IOP Conf. Series: Earth and Environmental Science **140** (2018) 012028 doi:10.1088/1755-1315/140/1/012028

The classification of PM_{10} concentrations in Johor Based on Seasonal Monsoons

Hazrul Abdul Hamid[1], Muhamad Hanafi Rahmat[2], Siti Aisyah Sapani[2]
[1]Mathematics Division, School of Distance Education, Universiti Sains Malaysia, 11800, Penang
[2]Department of Design Engineering and Architecture, Faculty of Civil and Environmental Engineering, Universiti Tun Hussein Onn Malaysia, 86400, Batu Pahat, Johor

E-mail: hazrul@usm.my

Abstract. Air is the most important living resource in life. Contaminated air could adversely affect human health and the environment, especially during the monsoon season. Contamination occurs as a result of human action and haze. There are several pollutants present in the air where one of them is PM_{10}. Secondary data was obtained from the Department of Environment from 2010 until 2014 and was analyzed using the hourly average of PM_{10} concentrations. This paper examined the relation between PM_{10} concentrations and the monsoon seasons (Northeast Monsoon and Southwest Monsoon) in Larkin and Pasir Gudang. It was expected that the concentration of PM_{10} would be higher during the Southwest Monsoon as it is a dry season. The data revealed that the highest PM_{10} concentrations were recorded between 2010 to 2014 during this particular monsoon season. The characteristics of PM_{10} concentration were compared using descriptive statistics based on the monsoon seasons and classified using the hierarchical cluster analysis (Ward Methods). The annual average of PM_{10} concentration during the Southwest Monsoon had exceeded the standard set by the Malaysia Ambient Air Quality Guidelines (50 $\mu g/m^3$) while the PM_{10} concentration during the Northeast Monsoon was below the acceptable level for both stations. The dendrogram displayed showed two clusters for each monsoon season for both stations excepted for the PM_{10} concentration during the Northeast Monsoon in Larkin which was classified into three clusters due to the haze in 2010. Overall, the concentration of PM_{10} in 2013 was higher based on the clustering shown for every monsoon season at both stations according to the characteristics in the descriptive statistics.

1. Introduction

Malaysia, which is located near the equator, has good climate characteristics and receives much rain. It also experiences a uniform temperature and high humidity despite the weak wind. Although Malaysia generally experiences weaker winds, there are periodic changes in wind flow patterns and rainfall [1]. Based on these changes, there are four seasons that can be distinguished namely the Southwest Monsoon (May to September), Northeast Monsoon (November to March), and two shorter transitional periods (April to May and October to November) [2]. However, the change in seasons also affect air pollution especially during the Southwest Monsoon as it receives less rainfall. Unhealthy air quality can have a negative impact on the environment [3]. The Environmental Protection Agency (EPA) has identified six major types of pollutants including sulphur dioxide, nitrogen oxides, carbon monoxide, ozone, lead and particulate matter (PM).

The first aim of this study is to compare the characteristics of PM_{10} concentrations in Johor (Larkin and Pasir Gudang) which is located in the southern part of Malaysia during the Northeast and Southwest Monsoon. Monitoring data and ambient air quality studies show that some of the air pollutants (PM_{10}) in several large cities in Malaysia are increasing with time and do not always comply to the acceptable levels set by the Malaysian Ambient Air Quality Guidelines (MAAQG) [4] (table 1).

Johor is one of the affected states in Malaysia due to haze, resulting from forest fires in Indonesia. Hazardous Air Pollution Index (API) readings were recorded in Larkin and Pasir Gudang in June 2013 during the Southwest Monsoon and October 2010 during the Northeast Monsoon respectively [5]. The hourly average of PM_{10} concentration for each year during the monsoon seasons was analyzed using descriptive statistics.

The second aim of this study is to classify the characteristics of PM_{10} concentration using hierarchical cluster analysis. The air pollution data is usually a random variable [6] and statistical techniques are useful for classifying air pollution data. Cluster analysis is one of the most commonly used statistical techniques [7–10]. It was designed to explore the structure within a dataset [11] where different algorithms and methods of grouping objects of similar kind are sorted into respective categories.

