The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models over particulate matter (PM$_{10}$) variability during haze and non-haze episodes: A decade case study

Ku Mohd Kalkausar Ku Yusof $^a$, Azman Azid $^{a,*}$, Muhamad Shirwan Abdullah Sani $^b$, Mohd Saiful Samsudin $^c$, Siti Noor Syuhada Muhammad Amin $^a$, Nurul Latiffah Abd Rani $^a$, Mohd Asrul Ku Mohd Kalkausar Ku Yusof $^a$, Azman Azid $^{a,*}$, Muhamad Shirwan Abdullah Sani $^b$, Mohd Saiful Samsudin $^c$, Siti Noor Syuhada Muhammad Amin $^a$, Nurul Latiffah Abd Rani $^a$, Mohd Asrul

$^a$ Faculty Bioresources and Food Industry, Universiti Sultan Zainal Abidin (UniSZA), Besut Campus, 22200 Besut, Terengganu, Malaysia
$^b$ Dr. F.A.S. Technologies, Blok D1, Tingkat 2, UniSZA Digital Hub, UniSZA Besut Campus, 22200 Besut, Terengganu, Malaysia
$^c$ Faculty of Environmental Studies, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia
$^d$ Corresponding author: azmanazid@unisza.edu.my

**Abstract**

The comprehensives of particulate matter studies are needed in predicting future haze occurrences in Malaysia. This paper presents the application of Artificial Neural Networks (ANN) and Multiple Linear Regressions (MLR) coupled with sensitivity analysis (SA) in order to recognize the pollutant relationship status over particulate matter (PM$_{10}$) in eastern region. Eight monitoring studies were used, involving 14 input parameters as independent variables including meteorological factors. In order to investigate the efficiency of ANN and MLR performance, two different weather circumstances were selected; haze and non-haze. The performance evaluation was characterized into two steps. Firstly, two models were developed based on ANN and MLR which denoted as full model, with all parameters (14 variables) were used as the input. SA was used as additional feature to rank the most contributed parameter to PM$_{10}$ variations in both situations. Next, the model development was evaluated based on selected model, where only significant variables were selected as input. Three mathematical indices were introduced ($R^2$, RMSE and SSE) to compare on both techniques. From the findings, ANN performed better in full and selected model, with both models were completely showed a significant result during haze and non-hazy. On top of that, UV, carbon monoxide were both variables that mutually predicted by ANN and MLR during hazy and non-hazy days, respectively. The precise predictions were required in helping any related agency to emphasize on pollutant that essentially contributed to PM$_{10}$ variations, especially during haze period.

**Keywords:** Haze, sensitivity analysis, artificial neural network, multiple linear regressions

---

**INTRODUCTION**

Air pollution is not a brand new issue amongst worldwide nation including Malaysia. It was produced before, and it will be remained to be emitted for hundred years ahead if it still has demand from industries, vehicles, agricultures and etc. Three major sources of air pollution in Malaysia are highly contributed from industries, motor vehicles and power plant. In 2015, Department of Statistics Malaysia has released a compendium of environment, whereas the emission of pollutants to the atmosphere from power plant and motor vehicles have been increased for about 20.0% (619,200 to 742,900 tonnes) and 14.3% (1,829,700 to 2,092,000 tonnes), respectively since 2010. In total, there are 26.3 million vehicles that registered up to 2016 (DOS 2016). Johor, Selangor, Perak and Pulau Pinang are the highest vehicle possession among other states in Malaysia. Hence, the deterioration of atmospheric circumstances is due to mostly by exhaust emissions from motor vehicles (Afroz et al. 2003).

Recently, the air pollution series become a severe problem where the transboundary pollution such as haze episodes are worse in terms of the severity, period, and the affected areas. Thus, a serious attention is needed by all parties, not only by government sector, but also more to individual responsibility (Azid et al. 2015a).

