IOP Conference Series: Earth and Environmental Science

The Spatial Influence of Environmental and Anthropogenic Factors on The Pattern of Air Pollution in Malaysia

IOP Conf. Series: Earth and Environmental Science **549** (2020) 012011 doi:10.1088/1755-1315/549/1/012011

The Spatial Influence of Environmental and Anthropogenic Factors on The Pattern of Air Pollution in Malaysia

Siti Hajar Ya'acob[1*] and Abdul Hamid Mar Iman[2]

[1] Faculty of Earth Science, Universiti Malaysia Kelantan, Malaysia
[2] Sustainable Environment and Conservation Cluster, Universiti Malaysia Kelantan, Malaysia

E-mail: hajar.y@umk.edu.my

Abstract. Air contaminant levels experience a large degree of spatial dimension. This study is an attempt to address the issue of spatially related human-made activities that cause distribution air pollutants over the geographic area. PM_{10} variable within 16-year annual observations (2000-2015) from 37 fixed monitoring stations across Peninsular Malaysia was analysed. Spatial analysis was performed using Exploratory Regression, Ordinary Least Square Regression, spatial autocorrelation, and kriging interpolation in ArcGIS software version 10.5. Generally, the variance inflation factor for all the exploratory variables was below 7.5, indicating the absence of multicollinearity among them. The adjusted R^2 was in the range of 0.3–0.4 for the selected sub-model. Only industrial land use and RH were significant predictors in the selected sub-model. The initial profiling of PM_{10} using geographic information system (GIS) was able to identify the relevant spatial relationship leading to the identification of monitoring stations area either belongs to a hotspot or cold spots.

1. Introduction

Air pollution has been a long-nagging phenomenon since the start of the industrial revolution [1,2], and it contributes to the deterioration of quality of human life, including pregnancy outcomes [3,4] and also affects other living organisms. It affects health and environment in particular related to the cardio-respiratory endpoint, water pollution, and vegetation, respectively [5–8]. Foremost, air pollution kills approximately 5.5 million people annually worldwide [9]. It also has economic consequences [10].

Causes of air pollution are mostly anthropogenic mainly due to growing urbanization and industrialization, with most of them originating from various sources such as factories, power plants, dry cleaners, vehicles, and even wind-blown dust and wildfires [11,12]. Afroz et al [13] reported that for the past five years, the three significant sources of air pollution in Malaysia are mobile (c.a. 70-75% of total air pollution), stationary sources (c.a. 20-25%), and open burning sources (c.a. 3-5%).

The primary sources of air pollution include emissions from vehicles and industrial activities that discharge greenhouse gases (GHG) into the environment [14]. Transboundary air pollution also has become a new phenomenon in the modern world [15–19]. According to [20], urbanization, industry, motor vehicles, and forest fires remain the main contributors to the deterioration of air quality in the Klang Valley. The most common air pollutants are carbon monoxide (CO), hydrocarbons (HCs), oxides of nitrogen (NOx) such as nitric oxide (NO) and nitrogen dioxide (NO_2), ozone (O_3), particulate matter such as PM_{10}, sulfur dioxide (SO_2), and ammonia (NH_3) [20,21].

Air pollution attracts attention from the relevant authorities around the globe leading to significant actions such as the 1979 Convention on Long-range Transboundary Air Pollution and its Protocols [22,23] and the Ten Research Priorities for Airborne Particulate Matter agenda [24,25]. Apart from that, research works on air pollution in Malaysia have included the level of air pollutant concentration at various locations and duration, source apportionment, trajectory pathways, modeling of particulate matter using satellite remote sensing and meteorological relationship [26]. Among others, the monitoring activities give the status of current air quality based on the Malaysia Air Pollution Index (MAPI) for only six air pollution parameters namely ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2) and particulate matter (PM).

Patterns of air pollution in Malaysia are yet to be more understood, especially because vulnerable populations are more likely to live closer to pollutant sources and, thus, closer to pollutant monitors [27]. Their spatiality and temporality are indicated in many past studies [28–32]. Awang et al. [33], for example, discovered that NO_2, CO, PM_{10}, and SO_2 emitted from industrial and urban areas had demonstrated two peaks in the diurnal variations whereby the morning rush-hour peak was mainly due to vehicle emissions while the late evening peak was attributed mostly to meteorological conditions, particularly atmospheric stability and wind speed.

This paper attempts to test the hypothesis that more industrialized areas and more populated cities tend to have a higher level of air pollution, specifically PM_{10} pollutants. If this is true, in the Malaysian case, the Peninsular Malaysia west coast's air can be expected to be more polluted than that of the east coast.

2. Theoretical background

The Malaysian states can be divided into two groups in the number of active vehicle populations – one with vehicle population exceeding one million units (Figure 1, panel a) and the other less than one million units (Figure 1, panel b). In Peninsular Malaysia, air pollution can be expected to be higher on the west coast, especially in the Klang Valley (F.T. of Kuala Lumpur and Selangor), Johor, Pulau Pinang, and Perak, compared to the east coast. The trend in the number of industries reflecting the intensity of industrialization in the Malaysian states is summarized in Figure 2. Similarly, these states can be subdivided into three groups with Selangor, Johor, and Pulau Pinang top the list followed by Perak, Kedah, Negeri Sembilan, Sarawak, and Sabah in the second group while Pahang, Terengganu, Federal Territory of Kuala Lumpur, Kelantan, Perlis, and Federal Territory of Labuan in are in the third group.

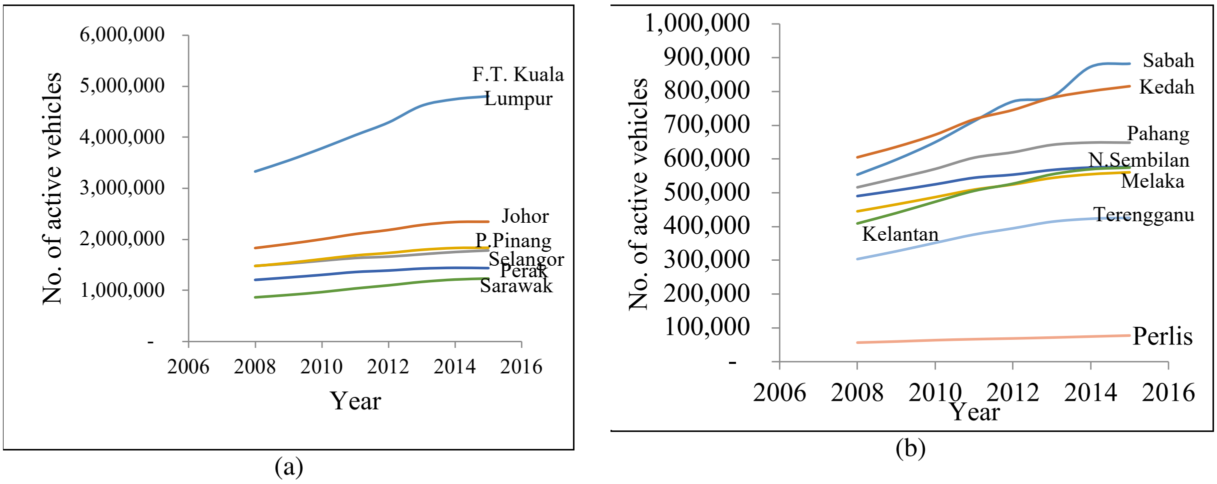
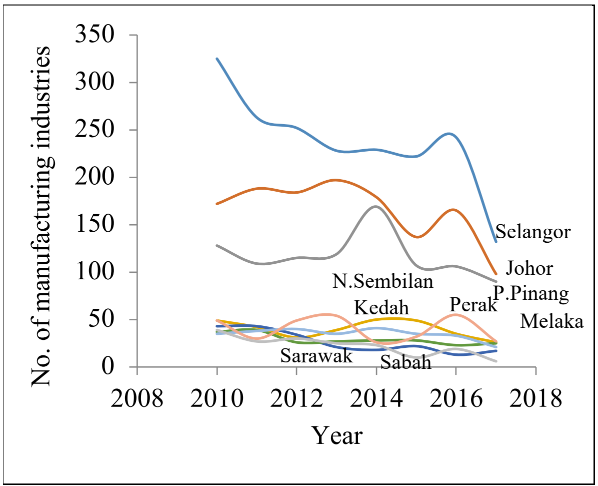


