Desalination and Water Treatment
www.deswater.com
doi: 10.5004/dwt.2021.26780

219 (2021) 31–39
April

Haze episodes: identification of air pollutants and meteorological factors in Borneo, Central, Eastern, Northern, and Southern regions of Malaysia

K.M.K.K. Yusof[a], K.C.A. Jalal[b,*], S. Abdullah[c], A. Azid[d], M.S.A. Sani[e], K. Yunus[b], N.M. Isa[f], M.A. Jamalani[f], Z.R. Zulkifli[g], N.L.A. Rani[h]

[a]*Faculty of Science and Marine Environment, Universiti Malaysia Terengganu, 21300, Kuala Terengganu, Terengganu, Malaysia, Tel. +6096683787; email: kukausar@umt.edu.my*
[b]*Kulliyyah of Science, International Islamic University Malaysia, Kuantan Campus, Kuantan, Pahang, Malaysia, emails: jkchowdhury@iium.edu.my (K.C.A. Jalal), kama@iium.edu.my (K. Yunus)*
[c]*Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, 21300, Kuala Terengganu, Terengganu, Malaysia, Tel. +6096683491; email: samsuri@umt.edu.my*
[d]*Faculty of Bioresources and Food Industry, Universiti Sultan Zainal Abidin (UniSZA), Besut Campus, 22200 Besut, Terengganu, Malaysia, Tel. +606993699; email: azman.azid@unisza.my*
[e]*International Institute for Halal Research and Training, International Islamic University Malaysia, Selangor, Malaysia, Tel. +60364216491; email: shirwansany@iium.edu.my*
[f]*Faculty of Environmental Studies, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia, Tel. +60397698498; email: mi.noorain@upm.edu.my (N.M. Isa)*
[g]*Department of Environment, Wisma Sumber Asli, Persiaran Perdana, Presint 4, 62574, Wilayah Persekutuan Putrajaya, Malaysia, Tel. +60388712341; email: zrz@doe.gov.my*
[h]*Excelwaves Sdn. Bhd., No. PT 35442-2, Dataran Austin, Gong Badak, Wakaf Tembesu, 21300, Kuala Terengganu, Malaysia, email: nurullatiffah1076@gmail.com*

Received 15 August 2020; Accepted 23 November 2020

ABSTRACT

This research was based on selected daily haze occurrences between the year 2006 until 2015 in five regions of Malaysia (Borneo, Central, Eastern, Northern, and Southern regions). The generalized linear model (GLM), principal component regression (PCR), incorporation of artificial neural network and sensitivity analysis (ANN-SA) techniques were applied in this study to generate respective models namely as MLP-HM-GLM, MLP-HM-PCR, and MLP-HM-LO and identify the relationship of air pollutants and meteorological factors to particulate matter (PM_{10}) variability. The performances of these models were compared based on coefficient of determination (R^2), root-mean-square error (RMSE), and squared-sum error (SSE). From the findings, ANN-SA that generated the MLP-HM-LO model was the most suitable technique to identify the main contributors to PM_{10} variability. Ultraviolet-b (UV_b) and sulphur dioxide (SO_2) were found as the most significant pollutants that affected the PM_{10} variation. UV_b also had consistently influenced PM_{10} variability over five regions. MLP-HM-LO model had rendered the highest R^2, with the lowest RMSE and SSE values compared with MLP-HM-GLM and MLP-HM-PCR models. Thus, the ANN-SA technique was highly practicable in determining future haze circumstances in Malaysia.

Keywords: Artificial neural network; Sensitivity analysis; Haze episode; Principal component regression; Generalized linear model

* Corresponding author.

1. Introduction

A series of continuous unresolved issues related to haze episodes are currently considered as one of the annual occasions in South East Asia with Malaysia as one of the most affected countries. The haze occurrences for the past 20 year were frequently concentrated on urban areas, specifically in the Central and Southern areas of Peninsular Malaysia. The slower and drier winds and anomalous weather phenomena such as El-Nino Southern Oscillation (ENSO) may have had accelerated the haze formation [1].

In addition to these factors, any prolonged transboundary haze or localized transient haze could deteriorate this condition. Localized transient haze within urban areas due to anthropogenic emissions under stagnant wind circumstances had been recorded during Southwest monsoon or inter monsoonal season as early as April 1983, October 1991, and between August and October 1994 [2]. Furthermore, in 1997, a massive open burning of biomass in Southern Sumatera and Kalimantan had caused a transboundary haze and became a catastrophic level due to its intensity and duration to the surrounding receptors [3,4]. Activities such as agricultural production, industrial, and transportation emission had contributed to this condition that imposed limited visibility and serious health concerns [5].

