



Assessment of PM_{2.5} Patterns in Malaysia Using the Clustering Method

Ezahtulsyahreen Ab. Rahman^{1,2}, Firdaus Mohamad Hamzah^{3*},
Mohd Talib Latif⁴, Doreena Dominick^{5,6}

¹Department of Chemical and Process Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

²Air Division, Department of Environment, Ministry of Environment and Water, 62574 Putrajaya, Malaysia

³Department of Engineering Education, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

⁴Department of Earth Sciences and Environment, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

⁵Centre for Atmospheric Chemistry, University of Wollongong, Wollongong, NSW 2522, Australia

⁶School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, NSW 2522, Australia

ABSTRACT

Particulate matter is the parameter of most concern in air quality monitoring in Malaysia. This study discusses the variations and clustering of PM_{2.5} recorded from 2018 to 2019 at 65 stations of the Continuous Air Quality Monitoring Network of the Malaysian Department of Environment. PM_{2.5} concentrations were recorded continuously using a tapered element oscillating microbalance. The cluster analysis was conducted using the Agglomerative Hierarchical Cluster (AHC) method. The results show that the daily average of PM_{2.5} concentrations ranged between 8 and 31 $\mu\text{g m}^{-3}$. The cluster regions were classified into High Pollution Regions (HPR), Medium Pollution Regions (MPR) and Low Pollution Regions (LPR) based on the AHC analysis. The mean concentration of PM_{2.5} recorded in HPR was significantly higher with 23.04 $\mu\text{g m}^{-3}$ followed by MPR and LPR. The results also showed that the highest concentration of PM_{2.5} was recorded during the 2019 haze episode for all three regions, with the air pollutant index indicating very unhealthy and dangerous levels.

Keywords: PM_{2.5}, Regional concentration, Agglomerative hierarchical cluster

OPEN ACCESS

Received: August 28, 2021

Revised: November 20, 2021

Accepted: December 5, 2021

* **Corresponding Author:**

fir@ukm.edu.my

Publisher:

Taiwan Association for Aerosol
Research

ISSN: 1680-8584 print

ISSN: 2071-1409 online

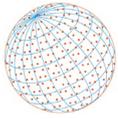
 **Copyright:** The Author(s).

This is an open access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

1 INTRODUCTION

Air pollution has been found to kill more people worldwide than other diseases such as breast cancer, malaria or tuberculosis (WHO, 2014). As described in Beelen *et al.* (2013), airborne particulate matter (PM) is especially detrimental to health and has previously been estimated to cause between three and seven million deaths every year, primarily by creating or worsening cardio-respiratory disease (Hoek *et al.*, 2013). The two main categories of particulate are fine and coarse. Coarse particulate has an aerodynamic parameter below 10 μm (PM₁₀) and fine particulate has an aerodynamic diameter below 2.5 μm (PM_{2.5}) (Shaylinda *et al.*, 2008; Wang and Ogawa, 2015). Most studies focus on PM_{2.5} due to its effects on the environment such as visibility and climate, and its ability to pass through the lungs and affect human health (Franceschi *et al.*, 2018).

Rapid development and urbanization have affected air quality and have led to an interest in studying the causes and effects of PM_{2.5}. Sinkemani *et al.* (2018) and Khalili *et al.* (2018) indicated that PM_{2.5} derives from fuel burning, vehicular exhaust, and some industrial activities, while Khan



et al. (2016) found that motor vehicle emissions, secondary inorganic aerosol, and coal-fired power plants are the predominant sources of PM_{2.5}. PM₁₀ and PM_{2.5} also originate from industrial and intensive commercial activities (Fava and Letizia Ruello, 2008). Li *et al.* (2020, 2021) suggested that airborne dust events contributed to high PM_{2.5} concentrations in Middle Eastern countries such as Kuwait, Iraq, Iran and Saudi Arabia. High concentrations of PM_{2.5} in Malaysia have been linked to Southeast Asian haze incidents, which are the consequence of the uncontrolled burning of forests in Indonesia (Rahman *et al.*, 2015). These trends show that PM_{2.5} is a critical issue requiring immediate development of regulation and policies to address the problem of PM_{2.5} globally.

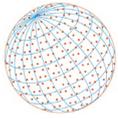
As a developing country, it is crucial for Malaysia to have a good air quality monitoring system to adopt the measurement of PM_{2.5} for its citizens. A recent improvement to the Malaysian Department of Environment (DOE) air quality monitoring network has been to incorporate continuous measurement of PM_{2.5} in the national environmental monitoring program. Standards and guidelines for PM_{2.5} were introduced in the middle of 2017. By monitoring PM_{2.5}, the actual situation concerning high particulate matter concentration due to combustion such as from biomass burning and vehicle emissions can be better represented compared with PM₁₀. Hence, the calculation of the Air Pollutant Index (API) in Malaysia has been improved with the addition of PM_{2.5} in the group of sub-index parameters. The API value is determined based on the sub-index of six types of air pollutants including ozone (O₃), carbon monoxide (CO), nitrogen dioxide (NO₂), sulfur dioxide (SO₂), and particulate matter sized under 10 μm (PM₁₀) and 2.5 μm (PM_{2.5}) (DOE, 2019). The higher the API number the more polluted the air is and the greater the health risk. The Malaysian Department of Environment (DOE) has been monitoring and reporting on air quality in Malaysia since 2014 under the Malaysia Ambient Air Quality Standard (MAAQS).

Clustering is an exploratory data analysis technique used for investigating the underlying structure in data. Since the 1980s, two well-known and widely used approaches, *k*-means and hierarchical agglomerative clustering, have been applied in air pollution research and have received a lot of attention (Govender and Sivakumar, 2020). Previous reviews on clustering applications, such as those by Gong and Richman (1995) and Jolliffe and Philipp (2010), have primarily concentrated on climate and precipitation, with a minor focus on air pollution. Given the dangers of PM exposure to humans, a better knowledge of the temporal and spatial behavior and dynamics of air pollutants is essential. Agglomerative Hierarchical Cluster (AHC) analysis is a technique for grouping things into clusters in which the objects (monitoring stations) inside a cluster are similar to each other while objects in other clusters are dissimilar (Pires *et al.*, 2008a, b). Therefore, this study aims to illustrate the overall trend of PM_{2.5} from 2018 to 2019 at the 65 monitoring stations in Malaysia based on the spatial classification of cluster analysis. The annual Air Pollutant Index (API), monthly and annual PM_{2.5} variations have also been explored in this study. The relationship between PM_{2.5} and other air pollutants and meteorological factors have been investigated according to the clustered groups.

2 MATERIALS AND METHOD

2.1 Study Area

Malaysia is situated between 2°30'N, 112°30'E, within 150 km of the Equator, in central Southeast Asia on the South China Sea. The country is comprised of Peninsular Malaysia (West Malaysia) and the states of Sabah and Sarawak (East Malaysia) on the island of Borneo. The area of Malaysia is approximately 330,000 square kilometers, the majority of which is on the island of Borneo, with Peninsular Malaysia accounting for only roughly 40% of the total. Tropical forests encompass around half of Malaysia, with the majority in Sabah and Sarawak. The Klang Valley, located in the middle of Peninsular Malaysia's west coast, is the most developed and fastest growing region (DOS, 2017a, b). Malaysia experiences a moderately uniform annual temperature which ranges from 26°C to 28°C. Malaysia has two monsoon seasons, the northeast monsoon and the southwest monsoon, which occur from November to March and from May to September. Occasionally, the northeast monsoon brings heavy rain while the southwest monsoon causes a lack of rain cloud formation resulting in less rainfall during the period. Despite the monsoons, Malaysia is safe from natural disasters such as volcanic eruptions and typhoons.



sub-index value is selected as the representative API for the hour. The data for API in this study were calculated and recorded by the Malaysian DOE. The API informs the public about air pollution levels, recommends actions and gives health advice. The index is numbered 0–500 and is divided into five breakpoints or bands providing air pollution levels in a simple way such as: 0–50 (good), 51–100 (moderate), 101–200 (unhealthy), 201–300 (very unhealthy), and > 300 (hazardous). The calculation of the API in different categories is presented in Table S2.