Figure 1. The trend in the number of vehicles in Malaysia, 2008-2015 (Source: Constructed from Department of Road Transport Malaysia, 2008-2015)

On the basis of industrialization, air pollution is expected to be higher on the west coast of Peninsular Malaysia, especially in the Klang Valley, Johor, and Pulau Pinang compared to the east coast.

Based on the trend of vehicle population and industrialization discussed above, it can be predicted that air pollution tends to demonstrate spatial patterns. However, the spatial pattern of air pollution is of special interest since it indicates some important spatial relationships that can help us to understand people's activities and environmental quality. It may also indicate people's demographic characteristics. For example, exposure to benzene pollution was related to low-side demographic characteristics of the population [34].



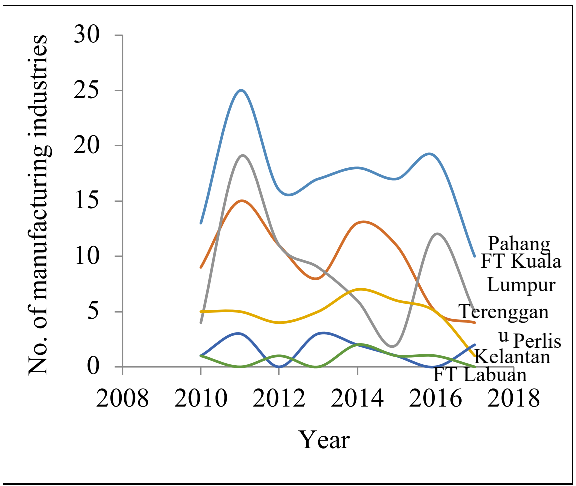


Figure 2. The trend in the number of manufacturing industries in Malaysia, 2010-2017 (Source: Malaysian Industrial Development Authority (MIDA), 2010-2017)

Profiling air pollution spatially will add to the understanding of air quality across a particular geographic locality so that monitoring activities can give the status of current air quality by geographic sub-region. In this context, potential accumulation of air pollutants can be analyzed for a specified duration of exposure, which represents the level of air pollution calculated spatially and temporally over several years.

3. Methodology

3.1 Study area

Peninsular Malaysia is situated between Thailand and Indonesia. It has a total population of nearly 32.6 million [35]. Up to the year 2019, Peninsular Malaysia has 37 Continuous Air Quality Monitoring (CAQM) stations situated at various locations and managed by the Department of Environment (DOE) Malaysia (Figure 3). This study used the recorded pollution data from all of these sites to represent air pollution exposure in Peninsular Malaysia. These stations are divided by DOE into industrial, urban, and suburban areas, whereas only one location – Jerantut – is considered as a background. These stations were regularly operated by Alam Sekitar Malaysia Sdn. Bhd (ASMA) until April 2017. All instruments at those stations were fully automated. All data gathered from these stations were analyzed using ArcGIS 10.5 software.

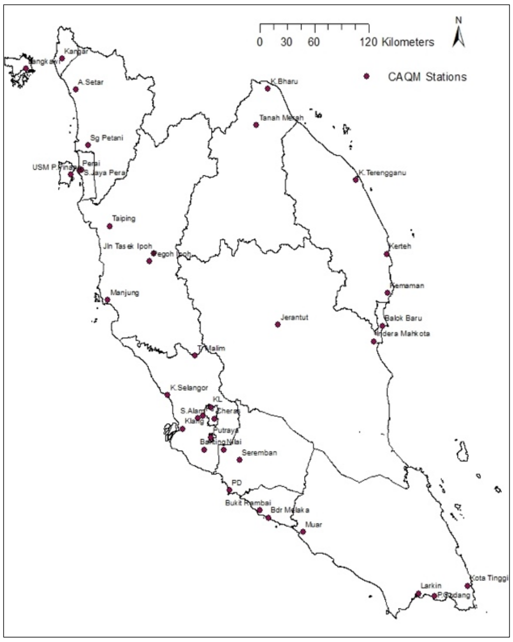


Figure 3. Locations of 37 CAQM stations across Peninsular Malaysia

3.2 Data sources and mining

The dataset on PM_{10} and three meteorological variables (wind speed, temperature, and humidity) covering a 16-year period from 2000 to 2015 were provided by the DOE. The data were originally provided in the hourly form but were then converted into yearly averages to simplify the annual trend analysis to meet the ArcGIS data mining procedure. We also included types of land-use (LU) area (for the year of 2015) and average daily traffic (ADT) (the year 2006 to 2015) to represent potential air pollutant sources. These two datasets were obtained from the Federal Department of Town and Country Planning (JPBD), Peninsular Malaysia, and constructed from the report of Transport Statistic Malaysia 2015 [36], respectively. Other predictor variables comprise climatic parameters, namely wind speed (WS), relative humidity (RH), and temperature also provided by the DOE. Altogether, a total of 407 data (37 observations points x 1 dependent variables x 11 independent variables) were prepared in Excel and spatially referenced using each location's longitude and latitude. The coordinates were projected in the World Grid System of 1984 (WGS84) during the process of converting the spreadsheet data into the GIS workspace.