Although a number of studies were reported in demonstrating the haze distribution, intensity, and health effect in Malaysia [6], the information on contributing parameters towards the haze occurrences is quite negligible. In 2012, air pollutants were successfully detected in eight monitoring stations using Hierarchical Agglomerative Cluster Analysis (HACA), principal components analysis (PCA) and multiple linear regression [7]. Meanwhile, the performance of the long-term air quality assessment has been analyzed using sensitivity analysis (SA), HACA, and principal component regression (PCR) from the background station [8]. Both studies succeed in interpreting the air quality assessment during normal conditions but not during the hazy period. Thus, to understand the inter-relationship between parameters during the haze, artificial neural network (ANN) coupled with SA was introduced. The main advantage of using ANN is its capability to mimic the human brain system by learning complex interaction between datasets (independent and dependent variables) and carry out prediction, clustering, and classification processes [9]. Neuron, which may in the form of single or multiple structures, holds and processes the information from the input. With this capability, ANN can solve non-parametric (non-linear) problems. Also, ANN-SA had provided a directional semi-quantitative and quantitative estimation of the emission changes [10]. Artificial neural networks, coupled with a sensitivity analysis (ANN-SA), will be then applied in order to extract the vital parameters from the Borneo, Central, Eastern, Northern, and Southern regions of Malaysia.

2. Methodology

2.1. Data collection

The air-quality dataset from January 2006 until December 2015 was provided by Air Quality Division, Department of Environmental (DOE), Ministry of Environment and Water. The data were obtained from 52 monitoring stations across Malaysia with a total of 11,207 sets of daily data, involving 15 variables. These stations entailed Borneo, Central, Eastern, Northern, and Southern regions of Malaysia, and all 52 stations covered rural, urban, suburban, and industrial areas [11].

Since this study was focusing on haze occurrences, only $PM_{10} \geq 150$ μg/m^3 and selected air pollutants were taken into consideration for data analysis. To understand the distribution of the air pollutants and their contributions to haze occurrences, the air pollutants were divided into three groups of independent variables: (1) air-pollutant-index (API) pollutants (sulphur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3), and carbon monoxide (CO)); (2) non-air-pollutant-index (NAPI) pollutants (nitrogen monoxide (NO), nitrogen oxides (NO_x), methane (CH_4), non-methane hydrocarbon (NmHC), and total hydrocarbon (THC)); and (3) meteorological factors (wind speed (WS), wind direction (WD), temperature (T), relative humidity (RH), and ultraviolet-b (UV_b)). From all, only particulate matter (PM_{10}) was selected as a dependent variable. No imputation procedure was employed for the treatment of missing data. The PM_{10} readings were logged using a BAM-1020 Beta Attenuation Mass Monitor (Met One Instrument Inc., USA) with the capability to log on the PM_{10} reading every 1- and 24-h basis. The readings of SO_2, NO_2, O_3, and CO were taken using Teledyne API Model 100A/100E, Teledyne API Model 200A/200E, Teledyne API Model 300/300E, and Teledyne API Model 400/400E, respectively (Teledyne Technologies Inc., USA) [12].

Methane, non-methane hydrocarbon, and total hydrocarbon were monitored by using the instrument, which was Teledyne API M4020, whilst NO and NO_x also shared the same instrument (Teledyne API Model 200A/200E) such as nitrogen dioxide (NO_2). In order to record the meteorological factors, Met One 010C, Met One 062, and Met One 083D were used in wind speed, temperature, and relative humidity, respectively [12]. The model equipment model for continuously monitoring program (CAQM) on each atmospheric pollutants and meteorological parameters [12] are listed in Table 1.

2.2. Statistical analysis

2.2.1. Development of air pollutant and meteorological-apportionment models

In this study, three types of models, namely as MLP-HM-GLM, MLP-HM-PCR, and MLP-HM-LO, were constructed based on generalized-linear-model (GLM), PCR, and ANN-SA techniques, respectively. The details of each technique were described as follows:

$$y = Y + \frac{\left\{g(u) - \mu\right\}}{\left[\left(\frac{d}{dY}\right) g\left\{f^{-1}(Y)\right\}\right]} \quad (1)$$

2.2.2. PCR technique

PCR is a combination of ordinal least square and PCA [8]. The PCR is very useful in identifying inter-relationship

Table 1
Model of equipment for continuous monitoring program

Parameter	Equipment model
Particulate matter (PM_{10}), μg/m^3	BAM-1020 Beta Attenuation
Wind speed (WS), km/h	Met One 010C
Air temperature (AT), °C	Met One 062
Relative humidity (RH), %	Met One 083D
Nitrogen oxides (NO_x), ppm	Teledyne API Model 200A/200E
Nitrogen monoxide (NO), ppm	Teledyne API Model 200A/200E
Ultraviolet-b (UV_b), J/m^{2}h	Not available
Methane (CH_4), ppm	Teledyne API M4020
Non-methane hydrocarbon (NmHC), ppm	Teledyne API M4020
Total hydrocarbon (THC), ppm	Teledyne API M4020
Sulphur dioxide (SO_2), ppm	Teledyne API Model 100A/100E
Nitrogen dioxide (NO_2), ppm	Teledyne API Model 200A/200E
Ozone (O_3), ppm	Teledyne API Model 400/400E
Carbon monoxide (CO), ppm	Teledyne API Model 300/300E

between air pollutants and meteorological factors by reducing any multicollinearity issues amongst independent variables. Thus, in this study, PCR identified any strong factor loading that equal and above 0.70 (with Kaiser-normalization) to be selected as significant to be compared with ANN [13].