Table 1. Malaysia Ambient Air Quality Guideline

Pollutant	Average Time	Concentration (µg/m^3)
Particulate Matter (PM_{10})	24-h	150
	1 year	50

Source: (Department of Environmental, 2015)

Due to the increasing population today, air pollution caused by construction work and, smoke released by vehicles is worsening. This causes significant harm to humans, animals, plants and the environment [12]. The pollutants can be emitted into the atmosphere through natural and anthropogenic sources [13]. The classification of the air pollutant depends on their sources, chemical composition, size and mode released into the environment [14].

Higher concentrations of PM_{10} can occur due to haze, biomass burning, industrial and vehicle emissions. This could pose a risk to human health, especially eye and throat irritation [15]. Back in 1994, Malaysia was stricken with serious haze problems. According to the Department of Environment, a haze occurs when particles present in the air are two times higher compared to normal conditions while the visibility is less than 1 kilometre (km) with air humidity measuring less than 95%. In terms of long term effects, children are more prone to mental and physical health risks when there is less than 10 µg Pb/dL of lead in blood. It can also cause disruption to the nervous system and the brain which can reduce intellectual ability, memory and cognitive ability [16].

The concentration of PM_{10} in industrial and residential areas is higher during the normal period as well as during the haze which can be influenced by the monsoon season [17]. The haze can negatively affect human health with long-term repercussions. The highest PM_{10} is expected to reach the peak during the Southwest Monsoon because it is a dry season [1]. The concentration of PM_{10} is higher in the morning, afternoon and at night at several locations in the city centre, sub-cities and suburban areas [18]. Higher concentrations of PM_{10} could contribute to the damage of respiratory organs such as lungs, heart disease, tuberculosis, and haemoptysis.

The classification using cluster analysis can be applied to many variables [9]. Cluster analysis detects objects which are in close proximity in a variable space. In the cluster, variable space contains many dimensions under consideration where there are variables. By using the sum of squares of different corresponding variable values, the distance between the cases in variable space would be defined. Based on the meteorological data, the value and composition of PM_{10} concentration are affected during the monsoon seasons [19]. The characteristics of PM_{10} concentration in ambient air are summarized using Ward's Method in clustering analysis.

2. Materials and Methods

In this study, the monitoring data was obtained from DoE Malaysia for two monitoring stations that are situated in Larkin and Pasir Gudang respectively. The hourly average of PM_{10} concentrations for 2010 until 2014 was obtained as well. Figure 1 shows the locations of the four monitoring stations in Johor but this study only focused on the monitoring stations in Larkin and Pasir Gudang. Both stations were located in an industrial area in South Peninsular Malaysia. These stations were expected to be highly polluted in terms of PM_{10} concentration due to industrialization, rapid development and economic growth accompanied by population growth that could also be affected by transboundary pollution from neighbouring countries.

2.1. Data Collection

Secondary monitoring data for PM_{10} concentrations was used in this study. The secondary data from 2010 until 2014 was obtained from the Department of Environment, Malaysia. The secondary data for PM_{10} concentrations were based on an hourly average. This secondary data was collected from air monitoring stations located in Johor. The mean top bottom method was used to replace the missing values where the data was filled with the average data available above and below the missing values [20-21].

Figure 1. Location of air quality monitoring stations in Johor
(Sources: Department of Environment, 2015)

2.2. Descriptive Statistics

Descriptive Statistics is defined as a statistic branch that includes the organization, display and description of data. Descriptive statistics contain constructions of graphs, tables, charts and the calculation of various measures such as average, percentiles and measure of dispersion [22]. Since this study focused only on monsoon seasons, the characteristics of PM_{10} were compared using descriptive statistics such as mean, median, mode, standard deviation and skewness.

2.3. Ward's Method

Ward's Method is defined as the distance between two clusters that are A and B [23] on how much the sum of squares will increase when merging them. Ward's method was determined using the following equation:

$$\Delta(A,B) = \sum_{i\in A\cup B} \|\overrightarrow{x_i} - \overrightarrow{m_{A\cup B}}\|^2 - \sum_{i\in A} \|\overrightarrow{x_i} - \overrightarrow{m_A}\|^2 - \sum_{i\in B} \|\overrightarrow{x_i} - \overrightarrow{m_B}\|^2$$

$$= \frac{n_A n_B}{n_A + n_B} \|m_A - m_B\|^2$$

IOP Conf. Series: Earth and Environmental Science **140** (2018) 012028 doi:10.1088/1755-1315/140/1/012028

Where: $\overrightarrow{m_j}$ = center of cluster j
n_j = number of point in it
Δ = merging cost of combining cluster A and B.