A number of haze series has been happened before, but only in 1997/1998, it was considered as the worst event in South East (SE) Asia historic records. Ignited by uncontrolled immense peat fires in Sumatera and Kalimantan (Sunderlin & Resosudarmo 1996; Fearnside 1997; Kartawinata et al. 2001; Koe et al. 2001; Alencar et al. 2006), the black-thickened haze was covered and blanketed few countries especially Malaysia for several months (Shaadan et al. 2015). The prolonged issues have already been discussed previously in ASEAN level, but the effort was considered to be effortless. The mega fires was not the only matter to be discussed, however, the anomaly in weather conditions like El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) exacerbated the scenario as well (Ashok et al. 2001; Yoo et al. 2006; Luo et al. 2010; Ash & Matyas 2012; Nayagam et al. 2013). Contrarily, the haze episodes are not only transported from the outside, but it has been proven to be locally produced especially during the southwest (SW) monsoon (Deni et al. 2015).
were depicted in Fig. 1. The air quality status over three states in eastern Malaysia and bordering to Southern Thailand on the north. Lengthwise, the coast lies more than 700km from Kelantan to eastern Johor, with openly exposed to South China Sea. Eastern region comprises of three states namely Kelantan, Terengganu, and Pahang. The total area of the region is 66,736 km² or 51% of the Peninsular Malaysia. Pahang is the biggest state amongst three states in the eastern with 35,840 km², followed by Kelantan (15,099 km²) and Terengganu (13,035 km²). Demographically, Kelantan has the highest population distribution with 5.7% of total population in Peninsular Malaysia. Pahang has the biggest state amongst three states in the eastern with 35,840 km², followed by Kelantan (15,099 km²) and Terengganu (13,035 km²). Demographically, Kelantan has the highest population distribution with 5.7% of total population in Malaysia, that is equal to 1.81 million, Pahang with 5.1% (1.62 million), while Terengganu is the lowest with 3.7% (1.17 million) (DOS, Malaysia). The region remains to be the least urbanized at 41.3%, compared to other regions in Peninsular Malaysia. The region holds over 51% of forest areas in the Peninsula and a high proportion of environmentally sensitive areas including highlands, islands, and wetlands (Bluian et al. 2012).

The air quality status over three states in eastern is varied. As Pahang is a newly developed state that driven by various industries especially in Kuantan and Gebeng. Terengganu is synonym to oil and gas sector in Kerteh and Paka. Contradictorily, Kelantan is least developed state in comparison to Pahang and Terengganu, with its economic structure is mostly based on agricultural and fisheries sectors. Therefore, it is vital to recognize and identify the relationship between air pollutant and PM10 at the eastern region during haze and non-haze period. All the details for studied area in the eastern region were depicted in Fig. 1.

Air pollutant and meteorological data

14 air-pollutant parameters were selected and used as independent variables, x (including meteorological factor) namely nitrogen oxide (NO), nitrogen monoxide (NOx), methane (CH4), non-methane hydrocarbon (NMHC), total hydrocarbon (THC), sulphur dioxide (SO2), nitrogen dioxide (NO2), ozone (O3), carbon monoxide (CO), wind speed (WS), wind direction (WD), air temperature (AT), relative humidity (RH), ultraviolet-b (UVs). At the same time, particulate matter (PM10) was assigned as sole parameter to be denoted as dependent variable, y. Together, as to evaluate the model performance, two situations were created, called hazy and non-hazy days. The datasets from both situations were daily averaged from hourly value, with hazy day data was specifically set to be differed from non-hazy day data. For hazy day, PM10 with equal or more than 150μg/m³ (PM10≥150 μg/m³) and any associated pollutant (including meteorological factor) were thoroughly screened to fit the objective's criteria. As to distinguish between hazy and non-hazy, Recommended Malaysia Air Quality Guideline (RMAQG) was integrated (DOE 1997). According to RMAQG, the stipulated limit for PM10 is 150μg/m³ for 24 hours duration. During January, 2006 – December, 2015 periods, a total of 27,543 days were perfectly included in this study, where only 1,502 days were considered as hazy days at three states in the eastern region. Thus, the remaining days for non-hazy days were 26,041 days. The air quality data was obtained from Air Quality Unit, Department of Environment Malaysia, as part of their Continuous Air Quality Monitoring (CAQM) program. Alam Sekitar Malaysia (ASMA), an environmental privatised company is responsible to do the installation works, as well as to operate and maintain the DOE’s instrument at all 52 monitoring stations across Malaysia. In order to continuously monitor the PM10 levels, b-ray attenuation mass monitor (BAM-1020), manufactured by Met One Instruments Inc. was used (Juneng et al. 2011; Latli et al. 2014). The models used for 14 parameters were presented in Table 1.