2.2.3. ANN-SA technique

Artificial neural network incorporated with sensitivity analysis (ANN-SA) was used to examine the relationship between independent variables (air pollutants and meteorological factors) and dependent variable (PM_{10}). All 15 variables were statistically analyzed using JMP10 (Ver. 2015, USA) and XLSTAT (Ver. 2014, USA) software. ANN structure comprises three main components: input, hidden node (neuron), and output. In this study, the air pollutants and meteorological factors were set as the input, while PM_{10} was selected as the output. The interaction between neurons provides a better interpretation of pattern recognition and prediction [14].

Sensitivity analysis (SA) is a powerful tool to identify the best parameters for model development of the environmental system [15]. In this study, SA was employed to estimate and examine the response of air-pollutant characteristics (independent variables) to the PM_{10} variability (dependent variable) using the leave-one-out technique [16]. The individual air-pollutant parameter was removed, and the remaining parameters were taken as the neuron inputs for the ANN model. The coefficient of determination (R^2) for each ANN model was then calculated and converted into percentage form [17]. The best definition to describe R^2 is the proportion of dependent variable (x), that is, contributed from the independent variable (y). To calculate the percentage contribution for individual leave-out parameter, the following formula was applied [18]:

$$\% \text{ Contribution} = \frac{b-a}{c} \times 100 \tag{2}$$

where a was the R^2 value after the leave-one-out calculation for each model, b was the reference R^2 value from multilayer perceptron-haze model-all parameter (MLP-HM-AP), and c was the sum of R^2 difference. This percentage represented the contribution of each air pollutant to the PM_{10} variability. Only data of air pollutants that gave >10% contribution will be taken for further analysis. The best model will fit the highest R^2 (near to 1) with the lowest RMSE (near to 0) [19]. The R^2, RMSE, and SSE can be calculated as follows:

$$R^2 = 1 - \frac{\sum\left(Y_i, \text{predicted} - \hat{Y}\right)^2}{\sum\left(Y_i, \text{observation} - \hat{Y}\right)^2} \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(t_i - td_i\right)^2} \tag{4}$$

3. Discussion

3.1. GLM and PCR

Two models, namely MLP-HM-GLM and MLP-HM-PCR, were developed using GLM and PCR to compare the main contributors to PM_{10} variation. Table 2 shows the effect test of GLM to select the significant parameters that contribute to the PM_{10}. Six parameters entailed WS, WD, RH, NO_x, NmHC, and SO_2 had demonstrated significant contributions ($p < 0.05$) to PM_{10} variability.

For PCR, the acceptance criterion for a significant parameter was to possess factor loading (FL) ≥ 0.70. As shown in Table 3, only T, NO_x, UV_b, NmHC, and THC fell into this criterion.

3.2. Development of model for air pollutant and meteorological apportionment by ANN-SA technique

The result from ANN-SA, which was used to identify the significant contributors to haze occurrence in Malaysia,

Table 2
Effect tests of generalized linear model

Source	DF	L-R Chi Square	Probability > Chi square value
Wind speed (WS)	1	11.35708	**0.0008**
Wind direction (WD)	1	3.84836	**0.0498**
Temperature (T)	1	0.37965	0.5378
Relative humidity (RH)	1	21.40204	**<0.0001**
Nitrogen oxides (NO_x)	1	21.86194	**<0.0001**
Nitrogen monoxide (NO)	1	0.00169	0.9672
Ultraviolet-b (UV_b)	1	3.34758	0.0673
Methane (CH_4)	1	3.79977	0.0513
Non-methane hydrocarbon (NmHC)	1	4.39644	**0.0360**
Total hydrocarbon (THC)	1	0.05474	0.815
Sulphur dioxide (SO_2)	1	25.84933	<0.0001
Nitrogen dioxide (NO_2)	1	0.52892	0.4671
Ozone (O_3)	1	13.83975	0.0002

Table 3
Factor loading by principal component regression

Parameter	Factor loading 1
Wind speed (WS)	–0.567
Wind direction (WD)	0.091
Temperature (T)	**–0.728**
Relative humidity (RH)	0.663
Nitrogen oxides (NO_x)	**0.718**
Nitrogen monoxide (NO)	0.580
Ultraviolet-b (UV_b)	**–0.701**
Methane (CH_4)	0.378
Non-methane hydrocarbon (NmHC)	**0.718**
Total hydrocarbon (THC)	**0.749**
Sulphur dioxide (SO_2)	0.305
Nitrogen dioxide (NO_2)	0.683
Ozone (O_3)	–0.532
Carbon monoxide (CO)	0.434
Statistics	*Value*
Eigenvalue	4.894
Variability (%)	34.959
Cumulative (%)	34.959

[a] Bold values had factor loading ≥0.70.

is shown in Table 4. The ANN technique had rendered several haze models, which consisted of a model of multilayer perceptron using all air-pollutant parameters (MLP-HM-AP) and a model of multilayer perceptron using leave-one-out parameter (MLP-HM-LO). The MLP-HM-AP model was the reference model, whilst the MLP-HM-LO models entailed 14 parameters that were individually removed for calculation of R^2 and R^2 difference. From Table 4, the MLP-HM-AP model possessed the highest R^2 (0.6974) compared with all MLP-HM-LO models since MLP-HM-AP is the reference model. Upon removal of individual air pollutants, the R^2 of each MLP-HM-LO model ranged between 0.312 and 0.616, and the R^2 difference was between 0.082 and 0.385. Among the MLP-HM-LO models, MLP-HM-LOUV$_b$ (ultraviolet-b) model exhibited the lowest R^2 (0.3124) with the highest R^2 difference (0.6156). Thus, it indicated that the removal of UV_b parameter from the reference model had significantly reduced the efficiency of the MLP-HM-AP model. This finding also signified UV_b as the main contributor to the haze occurrence in this study.

