INTERNATIONAL JOURNAL OF CONSERVATION SCIENCE

ISSN: 2067-533X Volume 13, Issue 3, July-September 2022: 1065-1078 www.ijcs.ro

MODELING OF PARTICULATE MATTER (PM_{10}) DURING HIGH PARTICULATE EVENT (HPE) IN KLANG VALLEY, MALAYSIA

Nursyaida Amila MOHD RIDZUAN[1,] , Norazian MOHAMED NOOR[1,2,*], Nur Alis Addiena A RAHIM[1,2], Izzati Amani MOHD JAFRI[1,2], Ahmad Zia UL SAUFIE[2,3], Mohd Remy Rozainy MOHD ARIF ZAINOL[2,4], Mohamad Anuar KAMARUDDIN[2,5], Gyorgy DEÁK[2,6]

[1]Faculty of Civil Engineering Technology, Universiti Malaysia Perlis, Kompleks Pengajian Jejawi 3, 02600 Arau, Perlis, Malaysia
[2]Sustainable Environment Research Group (SERG), Centre of Excellence Geopolymer and Green Technologies (CEGeoGTech), Universiti Malaysia Perlis, Kompleks Pengajian Jejawi 2, 02600 Arau, Perlis, Malaysia
[3]Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara (UiTM) 40450 Shah Alam, Selangor, Malaysia
[4]School of Civil Engineering, Universiti Sains Malaysia, Engineering Campus, Nibong Tebal 14300, Pulau Pinang, Malaysia
[5]School of Industrial Technology, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia.
[6]National Institute for Research and Development in Environmental Protection Bucharest (INCDPM), 294 Splaiul Independentei Street, 6th District, 060031 Bucharest, Romania

Abstract

Particulate matter (PM_{10}) is the main air pollutant during high particulate event (HPE) or also known as haze in Southeast Asia specifically in Malaysia. PM_{10} emanation is believed to cause the strongest harm to public health and environment during this time. Therefore, it is very important to develop good PM_{10} prediction model during these event that can be used to give the early warning to the public. A database with hourly PM_{10} concentration together with other trace gases and weather parameters were obtained from Department of Environment (DOE) Malaysia. The dataset was obtained from 2012 to 2016 at two study areas located in Klang Valley, namely, Petaling Jaya and Shah Alam. Three predictive models namely Multiple Linear Regression (MLR), Principle Component Regression (PCR) and Artificial Neural Network (ANN) were developed to predict the concentration of PM_{10} for the next-day, next-two-day and next-three-day. The predicted values were evaluated using several performance indicators i.e. Normalised Absolute Error (NAE), Root Mean Squared Error (RMSE), Prediction Accuracy (PA), Coefficient of Determination (R^2) and Index of Agreement (IA). ANN was selected as the best prediction model for PM_{10} concentration during HPE with the smallest average error (NAE = 0.11; RMSE = 9.69) and highest agreement with the observed values with the average of performances of R^2 = 0.97.

***Keywords**: Particulate matter; Haze; Air quality modeling; Linear regression; ANN*

Introduction

Air pollutant can be characterized as any form of particulates, biological molecules, or other harmful gases such as smoke, soot, fly ash, dust, dirt, fumes, vapours, odours that can be reason of illness, mortality and harm to other living being such as food crops, or the characteristic

* Corresponding author: norazian@unimap.edu.my

or built environment [1]. The most elevated air pollutant influenced by numerous nations can be distinguished as particulate matter (PM), which has aerodynamic diameter less than 10µm or can be known as PM_{10}. PM_{10} is produced in numerous pure and manmade forms which is therefore involved in chemical processes and physical aspects in the ambient air and could affect human well-being, climate and the natural environment. Prior to the actual outflow to the environment, PM_{10} remains in a coastal state to make a climate vaporized with various properties – depending on the particle's morphology, division, shape, surface and chemical composition. The chemical structure of particulate matter is diverse and depends on the type of emission sources as well as the time of residence within the air [2].

PM_{10} concentration can rise up tremendously due to smoke discharges from biomass burning. Biomass burning has played a recognizable part in unforeseen extreme haze episodes, known as high particulate event (HPE) that covered with the essential and auxiliary pollutants determined from motor debilitates and coal combustion [3]. Deficient combustion, especially from huge ranges of biomass burning ordinarily supplies high quantity of smoke and fine particles to the atmospheric conditions. Due to territorial winds, these fine particles can be transported across boundaries from their primary causes [4, 5].

Air pollutants emitted from haze in one country will have an effect on that country and the near and regional countries further. These events, mainly due to the very dry weather associated with the incidence of robust El Niño situations, resulted in the international transportation of unsafe pollutants from the original pollution sources from Indonesia's Borneo and Sumatra to the highly populated Malaysian Peninsular [6]. This haze events have been suffered by country at Southeast Asia (SEA), such as Singapore, Brunei Darussalam, Thailand, Indonesia and Malaysia. The creation of haze is particularly related to weather patterns, air pollutant emissions and conversion of gas to particles. Haze is structured under stable weather conditions such as low winds and reversed thermodynamic elements due to high particle concentrations and gas-to-particle conversion [7]. Meteorological variables such as atmospheric pressure, precipitation, humidity levels, atmospheric stabilization, physical and chemical relationships will determine the destiny and composition of suspended particles when they are transported [8-10]. Due to the perseverance of fine particles in the air and the ability of particles to be transferred over wider areas, the effect of haze is commonly geographic in scale.