Artificial Neural Network (ANN)

ANN is acted as mathematical analogue and mimicking human brain biological system (Azid et al. 2013; Azid et al. 2014; Nathan et al. 2017). The ANN capabilities are by perceiving the example of the intricate, fundamental and multi-dimensionality information with does not depend on any assumption before preparing the relationship among factors, including the non-parametric information (Amran et al. 2015). The ANN works as three basic types of layers; the input layer, hidden layer(s), and output layer (Munaf et al. 2013; Yusof et al. 2018; Rani et al. 2018). As mimicking the human brain system, the results from output layer are transmitted by hidden nodes from input layer (Bandyopadhyay & Chattopadhyay 2007). Fig. 2 shows the example of designing the ANN, which contains a series of equations for output calculations using the given input values (Haykin, 1999; Ozgoren et al. 2013; Yusof et al. 2013; Yusof et al. 2016). The hidden nodes can be in single layer or multilayers. Usually, the multiple layers neurons are called as Multiple Layer Perceptron (MLP).

Theoretically, the MLP is used to map the function between input and output, however the status of the relationship is unidentified. In MLP development, Jones et al. (1999) highlighted that development was based on back-propagation system. In this study, the feed forward back propagation was used as architecture structure. In this structure, the weights of each input, hidden and output were equal and back propagation process was not allowed (Caseli et al. 2009; Işıyaka & Azid 2015). Mathematically, ANN can be calculated using the following equations (Haykin, 1999; Tosun et al. 2016):

\[ \mathbf{u}_k = \sum_{j} w_{kj} \mathbf{x}_j \]
\[ y_k = \varphi(\mathbf{u}_k + b_k) \]  

(1)

Bias, denoted by \( b_k \), has the effect of increasing or lowering the net input of the activation function. \( x_1, x_2, \ldots, x_n \) are the inputs; \( w_{1k}, w_{2k}, \ldots, w_{nk} \) are the weights of the neuron \( k \); \( \mathbf{u}_k \) is the linear combiner output due to input signals; \( \varphi(.) \) is the activation function; \( y_k \) is the output signal of the neuron (Haykin, 1999; Tosun et al. 2016).
Fig. 1 The location for 12 study areas (N1 – N12) within eastern region.

Table 1 The CAQM model equipment for each parameter.

| Parameter                        | Model Equipment                  |
|----------------------------------|----------------------------------|
| Wind Speed (WS) (km/h)           | Met One 010C                     |
| Air Temperature (AT) (°C)        | Met One 062                      |
| Relative Humidity (RH) (%)       | Met One 083D                     |
| Nitrogen Oxide (NO\textsubscript{x}) (ppm) | Teledyne API Model 200A/200E     |
| Nitrogen Monoxide (NO) (ppm)     | Teledyne API Model 200A/200E     |
| Ultraviolet-b (UV\textsubscript{b}) (J/m\textsuperscript{2}/d\textsuperscript{-1}) | -                              |
| Methane (CH\textsubscript{4}) (ppm) | Teledyne API M4020               |
| Non-methane Hydrocarbon (NmHC) (ppm) | Teledyne API M4020               |
| Total Hydrocarbon (THC) (ppm)    | Teledyne API M4020               |
| Sulphur Dioxide (SO\textsubscript{2}) (ppm) | Teledyne API Model 100A/100E      |
| Nitrogen Dioxide (NO\textsubscript{2}) (ppm) | Teledyne API Model 200A/200E     |
| Ozone (O\textsubscript{3}) (ppm) | Teledyne API Model 400/400E      |
| Carbon Monoxide (CO) (ppm)       | Teledyne API Model 300/300E       |

Multiple linear regressions (MLR)

The MLR is a traditional methodology to examine the impact of dependent variable by identifying the relationship of each independent variables (Azid et al. 2015b; Azid et al. 2015c). MLR technique has been widely applied in environmental studies, especially in atmospheric pollution. The MLR model equation can be expressed as below:

\[ Y = a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n + \varepsilon \]  \hspace{1cm} (2)

where, \( Y \) represents dependent variables, \( a_i \) and \( X_i \) are the regression coefficient and independent variables, respectively (Shirsath & Singh 2010; Ördemir & Taner 2014). Whist, \( \varepsilon \) is regression conjectural error. In this study, MLR technique was applied to determine the air pollutant factor including meteorological on PM\textsubscript{10} behaviour for 10-consecutive years. JMP 10.0 (SAS Institute Inc.) was used to analyse MLR.