To study the effect of air pollutant and meteorological factor toward variability of PM_{10} by only depending on the R^2 and R^2 difference was insufficient. It would only tabulate the relationship between independent variables towards the dependent variable, but it cannot determine what the most significant variable to the PM_{10} variability during haze is. The previous study showed that the calculation of percentage contribution from SA had helped to identify the independent variables that contribute to the variation of the dependent variable [18].

From Table 4, MLP-HM-LOUV$_b$ and MLP-HM-LONO$_2$ had the highest and lowest percentage contributions to PM_{10} variability, respectively. The descending hierarchy of SA towards PM_{10} variation was as follows: UV_b > SO_2 > NO_x > NO > O_3 > RH > T > CO > THC > CH_4 > WD > W S > NmHC > NO_2. From SA result, the effective models were selected based on their percentage contributions that exceeded 10% [18]. Only MLP-HM-LOUV$_b$ and MLP-HM-LOSO$_2$ models had shown percentage contributions of 14.93% and 10.37%, respectively (Table 4), and since these models had percentage contribution >10%, UV_b and SO_2 were chosen as the strongest contributors to PM_{10} variability and haze occurrence in this study.

The UV_b had a strong association with PM_{10} that was by our result. Also, the PM_{10} was claimed to be one of the significant UV_b absorbers in the long run [20]. Likewise, as haze occurrence is an annual event, the UV radiation tends to decrease over the years. This claim was supported by the reduction of UV radiation from 36.90% during hazy days (250 μg/m^3 ≤ PM_{10} ≤ 350 μg/m^3) to 22.00% during normal non-hazy days (PM_{10} ≤ 150 μg/m^3) [21]. This finding had proven

Table 4
Result of ANN for pollutant-apportionment models

Model	(R^2)	R^2 difference	% Contribution to PM_{10} variability
MLP-HM-AP	0.6974[a]		
MLP-HM-LOUV$_b$	0.3124[b]	0.385	14.93
MLP-HM-LOSO$_2$	0.4300[b]	0.267	10.37
MLP-HM-LONO$_x$	0.4914[b]	0.206	7.99
MLP-HM-LONO	0.5111[b]	0.186	7.22
MLP-HM-LOO$_3$	0.5114[b]	0.186	7.21
MLP-HM-LORH	0.5163[b]	0.181	7.02
MLP-HM-LOT	0.5203[b]	0.177	6.87
MLP-HM-LOCO	0.5244[b]	0.173	6.71
MLP-HM-LOTHC	0.5317[b]	0.166	6.42
MLP-HM-LOCH$_4$	0.5501[b]	0.147	5.71
MLP-HM-LOWD	0.5503[b]	0.147	5.71
MLP-HM-LOWS	0.5504[b]	0.147	5.70
MLP-HM-LO NmHC	0.5691[b]	0.128	4.98
MLP-HM-LONO$_2$	0.6156[b]	0.082	3.17
Total	Nr	2.579	100.00

R^2 with the different superscript alphabet were significantly different ($p < 0.05$);
Nr = not related.

the gradual reduction of UV_b to 20% at Mexico City, which experienced haze occurrence [22].

In contrast to the UV_b circumstance, the SO_2 contribution was mostly driven by fossil-fuel consumption due to industrialization and personal-vehicle usage, which had distressed the ecosystem in urban areas. Time-series sampling was applied to study the relationship between daily-mortality rate, PM_{10}, and sulphate compound [23]. They found that sulphate appeared to have a strong correlation with PM_{10}. It was reported that inorganic contaminants derived from sulphate were suddenly increased during the haze period [24].

3.3. Performance evaluation of MLP-HM-LO, MLP-HM-GLM, and MLP-HM-PCR models

The MLP-HM-LO, MLP-HM-GLM, and MLP-HM-PCR models had identified significant parameters that contributed to PM_{10} variability for air pollutant data in Malaysia (Table 5).

These models also performed the same process on region basis: Borneo, Central, Eastern, Northern, and Southern regions. As a comparison, there were inconsistencies between MLP-HM-LO, MLP-HM-GLM, and MLP-HM-PCR identifications. For the MLP-HM-LO model, the ANN-SA technique had predicted UV_b as a significant contributor in all regions (Table 5) while the MLP-HM-GLM model had predicted T in Malaysia, Eastern, and Northern regions. The inconsistent prediction was also proven when the MLP-HM-PCR model had predicted NO_x as the main air pollutant in all regions. Since these models had identified different significant parameters, we had selected the best model based on their performances using R^2, RMSE, and SSE result in Table 6.