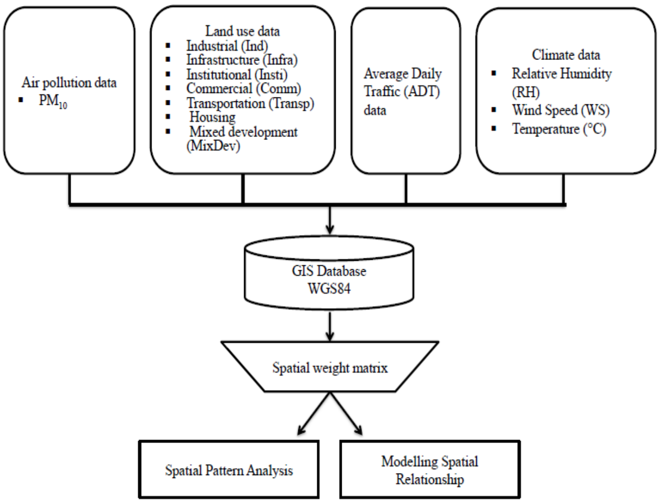


Figure 4. The flow of data mining and processing

3.3 Modelling of spatial relationship

3.3.1 Exploratory regression analysis

The modeling schema is summarized in Figure 4, consisting of PM_{10} as the independent variable and the other eleven independent variables as the predictor variables. Exploratory regression analysis was employed for model selection as well as to identify the relative importance of predictor variables [37]. Hence, this study used the exploratory regression tool as an initial step in modeling the statistical relationship between these variables. This approach provided a powerful method to find for the properly specified Ordinary Least Squares (OLS) model [37]. Through exploratory regression, all potential combinations of eleven candidates' predictor variables towards air pollution phenomena were assessed. This tool generated a summary report that comprised the maximum Variance Inflation Factor (VIF) value, the explanatory variables' significance and sign, the Jarque-Bera p-value, the spatial autocorrelation p-value, the Akaike's Information Criterion (AICc) and the adjusted R^2 value.

3.3.2 Ordinary least squares (OLS) regression

The OLS was first used to explore the possible spatial relationship of PM_{10} concentrations against pollution sources and meteorological factors across Peninsular Malaysia (see Figure 4). Specifically, the OLS technique was employed to answer the question: was there any spatial relationship between the concentrations of PM_{10} (y) and its explanatory variables (x) using the following equation:

3.4 Spatial pattern analysis

Spatial autocorrelation was used to assess the spatial dependence of PM_{10} concentration residual values. In the previous study, we compared two types of spatial autocorrelation statistics, namely Getis-Ord general G and global Moran's I [38]. That study aimed to test for the spatial clustering tendency of the particular air pollutants at 95% confidence level (p-value < 0.05) using the following equation:

$$\mathbf{y_i} = \boldsymbol{\beta}_\mathbf{0} + \textstyle\sum_\mathbf{k} \boldsymbol{\beta}_\mathbf{k} \mathbf{x_{ik}} + \boldsymbol{\varepsilon}_\mathbf{i} \qquad (1)$$

$$G = \frac{\sum_{i=1}^{n}\sum_{i=j}^{n} w_i x_i x_j}{\sum_{i=i}^{n}\sum_{j=1}^{n} x_i x_j}, \forall_{j \neq i} \qquad (2)$$

$$I_t = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - x)(x_j - x)}{\sum_{i=1}^{n}(x_i - x)} \qquad (3)$$

$$S_0 = \sum_{i=i}^{n}\sum_{j=1}^{n} w_{ij} \qquad (4)$$

Both spatial statistics generate z-score values that indicate either spatial dispersion or clustering of residual from PM_{10} regression.

4. Results

4.1 Descriptive and trend analysis for mean PM_{10} concentration

The results of the PM_{10} concentration profile at the 37 CAQM stations within 16 years of duration is shown in Table 1. Seven locations only started to operate recently than 16 years duration ranging from 6 to 14 years back. The comparison of PM_{10} concentration among the 37 CAQM stations was made by dividing them into four spatial categories, namely suburban, urban, industry, and background (see Figure 5). The results of the mean PM_{10} concentration level were also compared with the annual average exposure limit value of 50 $\mu g/m^3$, as outlined in the New Malaysia Ambient Air Quality Standard (MAAQS 2014). Four suburban locations have exceeded the above guideline value, namely Muar, Manjung, Kuala Selangor, Banting, and Seberang Jaya Perai. The urban areas have four (Kuala Terengganu, Cheras, Shah Alam, and Klang) while the industrial areas have seven (Pasir Gudang, Petaling Jaya, Tanah Merah, Perai, Balok Baru, Nilai, and Bukit Rambai) locations with yearly PM_{10} concentration level above 50 $\mu g/m^3$. In particular, Klang (urban area) and Bukit Rambai (industrial area) were two most polluted locations with the highest mean PM_{10} concentration level of 67.84±8.5 $\mu g/m^3$ and 65.88±10.6 $\mu g/m^3$, respectively. This suggests that urban and industrial areas have shown to have a higher level of air pollution. In our case here, Klang and Bukit Rambai were, respectively, about 1.36 times and 1.32 times more polluted than the national average or 1.83 times and 1.78 times more polluted than the background situation, i.e., Jerantut (PM_{10} concentration of 38 $\mu g/m^3$).

Table 1. Descriptive statistics of PM_{10} concentration

Location	Mean ($\mu g/m^3$)	Number of years (n)	Min.	Max.	SD
Klang	67.84	16	54.97	89.30	8.50
Bukit Rambai	65.88	16	49.51	83.40	10.65
Nilai	59.45	16	53.45	69.55	4.15
Balok Baru	57.34	16	44.42	71.79	7.27
S.Jaya Perai	57.32	16	47.85	75.70	7.91
Banting	55.75	6	46.48	71.16	8.84
Shah Alam	55.45	16	38.53	79.56	10.35
Perai	54.50	16	37.64	91.76	17.47
K. Selangor	54.05	16	40.23	76.61	10.80
Cheras	53.01	12	43.70	68.79	6.75
Tanah Merah	52.22	7	39.02	61.95	7.51
Petaling Jaya	52.21	16	36.68	64.29	8.28
Manjung	51.38	16	37.74	70.27	10.47
Pasir Gudang	51.02	16	45.16	64.67	4.80
Muar	50.83	16	42.16	62.62	4.76
Kuala Terengganu	50.63	16	28.66	56.92	6.30
Sungai Petani	49.82	16	37.78	71.15	8.52
Port Dickson	48.44	8	39.75	59.65	5.74
Pegoh Ipoh	48.35	16	36.66	56.85	6.49
K.L	48.31	7	42.05	61.42	7.21
Putrajaya	47.40	14	37.91	64.02	8.58
Jln Tasek Ipoh	47.20	16	36.25	59.76	6.38
Kota Tinggi	46.26	8	39.00	49.70	3.75
Seremban	45.73	16	39.03	58.17	4.82
Kemaman	45.25	16	32.73	61.51	8.62
Taiping	44.93	16	34.09	54.51	6.01
Bndr Melaka	44.21	16	36.13	58.25	7.57
Larkin	43.15	16	32.62	60.23	7.94
Kangar	42.57	16	34.74	60.20	7.07

Location	Mean (μg/m^3)	Number of years (n)	Min.	Max.	SD
Kota Bharu	42.42	16	36.24	67.19	7.26
Tg. Malim	39.25	16	33.78	46.71	4.35
Pulau Pinang	38.45	16	28.02	59.12	7.85
Jerantut	38.08	16	30.81	50.01	5.96
Kerteh	37.58	16	31.50	53.14	5.42
Alor Setar	35.92	16	28.51	54.10	6.04
Langkawi	35.42	16	28.30	47.76	6.27
Ind Mahkota	35.00	16	24.77	44.27	5.06