In hierarchical clustering, the sum of the square starts out at zero because each point is in its own cluster and then it grows into a merged cluster. Ward's Method makes this growth into a smaller cluster. If there are two pairs of clusters whose centers are far apart, Ward's method would prefer to merge the smaller ones.

Ward's Method gives a hint through the merging cost. If the merging cost increases a lot, it is probably going too far and loses a lot of structure. However, Ward's Method keeps reducing until the cost jumps and then uses the cost before the jump (merging cost take off).

4. Results and Discussion

The descriptive statistics for the hourly average of PM_{10} concentrations in Larkin and Pasir Gudang from 2010 to 2014 are shown in table 2 and table 3. The maximum concentrations of PM_{10} in Larkin and Pasir Gudang, which are also industrial areas, were 739µg/m^3 and 995µg/m^3 respectively. These readings were exceeded the Malaysian Ambient Air Quality Guidelines (MAAQG). These data were skewed to the right which means that the most readings we below the mean value except 2013 in Pasir Gudang where the standard division was more than the mean value.

The comparison was done to distinguish the characteristics of PM_{10} concentration in Larkin and Pasir Gudang between two monsoon seasons which are the Northeast Monsoon (wet season) and the Southwest Monsoon (dry season). Table 2 and table 3 show that the characteristics of PM_{10} concentration for the Southwest Monsoon were higher than the PM_{10} concentration of the Northeast Monsoon every year except for the skewness and the maximum value in 2010 which were lower than the Northeast Monsoon in Larkin. This may happen due to transboundary pollution which occurred in the neighboring country and resulted in haze in June 2013. In 2010 however, transboundary air pollution occurred in October in the southern part of Peninsular Malaysia. Thus, high level of PM_{10} were recorded in the southern part of the country including Larkin and Pasir Gudang. According to DoE Malaysia, the air quality status in Larkin and Pasir Gudang was categorised as unhealthy and hazardous based on the Malaysia Air Pollution Index scale.

Furthermore, the average air quality status during monsoon seasons was recorded below the acceptable level (50 µg/m^3) for PM_{10} concentration in accordance with Malaysian Ambient Air Quality Guideline (MAAQG) at a certain time every year, except during the Southwest Monsoon in 2011 and 2013 with readings of 54.12 µg/m^3 and 56.83 µg/m^3 respectively in Larkin and Pasir Gudang due to the dry season.

Since this study took into account both the wet season and the dry season in Johor, each set of data commenced from 16th of May until 31st of March. The clustering was classified using Ward's Method based on the characteristics in the descriptive statistics. The number of clustering was determined through the cutting from the longest length of the gap statistic methods in the dendrogram for each station.

Table 2. Characteristic of PM_{10} Concentration in Larkin based on Monsoonal Season

	2010		2011		2012		2013		2014
Monsoonal	NEM	SWM	NEM	SWM	NEM	SWM	NEM	SWM	SWM
No	3624	3312	3648	3312	3624	3312	3624	3312	3312
Mean	34.91	41.51	35.63	54.12	35.90	46.06	35.30	56.83	44.33
Median	33	39	34	51	35	43	35	43	41
Mode	27	31	32	47	37	42	37	41	39
Standard Deviation	15.40	17.94	13.76	19.92	13.86	18.95	15.17	54.28	20.30
Skewness	2.68	1.33	0.83	1.08	0.66	1.18	1.11	5.14	1.91
Minimum	5	6	5	6	7	5	5	8	7
Maximum	310	166	`118	171	101	192	174	739	190

Table 3. Characteristic of PM_{10} Concentration in Pasir Gudang based on Monsoonal Season

	2010		2011		2012		2013		2014
Monsoonal	NEM	SWM	NEM	SWM	NEM	SWM	NEM	SWM	SWM
No	3624	3312	3648	3312	3624	3312	3624	3312	3312
Mean	37.75	52.51	39.56	55.11	41.74	62.28	39.60	65.37	54.38
Median	36	43	38	49	39	58	37	51	49
Mode	36	40	38	42	37	57	33	53	49
Standard Deviation	15.63	43.49	15.77	26.59	18.93	25.31	21.51	66.75	26.37
Skewness	2.00	5.47	1.30	2.57	2.18	1.44	2.84	6.11	2.56
Minimum	5	8	7	8	5	12	5	7	11
Maximum	209	547	205	309	245	244	306	995	283