Model development and the analysis of the contribution of different predictor variables

In the model development, both statistical tools (ANN and MLR) have an input optimization by using “leave-one-out” technique offered by sensitivity analysis. However, in optimizing the process of the input, the best hidden nodes are required. The accurate selection of hidden nodes will denote by the highest R\textsuperscript{2}. The fit hidden nodes diagram can be seen in Fig 2. SA offers the information on the response of each network provided by pollutant involved (Le-Dimet et al. 2017). In this study, SA technique that required only a parameter would be purposely taken out from the input list in order to manually calculate the percentage contribution to the output (PM\textsubscript{10}). This
technique has already been proven and widely applied in atmospheric studies among researchers in Malaysia (Latif et al. 2014; Azid et al. 2016). The identifying process would be a repetitive backward elimination as suggested by Olden & Jackson (2002) and Olden et al. (2004) in order to recognize the most influencing or affecting factor to PM10 variations. According to Elangasinghe et al. (2014), in order to achieve the variation of each network, each input of parameters was locked at its mean value and plotted to the model response to distress the signal for each predictor variable. In other words, different models were constructed by removing different inputs while explaining the degree of importance of each variable into the variability of PM10. In the final phase, the result of each contribution (in % value) was compared and the key contributor to PM10 variation was clearly identified. The identification process would not only be executed in ANN, but in MLR as well. Thus, the coefficient of determination ($R^2$) was used to appraise the model performance. Two main purposes were set to compare and identify the mutual input prediction in each model development for specific scenario given. The goal in integration of ANN in the model development was to identify the most substantial model that would affect the particulate matter during haze and non-haze period. Two different types of models were developed and briefly described as below:

Full model : This model was based on all 14 parameters and recognized as ANN-HM-AP & MLR-HM-AP.

Selected Model : This model was developed by selection parameter based on the highest contribution to PM10 during haze and non-haze period (ANN-HM-LO & MLR-HM-LO).

Model performance evaluation

ANN and MLR performances were evaluated using three statistical indices during training and validation process: the coefficient of determination ($R^2$), Root Mean Square Error (RMSE) and sum of squared errors of prediction (SSE). The formulas are expressed as below:

$$ R^2 = 1 - \frac{\sum (x_i - y_i)^2}{\sum y_i^2} $$

$$ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2} $$

$$ SSE = \sum_{i=1}^{n} (x_i - y_i)^2 $$

where, in equation (3) – (5):

- $x_i$ = the observed data,
- $y_i$ = the predicted data, and
- $n$ = the observation number

The acceptable range of $R^2$ is in between 0.00 – 1.00. The ideal value for model development is $R^2$ value near or close to 1.00, whilst if the $R^2$ approaches 0.00, the model is considered weak and it is not suitable to be used for further analysis (Challoner et al. 2015).

RESULTS AND DISCUSSION

Sensitivity analysis: the comparison evaluation of ANN and MLR performance during hazy and non-hazy day

Sensitivity analysis coupled with ANN and MLR was used in this study to identify the most influential contributors amongst input variables (Azid et al. 2016) towards PM10 variability. Two different occasions (hazy and non-hazy) with 14 parameters were used as input variables, while PM10 was acted as output variable. Table 2 shows the $R^2$ results of full model (comprised of 14 parameters) for both techniques (ANN and MLR) in both occasions (hazy and non-hazy) with the value of 0.9999157, 0.650610 and 0.724085, 0.419122, respectively. The RMSE and SSE for ANN and MLR were 0.256023, 6.9445232 and 21.25249, 10.06634, respectively. The RMSE and SSE value was straightforwardly generated from the calculation of $R^2$. Therefore, any model developed that was significantly exhibited the $R^2$ near or close to 1.0, indicating that the erroneous produced by the model was closed to zero. From the result, ANN was overwhelmed MLR with highest $R^2$ value as well as lowest RMSE and SSE values recorded in both situations (hazy nor non-hazy day).