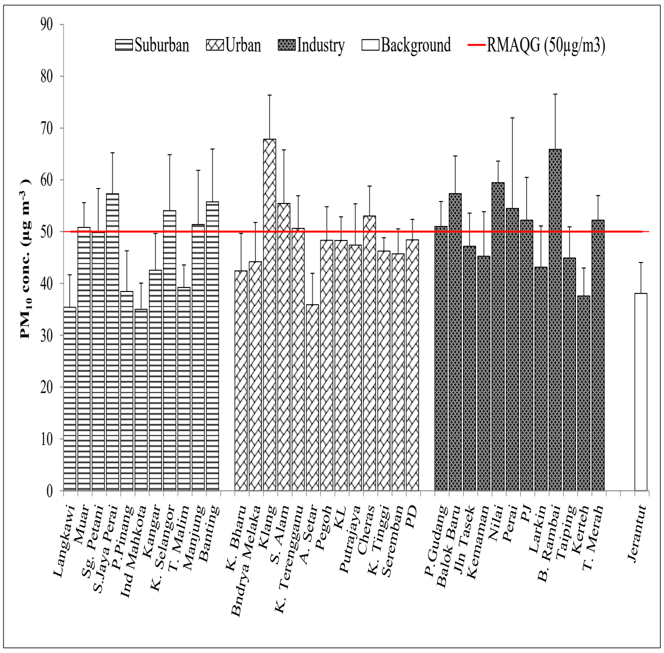


Figure 5. Comparison of average PM_{10} concentration at 37 CAQM stations

A similar pattern of air pollution (monthly mean PM_{10} concentration level) is depicted in Figure 6 on the basis of monsoonal seasons. The South-West monsoon, which usually occurs in June-September, shows a consistently higher monthly mean PM_{10} concentration across Peninsular Malaysia, ranging from 37.41 to76.89 μg/m^3. The North-East (November-March) and Inter monsoon (April, May, and October) recorded the mean PM_{10} concentration range of 31.25-63.34 and 33.74-63 μg/m^3, respectively.

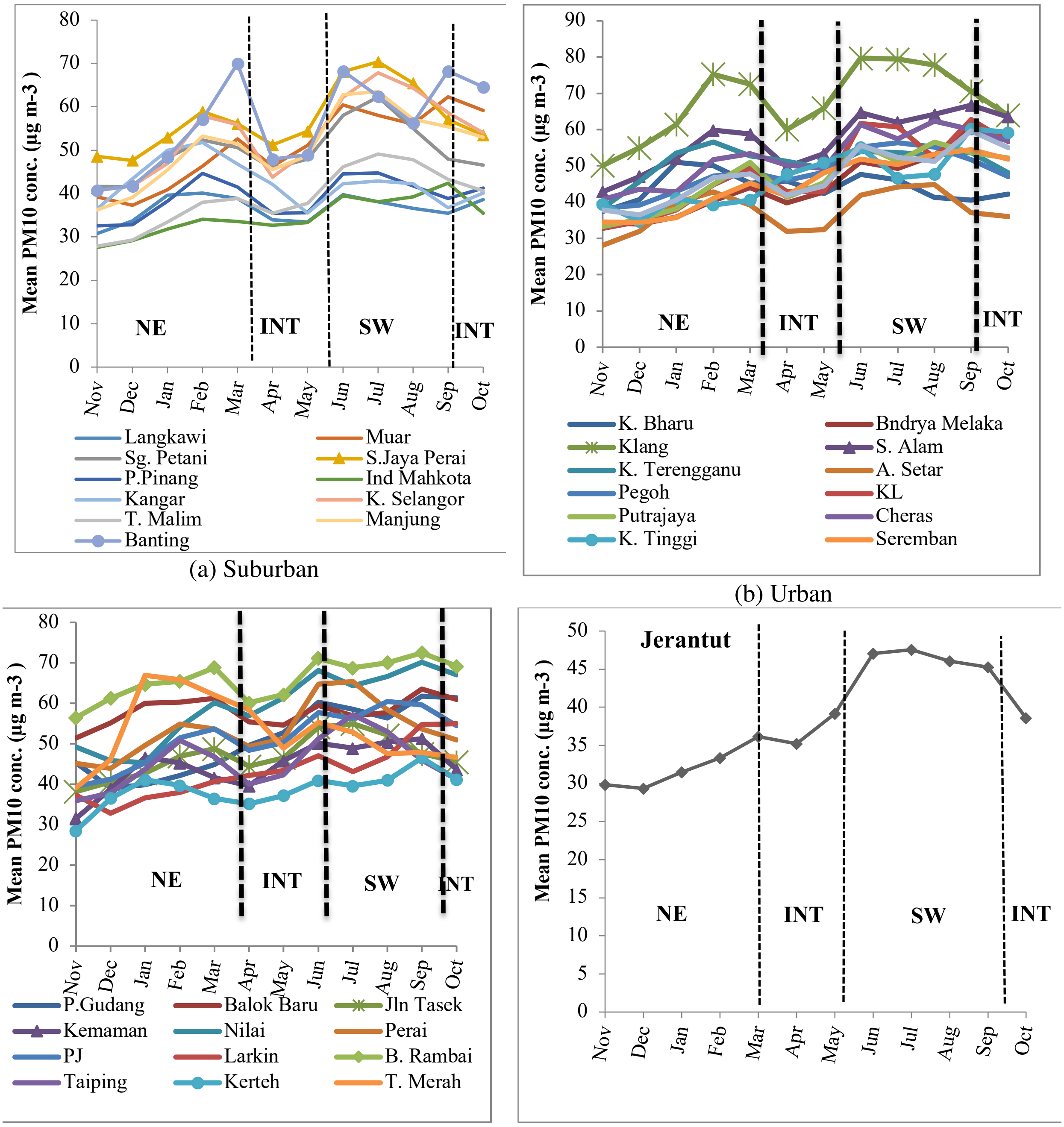


Figure 6. The trend of mean PM_{10} concentration level during different monsoon

Another data-set comprised land-use groups for the year of 2015. The land-use groups were used as a proxy for urbanization level, namely industry, infrastructure, institutional, commercial, transportation, housing, and mixed developments. They were measured in terms of the built-up area (ha.). The descriptive statistics for the land-use groups 'built-up areas (ha.) are shown in Figure 7.

The mean value for the built-up area was 7,923.8ha. Not surprisingly, Kuala Lumpur was the highest built-up area, followed by Ipoh (both Pegoh and Jln.Tasek). These locations were categorized as urban and industrial areas based on the DOE groupings for CAQM stations. Five industry locations have a built-up area of more than 5,000 ha. which also recorded high PM_{10} concentration (above the recommended limit), namely Pasir Gudang, Balok Baru, Nilai, Perai, Petaling Jaya, and Bukit Rambai.