Figure 2 shows the clustering of PM_{10} concentration for the Southwest Monsoon period in Larkin. There were two clusters based on the 24 hourly average of PM_{10} concentration. The first cluster included the years 2010, 2011, 2012, and 2014. Meanwhile, the second cluster was stated in 2013. 2013 was separated from the first cluster due to transboundary pollution that resulted in the serious haze problem in Central Sumatera, Indonesia and the short Southwest Monsoon in June 2013 in the southern peninsular country. Hence, the annual average PM_{10} concentration in 2013 was 56.83 µg/m^3 which exceeded the acceptable level of Malaysia Ambient Air Quality Guideline which is 50 µg/m^3.

Figure 3 presents the clustering using Ward's method for PM_{10} concentration for the Northeast Monsoon period in Larkin. From figure 4.2, there were three clusters based on the 24 hourly averages of PM_{10} concentrations. In 2012 and 2011, the PM_{10} concentrations were merged into one cluster while the PM_{10} concentrations in 2013 and 2010 were placed in the second cluster and the third cluster respectively. Based on the characteristics in the descriptive statistics, the hourly average of the PM_{10} concentration in 2010 and 2013 had exceeded the Malaysian Ambient Air Quality Guideline (150 µg/m^3) with the maximum values recorded as 310 µg/m^3 and 174 µg/m^3 respectively. 2010 had recorded the highest maximum value due to haze in Central Sumatera, Indonesia, that contributed to transboundary pollution during the Southwest Monsoon which lasted for a short period in October.

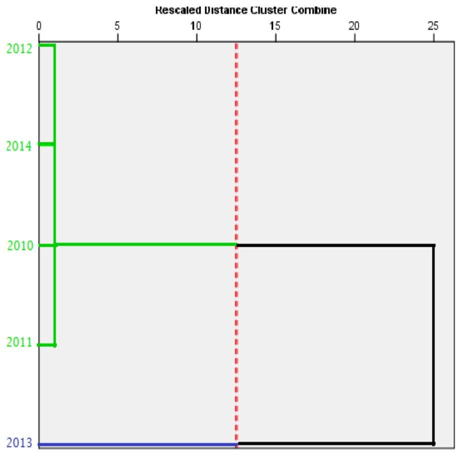

Figure 2. Dendrogram Using Ward's Method on South West Monsoon at Larkin

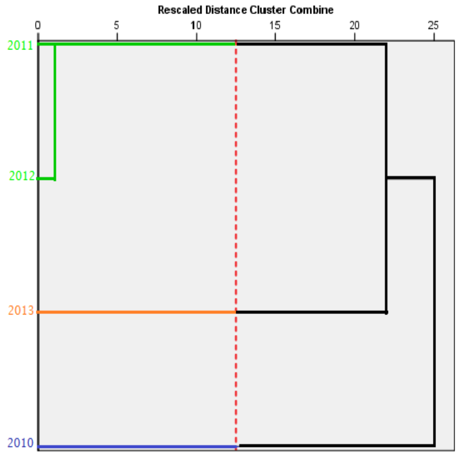

Figure 3. Dendrogram Using Ward's Method on North East Monsoon at Larkin

The hierarchical clustering using Ward's method had displayed the dendrogram of PM_{10} concentration during the Southwest Monsoon period at Pasir Gudang (figure 4). In general, there were two clusters based on the 24 hourly averages of PM_{10} concentration. The first cluster was stated in 2010 together with 2011, 2012 and 2014. The second cluster merged in 2013. From the dendrogram, 2013 had been separated from the first cluster due to the characteristics of PM_{10} in that year which recorded the highest concentrations based on the descriptive statistics. This may have happened due to the forest fires that resulted in transboundary pollution (haze problem) in Central Sumatera, Indonesia and the experience of a short Southwest Monsoon in June 2013 in the southern peninsular country. According to the information from DoE, the haze remained until September 2013.