2.4 Quality Control and Quality Assurance

All air quality data from the Continuous Air Quality Monitoring Network (CAQM) go through QA/QC procedures before submission to the DOE. Gas detection devices are examined manually every two weeks, while PM₁₀ and PM_{2.5} instruments are calibrated once a month in accordance with standard operating protocols. Insufficient data leads to data deletion (which results in negative values), and outliers trigger second-level QC tests. Some of the observations were confirmed using observable sources, while others were ruled out due to instrument malfunction.

2.5 Statistical and Multivariate Analysis

The air pollutants analyzed in this study are those listed in the MAAQS IT-2 as shown in Table S3. The daily average was calculated using hourly PM_{2.5} data, which totaled 47,450 data points (730 per station × 65 stations). The Pearson correlation test was used to determine the correlation or linear relationship between PM_{2.5} with other pollutants and meteorological factors. If the two variables have a perfect linear relationship with a positive slope, $r = 1$. If the two variables have a perfect linear relationship with a negative slope, $r = -1$. A correlation coefficient of 0 indicates that the variables have no linear relationship (Carslaw and Ropkins, 2012).

Agglomerative Hierarchical Cluster is one of the multivariate analyses used in this study. The AHC process begins with single observation clusters and gradually joins pairs of clusters, resulting in smaller clusters with more observations (Miligan, 1980; Myatt, 2009). Ward's approach was employed in this study and is one of several well-established AHC procedures (Cunningham and Ogilvie, 1972; Johnson and Wichern, 2002). The classification of the objects can be depicted in a dendrogram, which displays the degree of similarity between them, as measured by Euclidean distances using Ward's approach (Juahir *et al.*, 2011). The quotient of the linkage distance divided by the maximal distance is represented as $[(D_{link}/D_{max})100]$. Euclidean distance is based on a single linkage (also known as a nearest neighbor). The quotient is commonly multiplied by 100 to normalize the connectivity distance indicated by the y-axis (Singh *et al.*, 2004; Shrestha and Kazama, 2007). Euclidean distance can be defined by Eq. (1) (Sharma, 1996):

$$D_{ij}^2 = \sum_{k=1}^p (x_{ik} - x_{jk})^2 \quad (1)$$

where D_{ij}^2 is the squared distance between subjects i and j , x_{ik} is the value of the k th variable for the i th subject, x_{jk} is the value of the k th variable for the j th subject and p is the number of variables.

The daily average PM_{2.5} data from the 65 monitoring stations were analyzed using AHC analysis based on the characteristics of PM_{2.5} throughout Malaysia. Further analysis and discussion will focus on the cluster formation from AHC analysis. The AHC analysis, Pearson correlation and the other statistical analysis were performed using XLSTAT 2014 add-in software developed by Addinsoft.

3 RESULTS AND DISCUSSION

3.1 Overall Variation Pattern of PM_{2.5} Concentration

Descriptive statistics of daily average of PM_{2.5} for the 65 monitoring stations from 2018 to 2019 are presented in Table S4. The minimum, maximum, first quartile, median, third quartile, and mean values between the whiskers were represented using a boxplot (Fig. 2). Overall, the trend shows that PM_{2.5} concentrations were below the standard for all stations, yet some stations

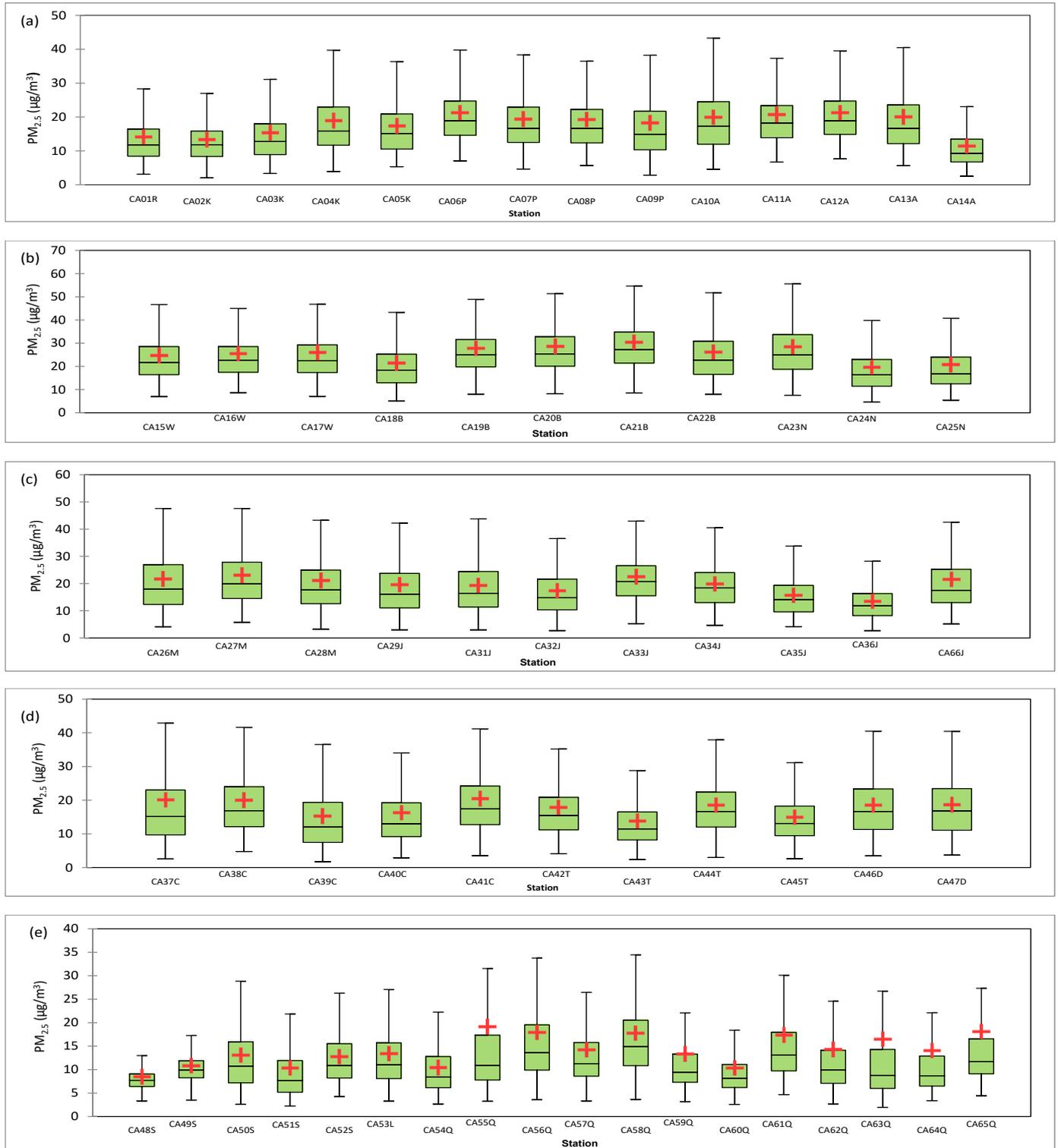
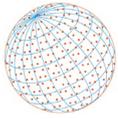
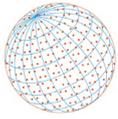


Fig. 2. Boxplot analysis on daily average concentrations of PM_{2.5} at 65 stations in Malaysia according to (a) Northern, (b) Central, (c) Southern, (d) Eastern and (e) Sabah, Sarawak and Labuan Region.

sometimes exceeded the MAAQS IT-2 of daily PM_{2.5} which is 50 µg m⁻³. The annual mean PM_{2.5} concentration recorded at all stations was between 8 and 31 µg m⁻³. The lowest concentration of 24-h averaged PM_{2.5} recorded was 2 µg m⁻³ in Jerantut, (CA39C) while the highest concentration was recorded in Sri Aman, (CA63Q) and followed by ILP Miri, (CA55Q) with 382.64 µg m⁻³ and 345.05 µg m⁻³ respectively.