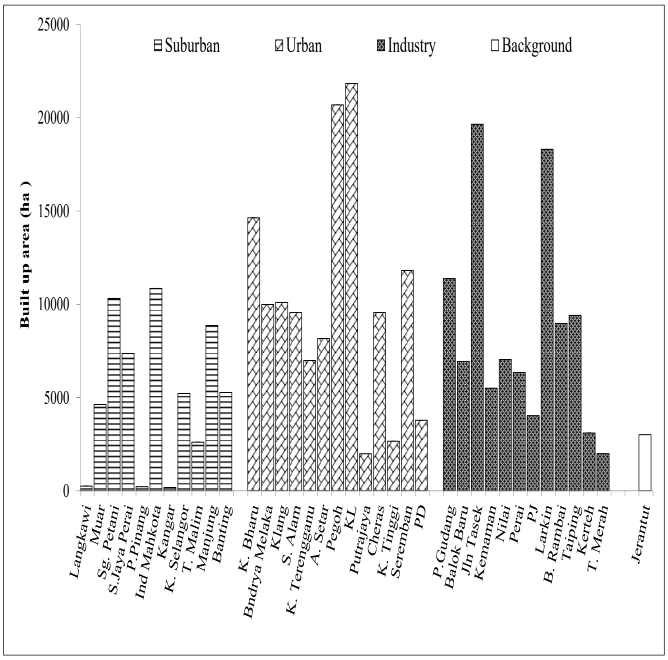
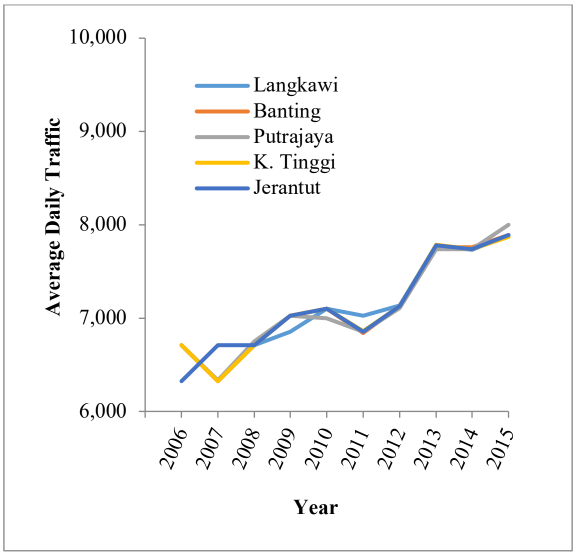
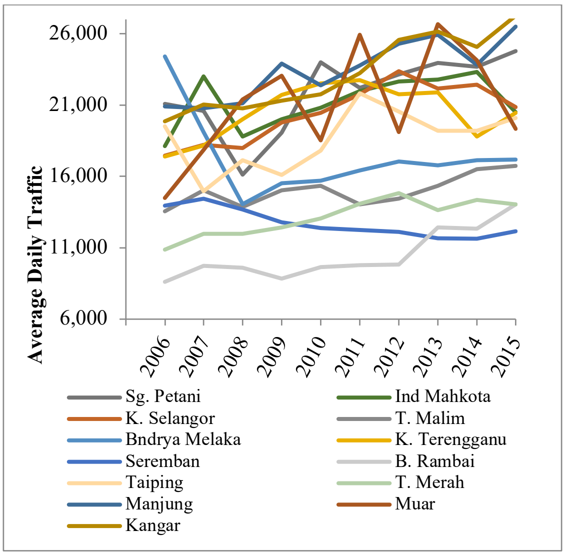


Figure 7. Built-up area (ha)

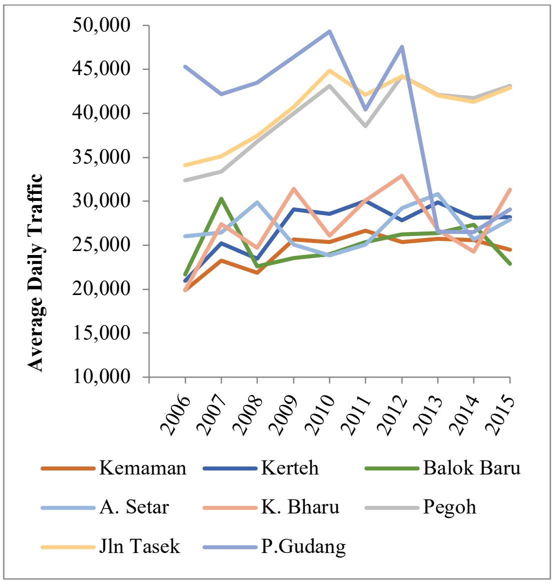
Apart from the land-use area as the potential sources for PM_{10} emission, we have also included Average Daily Traffic (ADT) to reflect traffic-related air pollutants. We analyzed the ADT data from 2006 to 2015 (see Figure 8). The results were divided into Figure (a) for ADT count less than 10,000, (b) for range ADT 10,000-25,000, (c) for range ADT 20,000-50,000 and (d) ADT above 50,000. It can be seen that Nilai recorded the highest ADT count with a value of 221,066.



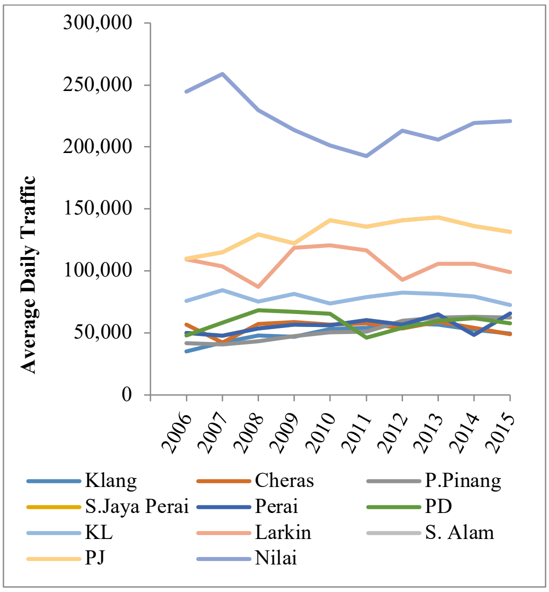
(a)



(b)



(c)



(d)

Figure 8. Trend Average Daily Traffic in 2006-2015

4.2 Determination of predictor importance

The selection of reliable predictors is crucial before the execution of any regression model. Scatterplot results indicate the relationship between two variables. Small adj R^2 values between two variables show a low correlation between them (Figures 9 and 10). Moreover, all VIF value was more modest than 7.5 by which indicates that there was no redundancy among the explanatory variables. In other words, no multi-collinearity issues exist in the data-set as a whole.