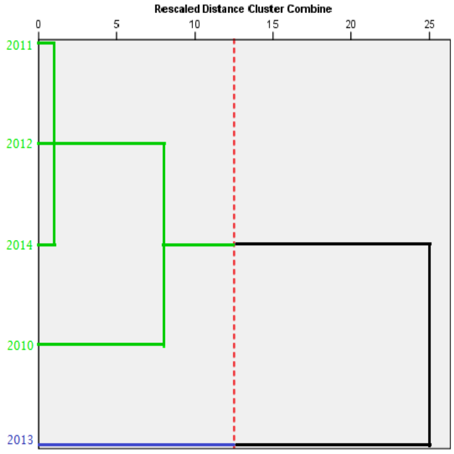

Figure 4. Dendrogram Using Ward's Method on South West Monsoon at Pasir Gudang

The dendrogram (figure 5) demonstrated the hierarchical clustering using Ward's method of PM_{10} concentration on the Northeast Monsoon season in Pasir Gudang. The dendrogram was based on two clusters of the 24 hourly average PM_{10} concentrations. The year of 2010, 2011 and 2012 were merged into one cluster while 2013 was included in the second cluster. From the overview of the dendrogram, 2013 had been separated far away from the first cluster. Referring to the characteristics of PM_{10} concentration in descriptive statistics, 2013 was spread far from the mean because the standard deviation and skewness were the largest values of the concentration. A high level of PM_{10} was recorded in the southern part of the country particularly in the state of Johor from June until September.

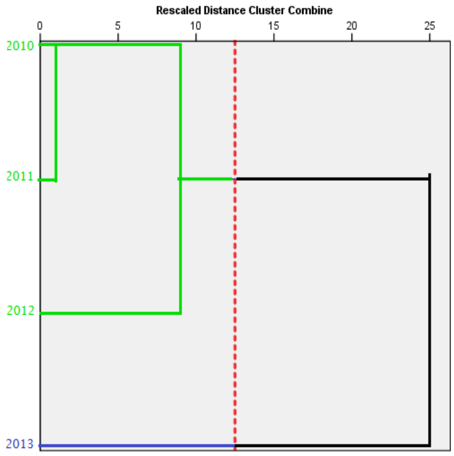

Figure 5. Dendrogram Using Ward's Method on North East Monsoon at Pasir Gudang

5. Conclusion

Based on the 24 hourly averages of PM_{10} concentration at both stations during the monsoon season, the highest mean was recorded during the Southwest Monsoon for two monitoring stations which were Larkin and Pasir Gudang as it was during the dry season. There were two clusters for every monsoon season except for Northeast Monsoon in Larkin which had been classified into three clusters. In 2013, the hierarchical clustering was separated from the other years for the Southwest and Northeast Monsoon at both stations based on the dendrogram. In other words, it means that the PM_{10} concentration was higher in 2013. This situation occurred because of the increasing of PM_{10} concentration due to transboundary pollution and the peatland fires which have resulted in serious haze problems in June 2013. High concentrations of the pollutant PM_{10} could affect human health. Thus, the government advised the people to reduce any outdoor activity during the haze until the PM_{10} concentration became stable. These monitoring stations were important for providing an assessment of the air quality as well as for comparing the level of air pollution with the surrounding area. It helped the respective parties to plan ahead for the benefit of local communities.