Table 5
Significant parameters identified by models

MLP-HM-LO	
Malaysia	UV_b and SO_2
Borneo	NO and UV_b
Central	UV_b, CO, SO_2, and WD
Eastern	WS, THC, RH, NO_x, and UV_b
Northern	UV_b, T, RH, and WS
Southern	UV_b, NO_2, and T
MLP-HM-GLM	
Malaysia	WS, T, UV_b, THC, SO_2, and O_3
Borneo	N/A
Central	CO
Eastern	T, RH, and SO_2
Northern	T
Southern	UV_b
MLP-HM-PCR	
Malaysia	NO_x and NO
Borneo	T, NO_x, UV_b, NmHC, and THC
Central	NO_x, NO, NmHC, and THC
Eastern	NO_x, NO, CH_4, NmHC, THC, and O_3
Northern	NO_x, CH_4, and THC
Southern	NO_x and NO

The R^2 result of MLP-HM-LO model had exhibited the highest values for Eastern, Central, Northern, and Southern regions (0.9009871, 0.4556180, 0.2445727, and 0.1908602), respectively, while MLP-HM-GLM and MLP-HM-PCR

Table 6
Overall performances of MLP-HM-LO, MLP-HM-GLM, and MLP-HM-PCR models for air pollutant apportionment in Malaysia, Borneo, Central, Eastern, Northern, and Southern regions

	MLP-HM-LO	MLP-HM-GLM	MLP-HM-PCR
Region	Coefficient of determination (R^2)		
Malaysia	0.1001773	**0.2631731**	0.0146705
Borneo	0.1840220	N/A	**0.9681503**
Central	**0.4556180**	0.1134654	0.1161727
Eastern	0.9009871	0.1419211	N/A
Northern	0.2445727	0.0034610	0.1676596
Southern	**0.1908602**	0.0380241	0.0439994
Region	Root-mean-square error (RMSE)		
Malaysia	38.505052	**33.194499**	46.240427
Borneo	25.002358	N/A	**4.004759**
Central	**20.920946**	45.875815	31.12534
Eastern	**28.212191**	39.799997	N/A
Northern	32.219578	43.878379	31.772294
Southern	**43.414998**	46.788598	47.398515
Region	Sum-square error (SSE)		
Malaysia	1,414,437.7	**794,767.68**	17,610,027
Borneo	53,135.024	N/A	272.64761
Central	**89,287.943**	539,614.24	**539,614.24**
Eastern	166,348.89	**272.64761**	N/A
Northern	**216,963.15**	497,672.97	497,672.97
Southern	**493,833.86**	4,084,353.8	4,084,353.8

models only demonstrated the highest R^2 values for the whole Malaysia (0.2631731) and Borneo (0.9681503) region, respectively.

Despite R^2 comparison, other indicators to determine the best model performance were RMSE and SSE values. From Table 6, the MLP-HM-LO model had the lowest RMSE value for three regions (Central, Eastern, and Southern regions) compared with only a region for MLP-HM-GLM (Malaysia) and two regions for MLP-HM-PCR (Borneo and Northern regions). The MLP-HM-LO model also had exhibited the lowest SSE value for Central, Northern, and Southern regions. In comparison, the MLP-HM-GLM model had the lowest SSE value for Malaysia and Eastern regions, while the MLP-HM-PCR model had shown the lowest SSE value for the Borneo region. This finding denoted that the MLP-HM-LO model was the most suitable model to explain the haze occurrence in high-density areas such as Central, Eastern, Northern, and Southern regions.

3.4. Evaluation of main contributors to PM_{10} variation in Borneo, Central, Eastern, Northern, and Southern regions using MLP-HM-LO model

By using the MLP-HM-LO model, the ANN-SA technique selected the main contributors to PM_{10} variability in Borneo, Central, Eastern, Northern, and Southern regions based on their percentage contributions >10% (Table 7). The summary of the main contributors in the five regions is illustrated in Fig. 1. In an overview, ultraviolet-b had percentage contribution >10% in all regions where Central, Southern, and Northern regions had percentage contribution >30% of UV_b.

The Borneo region had NO (48.840%) and UV_b (48.071%) as dominant contributors that made up to 97.00%. This huge gap of percentage contribution between these parameters could be due to limited monitoring stations to cover this large region. Since the Borneo region entailed Sabah and Sarawak states, these monitoring stations had produced insufficient and misrepresent datasets. For the record, Borneo only has 14 monitoring stations or 27% compared with 38 stations (73%) in peninsular Malaysia.

The Central region which comprised of Selangor and Negeri Sembilan states and Federal territories of Kuala Lumpur and Putrajaya had shown significant ($p < 0.05$) parameters that affected PM_{10} variability with the hierarchy of UV_b (25.380%) > CO (15.237%) > SO_2 (10.490%) > WD (10.126%). It might appear due to anthropogenic emission and rapid urbanization in this region, including industrialization [25] and high usage of vehicles [26]. Extensive application of biomass fuel, incomplete combustion, and endless exploitation of forest to cater to the need of urban development had worsened this situation [27,28].