It means that, the relationship between PM10 and other atmospheric pollutants was absolutely strong, whilst only 76.42% of air pollutant contributed towards PM10 variations during non-hazy days. Even though the relationship between other pollutants with PM10 was strong during hazy days, yet it was still hardly to elaborate the details of the highest contribution amongst all 14 parameters involved. Thus, the next phase was to identify the most pollutant involved during haze and non-haze by engaging sensitivity analysis in this study. By applying sensitivity analysis, it could rank the pollutant by using percentage of contribution as the calculation basis.

The sensitivity analysis calculation is based from formula as follows;

$$ \text{contribution} = \frac{b_i-a_i}{z_i} \times 100 $$

where;

- $a_i$ = the $R^2$ value for each input parameter,
- $b_i$ = the $R^2$ value for total parameter (ANN-HM-AP & MLR-HM-AP)
- $z_i$ = the sum of $R^2$ differences

Table 3 shows the sensitivity analysis calculation. In this study, only the contribution more than 10% from each pollutant would be accounted as the best input to redevelop the new model (Nasir et al. 2011). Thus, any pollutant that contributed lesser than 10% was considered as least affected or influenced by PM10 presence during haze and non-haze circumstances. The new ANN and MLR models were totally relied on the best input (pollutant with contribution more than 10%).

Wind speed, wind direction, relative humidity and ultraviolet-b were accounted as the best inputs for ANN during haze, whilst MLR was predicted by temperature, ultraviolet-b, sulphur dioxide and carbon monoxide as the inputs. On the other hand, during non-haze, ANN was predicted by wind direction, temperature, relative humidity and carbon monoxide, at the same time MLR has carbon monoxide as the sole parameter to be included in the new model development. The overall performances for $R^2$, RMSE and SSE for selected model for both techniques (ANN and MLR) were presented in the Table 4.
Table 2  The overall performance of $R^2$, RMSE and SSE for full model on ANN and MLR models during, hazy and non-hazy days in the eastern region.

| Model       | Artificial Neural Network (ANN) | Multiple Linear Regression (MLR) |
|-------------|---------------------------------|---------------------------------|
|             | Haze                           | Non-haze                        | Haze                           | Non-haze                        |
| ANN-HM-AP   | $R^2$ 0.99                     | 0.76                            | $R^2$ 0.65                      | 0.41                            |
| &           | RMSE 0.25                      | 6.94                            | RMSE 21.25                      | 10.06                           |
| MLR-HM-AP   | SSE 1.24                       | 142943.06                       | SSE 10388.38                    | 336318.20                       |

Table 3  The sensitivity analysis calculation during (a) hazy days and (b) non-hazy days based on ANN and MLR prediction.

(a)  

| Model               | $R^2$ | Difference $R^2$ | Contribution (%) | Model               | $R^2$ | Difference $R^2$ | Contribution (%) |
|---------------------|-------|------------------|------------------|---------------------|-------|------------------|------------------|
| ANN-HM-AP           | 1.00  |                  |                  | MLR-HM-AP           | 0.65  |                  |                  |
| ANN-HM-WS (km/h)    | 0.94  | 0.06             | **44.77**        | MLR-HM-SO$_2$ (ppm) | 0.27  | 0.38             | **35.31**        |
| ANN-HM-WD           | 0.98  | 0.02             | **15.43**        | MLR-HM-UV$_b$ (J/m$^2$/d$^1$) | 0.42  | 0.23             | **21.47**        |
| ANN-HM-UV$_b$ (J/m$^2$/d$^1$) | 0.98  | 0.02             | **13.37**        | MLR-HM-AT ($^\circ$C) | 0.47  | 0.18             | **16.37**        |
| ANN-HM-RH (%)       | 0.98  | 0.02             | **12.66**        | MLR-HM-THC (ppm)    | 0.52  | 0.13             | **12.24**        |
| ANN-HM-AT ($^\circ$C) | 1.00  | 0.00             | 3.48             | MLR-HM-RH (%)       | 0.59  | 0.06             | 5.69             |
| ANN-HM-NO$_2$ (ppm) | 1.00  | 0.00             | 3.30             | MLR-HM-O$_1$ (ppm)  | 0.62  | 0.03             | 2.75             |
| ANN-HM-CH$_4$ (ppm) | 1.00  | 0.00             | 3.25             | MLR-HM-WS (km/h)    | 0.63  | 0.02             | 1.93             |
| ANN-HM-THC (ppm)    | 1.00  | 0.00             | 1.80             | MLR-HM-CH$_4$ (ppm) | 0.63  | 0.02             | 1.64             |
| ANN-HM-NO (ppm)     | 1.00  | 0.00             | 1.61             | MLR-HM-NmHC (ppm)   | 0.64  | 0.01             | 1.00             |
| ANN-HM-CO (ppm)     | 1.00  | 0.00             | 0.38             | MLR-HM-THC (ppm)    | 0.64  | 0.01             | 0.94             |
| ANN-HM-O$_1$ (ppm)  | 1.00  | 0.00             | 0.02             | MLR-HM-NO (ppm)     | 0.65  | 0.00             | 0.23             |
| ANN-HM-SO$_2$ (ppm) | 1.00  | 0.00             | 0.01             | MLR-HM-NO$_X$ (ppm) | 0.65  | 0.00             | 0.22             |
| ANN-HM-NmHC (ppm)   | 1.00  | 0.00             | -0.01            | MLR-HM-NO$_2$ (ppm) | 0.65  | 0.00             | 0.18             |
| ANN-HM-NO$_X$ (ppm) | 1.00  | 0.00             | -0.08            | MLR-HM-WD           | 0.65  | 0.00             | 0.04             |
| Total              |       | 0.13             | 100.00           | Total               | 1.08  | 100.00           |