HPE events and its associated deterioration of air quality also affected Malaysia's economy, not to mention the agricultural production and biological diversity of the country. As detailed by *Samsuddin et al.* [11], episodes of haze have resulted in a decrease in production and development output and a decrease in tourist industry profit, particularly because of cancellation of flights. A study by *Manan et al.* [12] economic losses due to ambulatory expenses during haze episodes in this country during 2005, 2006, 2008 and 2009 resulted in losses of USD 91,000 per year with an average loss of USD 4,789 per haze day occurrence.

Aware of the incredible impacts caused by the air pollution phenomena, many researchers had been studied vatious air quality predictions that may give vital data in arrange to require actions for the public before the pollution happens. A broad range of administrative warning technologies for taking precautions before or during air pollution periods have been established using statistical and models. Commonly used techniques linear regression, and non-linear parametric (artificial neural networks (ANN)) or non-parametric models had been used by most previous research. Multiple linear regression method gave well to moderate performances as reported by many researchers [13-16]. Machine learning algorithm such as ANN was also among the popoular method in predicting PM_{10} concentration. From the studies conducted, ANN was stated to be a superior method compared to linear regression due to due to the ability to predict a non-linear pattern of the air pollutant [17-21]. Some of researchers combined two or more models in order to give better forecasting of PM_{10} concentration levels. To reduce the level of ambient air pollutants, *Ismail et al.* [22] conducted a study using statistical models through Multiple Linear Regression (MLR) and combination of Principal Component Analysis (PCA) and Multiple Linear Regression (MLR), can be known as Principal Component Regression (PCR) in industrial areas of Seberang Prai, Pasir Gudang, Kemaman and Nilai, Malaysia. Generally, MLR gives greater prediction model level of PM_{10} concentration at all stations compared to PCR based on performance indicators. The results obtained demonstrated that the percentage of model grow in accuracy with percent of 9.25, 12.53, 20.91, and 56.21% compared to the PCR model.

Among of these models developed by previous researchers mentioned earlier were not conducted during the high particulate event (HPE). Knowing the great impact of HPE to social, economic and environment, this study attempt to model PM_{10} level during HPE. Three statistical models including a hybrid model of particulate matter (PM_{10}) concentration during HPE were developed using Artificial Neural Network (ANN), Multiple Linear Regression (MLR) and Principal Component Regression (PCR) –a combination of Principle Component Analysis (PCA) and MLR. The performances of the prediction models were evaluated and the best model was selected using several performance measures.

Expermental Part

Study Location

The selected study areas are located in peninsular Malaysia specifically in Klang Valley, Malaysia that includes Kuala Lumpur and its subsurbs and neighbouring cities and towns in the state of Selangor. Titiwangsa Mountains geographically outline Klang Valley and the Straits of Malacca with 6.9 million of populations recorded in 2013. Klang Valley is now recognized as an economic city in Malaysia with all of extensive physical development of public infrastructure, industrialization and rapid urbanization that has gradually deteriorated the air quality [22]. The specific location of selected study areas is in Petaling Jaya and Shah Alam. The description of the selected location for this project are tabulated in Table 1 below.

Table 1. Geographical of the two air quality monitoring stations in Klang Valley

Location ID	Station Location	Coordinate		Background
		Latitude (E)	Longitude (N)	
CA016	Sek. Ren. Sri Petaling, Petaling Jaya	E 101 38.322	N 03 06.553	Industrial Areas
CA025	Sek. Keb. TTDI Jaya, Shah Alam, Selangor	E 101 33.375	N 03 06.281	Urban and Residential Areas

Monitoring dataset

Secondary dataset from Department of Environment (DOE), Malaysia consists of hourly data of trace gases including PM_{10}, SO_2, NO_2, O_3, CO concentration and weather parameters such as wind speed, wind direction, ambient temperature and humidity for 5 years from 2012 until 2016 were used in this study. The parameters and the units were tabulated in Table 2 below.

Table 2. Parameters included in this research

Parameter	Symbol	Unit
Particulate Matter	PM_{10}	$\mu g/m^3$
Sulphur Dioxide	SO_2	ppm
Nitrogen Dioxide	NO_2	ppm
Ozone	O_3	ppm
Carbon Monoxide	CO	ppm
Wind Speed	WS	km/hr
Ambient Temperature	T	°C
Humidity	H	%

Table 3 shows the statistical test for the hourly PM_{10} measurement records at the study areas. It can be seen that the skewness was > +1 indicating the dataset were skewed to the right. This shows that extreme event occurred and the higher value of skewness was recorded in Petaling Jaya. High range of PM_{10} level was observed (129 and 409$\mu g/m^3$ in Petaling Jaya and Shah Alam repsectively) showing high variability of PM_{10} dataset in these two areas. This study focuses on modeling PM_{10} concentration during high particulate event. According to Malaysian Ambient Air Quality Standard (MAAQS), the averaging time for PM_{10} for 24 hours is 150$\mu g/m^3$. Hence, only PM_{10} measurement that were >150$\mu g/m^3$ were used as an input to the predictive model. The total number of PM_{10} measurement records used as an input (independent variables) were 1863 and 2328 for Petaling Jaya and Shah Alam respectively.

Table 3. Descriptive statistics for PM_{10} measurement record from 2012 to 2016

Study Area	Mean	Median	Maximum	Minimum	Range	Skewness	N >150 $\mu g/m^3$
Petaling Jaya	51.12	44.6	390	5	385	3.27	1863
Shah Alam	53.23	46.2	426	5	421	3.07	2328

Model Development

Overall, three models were developed, i.e., Multiple Linear Regression (MLR), Principle Component Regression (PCR) and Artificial Neural Network (ANN). For MLR and ANN, the measurement records of hourly particulate matter (PM_{10}), weather parameters (wind speed (WS) ambient temperature (T), humidity (H), and other pollutants (O_3, SO_2, NO_2, and CO) were used as input. In the PCR model, the principal components (PCs) were used as input. The output for this study is the prediction value of PM_{10} concentration for the next-day ($PM_{10\ (t+24)}$), the next-two-day ($PM_{10\ (t+48)}$) and the next-three-day ($PM_{10\ (t+72)}$). Random partition of the dataset was conducted using SPSS where 80% of the data were used for model development and the remaining data (20%) were used for model validation.