Most of the stations recorded a mean value of $PM_{2.5}$ higher than the median. This shows that high concentrations of $PM_{2.5}$ influence the normal distribution of the particulate matter parameter. By chance, Malaysia was hit by a severe haze event caused by local and transboundary haze from neighboring countries in mid-September 2019 where the maximum concentration of $PM_{2.5}$ was recorded. Open burning locally and transboundary haze from Sumatra and Kalimantan, Indonesia are among the reasons for the high concentration of $PM_{2.5}$ (Latif *et al.*, 2018). This is in line with Heil and Goldammer (2001) who found that an increase in $PM_{2.5}$ caused the majority of the particle loading in the atmosphere during the smoke haze episode. This demonstrates that higher $PM_{2.5}$ affected suburban and rural areas, due to open burning during the hot season (DOE, 2019).

Compared with other ASEAN countries, the daily concentration of $PM_{2.5}$ is relatively low in Malaysia. Fold *et al.* (2020) reported that during the dry season in Thailand (November to April) between 2016–2019, the $PM_{2.5}$ 24-h standard ($50 \mu g m^{-3}$) and $PM_{2.5}$ maximum daily concentration ($100 \mu g m^{-3}$) was exceeded on approximately 50 days per year. While Kusumaningtyas *et al.* (2018) reported the daily average of $PM_{2.5}$ concentration during festivals in 2016 and 2017 ranged from $36.41 \mu g m^{-3}$ to $65.92 \mu g m^{-3}$ and from $3.66 \mu g m^{-3}$ to $59.16 \mu g m^{-3}$ in Jakarta. IQAir (2019) revealed that Indonesia had the highest annual concentration of $PM_{2.5}$ in Southeast Asia with $51.7 \mu g m^{-3}$ followed by Vietnam ($34.1 \mu g m^{-3}$), Myanmar ($31.0 \mu g m^{-3}$), Thailand ($24.3 \mu g m^{-3}$), Laos ($23.1 \mu g m^{-3}$), Cambodia ($21.1 \mu g m^{-3}$), Malaysia ($19.4 \mu g m^{-3}$), Singapore ($19.0 \mu g m^{-3}$) and the Philippines ($17.6 \mu g m^{-3}$).

3.2 $PM_{2.5}$ and Air Pollutant Index (API) Based on Clustering

In this study, the AHC method was used to cluster daily average concentrations of $PM_{2.5}$ parameter pollutants collected from the 65 monitoring stations with different backgrounds from 2018 to 2019. Three significant clusters that share the characteristic of homogeneity are shown in the dendrogram in Fig. S5. The AHC results shown in the dendrogram reveal the dissimilarity between the clusters involved. The three clusters termed High Pollution Regions (HPR), Medium Pollution Regions (MPR) and Low Pollution Regions (LPR) are shown in Table S6.

The classification of stations using AHC (HPR, MPR and LPR) based on $PM_{2.5}$ concentrations in Malaysia is shown in Fig. 3. Overall, there are 19 stations in HPR, 37 stations in MPR and nine stations in LPR. Most HPR stations are in the central region and a part of the southern region in Peninsular Malaysia, with one station being located in Sarawak. All the stations in the state of Sabah,

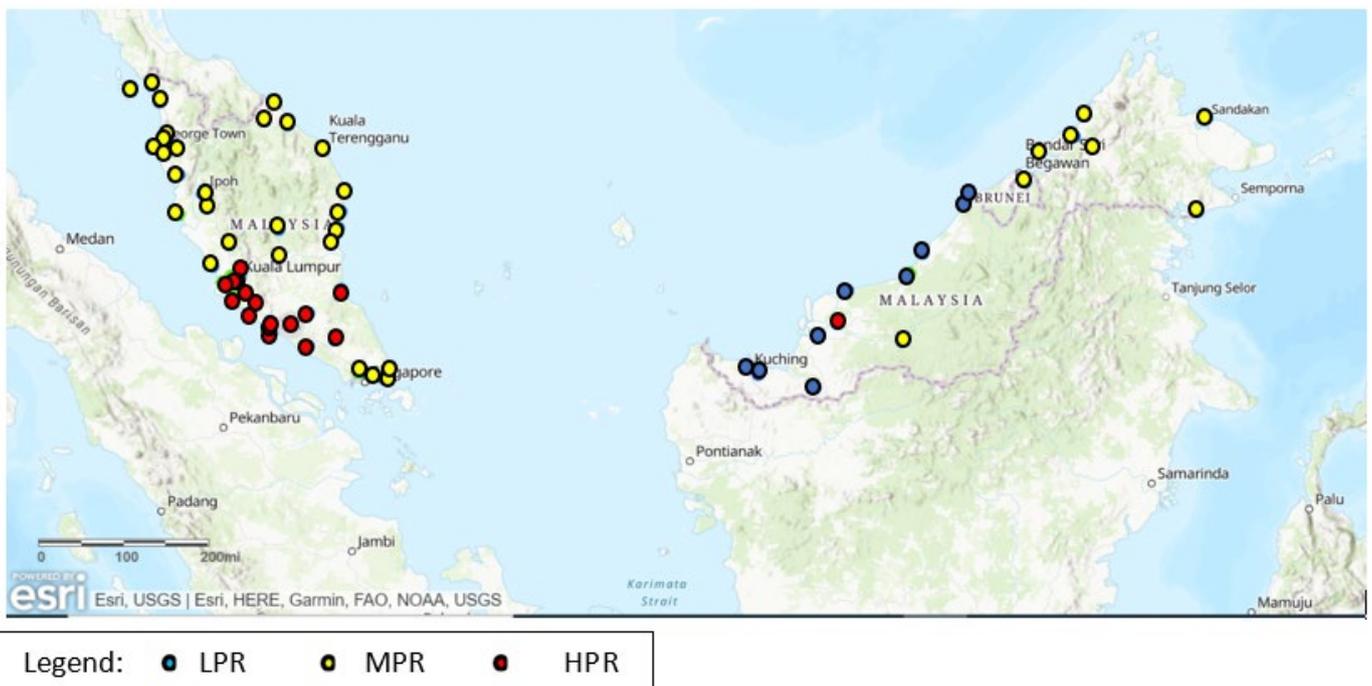
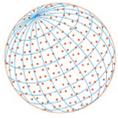


Fig. 3. Classification of stations using Agglomerative Hierarchical Cluster (AHC) based on daily average of $PM_{2.5}$ concentrations.



one station located in Sarawak, and some stations located in the east, southern and northern region of Peninsular Malaysia were classified as MPR. All the LPR stations are located in Sarawak.

Fig. S7 shows the annual average concentration of $PM_{2.5}$ from 2018 to 2019 for HPR, MPR and LPR. The trend was higher in 2019 than in 2018 due to annual haze events from Sumatera and Kalimantan, Indonesia and as well as open burning from bush fires and agricultural clearance in Malaysia. The events have affected most rural areas in LPR such as Sri Aman and ILP Miri in Sarawak leading to the annual average concentration of $PM_{2.5}$ being greater than in MPR in 2019.

The overall trend in the annual average $PM_{2.5}$ concentration in ambient air for all regions in 2018 and 2019 was within the limit of the MAAQS which is $25 \mu\text{g m}^{-3}$, except for HPR which exceeded the MAAQS in 2019 with $26.158 \mu\text{g m}^{-3}$. In general, HPR recorded the highest annual average $PM_{2.5}$ concentrations compared with MPR and LPR. The boxplots in Fig. 4 show the daily $PM_{2.5}$ average from 2018 to 2019 for the HPR, MPR and LPR clusters.

The mean value of $PM_{2.5}$ was $23.04 \mu\text{g m}^{-3}$ in HPR, and $16.41 \mu\text{g m}^{-3}$ and $16.18 \mu\text{g m}^{-3}$ in MPR and LPR respectively. $PM_{2.5}$ presented a high mean and median concentration value where the value of the mean is higher than the median in each of the regions. Table S8 shows that the distribution of $PM_{2.5}$ was skewed to the right in all regions. Positive skew suggests the presence of significant pollution levels (Abd Wahab *et al.*, 2016; Sansuddin *et al.*, 2011). In this study, LPR recorded the highest skewness value therefore LPR were the most affected during the study period due to the concentration of $PM_{2.5}$ in the atmosphere.