Values in bold are selected with a significance of greater than 10% contribution (>10%).
Values in bold were selected with a significance of greater than 10% contribution (>10%)
Based on Table 4, the $R^2$ during haze and non-haze for ANN and MLR were 0.7889337, 0.4073319 and 0.468328, 0.24339, respectively. The RMSE and SSE values for both approaches during haze and non-haze were 13.8025297, 12.776663, 323.92807, 2202802 and 35.97362, 14.4979, 102584.18, 3689871.60, respectively. From the result, ANN showed a great significance on both situations, similar to what had been predicted and calculated by having all 14 parameters as the input. However, it could be seen that the overall performance in the selected model was quite lower than full model in terms of the $R^2$, RMSE and SSE results. The huge difference between those two models was the selection of the input used. As full model was fully utilized 14 parameters as input, the selected model was determined by up to four input variables in ANN and MLR during haze and non-hazy days. Thus, selected model could be utilized to portray the given scenarios. In this case, this study has proved that the relationship between PM$_{10}$ with other pollutants was expressively robust during haze, with both developed models (full and selected models) for ANN and MLR were capable to predict higher $R^2$ result with lower RMSE and SSE values than non-hazy days.

In term of best input predicted by both techniques, ANN predicted more pollutant involvement than MLR. In other words, ANN explained that more pollutants were actively reacted to PM$_{10}$ variations during hazy days rather than during non-hazy days.

**ANN vs. MLR: The establishment of relationship between PM$_{10}$ and atmospheric pollutant during hazy and non-hazy days**

Based on this study, ANN was a better model than MLR in term of the overall performance. However, to portray the interpretation by both techniques, the best input predicted by selected models in both situations was finally assayed. This study showed that the inconsistency prediction between ANN and MLR was due to one reason, where ANN has better capability in interpreting the environmental non-linearity dataset, while MLR was merely based on the direct relationship between dependent and independent variables. Therefore, it was unsurprisingly fact that ANN could predict more pollutants than MLR.

During haze, both ANN and MLR have four pollutants in prediction line, but with dissimilar pollutant type. Only in non-haze, the scenario was completely deviated than hazy days. However, this study showed that during non-haze, ANN could still predict four pollutants. MLR could predict carbon monoxide as not the only pollutant that affected by PM$_{10}$ fluctuation, also with huge percentage contribution amongst other pollutants as well. Even there were some alternations between these two approaches, ultraviolet-b (UV-b) and carbon monoxide (CO) were two pollutants that perfectly predicted by ANN and MLR during hazy and non-hazy days, respectively. Eastern region is the only region that mostly affected by northeast monsoon (NEM), where NEM is usually happened during December to March each year. Unlike Peninsula’s west coast lies along Straits of Malacca, eastern is exposed by an open sea; South China Sea. The wind speed over eastern region seems windier than at the west coast side. Hence, wind speed as the highest contributor to PM$_{10}$ variations during haze was as expected. With 44.77%, wind speed was predicted as the foremost contributor by ANN during hazy day, followed by wind direction (15.43%), ultraviolet-b (13.37%) and relative humidity (12.66%).