Multiple Linear Regression (MLR)

The MLR is a linear procedure in which an implication is made with one or more reference sites about the spatial link between the concentrations of the target site. The obtainable dataset are being modelled with the use of linear predictor functions. Linear regression is generally the most utilized analysis in statistical applications and involves training in order to estimate the values of definite coefficients, following the equation [23]:

$$z_i = b_0 + b_1x_1 + b_2x_2 + \cdots + \varepsilon, \qquad (1)$$

where: z_i is the target site data, x is the reference site data, b is the regression coefficients that correlate the response variable (air pollutant concentrations at the target site) and ε is the error.

In this study, $PM_{10\ (t+24)}$, $PM_{10\ (t+48)}$ and $PM_{10\ (t+72)}$ was the dependent variable and PM_{10} and the trace gases (O_3, SO_2, NO_2, and CO concentration) and weather parameters (T, H and WS) were the explanatory variables.

Principle Component Regression (PCR)

Principle Component Regression (PCR) is a hybrid model where it is the combination of principle component analysis (PCA) and multiple linear regression (MLR). In this study, Principle Component Analysis was first developed as a determination method for grouping the descriptors and selecting the best group of them as input for the MLR. Prior to conducting Principal Component Analysis (PCA), the Kaiser–Meyer–Olkin (KMO) and Bartlett's test of sphericity tests needed to be performed. KMO test was used to measure sampling adequacy for each variable in the model. The value of KMO must be greater than 0.5, showing that the data are adequate [24]. In addition, Bartlett's test of sphericity was applied to show a high degree of relationship between the parameters and that the data are suitable for factor analysis ($p < 0.001$). These re-quirements had been completed before the Principal Component Analysis. PCA is generally written as below [24]:

$$PC_i = A_{1i}X_{1j} + A_{2i}X_{2j} + \cdots + A_{ni}X_{nj}, \qquad (2)$$

where: PC_i is ith principal component, A_{ji} is the loading of the observed variable, X is the measured value of variables, i is the component number, j is the sample number, and n is the total number of variables.

The principal components (PCs) generated by PCA is advisable to rotate them using varimax rotation with the eigenvalues greater than 1 [25]. Varimax factors (VFs) coefficient with a correlation from 0.75 are considered as a strong significant factor loading; those that range from 0.50–0.74 are moderate, while 0.30–0.49 are classified as weak significant factor loading [24-27]. The principle regression analysis (PCR) of these PCs as independent variables will yield appropriate estimation of the parameter. Architecture of the PCR model for prediction of $PM_{10\ (t+24)}$ for the next day is shown in figure 1.

Artificial Neural Network (ANN)

The structure of neural network developed includes of interrelated connections of artificial processing units, namely neurons, and they process information by error minimization within a finite computation loop. A neural network can be trained to digest a complex relationship between two or more distribution of datasets. The feedforward back-propagation (FFBP) neural network is the most frequent neural network used to predict PM_{10} concentrations where the input has a non-linear pattern and consisted of the pattern of its neurons. The input signal is commanded from the input layer after it is processed to the output layer. Each of the neurons in the hidden layer and output layer is functioned by a non-linear function, which can be known as activation function where it depends on a weighted total of its inputs and neuron-specific parameter, namely bias. The number of neurons used in this study were from 1 to 20. For the transfer function, three commonly applied transfer function were tested namely log-sigmoid, tangent-sigmoid and purelin functions.

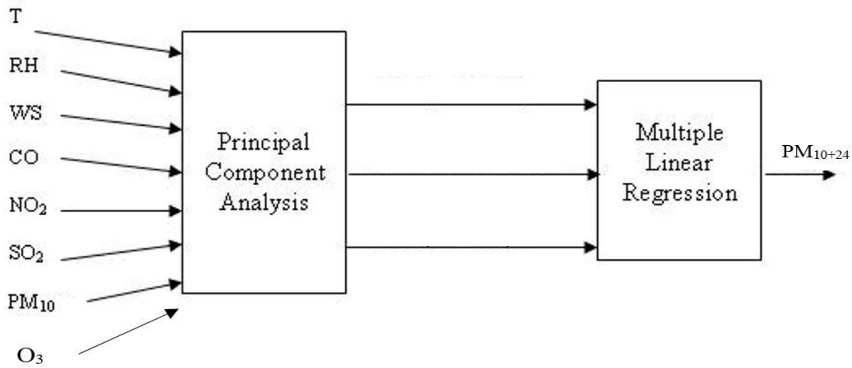

Fig. 1. Architecture of PCR model for prediction of PM_{10} for the next day

Performance Measure

After all the models had been developed, the performances of the model were evaluated using model performance assessment. In this research, five performance measures had been calculated to obtain the agreement between predicted and observed hourly PM_{10} concentration during high particulate event. The tests are Normalized Absolute Error (NAE), Root Mean Squared Error (RMSE), Index of Agreement (IA), Prediction Accuracy (PA) and coefficient of determination (R^2). The formula of all the performance indicators mentioned above are given as in Table 4 [28].