Unlike Borneo and Central regions, the annual Northeast Monsoon (NEM) affects the Eastern region between

Table 7
ANN-SA for haze prediction in Malaysia, Borneo, Central, Eastern, Northern, and Southern regions

Model	Percentage contribution (%)					
	Malaysia	Borneo	Central	Eastern	Northern	Southern
MLP-HM-LOWS	5.700	0.001	1.436	**23.017**	**11.127**	–1.070
MLP-HM-LOWD	5.705	0.000	**10.126**	2.065	–0.897	9.912
MLP-HM-LOT	6.867	0.000	5.721	–0.221	**25.204**	**12.937**
MLP-HM-LORH	7.022	0.000	4.489	**17.517**	**21.451**	–0.783
MLP-HM-LONO$_x$	7.986	0.000	3.066	**12.653**	6.920	4.368
MLP-HM-LONO	7.223	**48.840**	4.738	5.929	1.417	–5.720
MLP-HM-LOUV$_b$	**14.929**	**48.071**	**25.380**	**10.982**	**32.883**	**44.907**
MLP-HM-LOCH$_4$	5.712	2.977	5.900	4.593	–2.019	0.471
MLP-HM-LONmHC	4.975	0.000	4.607	0.185	–3.782	0.771
MLP-HM-LOTHC	6.423	0.000	2.573	**18.098**	–3.189	0.577
MLP-HM-LOSO$_2$	**10.369**	0.000	**10.490**	–0.440	–4.119	6.387
MLP-HM-LONO$_2$	3.172	0.002	1.602	–0.442	2.069	**26.044**
MLP-HM-LOO$_3$	7.211	0.010	4.635	4.413	8.645	–5.274
MLP-HM-LOCO	6.706	0.100	**15.237**	1.652	4.290	6.473

Bold values had percentage contribution >10%;
ANN-SA data of air pollutants in Malaysia were used as reference.

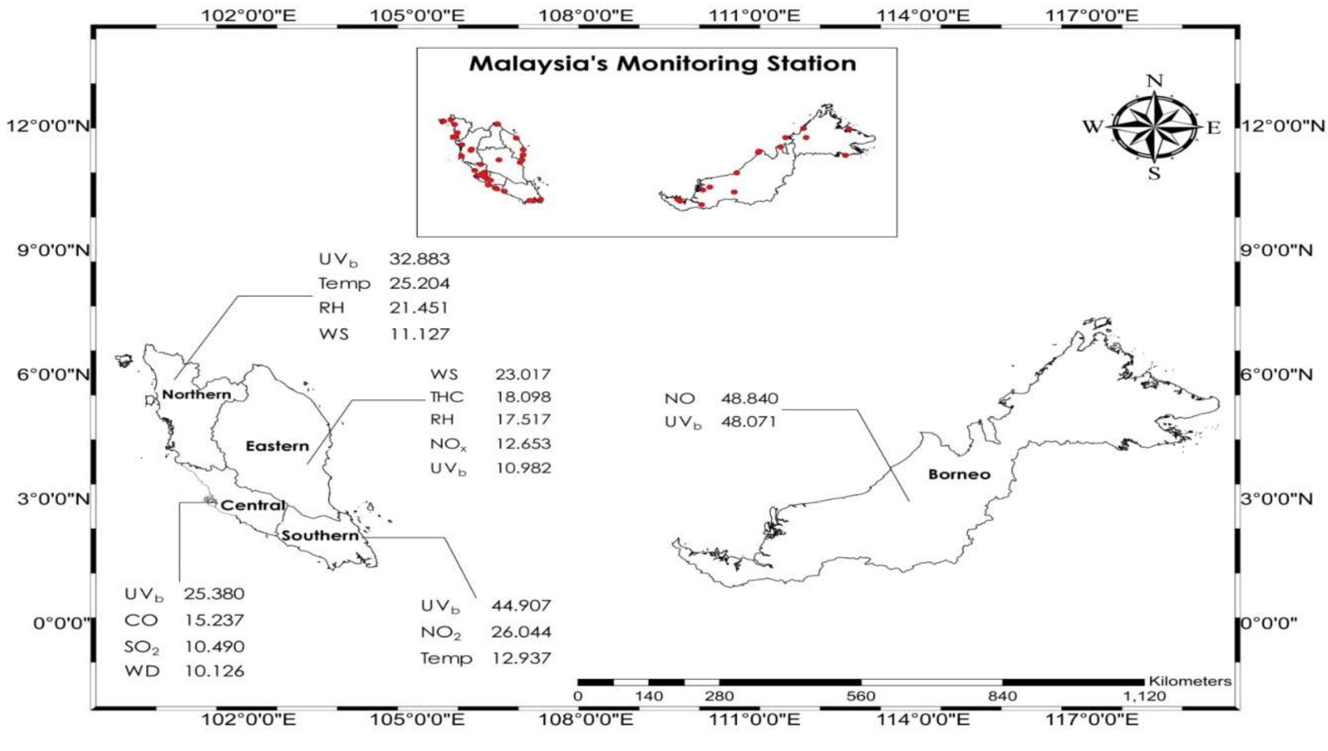

Fig. 1. Main contributors to PM_{10} variation (percentage contribution >10%) in Borneo, Central, Eastern, Northern, and Southern region.

November and February [29], in which it had caused additional parameters that contributed to PM_{10} variation. Since the Eastern region faces the South China Sea, WS became the highest contributor (23.020%) and was followed by THC (18.098%), RH (17.517%), NO_x (12.653%), and UV_b (10.982%). Besides, cloudy weather during NEM had limited the sunlight and UV_b from reaching the Eastern region and thus, supported the low percentage contribution of UV_b [30]. Nonetheless, T and two API pollutants, which were SO_2 and NO_2 had negative prediction value, indicated that the higher T, SO_2, and NO_2, do not have any interaction between PM_{10} during the haze period.