Air quality status for each region is shown in the API. Figs. 5 and 6 show the frequency of good, moderate, unhealthy, very unhealthy and hazardous air quality status in all three regions. It shows that moderate API levels have the highest incidence followed by a good API level from 2018 to 2019 in HPR and MPR. Meanwhile, good and moderate API levels in LPR did not show significant changes in 2018 and 2019.

Meanwhile, unhealthy API levels were recorded in 2018 and 2019 in all three regions. In 2019, unhealthy and very unhealthy API levels became dominant compared with 2018. HPR, MPR and LPR recorded 222, 342 and 111 occurrences of unhealthy levels and 10, 23 and 40 very unhealthy levels. Hazardous API levels were also observed in 2019 in LPR with 13 occurrences. The high number of occurrences of unhealthy, very unhealthy and hazardous API levels in 2019 was due to the prolonged and massive haze from Kalimantan and Sumatra, Indonesia in that year.

3.3 Comparison of $PM_{2.5}$ in HPR, MPR and LPR

The number of stations with different classifications based on HPR, MPR and LPR is shown in Table 1. Each cluster represents the stations located in industrial (I), urban (U), suburban (SU) and

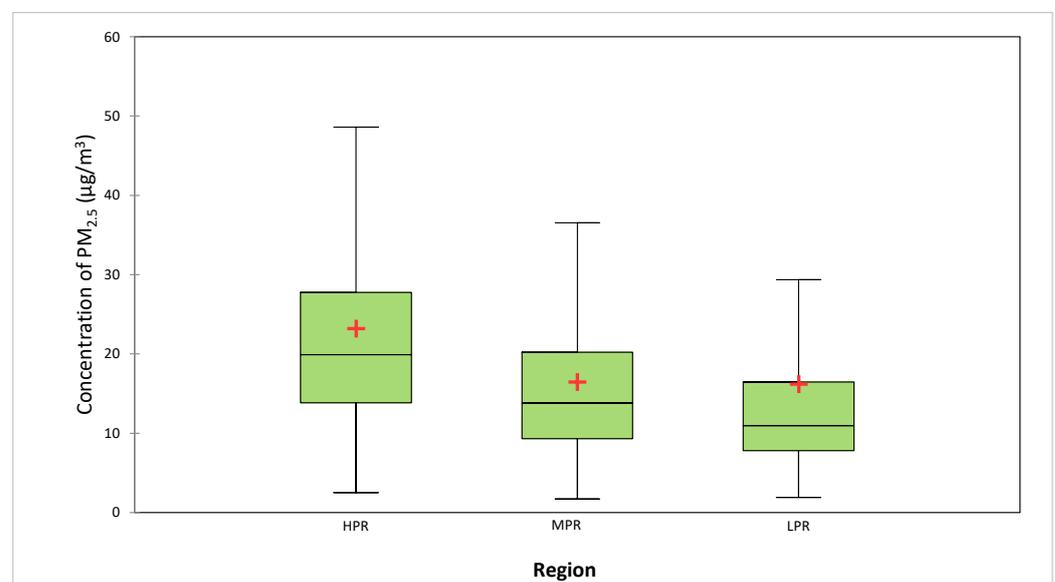


Fig. 4. Boxplot for daily average of $PM_{2.5}$ from 2018 to 2019 for High Pollution Regions (HPR), Medium Pollution Regions (MPR) and Low Pollution Regions (LPR).

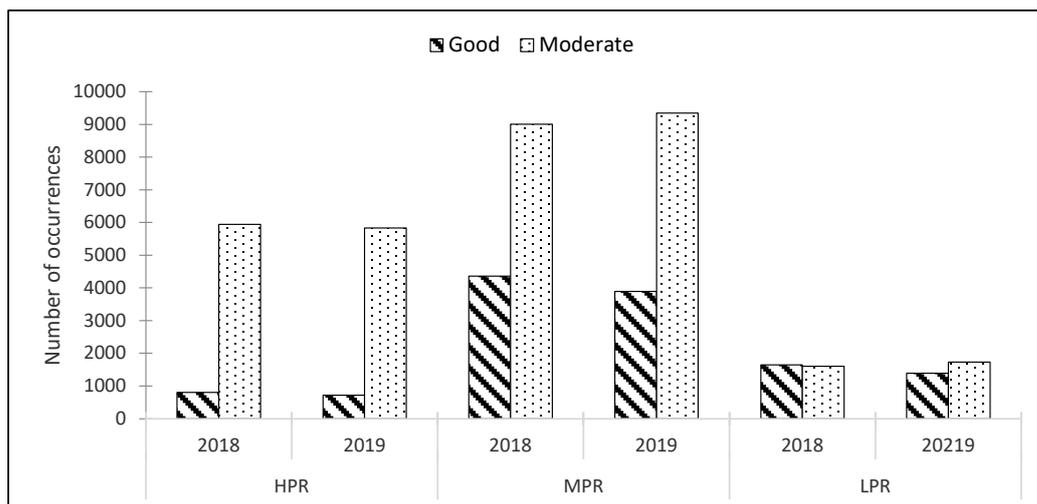
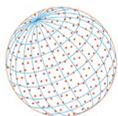


Fig. 5. Number of occurrences of Good and Moderate Air Pollutant Index for High Pollution Regions (HPR) Medium Pollution Regions (MPR) and Low Pollution Regions (LPR).

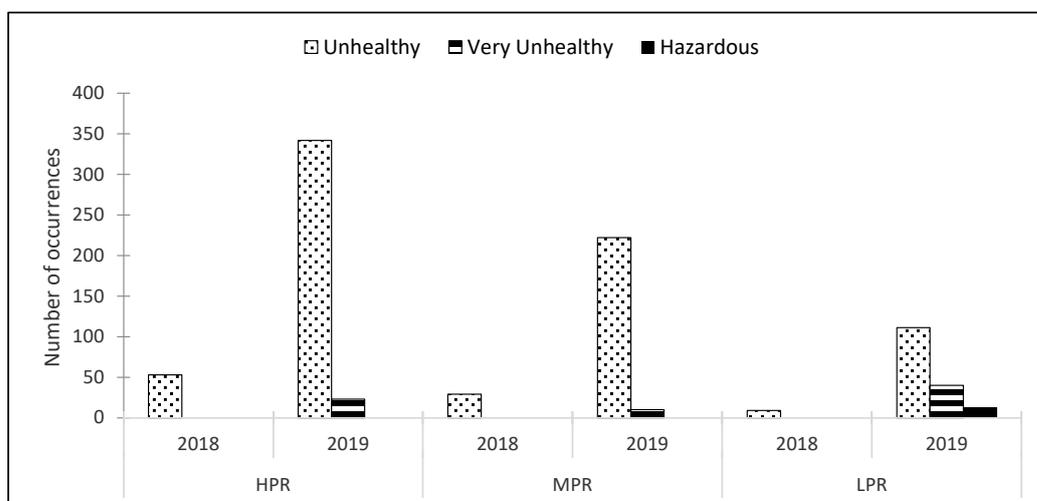


Fig. 6. Number of occurrences of Unhealthy, Very Unhealthy and Hazardous Air Pollutant Index for High Pollution Regions (HPR), Medium Pollution Regions (MPR) and Low Pollution Regions (LPR).

Table 1. Number of stations according to area based on regional classification.

Region	Classification of Stations					Total
	Industrial	Rural	Urban	Suburban	Background	
HPR	0	3	4	12	0	19
MPR	6	4	6	20	1	37
LPR	1	5	1	2	0	9

rural (R) areas while the background (B) represents in MPR. The trend of PM_{2.5} based on U, SU, I, R and B areas is shown in Fig. 7 for HPR, MPR and LPR.