Table 4. Performance Measure [28]

Performance Measure	Formula	Description
Normalised Absolute Error (NAE)	$\frac{\sum_{i=1}^{n} \lvert P_i - O_i \rvert}{\sum_{i=1}^{n} O_i}$	Closer to 0 indicates better method
Root Mean Squared Error (RMSE)	$\frac{1}{n-1}\sum_{i=1}^{n}(P_i - O_i)^2$	
Index of Agreement (IA)	$1 - \left[\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(\lvert P_i - \bar{O}\rvert + \lvert O_i - \bar{O}\rvert)^2}\right]$	Closer to 1 indicates better method
Prediction Accuracy (PA)	$\frac{\sum_{i=1}^{n}(P_i - \bar{O})^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2}$	
Coefficient of Determination (R^2)	$\frac{\sum_{i=1}^{n}(P_i - \bar{P})\,(O_i - \bar{O})}{n.S_{pred}S_{obs}}$	

Where: n is the sum of hourly measurements at particular site, P_i is the predicted values of a set of hourly monitoring record, O_i is the observed values of a set of hourly monitoring record, $\bar{P}$ is the mean predicted value of a set of hourly monitoring record, $\bar{O}$ is the mean observed value of a set of hourly monitoring record, and S_{pred} is the standard deviation of the predicted value of a set of hourly monitoring record and S_{obs} is the standard deviation of the observed value of a set of hourly monitoring record between the input and output vectors.

Results and discussion

Principle Components (PCs) as the input of PCR Model

Table 5 shows the results for Kaiser–Meyer–Olkin (KMO) and Bartlett's test of sphericity. The KMO values were greater than 0.5 and the significant p-value for Bartlett's Test were smaller than 0.001 for both stations. Hence, these datasets were suitable for PCA.

Table 5. Parameters included in this research

Station	KMO Measure of Sampling Adequacy	Bartlett's Test of Sphericity	
		Approximate Chi-Square	p-Value
Petaling Jaya	0.688	8292.85	0.000
Shah Alam	0.577	14109.95	0.000

After the extraction of PCA was applied, factors that were taken as the principal component (PCs) were based on eigenvalues of more than 1 (>1.0) and varimax rotation was used as a criterion. The eigenvalues for all linear components before extraction, after extraction, and after rotation are shown in Table 6. Based on the percentages of the eigenvalues, the most significant principal component in explaining the amount of variance for Petaling Jaya is the first, followed by the second and third principal components whereas for Shah Alam, only two principle componets were included.

Table 6. Total Variance Explained for Petaling Jaya and Shah Alam

Component	Station	Initial Eigenvalues		
		Total	Variance (%)	Cumulative (%)
1	Petaling Jaya	3.424	42.803	42.803
2		1.582	19.776	62.579
3		1.050	13.120	75.699
1	Shah Alam	3.805	47.567	47.567
2		1.744	21.804	69.371

Table 7. Sub model for PCR model in Petaling Jaya and Shah Alam

Area	Principle Component (PCs)	Sub-model
Petaling Jaya	PC1	$0.955T – 0.901H +0.819 WS +0.805O_3$
	PC2	$0.876PM_{10} + 0.867 CO$
	PC3	$0.832SO_2 – 0.588NO_2$
Shah Alam	PC1	$0.970T + 0.902H +0.897WS – 0.856O_3 – 0.656PM_{10}$
	PC2	$0.866CO + 0.842SO_2 +0.507NO_2$

The scores of high loadings components with an eigenvalue greater than or equal to 1 were selected as an input to the modified models. The sub-models of each principal component according to the study areas is given in Table 7. The moderate factor loading of the Varimax factors (VFs) and coefficient (≥0.50) are considered as the components of each principal component (PCs).

Prediction Model

Table 8 presents the MLR and PCR prediction models for PM_{10} level in Petaling Jaya and Shah Alam for the next-day, the next-two-days and the next-three-days. From the MLR equations obtained, PM_{10} consists of positive contribution from CO, NO_2, and O_3 for all the predicted days in both study areas. While there was negative contribution from all meteorological parameters with PM_{10} such as temperature (T), wind speed (WS) and humidity (H). PCR was developed in order to enhance the prediction of MLR model and improve the extrapolative ability of MLR by using PCs as the inputs for PM_{10} concentrations forecasting. For PCR model, significant high coefficient value of PC3 can be observed in Petaling Jaya which mainly consist of SO_2 and NO_2 concentration (Table 8). In Shah Alam, only 2 PCs were included as input for MLR.

Table 8. Predictive Models using MLR and PCR in Petaling Jaya and Shah Alam

Area	Forecasting Day	Method	Prediction Model
Petaling Jaya	Next day	MLR	$PM_{10\,(t+24)}$ = 9.908 + 3.443 CO - 0.082 H + 9.363 NO_2 + 50.807 O_3 + 0.975 PM_{10} + 50.408 SO_2 - 0.219 T - 0.403 WS
		PCR	$PM_{10\,(t+24)}$ = -0.182 + 0.063 PC1 + 1.968 PC2 - 45.728 PC3
	Next two days	MLR	$PM_{10\,(t+48)}$ = 21.708 + 6.614 CO - 0.163 H + 16.99 NO_2 + 94.248 O_3 + 0.947 PM_{10} + 20.428 SO_2 – 0.436 T - 0.862 WS
		PCR	$PM_{10\,(t+48)}$ = 0.76 + 0.068 PC1 + 1.947 PC2 - 122.595 PC3
	Next three days	MLR	$PM_{10\,(t+72)}$ = 37.466 + 9.541 CO - 0.246 H + 27.919 NO_2 + 137.11 O_3 + 0.916 PM_{10} - 43.435 SO_2 – 0.778 T - 1.201 WS
		PCR	$PM_{10\,(t+72)}$ = 2.23 + 0.056 PC1 + 1.919 PC2 - 194.894 PC3
Shah Alam	Next day	MLR	$PM_{10\,(t+24)}$ = -0.948 + 4.628 CO - 0.048 H + 84.333 NO_2 + 54.739 O_3 + 0.963 PM_{10} + 115.577 SO_2 + 0.034 T - 0.313 WS
		PCR	$PM_{10\,(t+24)}$ = 0.993 + 0.089 PC1 + 2.496 PC2
	Next two days	MLR	$PM_{10\,(t+48)}$ = -0.954 + 8.553 CO-0.103 H + 144.744 NO_2 + 91.117 O_3 + 0.927 PM_{10} + 208.419 SO_2 + 0.134 T - 0.7 WS
		PCR	$PM_{10\,(t+48)}$ = 3.033 + 0.154 PC1 + 2.468 PC2
	Next three days	MLR	$PM_{10\,(t+72)}$ = -13.025 + 12.021 CO -0.097 H + 194.902 NO2 + 107.144 O_3 + 0.893 PM_{10} + 171.573 SO_2 + 0.563 T - 1.195 WS
		PCR	$PM_{10\,(t+72)}$ =5.079 + 0.181 PC1 + 2.434 PC2