For the Northern region, only meteorological factors such as UV_b (32.883%), T (25.204%), RH (21.451%), and WS (11.127%) had influenced the PM_{10} variability while CH_4, THC, NMHC, WD, and SO_2 had a negative contribution (Table 7). The T and RH that affected haze occurrence in this region was probably because of dry and humid weather and its location near the dry area of Thailand. The MLP-HM-LO model also had identified UV_b, NO_2, and T as dominant contributors over the Southern region. As been discussed in the Central region, UV_b showed the highest rank with 44.91% percentage contribution, while NO_2 and T were 26.04% and 12.94%, respectively. Besides, other factors such as WD, CO, and NO (4.37%–9.91%) had a slight effect on the PM_{10} variability.

4. Conclusions

This study had successfully proven the applicability of ANN-SA technique in two aspects: (1) identification on main contributors to PM_{10} viability or haze circumstance in Malaysia between 2006 and 2015, and (2) identification on main contributors during haze episodes in Borneo, Central, Eastern, Northern, and Southern regions. Statistically, ANN-SA able to demonstrate a better predicting technique than PCR and GLM techniques as proven with higher R^2 and lower RMSE and SSE values in the five regions. Hence ANN-SA was selected as the most suitable technique in this study. Our results had demonstrated the dissimilarities of contributing parameters in each region except UV_b. The UV_b also was the highest contributor in the Northern, Central, and Southern regions but appeared as the least contributor as compared with WS, THC, RH, and NO_x in the Eastern region. The temperature was also among the dominant parameters in the Northern and Southern regions. This finding could assist the Department of Environment (DOE) in tackling haze-related issues in the near future. However, the deterioration in the quality of natural water due to haze and other natural mining activities could be key issues in any environmental problems as observed in the Eastern region especially Kuantan water bodies (especially in the Balok and Tunggak rivers Kuantan in the state of Pahang) in 2016. The main problems caused due to bauxite mining activities along with haze pollution could be correlated with severe water pollution. In fact it was occurred in 2016 at Kuantan riverine areas which needed further details studies in the near future.

Acknowledgements

A special appreciation to Air Quality Division, Department of Environment, Ministry of the Environment and Water (MEWA) for supplying and permitting us air-quality data to be included in this publication. The authors wish to express their appreciation to the Faculty of Science and Marine Environment, Universiti Malaysia Terengganu for the possibility offered during the drafting process. The authors are also grateful to the Ministry of Higher Learning Malaysia, (MOHE) and Research Management Centre, International Islamic University Malaysia for provided Research Grant [FRGS: 19-042-0650], which was assisting the authors for processing the publication in this journal.