There are 12 stations located in suburban areas followed by urban areas (four stations) and rural areas (three stations), and none in industrial areas in HPR. The highest number of stations is found in MPR, with 20 stations located in suburban areas, six in urban areas, six in industrial areas and four in rural areas. In contrast, there are five stations located in rural areas, two in suburban areas and one each in industrial and urban areas in LPR.

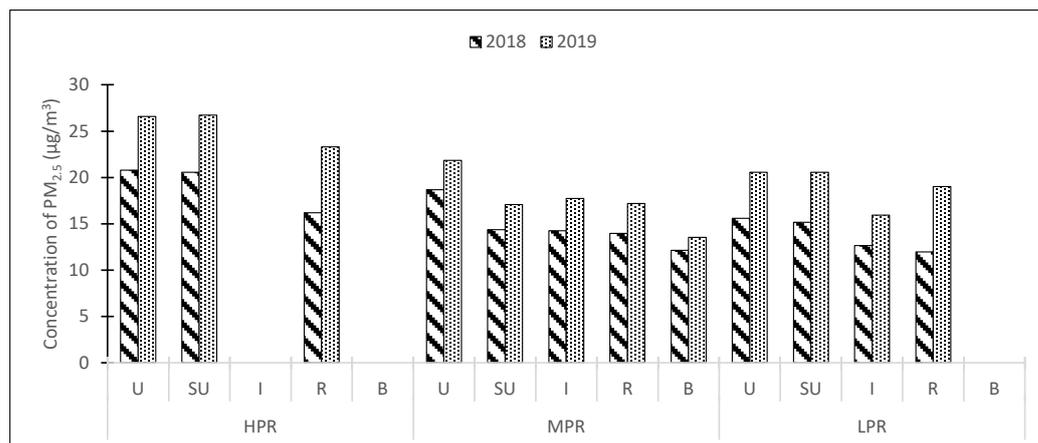
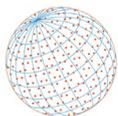


Fig. 7. Annual average concentration of PM_{2.5} for High Pollution Regions (HPR), Medium Pollution Regions (MPR) and Low Pollution Regions (LPR) according to urban (U), sub urban (SU), industrial (I), rural (R) and background (B) areas.

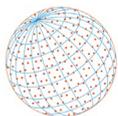
The highest annual average of PM_{2.5} in HPR was observed in urban and suburban areas in 2019 with 26.3 µg m⁻³ and 26.7 µg m⁻³ respectively, which exceeded the MAAQS. For industrial areas in MPR and LPR, the highest reading was recorded in 2019 and the same trend applies to the other areas in MPR and LPR, but the readings fall within the MAAQS limit.

Several stations located mostly in the Klang Valley namely Batu Muda, Cheras, Klang, Shah Alam and Petaling Jaya were classified as HPR. The rapid transformation of the Klang Valley into an expansive urban region during the 1990s has resulted in air pollution from motor vehicles, industrial activities and the urbanization process. Particulate matter in the ambient air is one of the main contaminants in urbanized environments (Mahapatra *et al.*, 2018). Furthermore, in most developing countries, motor vehicles are the primary mobile source of air pollution in metropolitan areas (Azmi *et al.*, 2010; Zakaria *et al.*, 2010; Ishii *et al.*, 2007).

Leh *et al.* (2012) stated that air pollutants are produced at a higher rate in urban areas, than in less developed areas and the natural environment. Therefore, many stations being situated in urban areas is a factor in their classification as HPR. Whereas forest fires in Sumatra are one of the events that promote long range transportation of haze to Malaysia during the northeast monsoon season and affect most of the suburban stations classified as HPR located along the west coast of Peninsular Malaysia in the states of Selangor, Negeri Sembilan, Melaka and Johor. While three rural stations classified as HPR recorded a high concentration of PM_{2.5} due to the local bush fires and open burning activities in peatland areas. A fire in peat soil can burn for a long time, giving it enough time to spread deep underground (Zaccone *et al.*, 2014). These prolonged peat fires can cause a hazy condition with high PM_{2.5} concentrations.

Most of the stations located in the northern, southern, and eastern areas of Peninsular Malaysia and Sabah were classified as MPR. Most MPR are suburban which vary greatly in terms of commercial and industrial development, motor vehicles and transboundary haze, which are similar conditions to those in the studies of Leh *et al.* (2012), Zakaria *et al.* (2010), Afroz *et al.* (2003) and Makmom *et al.* (2012). Power plants, industrial waste incinerators, dust from urban construction works and quarries, as well as open burning, are all sources of air pollution in suburban areas (Dominick *et al.*, 2012). Meteorology is a crucial factor influencing particles of various sizes in different ways in MPR. Regardless of the impact of changes in sources, a changing climate which alters local and global climatic factors might have an impact on particle properties (de Jesus *et al.*, 2020). Latif *et al.* (2014) concluded that wind direction was a factor in the transport of particulate matter from more urbanized and industrial areas to rural areas.

Meanwhile, most of the stations classified as LPR were less polluted due to the low PM_{2.5} concentrations in the industrial, urban and suburban areas located there. However, these areas are sometimes affected by high PM_{2.5} concentrations due to forest and bush fires, PM_{2.5} long range transportation from Kalimantan, Indonesia during forest fires, and local open burning for clearing agricultural land in rural areas during the northeast monsoon season. According to



Leewe *et al.* (2016), most forest fires are caused by human activities during prolonged dry and hot weather. Forest fires can be detected by satellites using hotspots. DOE (2019) reported a total of 520 hotspots in Malaysia and 174 in Sarawak, which was the highest recorded among Malaysian states in 2018. Frequent open burning incidents result from activities such as burning of garbage in residential areas, garbage burning by the roadside and the burning of brush and agricultural land, which normally occurs during the hot and dry period.

Besides the transboundary haze affecting air quality in Malaysia, local burning in urban, suburban and rural areas also contributes to the presence of pollutants. Motor vehicles are one of the primary contributors to air pollution in urban areas, which includes dust, suspended particulate matter, and lead (Awang *et al.*, 2000). Air pollution has a variety of sources including automobiles, industrial waste incinerators, power plants and airborne dust from construction and quarries in urban areas (Rahman *et al.*, 2015). The accelerated urbanization process has resulted in significant levels of pollution in each region. Azmi *et al.* (2010) found that extremely poor air quality occurs in heavily populated areas. However, in this study, industrial areas are not categorized as HPR based on $PM_{2.5}$, while suburban and rural areas recorded high annual concentrations of $PM_{2.5}$. Industrial areas showed good air quality in terms of low $PM_{2.5}$ concentration. This is supported by DOE (2019), which reported that there were 31 categories of industries achieving 100% compliance subject to the Environmental Quality (Clean Air) Regulation, 2014.

According to Rosofsky *et al.* (2018), inequality in exposure to pollutants recorded at the stations located in urban, suburban, industrial as well as rural areas might be due to spatiotemporal shifts in air pollution. This is due to the presence of a high background concentration of contaminants, which is common in urban, suburban, and industrial areas. It has been shown that the variety of station locations gives different readings for each region (Amran *et al.*, 2015). Therefore, the various pollutant sources present in urban, suburban, industrial and background areas promote inequitable exposure to pollutants.

Fig. 8 shows the monthly average concentration of $PM_{2.5}$ from 2018 to 2019 by region at monitoring stations in Malaysia. The highest monthly concentration of $PM_{2.5}$ is found in HPR compared with MPR and LPR, except in August where the monthly concentration of $PM_{2.5}$ is slightly lower than LPR. Meanwhile, the concentration of $PM_{2.5}$ in LPR is slightly lower than in MPR throughout the year except for July, August and September where the concentration of $PM_{2.5}$ increased drastically.