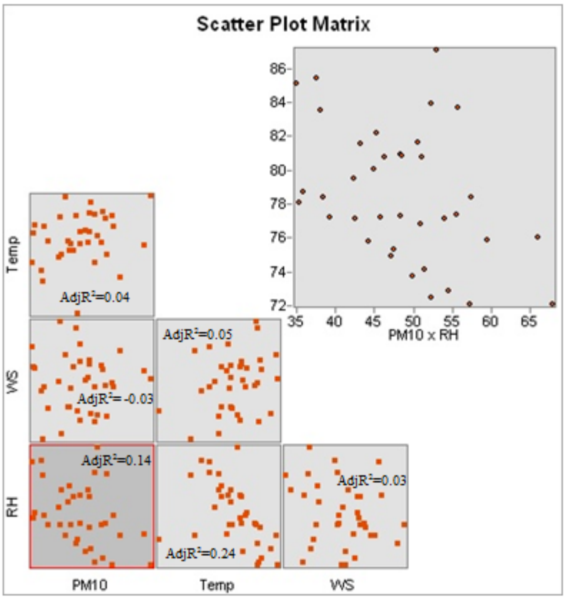


Figure 9. Scatterplot of PM_{10} and climatic parameters

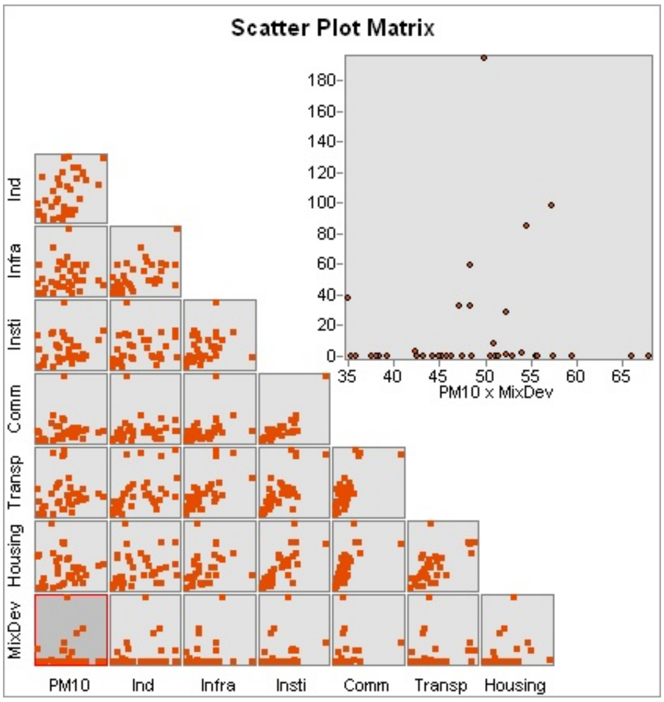


Figure 10. Scatterplot of PM_{10} and six types of land-use area

4.3 Spatial model selection

Exploratory regression in ArcGIS initially generated 15 sub-model outputs. However, there were only 11 sub-models chosen in this paper, which have the value of adjusted R^2 higher than 0.30, except for the first sub-model, as shown in Table 1. Overall, the degree of explanatory of air pollution using the selected predictor variables was rather low. Despite that, two IVs consistently appeared as significant variables, namely RH and Industry area. The last sub-model was chosen for further analysis based on the combination of the highest possible value of adj R^2 with the lowest possible AICc value (0.34 and 254.43, respectively).

Table 1. Comparison of the 11 sub-models

Variables and expected sign					$AdjR^2$	AICc
-RH*	+Builtup				0.13	259.54
-RH**	+ Ind***				0.31	251.08
-RH**	+Ind**	+ADT			0.30	252.73
-RH**	+Ind***	-Infra			0.31	252.59
-RH**	+Ind***	-House			0.31	252.59
-RH**	+Ind***	+ADT**	- Infra		0.32	253.96
-RH**	+Ind***	+Comm	-Transp		0.31	253.96
-RH**	+Ind***	–House*	+Insti		0.34	
-RH**	+Ind***	+ADT	-House*	+Instit	0.33	254.86
-RH**	+Ind***	-House	+ Insti*	-Transp	0.34	254.35
-RH**	+Ind***	-House**	+Insti***	-Infra	0.34	254.43

* = p-value < 0.10, ** = p-value < 0.05, *** = p-value < 0.01

4.4 Modeling spatial relationship

Ordinary least square (OLS) regression was performed to estimate the level of PM_{10} concentration. Generally, the output from the exploratory regression analysis was used to find the most potential suitable predictor variables. We then proceeded with a final prediction model that used only four types of land-use (Industry, Housing, Institutional, and Infrastructure) as the proxy for urbanization level. Also, based on results from the exploratory regression, only RH was included in the regression model as the meteorological influence factor. Table 2 shows the regression coefficients of the final model.

The adj. R^2 was 0.34, indicating that the selected IVs were able to explain approximately 34% of the PM_{10} variation. Regression coefficient of RH showed a negative relationship (-0.64) with PM_{10}, p-value <0.05. Only industry and institutional have a positive association with the variation of PM_{10} level; however, the coefficient values for both were less than 0.01.

Subsequently, the OLS regression output for PM_{10} was mapped using kriging interpolation, as shown in Figure 11. Generally, four areas showed a predicted concentration level of PM_{10} above 50 $\mu g/m^3$. Most of the locations on the East coast have prediction PM_{10} levels below 50 $\mu g/m^3$ except for Balok Baru and Kemaman. As expected, Selangor has the most significant area that predicted to have levels of PM_{10} higher than the limit, followed by Ipoh-Manjung and Perai-Seberang Jaya Perai.

The spatial pattern was also conducted using Getis-Ord General G and Morans I. We found that z-score value was 1.3405 and 0.2924 for Getis-Ord General G and Morans I, respectively. Both spatial autocorrelations for the OLS regression residuals resulted in an insignificant p-value (Getis-Ord General G = 0.1800 and Moran's I = 0.7699).

Table 2. OLS regression coefficient of PM_{10}

Explanatory variables	b	SEs	p-value
Constant	94.8630	22.2681	0.0001**
RH	-0.6422	0.2795	0.0285*
Industry	0.0074	0.0022	0.0024**
Infrastructure	-0.0043	0.0042	0.3213
Institutional	0.0040	0.0023	0.0914
Housing	-0.0014	0.0008	0.0860
F	4.7011		
Prob >F (5,31)	0.0026*		
Adjusted R^2	0.3395		

*p-value < 0.05, **p-value < 0.01

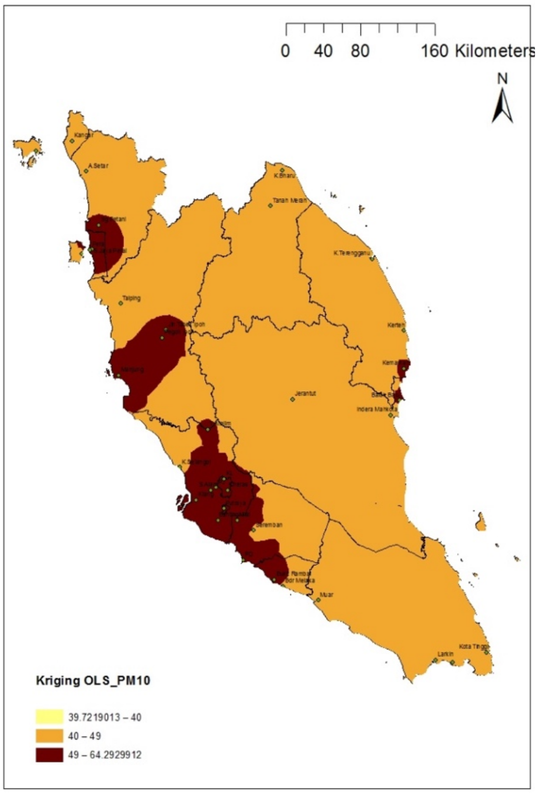


Figure 11. Kriging mapping for OLS PM_{10} prediction

5. Discussion

The 16-year average PM_{10} concentration observed based on the DOE data at 37 air quality monitoring stations, performed using GIS application, was analyzed spatially by taking into consideration the potential pollution sources, namely land-use, average daily traffic, and climatic conditions. The spatial relationship between PM_{10} and the pollution sources was analyzed in this study using exploratory regression, OLS regression, and kriging interpolations.