References

[1] P.A. Pillai, K. Mohankumar, Individual and combined influence of El Nino - Southern oscillation and Indian Ocean Dipole on the Tropospheric Biennial Oscillation, Q. J. Royal Meteorol. Soc., 136 (2010) 297–304.
[2] A.M. Radzi, D.R. Oros, B.R.T. Simoneit, Biomass burning as the main source of organic aerosol particulate matter in Malaysia during haze episodes, Chemosphere, 55 (2004) 1089–1095.
[3] K.M.K. Ku Yusof, A. Azid, M.S. Samsudin, M. Asrul Jamalani, An overview of transboundary haze studies: the underlying causes and regional disputes on Southeast Asia Region, Malaysian J. Fundam. Appl. Sci., 13 (2017) 747–753.
[4] M. Radojevic, Chemistry of forest fires and regional haze with emphasis on Southeast Asia, Pure Appl. Geophys., 160 (2003) 157–187.
[5] M.D. Keywood, G.P. Ayers, J.L. Gras, R. Boers, C.P. Leong, Haze in the Klang Valley of Malaysia, Atmos. Chem. Phys., 3 (2003) 615–653.
[6] N.A. Sulong, M.T. Talib, M.F. Khan, N. Amil, M.J. Ashfold, M.I.A. Wahab, K.M. Chan, M. Sahani, Source apportionment and health risk assessment among specific age groups during haze and non-haze episodes in Kuala Lumpur, Malaysia, Sci. Total Environ., 601–602 (2017) 556–570.
[7] D. Dominick, H. Juahir, M.T. Latif, S.M. Zain, A.Z. Aris, Spatial assessment of air quality patterns in Malaysia using multivariate analysis, Atmos. Environ., 60 (2012) 172–181.
[8] M.T. Latif, D. Dominick, F. Ahamad, M.F. Khan, L. Juneng, F.M. Hamzah, M.S.M. Nadzir, Long term assessment of air quality from a background station on the Malaysian Peninsula, Sci. Total Environ., 482 (2014) 336–348.
[9] L. Petzold, S.T. Li, Y. Cao, R. Serban, Sensitivity analysis of differential-algebraic equations and partial differential equations, Comput. Chem. Eng., 30 (2006) 1553–1559.
[10] A. Dunker, G. Yarwood, The decoupled direct method for sensitivity analysis in a three-dimensional air quality model implementation, accuracy, and efficiency, Environ. Sci. Technol., 36 (2002) 2965–2976.
[11] A. Azid, H. Juahir, E. Ezani, M.E. Toriman, A. Endut, M.N.A. Rahman, K. Yunus, M.K.A. Kamarudin, C.N.C. Hasnam, A.S.M. Saudi, R. Umar, Identification source of variation on regional impact of air quality pattern using chemometrics, Aerosol Air Qual. Res., 15 (2015) 1545–1558.
[12] L. Juneng, M.T. Latif, F. Tangang, Factors influencing the variations of PM_{10} aerosol dust in Klang Valley, Malaysia during the summer, Atmos. Environ., 45 (2011) 4370–4378.
[13] A. Azid, H. Juahir, M.E. Toriman, M.K.A. Kamarudin, A.S.M. Saudi, C.N.C. Hasnam, N.A.A. Aziz, F. Azaman, M.T. Latif, S.F.M. Zainuddin, M.R. Osman, Prediction of the level of air pollution using principal component analysis and artificial neural network techniques: a case study in Malaysia, Water Air Soil Pollut., 225 (2014) 1–14.
[14] A. Azid, H. Juahir, M.T. Latif, S.M. Zain, M.R. Osman, Feed-forward artificial neural network model for air pollutant index prediction in the Southern Region of Peninsular Malaysia, J. Environ. Prot., 4 (2013) 1–10.
[15] Y.J. Yang, J.G. Wilkinson, A.G. Russell, Fast, direct sensitivity analysis of multidimensional photochemical models, Environ. Sci. Technol., 31 (1997) 2859–2868.
[16] D.S. Cohan, A. Hakami, Y. Hu, A.G. Russell, Nonlinear response of ozone to emissions: source apportionment and sensitivity analysis, Environ. Sci. Technol., 39 (2005) 6739–6748.
[17] D. Hwang, D.W. Byun, M.T. Odman, An automatic differentiation technique for sensitivity analysis of numerical advection schemes in air quality models, Atmos. Environ., 31 (1997) 879–888.
[18] A. Azid, H. Juahir, M.E. Toriman, A. Endut, M.N.A. Rahman, M.K.A. Kamarudin, M.T. Latif, A.S.M. Saudi, C.N.C. Hasnam, K. Yunus, Selection of the most significant variables of air pollutants using sensitivity analysis, J. Test. Eval., 44 (2016) 1–10.
[19] D. Arslan, M.M. Özcan, H.O. Mengeş, Evaluation of drying methods with respect to drying parameters, some nutritional

and colour characteristics of peppermint (Mentha x piperita L.), Energy Convers. Manage., 51 (2010) 2769–2775.
[20] W.F. Barnard, V.K. Saxena, B.N. Wenny, J.J. DeLuisi, Daily surface UV exposure and its relationship to surface pollutant measurements, J. Air Waste Manage. Assoc., 53 (2003) 237–245.
[21] H.K. Elminir, Sensitivity of ultraviolet solar radiation to anthropogenic air pollutants and weather conditions, Atmos. Res., 84 (2007) 250–264.
[22] L.R. Acosta, W.F.J. Evans, Design of the Mexico City UV monitoring network: UV-B measurements at ground level in the urban environment, J. Geophys. Res. Atmos., 105 (2000) 5017–5026.
[23] R.M. Harrison, J. Yin, Particulate matter in the atmosphere: which particle propoerties are important for its effects on health?, Sci. Total Environ., 249 (2000) 85–101.
[24] X. Liu, K. Sun, Y. Qu, M. Hu, Y. Sun, F. Zhang, Y. Zhang, Secondary formation of sulfate and nitrate during a haze episode in megacity Beijing, China, Aerosol Air Qual. Res., 15 (2015) 2246–2257.
[25] S.A. Rahman, M.S. Hamzah, A.K. Wood, M.S. Elias, N.A.A. Salim, E. Sanuri, Sources apportionment of fine and coarse aerosol in Klang Valley, Kuala Lumpur using positive matrix factorization, Atmos. Pollut. Res., 2 (2011) 197–206.
[26] R. Shokoohi, A. Nikitas, Urban growth, and transportation in Kuala Lumpur: Can cycling be incorporated into Kuala Lumpur's transportation system?, case studies on transport policy, 5 (2017) 615–626.
[27] L.K. Sahu, V. Sheel, M. Kajino, P. Nedelec, Variability in tropospheric carbon monoxide over an urban site in Southeast Asia, Atmos. Environ., 68 (2013) 243–255.
[28] M. Yang, T. Ma, C. Sun, Evaluating the impact of urban traffic investment on SO_2 emissions in China cities, Energy Policy, 113 (2018) 20–27.
[29] M.J. Ashfold, M.T. Latif, A.A. Samah, M.I. Mead, N.R.P. Harris, Influence of Northeast Monsoon cold surges on air quality in Southeast Asia, Atmos. Environ., 166 (2017) 498–509.
[30] S. Nair, E.A. Resmi, G. Kulkarni, N. Malap, S. Patade, T.V. Prabha, Thermodynamical and cloud microphysical response during the transition from southwest to northeast monsoon, Atmos. Res., 166 (2015) 182–194.