September and December had the highest and lowest monthly $PM_{2.5}$ values, respectively. The southwest monsoon occurs between May and September, whereas the northeast monsoon occurs between November and March. As a result, the monsoon event is most likely responsible for the highest and lowest monthly $PM_{2.5}$ concentrations in September and December. $PM_{2.5}$ concentrations were generally greater during the southwest monsoon (May–September) than

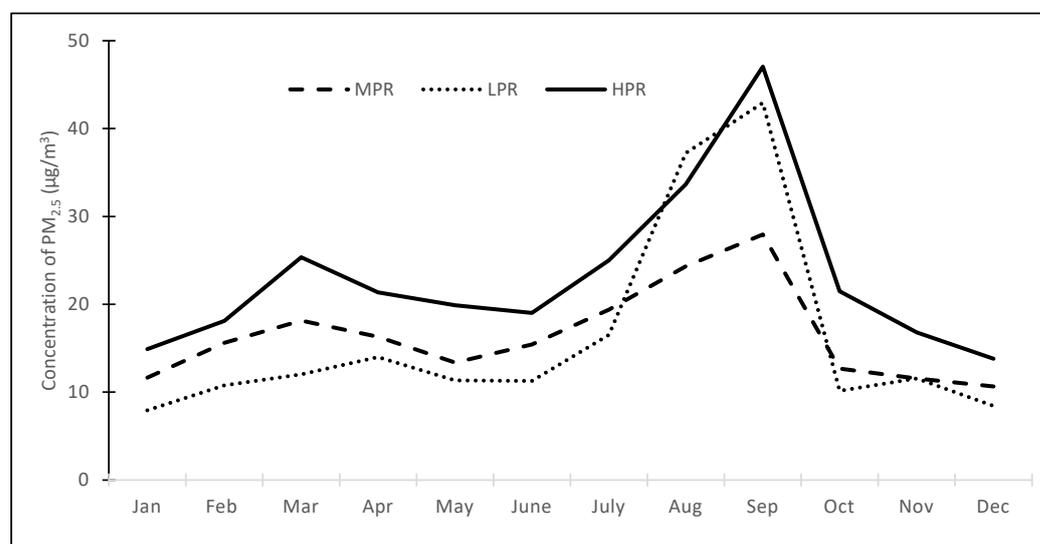
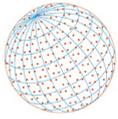


Fig. 8. Monthly average concentration of $PM_{2.5}$ from 2018 to 2019 based on region.



during the northeast monsoon (November–March) (Abdullah *et al.*, 2017). The higher PM_{2.5} concentrations during this period is mostly due to drier weather conditions, a stable atmosphere, local effects, and transboundary movement of air pollution from biomass burning in neighboring countries (Abdullah *et al.*, 2011). According to Asif *et al.* (2018), low rainfall and steady meteorological conditions caused the high PM_{2.5} concentrations, as stagnant meteorological conditions accelerate the accumulation of PM_{2.5}. Results from this study also show that PM_{2.5} levels remained quite high during the southwest monsoon.

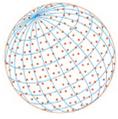
3.4 The Relationship between Air Contaminants and Weather Conditions

Correlation analysis can be used to determine the strength of a link between two variables. The relationship between PM_{2.5} and other pollutant parameters in Malaysia in HPR, MPR and LPR is shown in Table S9. CO and PM₁₀ were shown to have a strong and significant relationship with PM_{2.5} in HPR. The correlation coefficient (*r*) between PM_{2.5} and PM₁₀ is 0.981 and PM_{2.5} and CO is 0.777. Meanwhile for MPR, CO, PM₁₀ and O₃ were correlated with PM_{2.5} with the correlation as follows: PM_{2.5} and PM₁₀ (*r* = 0.987), CO (*r* = 0.786) and O₃ (*r* = 0.556). For LPR, CO, PM₁₀ and O₃ were correlated with PM_{2.5} with the correlation as follows: PM_{2.5} and PM₁₀ (*r* = 0.978), CO (*r* = 0.790) and O₃ (*r* = 0.606). The strong correlation of PM_{2.5} and PM₁₀ in HPR, MPR and LPR indicates that PM_{2.5} is significantly associated with PM₁₀. Dominick *et al.* (2012) suggested that CO is the major pollutant creating high particulate concentrations due to the combustion process in motor vehicles based on the relationship between particulate matter (PM₁₀ and PM_{2.5}) and CO. While the precursors of pollutants (NO_x and VOCs) are shown by the correlation between particulate matter (PM₁₀ and PM_{2.5}) and O₃ in MPR and LPR and connected to particulate matter from the same sources.

Table S10 presents the relationship between PM_{2.5} and meteorological parameters in Malaysia for HPR, MPR and LPR. For each region, there was a moderate to poor association between PM_{2.5} and meteorological indicators such as wind speed, temperature, and humidity. However, they remained significant and had an indirect impact on pollution concentrations. Wind speed is the primary meteorological characteristic that dilutes pollutants (Akhtar *et al.*, 2018). Meanwhile, rising temperatures aid chemical reactions that generate finely divided particulate matter (Afzali *et al.*, 2014). As a result, the concentration of PM_{2.5} is indirectly increased. Although this study showed a low correlation between PM_{2.5} and meteorological parameters, atmospheric dynamics and meteorological conditions are assumed to be an essential part of controlling air pollution. The Pearson correlation enables the determination of a relationship between PM_{2.5} and each of the meteorological parameters. In addition, a modeling technique such as a mixed-effect regression model may further explain the variation of PM_{2.5} as it allows both fixed and random effects in the model. The changes in PM_{2.5} due to other pollutants and meteorological factors could be determined using a mixed-effect regression model which provides a better understanding of the effect of meteorological parameters and pollutants on PM concentrations (Al-Hemoud *et al.*, 2021).

4 CONCLUSION

PM_{2.5} is of most concern as it is more harmful to human health than PM₁₀. In this study, the average concentration of PM_{2.5} recorded in Malaysia was between 8 µg m⁻³ to 31 µg m⁻³. Using AHC analysis, 65 monitoring stations in Malaysia were classified into HPR, MPR and LPR with 19, 9 and 37 stations respectively based on the parameter PM_{2.5}. The results show that the concentration of PM_{2.5} in HPR was the highest with the averaged concentration of 23.04 µg m⁻³, and higher concentrations were recorded in urban, suburban and rural areas compared with industrial areas in HPR and MPR. Meanwhile, LPR was most affected when extreme events occurred (haze) with the greatest skew compared with HPR and MPR. In 2019, the highest levels of PM_{2.5} were observed for HPR, MPR, and LPR with API levels recorded both unhealthy and dangerous (40 and 13 occurrences, respectively). The annual concentration of PM_{2.5} also exceeded the standard in HPR in 2019. During the southwest monsoon, PM_{2.5} levels were extremely high due to the hot and dry season. From the correlation analysis, PM₁₀ and CO were significantly correlated with PM_{2.5} while meteorological parameters had moderate and poor correlation with PM_{2.5} in all regions. For future work, other air pollutants such as SO₂, NO₂, CO and O₃ should be investigated



in detail together with PM_{2.5} concentration based on different regions. Mixed-effect regression could be considered in the future to give a better understanding of the relationship between PM_{2.5} and other pollutants and meteorological factors.

ACKNOWLEDGMENTS

The authors are thankful to the Universiti Kebangsaan Malaysia for the financial support of this research through research grant no. GUP-2020-013. The authors would also like to acknowledge their gratitude to the Department of Environment (DOE) of Malaysia for providing the air quality ambient data. This research is partly support by the Newton-Ungku Omar Fund administered by the Malaysian Industry-Government Group for High Technology (MIGHT) and Innovate UK (XX-2017-002).