Table 9 summarizes the best ANN model for predicting PM_{10} level in Petaling Jaya and Shah Alam. Two parameters were varied i.e. the transfer function (log-sigmoid, tangent-sigmoid and purelin functions) and the number of hidden notes (from 1 to 20). The selection of the best ANN models was done using performance measures as depicted in Table 4.

For transfer function, transig-purelin was the best for both of the areas. Mean while, for number of hidden notes, higher numbers gave better prediction of PM_{10} level.

Table 9. Summary of the best ANN model in Petaling Jaya and Shah Alam

Area	Forecasting Day	Transfer Function	Number of hidden notes
Petaling Jaya	Next day		6
	Next two days	Tansig-Purelin	20
	Next three days		14
Shah Alam	Next day		17
	Next two days	Tansig-Purelin	17
	Next three days		15

Performance Assesment of the Prediction Model

Table 10 shows the calculated value of performance assessments of the three prediction models for the next-day, the next-two-day and the next-three-day.

Table 10. Predictive Models using MLR and PCR in Petaling Jaya and Shah Alam

Area	Forecasting Day	Method	Performance Measures				
			NAE	RMSE	IA	PA	R^2
Petaling Jaya	Next day	MLR	0.0257	6.4615	0.9980	0.9956	0.9866
		PCR	0.0268	6.6047	0.9979	0.9958	0.9863
		ANN	**0.0242**	**6.0905**	**0.9982**	**0.9964**	**0.9876**
	Next two days	MLR	0.0484	11.1376	0.9939	0.9879	0.9707
		PCR	0.0509	11.6173	0.9933	0.9868	0.9686
		ANN	**0.0444**	**10.2152**	**0.9948**	**0.9899**	**0.9747**
	Next three days	MLR	0.0700	15.6931	0.9877	0.9758	0.9471
		PCR	0.0737	16.5094	0.9862	0.9732	0.9420
		ANN	**0.0634**	**13.7365**	**0.9905**	**0.9817**	**0.9586**
Shah Alam	Next day	MLR	0.0238	6.0483	0.9980	0.9960	0.9878
		PCR	0.0239	6.3419	0.9978	0.9956	0.9870
		ANN	**0.0230**	**5.7976**	**0.9982**	**0.9963**	**0.9884**
	Next two days	MLR	0.0427	9.9477	0.9945	0.9890	0.9740
		PCR	0.0451	10.7766	0.9935	0.9871	0.9703
		ANN	**0.0392**	**9.1886**	**0.9953**	**0.9906**	**0.9772**
	Next three days	MLR	0.0592	13.8466	0.9891	0.9787	0.9537
		PCR	0.0644	15.2131	0.9867	0.9743	0.9452
		ANN	**0.0573**	**13.1267**	**0.9902**	**0.9808**	**0.9579**

ANN had rather performed better and more efficiency to predict better future PM_{10} concentration during high particulate event especially in Petaling Jaya and Shah Alam. This can be observed as the prediction accuracy (IA, PA and R^2) were the highest and the least calculated error (NAE and RMSE) found in ANN models for both study areas. Subsequently, the prediction for one day ahead indicated a decrement trend in performances in each location. Hence, next day prediction was the best day to predict the PM_{10} specifically during high particulate event in

Petaling Jaya and Shah Alam. The model performances showed a good agreement between the predicted and observed PM_{10} concentration for one day ahead in Petaling Jaya and Shah Alam.

ANN also performed better in forecasting PM_{10} concentration for the next-two-day at both study areas. In instance, the result from performance indicators indicated that ANN had the least calculated error and the highest value of accuracy measure making ANN the best method for the next-two-day prediction. The result obtained also proved that ANN are able in estimating PM_{10} concentration for next-three-day as the result of the best calculated value of performance measure than other models in both of the study areas.

Most of previous study, obtained that PCR model was satisfactory compared to MLR. However, once comparing MLR with PCR, the MLR model performed better than PCR for both study areas. This result was corresponded with *Ismail et al.* [22] that this might because the dataset used for this analysis were not sufficient because only the dataset of high concentration of PM_{10} during high particulate event (HPE) were used in this study. In addition, the performance of PCR was not good as compared to MLR because the frequency distribution of PM_{10} in the dataset may intensely influence the modelling outcomes, and the application of PM_{10} data with uniform distribution may lead to an appropriate model for the forecasting of extreme events. However, utilization of this training database reduces the accuracy of low and normal PM_{10} concentration forecasts. Accordingly, combining two PM_{10} forecasting models, which had been developed using training datasets with different frequency distributions, may lead to a suitable model for predicting with low to high PM_{10} concentrations [29, 30]. MLR model could performed better than PCR because the selection of high PM_{10} concentration where only the peak concentration of PM_{10} during hourly was selected. It means that during high particulate event, simple model can be used to predict the high concentration of PM_{10} for one day ahead as it is easier compared to its hybrid model that is more complicated.