Specifically, ANN predicted meteorological factors as the foremost contributor. Aside ANN, MLR also came with four possibilities that could cause deterioration during haze with sulphur dioxide as the highest contributor with 35.31%, ultraviolet-b (21.48%), ambient temperature (16.37%) and lastly carbon monoxide (12.24%). MLR predictions were varied than ANN, where it predicted two out of four pollutants that were listed as Air Pollution Index (API), at the same time another two pollutants were meteorological parameters.

In contradiction to hazy days, the prediction interpretation by ANN was quite consistent in non-hazy, with the meteorological parameters were primarily led the contribution, however ANN added some additional features. In this case, ANN added carbon monoxide as another factor. For MLR, none of atmospheric pollutants was strongly related to PM$_{10}$ except carbon monoxide. Carbon monoxide obviously has strong relationship to PM$_{10}$ which in accordance to MLR prediction with 66.32%. Overall, even though there was few pollutants that were proven as highly associated to PM$_{10}$ during haze and non-haze, other pollutants that have not been described in this subchapter still have their contribution, but with small percentage. Therefore, even with minor difference predictions, both strategies have proved that meteorological factors were the significant response to haze occurrence (Tangang et al. 2010; Payus et al. 2013; Ramsey et al. 2014).

**CONCLUSION**

The objective of this study was to investigate the capabilities of ANN and MLR techniques onto PM$_{10}$ predictions during haze and non-haze periods. With an improvement of hidden node selection used, it indirectly generated a better performance in ANN prediction. Various statistical performance indices ($R^2$, RMSE and SSE) were used to evaluate the performance on each model. From the findings, ANN performed better in both models development; full model and selected model, compared to MLR either in $R^2$ evaluation or in determining the RMSE and SSE values. During hazy and non-hazy days, ANN has successfully identified that meteorological factors were the main contributors towards PM$_{10}$ variability, whilst MLR was mostly focused on API pollutant as additional contributor. In scientific perspective, ANN was practical than MLR. ANN was not only being precisely distinguished the pollutant contribution during haze and non-haze, it also described a real situation behind the model developed. Practically, ANN concisely predicted meteorological parameters were dominating on both hazy and non-hazy circumstances, however MLR was quite inconsistent. Thus, with a better performance that showed by ANN in all situations in this study, ANN proved that it could fully utilize in environmental studies especially in haze circumstances.

**ACKNOWLEDGEMENT**

This manuscript has been funded through University Research Fund (Grant no.: UniSZA/2016/DPU/01) of Universiti Sultan Zainal Abidin and Research Initiative Grant (RIGS16-363-0527) of International Islamic University Malaysia, Gombak, Selangor, Malaysia.

**REFERENCES**

Afroz, R., Hassan, M. N. & Ibrahim, N. A. 2003. Review of air pollution and health impacts in Malaysia. Environmental Research, 92(2), 71–77.
Reason, C. J. C., Behera, S. K. & Yasunari, T. 2008. On the roles of the northeast cold surge, the Borneo vortex, the Madden-Julian Oscillation, and the Indian Ocean Dipole during the extreme 2006/2007 flood in Southern Peninsular Malaysia. *Geophysical Research Letters*. 35(14), 1–6.

Tangang, F., Latif, M. T. & Juneng, L. 2010. The roles of climate variability and climate change on smoke haze occurrences in the Southeast Asia region. London: LSE IDEAS.

Tosun, E., Aydin, K. & Bilgili, M. 2016. Comparison of linear regression and artificial neural network model of a diesel engine fueled with biodiesel-alcohol mixtures. *Alexandria Engineering Journal*. 55(4), 3081-3089.

Yoo, S. H., Yang, S. & Ho, C. H. 2006. Variability of the Indian Ocean sea surface temperature and its impacts on Asian-Australian monsoon climate. *Journal of Geophysical Research Atmospheres*. 111(3), 1–17.

Yusof, K.M.K.K, Azid, A. & Jamalani, M.A. 2018. Determination of significant variables to particulate matter (PM$_{10}$) variations in northern region, Malaysia during haze episodes (2006-2015). *Journal of Fundamental and Applied Sciences*, 10(1S), 300-312.