DISCLAIMER

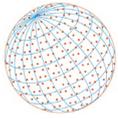
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

SUPPLEMENTARY MATERIAL

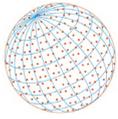
Supplementary material for this article can be found in the online version at <https://doi.org/10.4209/aaqr.210161>

REFERENCES

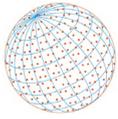
- Abd Wahab, N., Kamarudin, M.K.A., Rahim, K.A. (2016). Prediction of damage cost of bronchitis due to haze in Malaysia. *Malaysian J. Appl. Sci.* 1, 1–8. <http://orcid.org/0000-0002-1713-5679>
- Abdullah N.A., Shuhaimi, S.H., Ying, T.Y., Shapee, A.H. (2011). The study of seasonal variation of PM₁₀ concentration in Peninsular, Sabah and Sarawak. *Malaysian Meteorological Department*, 9, 1–28.
- Abdullah, S., Ismail, M., Fong, S.Y. (2017). The relationship between daily maximum temperature and daily maximum ground level ozone concentration. *Pol. J. Environ. Stud.* 26, 517–523. <https://doi.org/10.15244/pjoes/65366>
- Afroz, R., Hassan, M.N., Ibrahim, N.A. (2003). Review of air pollution and health impacts in Malaysia. *Environ. Res.* 92, 71–77. [https://doi.org/10.1016/S0013-9351\(02\)00059-2](https://doi.org/10.1016/S0013-9351(02)00059-2)
- Afzali, A., Rashid, M., Sabariah, B., Ramli, M. (2014). PM₁₀ pollution: Its prediction and meteorological influence in Pasir Gudang, Johor. *IOP Conf. Ser.: Earth Environ. Sci.* 18, 012100. <https://doi.org/10.1088/1755-1315/18/1/012100>
- Akhtar, A., Masood, S., Gupta, C., Masood, A. (2018). Prediction and Analysis of Pollution Levels in Delhi Using Multilayer Perceptron, in: Satapathy, S.C., Bhateja, V., Raju, K.S., Janakiramaiah, B. (Eds.), *Data Engineering and Intelligent Computing*, Springer, Singapore, pp. 563–572. https://doi.org/10.1007/978-981-10-3223-3_54
- Al-Hemoud, A., Al-Khayat, A., Al-Dashti, H., Li, J., Alahmad, B., Koutrakis, P. (2021). PM_{2.5} and PM₁₀ during COVID-19 lockdown in Kuwait: Mixed effect of dust and meteorological covariates. *Environ. Challenges* 5, 100215. <https://doi.org/10.1016/j.envc.2021.100215>
- Amran, M.A., Azid, A., Juahir, H., Toriman, M.E., Mustafa, A.D., Hasnam, C.N., Azaman, F., Kamarudin, M.K.A., Saudi, A.S.M., Yunus, K. (2015). Spatial analysis of the certain air pollutants using envirometric techniques. *J. Teknol.* 75, 241–249. <https://doi.org/10.11113/jt.v75.3977>
- Asif, Z., Chen, Z., Han, Y. (2018). Air quality modelling for effectiveness environmental management in the mining region. *J. Air Waste Manage. Assoc.* 68, 1001–1014. <https://doi.org/10.1080/10962247.2018.1463301>
- Awang, M., Jaafar, A.B., Abdullah, A.M., Ismail, M., Hassan, M.N., Abdullah, R., Johan, S., Noor, H. (2000). Air quality in Malaysia: impacts, management issues and future challenges. *Respirology* 5 183–196. <https://doi.org/10.1046/j.1440-1843.2000.00248.x>



- Azmi, S.Z., Latif, M.T., Ismail, A.S., Juneng, L., Jemain, A.A. (2010). Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. *Air Qual. Atmos. Health* 3, 53–64. <https://doi.org/10.1007/s11869-009-0051-1>
- Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z.J., Weinmayr, G. (2013). Effects of long-term exposure to air pollution on natural-cause mortality: An analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet* 383, 785–795. [https://doi.org/10.1016/S0140-6736\(13\)62158-3](https://doi.org/10.1016/S0140-6736(13)62158-3)
- Carslaw, D.C., Ropkins, K. (2012). Open air—An R package for air quality data analysis. *Environ. Model. Softw.* 27, 52–61. <https://doi.org/10.1016/j.envsoft.2011.09.008>
- Cunningham, K.M., Ogilvie, J.C. (1972). Evaluation of hierarchical grouping techniques: A preliminary study. *Comp. J.* 15, 209–213. <https://doi.org/10.1093/comjnl/15.3.209>
- de Jesus, A.L., Thompson, H., Knibbs, L.D., Kowalski, M., Cyrus, J., Niemi, J.V., Kousa, A., Timonen, H., Luoma, K., Petäjä, T., Beddows, D. (2020). Long-term trends in PM_{2.5} mass and particle number concentrations in urban air: The impacts of mitigation measures and extreme events due to changing climates. *Environ. Pollut.* 263, 114500. <https://doi.org/10.1016/j.envpol.2020.114500>
- Department of Statistics (DOS) (2017a). Department of statistics: Current population estimates 2017. Department of Statistics, M. Department of Statistics, Malaysia.
- Department of Statistics (DOS) (2017b). Department of statistic: Labour force survey report 2016, Malaysia. Department of Statistics, M. Department of Statistics, Malaysia.
- Department of Environment (DOE) (2019). Malaysia Annual Report. Department of Environment, Ministry of Environment and Water, Malaysia.
- Dominick D., Juahir H., Latif M.T., Zain S.M., Aris A.Z. (2012). Spatial assessment of air quality patterns in Malaysia using multivariate analysis. *Atmos. Environ.* 60, 172–181. <https://doi.org/10.1016/j.atmosenv.2012.06.021>
- Fava, G., Letizia Ruello, M. (2008). Air pollution from traffic, ships and industry in an Italian port, Presented at the Air Pollution 2008, Skiathos, Greece, pp. 271–279. <https://doi.org/10.2495/AIR080281>
- Fold, N.R., Allison, M.R., Wood, B.C., Thao, P.T., Bonnet, S., Garivait, S., Kamens, R., Pengjan, S. (2020). An assessment of annual mortality attributable to ambient PM_{2.5} in Bangkok, Thailand. *Int. J. Environ. Res. Public Health.* 17, 7298. <https://doi.org/10.3390/ijerph17197298>
- Franceschi, F., Cobo, M., Figueredo, M. (2018). Discovering relationships and forecasting PM₁₀ and PM_{2.5} concentrations in Bogotá Colombia, using artificial neural networks, principal component analysis, and k-means clustering. *Atmos. Pollut. Res.* 9, 912–922. <https://doi.org/10.1016/j.apr.2018.02.006>
- Gong, X., Richman, M.B. (1995). On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *J. Clim.* 8, 897–931. [https://doi.org/10.1175/1520-0442\(1995\)008<0897:OTAOCA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<0897:OTAOCA>2.0.CO;2)
- Govender, P., Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmos. Pollut. Res.* 11, 40–56. <https://doi.org/10.1016/j.apr.2019.09.009>
- Heil, A. and Goldammer, J. (2001). Smoke-haze pollution: A review of the 1997 episode in Southeast Asia. *Reg. Environ. Change*, 2(1), 24–37. <https://doi.org/10.1007/s101130100021>
- Hoek, G., Krishnan, R.M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., Kaufman, J.D. (2013). Long-term air pollution exposure and cardiorespiratory mortality: A review. *Environ. Health* 12, 43–45. <https://doi.org/10.1186/1476-069X-12-43>
- IQAir (2019). 2019 World Air Quality Report, Region and City PM_{2.5} Ranking. Southeast Asia. IQAir. 14. <https://www.greenpeace.org/static/planet4-thailand-stateless/2020/02/91ab34b8-2019-world-air-report.pdf>
- Ishii, S., Bell, J.N.B., Marshall, F.M. (2007). Phytotoxic risk assessment of ambient air pollution on agricultural crops in Selangor State, Malaysia. *Environ. Pollut.* 150, 267–279. <https://doi.org/10.1016/j.envpol.2007.01.012>
- Jain, A.K., Murty, M.N., Flynn, P.J. (1999). Data clustering: A review. *ACM Comput. Surveys.* 31, 264–323. <https://doi.org/10.1145/331499.331504>
- Johnson, R.A., Wichern, D.W. (2002). Applied multivariate statistical analysis. Upper Saddle River, Prentice Hall, NJ.
- Jolliffe, I.T., Philipp, A. (2010). Some recent developments in cluster analysis. *Phys. Chem. Earth.*