In general, a higher level of mean PM_{10} concentration was associated with a larger built-up area of land use. Figures 5 and 7 show that 15 CAQM stations situated on the built-up area of more than 4,000 ha have exceeded the exposure limit of RMAQG. The stations were Muar, Sungai Petani, Seberang Jaya Perai, Kuala Selangor, Manjung, Banting Klang, Shah Alam, Kuala Terengganu, Cheras, Pasir Gudang, Balok Baru, Nilai, Perai, PJ, Bukit Rambai and Tanah Merah. However, the mean PM_{10} concentration at the other seven stations located in Indera Mahkota, Kota Bharu, Pegoh, Kuala Lumpur, Seremban, Jalan Tasek, and Larkin with a built-up area of more than 10,000 ha, did not exceed the RMAQG. This suggests that other predictor variables could be counted instead of using the total size of built-up areas such as specific types of land-use areas in a particular industry area and volume of traffic. Both later predictor variables are discussed in the subsequent paragraphs.

Our results also revealed that Petaling Jaya and Tanah Merah have small built-up areas, but their mean concentration levels of PM_{10} were higher than the recommended guideline. This could be because both locations were situated in the industrial sectors, which have a variety of industrial processes potentially emitting particles [11]. On top of that, the mean concentration of PM_{10} recorded at the CAQM stations within the Klang Valley (including Klang, Shah Alam, KL, and Cheras) was also relatively high. Air pollution researcher in Malaysia usually regards the Klang Valley as a continuous hot spot location throughout Peninsular Malaysia [39–41]. The Klang Valley is considered to be a highly polluted area due to a mixture of local air pollutant emission originated from rapid development, a massive volume of traffic, increasing urban migration population added with trans-boundary pollution, in particular, recurrent haze episodes [41–43].

Moreover, a previous study has revealed that the regional and transboundary haze in 2015, El Niño, significantly affected the southern and western parts of Peninsular Malaysia [44]. We also found that the insignificant predicted higher PM_{10} concentration level using the OLS regression model also occurred mainly in the western part of Peninsular Malaysia (Figure 11). This finding was supported by the result of spatial autocorrelation for the regression residuals that was by which means the residual exhibit random spatial pattern. Hence, this study was able to produce a correctly specified model. Another potential pollution source for PM_{10} that was included in this study was the average daily traffic (ADT). As can be seen in Figure 8, ADT counts that were higher than 50,000 were also associated with a higher PM_{10} concentration level. This suggests that ADT was a possible predictor variable air pollution.

The average PM_{10} concentration was also substantially influenced by climatic conditions [45]. Our finding has also shown that a consistent trend of high concentration levels of PM_{10} was recorded during the southwest (SW) monsoon for all stations (Figure 6). The finding was supported by a previous study which stated that most days during the SW monsoon have readings exceeded the value of $50\mu g/m^3$ [46]. In Figure 6, Klang and Bukit Rambai have shown the highest readings consecutively during SW, NE, and inter (INT) monsoons. The results suggested that despite the different monsoonal seasons, the locational aspect remained as the essential PM_{10}-variation factor. Moreover, the monsoon seasons play an additional important role in intensifying the concentration levels of $PM_{10,}$ notably during the dry SW monsoon [47]. This was due to the smoke from biomass burning from regional sources during that season [46].

We also determined the importance of each of the specific predictors that comprised six individual types of land-use area and climatic parameters. They were RH, WS and temperature and Industry, Infrastructure, Institutional, Commercial, Transportation, Housing, Mixed Development, as shown in Figures 9 and 10, respectively. The importance of each potential predictor variable helped to resolve multicollinearity issues that were potentially experienced in the model development. All of the VIF values were lower than 7.5, indicating the absence of redundancy among the explanatory variables.

To find the key variables using the scatterplot matrix, the selection of the best model also was being performed using the exploratory regression tool in ArcGIS. Identifying a correctly specified OLS model is usually an iterative process; therefore, execution of the exploratory regression analysis beforehand has shortened the model building process. The output from the analysis showed that RH and Ind were the only consistent and significant variables. Both the predictor variables also have the expected sign of coefficients. Our findings revealed that RH has a meaningful negative relationship with the PM_{10} concentration level, similar to a previous study [45]. Particles tend to grow densely in high humidity settings that lead to dry deposition occurrence resulted in decreasing particle concentration in the air [48].

Table 1 also shows one of the candidate variables, namely the total built-up area, which has the lowest adjusted R^2 as compared to other variables. Specific types of land-use promote better explanation towards air pollution phenomena because of every kind of land-use associate differently with air pollutants emission. For example, industry land-use accounted for significant air pollutants emission since during processing and manufacturing process because it requires fuel burning to generate energy. By contrast, housing land-use has a lesser effect on air pollution. The use of coal-burning for home cooking is insignificant in our country, unlike in other countries such as China. China reported that most of its residents depend on coal-burning for cooking and energy production, especially during winter [49].

OLS is a global model that employs assumptions like consistent and static relationships across the whole study area [50]. Several diagnostic tests can increase the reliability of an OLS model. They include checking for the coefficients with the correct expected sign, statistically significant explanatory variables, customarily distributed residuals, and the absence of spatial autocorrelation of residuals, no redundancy among explanatory variables, and strong adjusted R^2. Table 2 shows an adjusted R^2 value of 0.34, indicating a low model's interpretative ability in the variation of PM_{10} concentration. Spatial autocorrelation of the regression residuals generated statistically insignificant z-score resulting in a random spatial pattern. It has indicated that we have a correctly specified model. An adequately defined model is needed to ensure the prediction of air pollution phenomena becomes highly trusted and less biased.

6. Conclusion

There was evidence about the geographic location as a significant attribute affecting air pollution originated from different levels of urbanization as well as types of human-made activities. Compared to the west coast, the east coast was a cleaner sub-region to live in. The overall finding of this study was confirmatory to some past studies that air pollution could be profiled on the basis of location and economic activities. Locations with a higher level of industrialization and urbanization can expect to be related to a higher level of air pollution compared to other areas. We found increasing levels of PM_{10} concentration was proportionate to the increased size of the built-up area; however, specific types of the land-use area was a better predictor than the total built-up area. Moreover, the industry area has appeared with a significant and positive association with the PM_{10} concentration level. Besides, being contributed from the industrial regions, PM_{10} variation has also fluctuated following RH. In this study, the locational influence remained as a prominent contributor to PM_{10} distribution despite different monsoonal seasons. Notwithstanding this, the highest concentration level occurred during the SW monsoon for all locations. There was also evidence of clustering PM_{10} concentration levels in this study using the regression model. Our finding revealed that the global model was able to predict a 34% variation of PM_{10} level spatially. The prediction by kriging interpolation in the geographic information system (GIS) software has identified the monitoring areas belonging to hotspots or cold spots. Therefore, the hot spots and cold spots identification should be taken into consideration by the authority and decision-makers in formulating measures to mitigate air pollution issues, minimizing the loss of air pollution data quality as well as on guard the public health concern without significant financial burden.