Figure 2 shows the scatter plot of the predicted PM_{10} concentration and the observed PM_{10} concentration using ANN method in Petaling Jaya and Shah Alam.

The correlation coefficient was significantly high, where the R^2 value were 0.9929, 0.98 and 0.9638 for the next day, next two days and next three days prediction in Petaling Jaya. Moreover, Shah Alam also recorded the successful model developed, as the value tend to be closer to 1 for the next day (R^2 = 0.9929), next two days (R^2 = 0.9814) and the next three days prediction (R^2 = 0.962). Hence, it can be summarized that the ANN model is the most satisfactory model to predict the PM10 concentration specifically during the high particulate event (HPE) in Klang Valley. *Wong et al.* [19] has developed models such as MLR, PCR, ANN and PCA-ANN showing the predictive ability for 24-hours PM_{10} concentration in urban-industrial areas in Shah Alam, Kuala Terengganu and Melaka by using 5 years interval dataset (2008-2012) consisting of PM_{10}, SO_2, NO_X, O_3, wind speed, humidity and temperature. The author summarized that the obtained results were rather satisfactory with ANN models, with values of R^2 for the test sets ranging between 0.32 to 0.62 for three sites and the values of IA ranging from 0.69 to 0.87. Additionally, the author stated that the performance of the examined ANN model was superior in comparison with the other models developed in parallel. However, it must be highlighted that the author analysed the model not during the high particulate event (HPE).

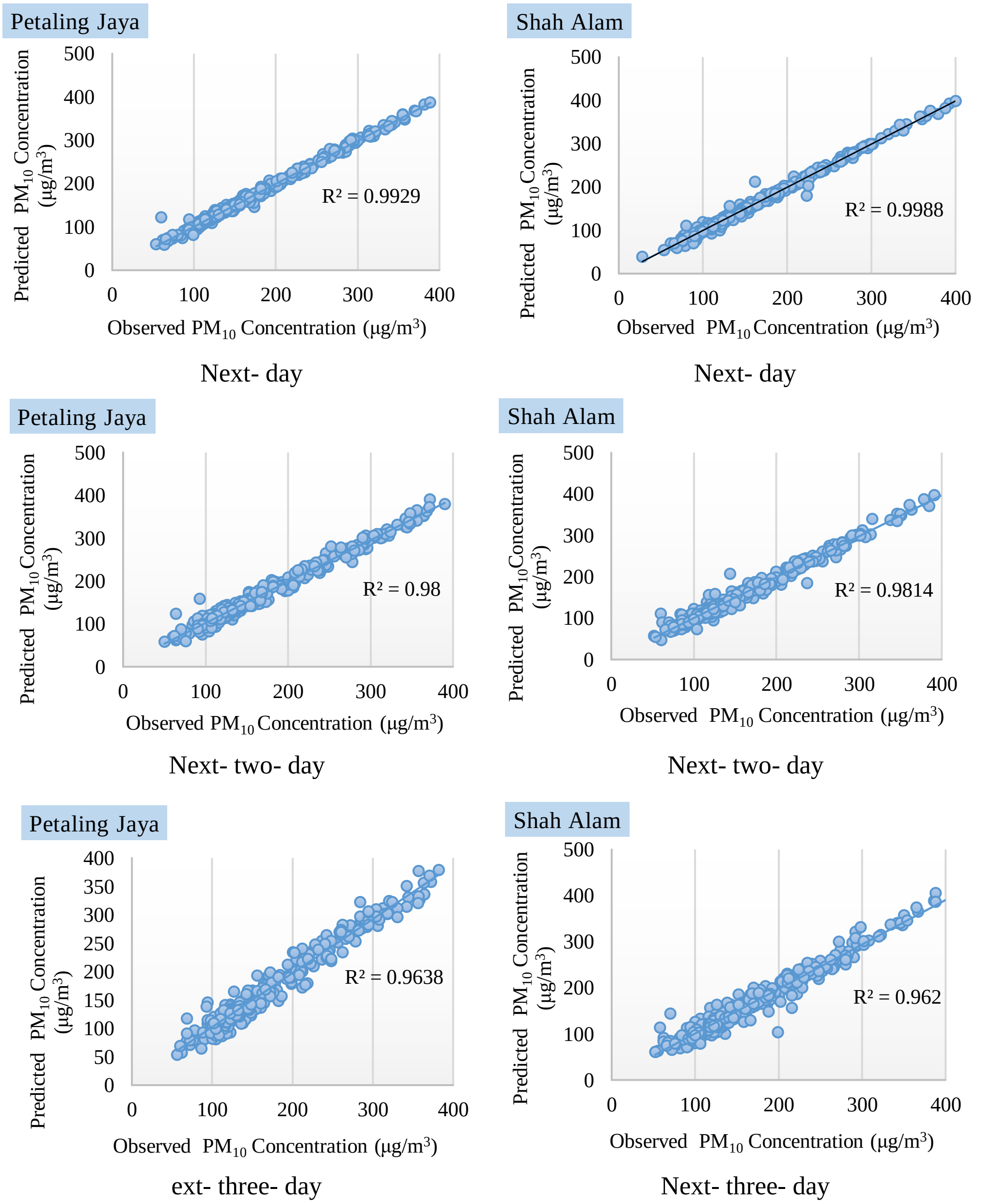

Fig. 2. Scatter plot of the predicted against observed PM_{10} concentration at Petaling Jaya and Shah Alam