- A/B/C 35, 309–315. <https://doi.org/10.1016/j.pce.2009.07.014>
- Juahir, H., Zain, S.M., Yusoff, M.K., Hanidza, T.T., Armi, A.M., Toriman, M.E., Mokhtar, M. (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environ. Monit. Assess.* 173, 625–641. <https://doi.org/10.1007/s10661-010-1411-x>
- Khalili, R., Bartell, S.M., Hu, X., Liu, Y., Chang, H.H., Belanoff, C., Strickland, M.J., Vieira, V.M. (2018). Early-life exposure to PM_{2.5} and risk of acute asthma clinical encounters among children in Massachusetts: A case-crossover analysis. *Environ. Health* 17, 20. <https://doi.org/10.1186/s12940-018-0361-6>
- Khan, M.F., Sulong, N.A., Latif, M.T., Nadzir, M.S.M., Amil, N., Hussain, D.F.M., Lee, V., Hosaini, P.N., Shaharom, S., Yusoff, N.A.Y.M., Hoque, H.M.S. (2016). Comprehensive assessment of PM_{2.5} physicochemical properties during the Southeast Asia dry season (southwest monsoon). *J. Geophys. Res.* 121, 14–589. <https://doi.org/10.1002/2016JD025894>
- Kusumaningtyas, S.D.A., Aldrian, E., Wati, T., Atmoko, D., Sunaryo, S. (2018). The recent state of ambient air quality in Jakarta. *Aerosol Air Qual. Res.* 18, 2343–2354. <https://doi.org/10.4209/aaqr.2017.10.0391>
- Latif, M.T., Dominick, D., Ahamad, F., Khan, M.F., Juneng, L., Hamzah, F.M., Nadzir, M.S.M. (2014). Long term assessment of air quality from a background station on the Malaysia Peninsular. *Sci. Total Environ.* 482–483, 336–348. <https://doi.org/10.1016/j.scitotenv.2014.02.132>
- Latif, M.T., Othman, M., Idris, N., Juneng, L., Abdullah, A.M., Hamzah, W.P., Khan, M.F., Nik Sulaiman, N.M., Jewaratnam, J., Aghamohammadi, N., Sahani, M., Xiang, J.C., Ahamad, F., Amil, A., Darus, M., Varkkey, H., Tangang, F., Jaafar, A.B. (2018). Impact of regional haze towards air quality in Malaysia: A review. *Atmos. Environ.* 177, 28–44. <https://doi.org/10.1016/j.atmosenv.2018.01.002>
- Leewe, Y., Ahmad, A.N., Ismail, A., Sheriza, M.R. (2016). Analysis of hotspot pattern distribution at Sabah, Malaysia for forest fire management. *J. Environ. Sci. Technol.* 9, 291–295. <https://doi.org/10.3923/jest.2016.291.295>
- Leh, O.L.H., Ahmad, S., Aiyub, K., Jani, Y.M., Hwa, T.K. (2012). Urban air environmental health indicators for Kuala Lumpur city. *Sains Malaysiana* 41, 179–191.
- Li, J., Garshik, E., Al-Hemoud, A., Huang, S., Koutrakis, P. (2020). Impacts of meteorology and vegetation on surface dust concentrations in Middle Eastern Countries. *Sci. Total Environ.* 712, 136597. <https://doi.org/10.1016/j.scitotenv.2020.136597>
- Li, J., Garshik, E., Hart, J.E., Li, L., Shi, L., Al-Hemoud, A., Huang, S., Koutrakis, P. (2021). Estimation of ambient PM_{2.5} in Iraq and Kuwait from 2001 to 2018 using machine learning and remote sensing. *Environ. Int.* 151, 106445. <https://doi.org/10.1016/j.envint.2021.106445>
- Mahapatra, P.S., Sinha, P.R., Boopathy, R., Das, T., Mohanty, S., Sahu, S.C., Gurjar, B.R. (2018). Seasonal progression of atmospheric particulate matter over an urban coastal region in Peninsular India: Role of local meteorology and long-range transport. *Atmos. Res.* 199, 145–158. <https://doi.org/10.1016/j.atmosres.2017.09.001>
- Makmom, A., Abu Samah, A.M., Jun, Y.T. (2012). An overview of the air pollution trend in Klang Valley, Malaysia. *Open Environ. Sci.* 6, 13–19. <https://doi.org/10.2174/1876325101206010013>
- Milligan, G.W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika* 45, 325–342. <https://doi.org/10.1007/BF02293907>
- Myatt, G.J. (2009). *Making sense of data II*. Wiley, Canada.
- Pires, J., Sousa, S., Pereira, M., Alvim-Ferraz, M., Martins, F. (2008a). Management of air quality monitoring using principal component and cluster analysis—Part I: SO₂ and PM₁₀. *Atmos. Environ.* 42, 1249–1260. <https://doi.org/10.1016/j.atmosenv.2007.10.044>
- Pires, J., Sousa, S., Pereira, M., Alvim-Ferraz, M., Martins, F. (2008b). Management of air quality monitoring using principal component and cluster analysis—Part II: CO, NO₂ and O₃. *Atmos. Environ.* 42, 1261–1274. <https://doi.org/10.1016/j.atmosenv.2007.10.041>
- Rahman, S.R.A., Ismail, S.N., Ramli, M.F., Latif, M.T., Abidin, E.Z., Praveena, S.M. (2015). The assessment of ambient air pollution trend in Klang Valley, Malaysia. *World Environ.* 5, 1–11. <https://doi.org/10.5923/j.env.20150501.01>
- Rosofsky, A., Levy, J.I., Zanobetti, A., Janulewicz, P., Fabian, M.P. (2018). Temporal trends in air pollution exposure inequality in Massachusetts. *Environ. Res.* 161, 76–86. <https://doi.org/10.1016/j.envres.2017.10.028>
- Sansuddin, N., Ramli, N.A., Yahaya, A.S., Yusof, N.F.F.M., Ghazali, N.A., Al Madhoun, W.A. (2011).



- Statistical analysis of PM₁₀ concentrations at different locations in Malaysia. *Environ. Monit. Assess.* 180, 573–588. <https://doi.org/10.1007/s10661-010-1806-8>
- Sharma, S. (1996). *Applied Multivariate Techniques*. John Wiley & Son. 187.
- Shaylinda, N., Saphira, R.M., Zarina, Roshidi, M., Arif, M., Hazreek, M. (2008). Study on concentration of PM₁₀ and PM_{2.5} particulate matter in UTHM Campus by using E-Sampler. *International Conference on Environment 2008 (ICENV 2008)*, 2008, Penang, Malaysia.
- Shrestha, S., Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environ. Model. Softw.* 22, 464–475. <https://doi.org/10.1016/j.envsoft.2006.02.001>
- Singh, K.P., Malik, A., Mohan, D., Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study. *Water Res.* 38, 3980–3992. <https://doi.org/10.1016/j.watres.2004.06.011>
- Sinkemani, R., Sinkemani, A., Li, X., Chen, R. (2018). Risk of cardiovascular disease associated with the exposure of particulate matter (PM_{2.5}). *J. Environ. Prot.* 9, 607–618. <https://doi.org/10.1093/ije/dyz114>
- Wang, J., Ogawa, S. (2015). Effect of meteorological conditions on PM_{2.5} concentrations in Nagasaki, Japan. *Int. J. Environ. Res. Public Health* 12, 9089–9101. <https://doi.org/10.3390/ijerph120809089>
- World Health Organization (WHO) (2014). WHO methods and data sources for global causes of death 2000–2012. *Global Health Estimates Technical Paper WHO/HIS/HSI/GHE/2014.7*. https://www.who.int/healthinfo/global_burden_disease/GlobalCOD_method_2000_2012.pdf
- Zaccone, C., Rein, G., D'Orazio, V., Hadden, R.M., Belcher, C.M. (2014). Smouldering fire signatures in peat and their implications for palaeo environmental reconstructions. *Geochim. Cosmochim. Acta.* 137, 134–146. <https://doi.org/10.1016/j.gca.2014.04.018>
- Zakaria, J., Lye, M.S., Hashim, J.H., Hashim, Z. (2010). Allergy to air pollution and frequency of asthmatic attacks among asthmatic primary school children. *Am. Eurasian J. Toxicol. Sci.* 2, 83–92.