Acknowledgments
The authors would like to thank the Ministry of Education Malaysia for funding this research (R/FRGS/A08.00/012281/001/2016/000373). The authors gratefully acknowledge the Department of Environment and Federal Department of Town and Country Planning for providing the required data.

References

[1] Akatsu M 2015 *Economic History of Energy and Environment* p. 134
[2] Parrish DD and Stockwell WR 2015 *Eos* **96**(1):10–5
[3] Shah PS and Balkhair T 2011*Environ Int* **37**(2):498–516
[4] Šrám RJ, Binková B, Dejmek J and Bobak M 2005 *Environ Health Perspect* **113**(4):375–82
[5] Emmerechts J, Jacobs L and Hoylaerts MF 2011 *The Impact of Air Pollution on Health, Economy, Environment and Agricultural Sources*
[6] Fortoul TI, Rojas-Lemus M, Rodriguez-Lara V, Cano-Gutierrez G, Gonzalez-Villalva A, Ustarroz-Cano M, et al 2011. *The Impact of Air Pollution on Health, Economy, Environment and Agricultural Sources*
[7] Gheorghe IF and Ion B 2011 *The Impact of Air Pollution on Health, Economy, Environment and Agricultural Sources*
[8] Lin M, Chan I-N, Chan C-Y, Wang X-M and Dong H-Y 2011 *The Impact of Air Pollution on Health, Economy, Environment and Agricultural Sources*
[9] Brauer M 2016 *Poor air quality kills 5.5 million worldwide annually* p. 7–9
[10] WHO 2016 *Global Urban Ambient Air Pollution Database*
[11] Popescu F and Ionel I 2010 *Air Quality*
[12] Singh RP, Prasad AK, Chauhan SS and Singh S. 2005 *IAUC Newsl* (14):1–21
[13] Afroz R, Hassan MN and Ibrahim NA 2003 *Environ Res* **92**(2):71–7
[14] Mustafa M, Syed Abdul Kader SZ and Sufian A 2012 *Int Conf Environ Energy Biotechnol* **33**:101–5
[15] Forsyth T 2014 *Glob Environ Chang* 25(1):76–86
[16] Hayasaka H, Noguchi I, Putra EI ndr., Yulianti N and Vadrevu K *2014 Environ Pollut* **195**:257–66
[17] Fujii Y, Kawamoto H, Tohno S, Oda M, Iriana W and Lestari P 2015 *Atmos Environ* **110**:1–7
[18] Lee JSH, Jaafar Z, Tan AKJ, Carrasco LR, Ewing JJ, Bickford DP, et al 2016 *Environ Sci Policy* **55**:87–95
[19] Ku Yusof KMK, Azid A, Samsudin MS and Jamalani MA *Malaysian J Fundam Appl Sci* **13**(4):747–53
[20] Abdullah AM, Abu Samah MA and Jun TY 2012 *Open Environ Sci* **6**(1):13–9
[21] Shuhaili AFA, Ihsan SI and Faris WF 2013 *WSEAS Trans Syst* **12**(2):67–84
[22] UNECE 2006 *Handbook for the 1979 Convention on Long-range Transboundary Air Pollution and Its Protocols*
[23] UNECE 2015 *Updated Handbook for the 1979 Convention of Long-range Transboundary Air Pollution and its Protocols* **2015**;255
[24] Samet J, Wassel R, Holmes K., Abt E and Bakshi K 2005 *Environ Sci Technol* **39**(14):299A–304A
[25] National Research Council 1998 *Research Priorities for Airborne Particulate Matter I Immediate Priorities and a Long-Range Research Portfolio*
[26] Kanniah KD, Kaskaoutis DG, San Lim H, Latif MT, Kamarul Zaman NAF and Liew J 2016 *Atmos Res* **182**:302–18
[27] Geer LA 2014 *Environ Res Lett* **9**(10)
[28] Wong DW, Yuan L and Perlin SA 2004 *J Expo Anal Environ Epidemiol* **14**(5):404–15

[29] Jiang H, Li H, Yang L, Li Y, Wang W and Yan Y 2014 *J Environ Qual* **43**(6):1853–63
[30] Kheirbek I, Ito K, Neitzel R, Kim J, Johnson S, Ross Z, et al 2014 *J Urban Heal*. **91**(3):415–31
[31] Liu L, Yang X, Liu H, Wang M, Welles S, Márquez S, et al 2016 *Front Public Heal*. **4**(July):1–13
[32] Dai YH and Zhou WX 2017 *PLoS One*. **12**(8):1–24
[33] Awang MB, Jaafar AB, Abdullah AM and Ismail MB 2000 *Air quality in Malaysia : Impacts , management issues and future challenges* (**December 2001**):183–96
[34] Zou B, Peng F, Wan N, Mamady K and Wilson GJ 2014 *PLoS One*. **9**(3)
[35] Department of Statistics Malaysia 2019 *Press Release Current Population Estimates, Malaysia , 2018-2019*
[36] Ministry of Transport Malaysia 2016 *Statistik Pengangkutan Malaysia 2015*
[37] Braun MT and Oswald FL 2011 *Behav Res Methods*. **43**(2):331–9
[38] Siti Hajar Ya'acob and Iman AHM 2018 *ASM Sc J*. **2018**;(3):40–7
[39] Rahman SR, Ismail SN, Raml MF, Latif MT, Abidin EZ and Praveena SM 2015 *World Environ*. **5**(1):1–11
[40] Jamal HH, Pillay MS, Zailina H, Shamsul BS, Sinha K, Zaman Huri Z, et al. 2004 *A Study of Health Impact and Risk Assessment of Urban Air Pollution in the Klang Valley , Malaysia* **2004**;1–100
[41] Sahani M, Zainon NA, Wan Mahiyuddin WR, Latif MT, Hod R, Khan MF, et al. 2014 *Atmos Environ* **96**:257–65
[42] Latif MT, Othman M, Idris N, Juneng L, Abdullah AM, Hamzah WP, et al. 2018 *Atmos Environ* **177**:28–44
[43] Fadzil M and Rashid A 2017 *J Urban Planning, Landsc Environ Des*. **2**(2):107–27
[44] Samsuddin NAC, Khan MF, Maulud KNA, Hamid AH, Munna FT, Rahim MAA, et al. 2018 *Sci Total Environ* **630**:1502–14
[45] Amil N, Latif MT, Khan MF and Mohamad M 2015 *Atmos Chem Phys Discuss* **15**(18):26423–79
[46] Juneng L, Latif MT and Tangang F 2011 *Atmos Environ* **45**(26):4370–8
[47] Khan MF, Latif MT, Lim CH, Amil N, Jaafar SA, Dominick D, et al. 2015 *Atmos Environ*. **106**:178–90
[48] Wang J and Ogawa S 2015 *Int J Environ Res Public Heal*. **12**:9089–101
[49] Zhang Y, Tao S, Cao J and Coveney RM 2007 *Environ Sci Technol*. **41**(3):683–7
[50] Scoot LM and Janikas M V 2010 *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications* p. 27–41