Conclusions

PM_{10} is among of the main air pollutants in Malaysia and seems to be crucial to the Malaysian Air Pollution Index (MAPI) calculation especially during High Particulate Event (HPE). A dataset of PM_{10} concentration in Petaling Jaya and Shah Alam for five years (2012-2016) was thus obtained from Department of Environment, Malaysia. A total of eight parameters, consists of trace gases, i.e. particulate matter (PM_{10}), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), ozone (O_3) and carbon monoxide (CO) and weather parameters i.e. wind speed, ambient temperature and humidity had been used in this research. This study focuses on modeling PM_{10} level during High Particulate Event (HPE). Three models were employed to predict next-day ($PM_{10\ (t+24)}$), next-two-day ($PM_{10\ (t+48)}$), and next-three-day ($PM_{10\ (t+48)}$) of PM_{10} air pollutant concentrations and the best model was selected. Based on the performance comparison, ANN models showed highest R^2, IA and PA which the values nearest to 1 while it showed least error that closest to zero for NAE and RMSE. ANN model successfully trained by different transfer function. Hybrid model of PCR did not perform better as claimed by previous research. PCR could not perform better due to insufficient dataset as only the data of high concentration (>150µg/m^3) of PM_{10} concentration were used in this study. Besides, the performance of PCR was not good as compared to MLR because the frequency distribution of PM_{10} in the training dataset may intensely influence the modelling outcomes. It means that during high particulate event, simple model can be used to predict the high concentration of PM_{10} for one day ahead as it is much easier compared to its hybrid model which is more complicated. Thus, it is concluded that ANN models were most suitable for estimating the PM_{10} concentration during high particulate event in Klang Valley. The models can be effectively developed for the protection of public health by providing the respective population with way earlier warnings.

Acknowledgments

The authors would like to thank Ministry of Higher Education, Malaysia for Fundamental Research Grant Scheme (FRGS 9003-00837) and Department of Environment, Malaysia for providing the air pollutants and meteorology dataset.

References

[1] A.C. Pinheiro, C. Viegas, S. Viegas, C. Veríssimo, J. Brandão, M.F. Macedo, *Particulate Matter Distribution in Selected Portuguese Archives: A Preliminary Study,* **International Journal of Conservation Science, 5**(2), 2014, pp. 139-150.

[2] A.C. Styś, I. Sówka, D. Kobus, Ł. Pachurka, *Analysis of concentrations trends and origins of PM10 in selected European cities*, **E3S Web of Conferences**, **17**, 2017, Article Number: 00013.

[3] J. Chen, C. Li, Z. Ristovski, A. Milic, Y. Gu, M.S. Islam, S. Wang, J. Hao, H. Zhang, C.H. Guo, H. Fu, B. Miljevic, L. Morawska, P. Thai, Y.F. Lam, G. Pereira, A. Ding, X. Huang. U.C. Dumka, *A review of biomass burning: Emissions and impacts on air quality, health and climate in China*, **Science of the Total Environment, 579**, 2017, pp. 1000–1034.

[4] S.D.A. Kusumaningtyas, E. Aldrian, *Impact of the {June} 2013 {Riau} province {Sumatera} smoke haze event on regional air pollution,* **Environmental Research Letters, 11**(7)**,** 2016, 75007.

[5] W.N. Shaziayani, F.D. Harun, A.Z. Ul-Saufie, N. Samsudin, N.M. Noor, *Three-Days Ahead Prediction of Daily Maximum Concentrations of PM_{10} using Decision Tree Approach*, **International Journal of Conservation Science, 12**(1), 2021, pp. 217-224.

[6] M.I. Mead, S. Castruccio, M.T. Latif, M.S.M. Nadzir, D. Dominick, A. Thota, P. Crippa, *Impact of the 2015 wildfires on Malaysian air quality and exposure: A comparative study of observed and modeled data*, **Environmental Research Letters, 13**(4), 2018, 044023**.**

[7] A. Timofte, L. Belegante, M.M. Cazacu, B. Albina, C. Talianu, S. Gurlui, *Study of planetary boundary layer height from LIDAR measurements and ALARO model*, **Journal of Optoelectronics and Advanced Materials, 17**(7-8), 2015, pp. 911-917.

[8] A. Banica, E.D. Bobric, M.M. Cazacu, A. Timofte, S. Gurlui, I.G. Breaban, *Integrated Assessment of Exposure to Traffic-Related Air Pollution in Iasi City, Romania*, **Environmental Engineering and Management Journal, 16**(9), 2017, pp. 2147-2163.

[9] M.T. Latif, M. Othman, N. Idris, L. Juneng, A. Makmom Abdullah, W.P. Hamzah, M.F. Khan, N.M. Nik Sulaiman, J. Jewaratnam, N. Aghamohammadi, M. Sahani, C.J. Xiang, F. Ahamad, N. Amil, M. Darus, H. Varkkey, F. Tangang, A.B. Jaafar, *Impact of regional haze towards air quality in Malaysia: A review*, **Atmospheric Environment, 177,** 2018, pp. 28–44.

[10] M.F. Khan, Y. Shirasuna, K. Hirano, S. Masunaga, *Urban and suburban aerosol in Yokohama, Japan: a comprehensive chemical characterization*, **Environmental Monitoring and Assessment, 171**(1-4), 2010, pp. 441-456.