References

[1] Md Yusof N F, Ramli N A, Yahaya A S, Sansuddin N, Ghazali N A, and Al Madhoun W 2010 Monsoonal Differences and Probability Distribution of PM_{10} Concentration *Environ. Monit Assess* 655
[2] Meteorological Department 2008 Monsoon Ministry of Science Technology & Innovation
[3] Mohd Zin N S, Radin Mohamad R M S, Mad Ali Z, Ismail M R, Ahmad Nazri M A, and Zainal Abidin M H 2008 Study on Concentration of PM_{10} and $PM_{2.5}$ Particulate Matter in UTHM Campus by using E- SAMPLER in *International Conference on Environment* no. Icenv p 3-6
[4] Department of Environment, Malaysia 2015 *MALAYSIA Environmental Quality Report 2015* Kuala Lumpur: Department of Environment, Ministry of Science Technology and the Environment, Malaysia
[5] Department of Environment, Malaysia 2010 *MALAYSIA Environmental Quality Report 2010* Kuala Lumpur: Department of Environment, Ministry of Science Technology and the Environment, Malaysia
[6] Lu H C 2002 The Statistical Characteristics of PM_{10} Concentration in Taiwan Area *Atmos. Environ.* **36**(3) 491
[7] Giri D, Murthy V K, Adhikary P R, and Khanal S N 2007 Cluster Analysis Applied to Atmospheric PM_{10} Concentration Data for Determination of Sources and Spatial Patterns in Ambient Air-Quality of Kathmandu Valley *Curr. Sci.* **93**(5) 684
[8] Jihyun H, Meehye L, and Young S G 2008 Cluster Analysis of PM_{10} Concentration from Urban Monitoring Network in Korea during 2000 to 2005 *J. Korea Soc. Atmos. Environ. (J. KOSAE)* **24**(3) 300
[9] Saithan K and Mekparyup J 2012 Clustering of Air Quality and Meteorological Variables Associated with High Ground Ozone Concentration in the Industrial Areas, at the East of Thailand *Int. J. Pure Appl. Math.* **81**(3) 505
[10] Jorba O, Pérez C, Rocadenbosch F, and Baldasano J 2004 Cluster Analysis of 4-Day Back Trajectories Arriving in the Barcelona Area, Spain, from 1997 to 2002 *J. Appl. Meteorol.* **43**(6) 887
[11] Anderberg M R 2014 *Cluster analysis for applications* Academic Press/Elsevier Science
[12] Kampa M and Castanas E 2008 Human Health Effects of Air Pollution *Environ. Pollut.* **151**(2) 362
[13] Shah R 2016 Air Pollutants: Classification and Adverse Effects Retrieved on 01st May 2017 from http://www.biologydiscussion.com/pollution/air-pollutants/air-pollutants-classification-and-adverse-effects/16725
[14] Bernstein A J and MD 2004 Health Effects of Air Pollution *Environmental and Occupational Respiratory Disorders* **114**(5) 1116
[15] Amin N A M, Adam M B, and Aris A Z 2015 Bayesian Extreme for Modeling High PM_{10} Concentration in Johor *Procedia Environ. Sci.* **30** p 309–314
[16] Shamsul S B Zailina H Jamal H H and Ariffin O 2002 Kepekatan Partikel Ternafas (PM_{10}) dan Plumbum Ternafas dalam Udara Ambien di Kawasan Bandar dan Luar Bandar *Life Sci.* **31** 223
[17] Sansuddin N, Ramli N A, Yahaya A S, Md Yusof N F F, Ghazali N A, Al Madhoun W A 2011 Statistical Analysis of PM_{10} Concentrations at Different Locations in Malaysia *Env. Monit Assess* 573
[18] Ibrahim M H, Che Leh F, Adnan M, and Mohd Isa N K 2016 Pencemaran Habuk di Malaysia: Mengesan Taburan Konsentrasi PM_{10} di Pusat Bandar, Sub Bandar dan Pinggir Bandar di Ipoh, Perak Dust pollution in Malaysia: An Investigation of the Distribution of PM_{10} in the City Centre, Sub-Cities and Suburbs of Ipoh *Malaysian J. Soc. Sp.* **5**(5) 104

[19] Fujimori E, Kobayashi T, Aoki M, Sakaguchi M, Saito T, Fuka I T and Haraguchi H 2007Annual Variations of the Elemental Concentrations of PM_{10} in Ambient Air of Nagoya City as Determined by ICP-AES and ICP-MS *Anal. Sci.* **23**(12) 1359
[20] Hamid H A, Yahaya A S, Ramli N A, Ul-Saufie A Z and Yasin M N 2016 Short Term Prediction of PM_{10} Concentrations using Seasonal Time Series Analysis *3rd Int. Conf. Civ. Environ. Eng. Sustain. IConCEES 2015, December 1, 2015 - December 2, 2015* **47** p 1051
[21] Yahaya A S, Ramli N A and Yusof N F F M 2005 Effects of Estimating Missing Values on Fitting Distributions in *Proceeding International Conference on Quantitative Sciences and Its Application 2005 (ICOQSIA)* Penang
[22] Isotalo J 2009 Basics of Statistics Finland: University of Tamperepp pp. 82
[23] Shalizi C 2009 Distances between Clustering, Hierarchical Clustering *Data Mining* pp. 36–350