wilson, J. G., & Epton, M. (2006). Winter comparison of TEOM, MiniVol and DustTrak PM 10 monitors in a woodsmoke environment. *Atmospheric environment, 40*(2), 338-347.

Latif, M. T., Dominick, D., Ahamad, F., Khan, M. F., Juneng, L., Hamzah, F. M., & Nadzir, M. S. (2014). Long term assessment of air quality from a background station on the Malaysian Peninsula. *Sci Total Environ, 482-483*, 336-348. doi: 10.1016/j.scitotenv.2014.02.132

Majumder, A. K., Islam, K. N., Bajracharya, R. M., & Carter, W. S. (2012). Assessment of occupational and ambient air quality of traffic police personnel of the Kathmandu valley, Nepal; in view of atmospheric particulate matter concentrations (PM 10). *Atmospheric Pollution Research, 3*(1), 132-142.

Nejadkoorki, F., & Baroutian, S. (2012). Forecasting extreme PM10 concentrations using artificial neural networks. *Int. J. Environ. Res, 6*(1), 277-284.

Özbay, B. (2012). Modeling the effects of meteorological factors on SO2 and PM10 concentrations with statistical approaches. *CLEAN–Soil, Air, Water, 40*(6), 571-577.

Pant, P., & Harrison, R. M. (2012). Critical review of receptor modelling for particulate matter: a case study of India. *Atmospheric environment, 49*, 1-12.

Papanastasiou, D., 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., Martins, F., Sousa, S., Alvim-Ferraz, M., & Pereira, M. (2008). Selection and validation of parameters in multiple linear and principal component regressions. *Environmental Modelling & Software, 23*(1), 50-55.

Sahani, M., Zainon, N. A., Mahiyuddin, W. R. W., Latif, M. T., Hod, R., Khan, M. F., . . . Chan, C.-C. (2014). A case-crossover analysis of forest fire haze events and mortality in Malaysia. *Atmospheric environment, 96*, 257-265.

Sfetsos, A., & Vlachogiannis, D. (2010). A new methodology development for the regulatory forecasting of PM10. Application in the Greater Athens Area, Greece. *Atmospheric environment, 44*(26), 3159-3172.

Taşpınar, F. (2015). Improving Artificial Neural Network Model Predictions of Daily Average PM10 Concentrations by Applying PCA and Implementing Seasonal Models. *Journal of the Air & Waste Management Association*(just-accepted).

Thomas, S., & Jacko, R. B. (2007). Model for forecasting expressway fine particulate matter and carbon monoxide concentration: application of regression and neural network models. *Journal of the Air & Waste Management Association, 57*(4), 480-488.

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.

Wai, K., & Tanner, P. A. (2005). Extreme particulate levels at a western pacific coastal city: the influence of meteorological factors and the contribution of long-range transport. *Journal of atmospheric chemistry, 50*(2), 103-120.