[11] N.A.C. Samsuddin, M.F. Khan, K.N.A. Maulud, A.H. Hamid, F.T. Munna, M.A. Ab Rahim, M. Akhtaruzzaman, *Local and transboundary factors' impacts on trace gases and aerosol during haze episode in 2015 El Niño in Malaysia*, **Science of the Total Environment, 630,** 2018, pp. 1502-1514.

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

[13] A. Nazif, N.I. Mohammed, A. Malakahmad, M.S. Abualqumboz, *Regression and multivariate models for predicting particulate matter concentration level*, **Environmental Science and Pollution Research, 25**(10), 2017, pp. 283-289.

[14] K.Y. Ng, N. Awang, *Pyrolysis - GC/MS Data Book of Synthetic Polyme Multiple linear regression and regression with time series error models in forecasting PM10 concentrations in Peninsular Malaysia*, **Environmental Monitoring and Assessment, 190,** 2018, Article Number: 63.

[15] N. M. Hashim, N. M. Noor, A. Z. Ul-Saufie, A.V. Sandu, P. Vizureanu, G. Deák, M. Kheimi, *Forecasting Daytime Ground-level Ozone Concentration in Urbanized Areas of Malaysia Using Predictive Model*, **Sustainability, 14(13)**, 2022, art. 7936, https://doi.org/10.3390/su14137936.

[16] F. Taşpınar, Z. Bozkurt, *Application of artificial neural networks and regression models in the prediction of daily maximum PM10 concentration in Düzce, Turkey*, **Fresenius Environmental Bulletin, 23**(10), pp. *2450*-2459.

[17] G. Deak, N. Raischi, M. Matei, M. Boboc, G. Cornateanu, M. Raischi, S. Matei, S.Y. Yusuf, *Meteorological Parameters and Air Pollution in Urban Environments in the Context of Sustainable Development*, **IOP Conference Series-Earth and Environmental Science, 616,** 2020**,** Article Number: 012003, DOI10.1088/1755-1315/616/1/012003.

[18] K. Bodor, R. Boga, T. Pernyeszi, S. Tonk, G. Deak, *Variation of PM10 Concentration Depending on the Meteorological Parameters in Two Bucharest Monitoring Stations (in Green Areas)*, **Present Environment and Sustainable Development, 14**(1), 2020, pp. 277-292, DOI: 10.15551/pesd2020141022.

[19] Y.K. Wong, N.M. Noor, N.I.M. Hashim, *Temporal Variation of Ambient PM10 Concentration within an Urban-Industrial Environment*, **E3S Web of Conferences, 34,** 2018, Article Number: 02002.

[20] Z. Ceylan, S. Bulkan, *Forecasting PM10 levels using ANN and MLR: A case study for Sakarya City*, **Global Nest Journal, 20**(2), 2018, pp. 281-290.

[21] M. Ismail, S. Abdullah, A.D. Jaafar, T.A.E. Ibrahim, M.S.M. Shukor, *Statistical modeling approaches for PM10 forecasting at industrial areas of Malaysia*, **AIP Conference Proceedings, 2020**(1), 2018, 020044.

[22] M.P. Lehadus, V. Nedeff, N. Barsan, A.V. Sandu, E. Mosnegutu, C. Tomozei, O. Irimia, G. Andrioai, I. Sandu, *Monitoring the Particulate Matter (PM10) Emissions from Bacau City Termo-Energetic Industry,* **Revista de Chimie, 70**(8), 2019, pp. 2869-2872.

[23] A. Asmat, S. Norhayati, M. Tarmizi, N.H. Zakaria, *Seasonal Particulate Matter (PM 10) Concentration in Klang Valley, Malaysia*, **International Journal of Engineering & Technology, 7**(3), 2018, pp. 162–167.

[24] M. Aljanabi, M. Shkoukani, M. Hijjawi, *Ground-level Ozone Prediction Using Machine Learning Techniques: A Case Groundlevel Ozone Prediction Using Machine Learning Techniques: A Case Study in Amman, Jordan*, **International Journal of Automation and Computing, 17,** 2020, pp. 667–677.

[25] A.Z. Ul-Saufie, N.H. Hamzan, Z. Zahari, W.N. Shaziayani, N.M. Noor, M.R.R.M.A. Zainol; A.V. Sandu, G. Deak, P. Vizureanu, *Improving Air Pollution Prediction Modelling Using Wrapper Feature Selection*, **Sustainability, 14**(18), 2022, Art. 11403, DOI: https://doi.org/10.3390/su141811403.

[26] D. Capsa, N. Barsan, D. Felegeanu, M. Stanila, I. Joita, M. Rotaru, C. Ureche, *Influence of Climatic Factors on the Pollution with Nitrogen Oxides (Nox) in Bacau City, Romania*, **Environmental Engineering and Management Journal, 15**(3), 2016, pp. 655-663.

[27] A.Z. Ul–Saufie, A.S. Yahaya, N.A. Ramli, N. Rosaida, H.A. Hamid, *Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA),* **Atmospheric Environment, 77,** 2013, pp. 621–630.

[28] H.F. Kaiser, *An index of factorial simplicity*, **Psychometrika, 39,** 1974, pp. 31–36.

[29] A.Z. Ul-Saufie, A.S. Yahaya, N.A. Ramli, H.A. Hamid, *Performance of Multiple Linear Regression Model for Longterm PM10 Concentration Prediction based on Gasesous and Meteorological Parameters*. **Journal of Applied Sciences, 12,** 2012, pp. 1488–1494.

[30] H.T. Shahraiyni, S. Sodoudi, *Statistical Modeling Approaches for PM10 Prediction in Urban Areas; A Review of 21st-Century Studies,* **Atmosphere, 7**(2), 2016, 7020015.

Received: January 02, 2022
Accepted: August 24, 2022