Available online freely at www.isisn.org

Bioscience Research

Print ISSN: 1811-9506 Online ISSN: 2218-3973

*Journal by Innovative Scientific Information & Services Network*

**RESEARCH ARTICLE** BIOSCIENCE RESEARCH, 2022 19(SI-1): 126-134. **OPEN ACCESS**

# Air Quality Pattern in Central Region of Malaysia using Multivariate Technique: A New Approach

**Nurul Alia Azizan[1], Azman Azid[2*], Asheila AK Meramat[1], Ahmad Syibli Othman[1], and Siti Noor Syuhada Muhammad Amin[3]**

[1]School of Biomedical Science, Faculty of Health Sciences, Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Kuala Nerus, Terengganu, **Malaysia**
[2]School of Animal Science, Aquatic Science and Environment, Faculty of Bioresources and Food Industry, Universiti Sultan Zainal Abidin, Kampus Besut, 22200 Besut, Terengganu, **Malaysia**
[3]Universiti Sultan Zainal Abidin Science and Medicine Foundation Centre, Gong Badak Campus, 21300 Kuala Nerus, Terengganu, **Malaysia**

***Correspondence:** azmanazid@unisza.edu.my **Received 25-07-2022, Revised: 31-08-2022, Accepted: 01-09-2022 e-Published: 20-10-2022**

Assessment of air quality patterns require multiple variables to be analysed. It becomes a multivariate problem that requires novel methodologies as agglomerative hierarchical cluster analysis (AHC) is used primarily to determine cluster patterns. Therefore, this study aimed to demonstrate a new approach to cluster analysis in air quality monitoring stations in the central region of Peninsular Malaysia to provide a better spatial cluster distribution. The data sets was obtained by the Department of Environment (DOE) for the years 2018 to 2019. Six air quality pollutants were involved in this study. Before clustering the data, multivariate techniques such as principal component analysis (PCA) were used to summarize the information to better understand of the variables by reducing dimensionality. The AHC was then created using the PCA factor scores. The factor scores were employed in a discriminant analysis (DA) to verify the clusters. PCA factor scores revealed that 10 of 11 stations needed to be further investigated using AHC. High Polluted Region (HPR=five stations) and Low Polluted Region (LPR=five stations) were established from AHC. Each class was distinguished using discriminant analysis (DA). The researchers discovered that each class has its own set of variables. According to the DA, confusion matrix for the clusters is 94.58 %. The framework provided here offers a new approach for identifying and classifying stations based on air quality variables. This demonstrates that the hybrid cluster analysis technique utilized in this study is capable of producing more precise pollutant distributions.

**Keywords: air quality, multivariate technique, a new approach**

## INTRODUCTION

The central region of Malaysia consisting of Kuala Lumpur, Selangor, Putrajaya and Negeri Sembilan is known as Malaysia's main economic hub and has undergone extensive physical development in terms of infrastructure, industrialization and urbanisation, all of which have significantly worsened the country's air quality (Zakaria et al. 2017). Pollutants in this area come from a wide range of sources, including transboundary haze, motor vehicles, and commercial and industrial development (Afroz et al. 2003; Zakaria et al. 2017). Typically, mobile sources especially in urban areas are the main causes of air pollution in most developing nations (Mohamad et al. 2015). Due to an increase in vehicles and industry, most emerging nations produce a lot of these harmful pollutants such as particulate matter ($PM_{10}$), $SO_2$, $CO_2$ and $NO_2$ are the primary pollutants contributing to the deterioration of air pollution in Klang Valley (Azmi et al. 2010; Jamhari et al. 2014; Mohamad et al. 2015; Sahrir et al. 2019). Rising populations, rapid industrialisation and socioeconomic activities involving heavy traffic with motor vehicles, urbanisation or population growth may indirectly be one of the potential causes of rising air pollution. According to previous research, CO was the main cause of ambient air pollution in the Klang Valley. It was mostly produced by incomplete combustion of the fuel in motor vehicles (Dominick et al. 2012). As a result, urban areas in Malaysia's central region have been the subject of the majority of studies on air quality since they physically exhibit a large number of population, transportation, industrial, commercial, and residential variables (Azmi et al. 2010; Sahrir et al. 2019).

According to the Department of Statistics (DOS, 2020), most states in Malaysia recorded rapid growth in the economy that indirectly contributed to the degradation of the urban environment and air quality. Low air quality is one of the world's largest environmental and health problems with an increasing number of acute air pollution episodes in many cities worldwide. It affects all regions, settings, socioeconomic groups and age groups (WHO, 2021). Scientific studies have shown that some pollutants can harm public health and welfare even at very low levels and affect economies and people's quality of life. The main causes of air quality deterioration could be contributed emission from traffics and industries as well as volcanic emission and forest fires as transboundary pollution (Sahrir et al. 2019). According to the World Health Organization (WHO, 2018), air pollution was identified as the second largest risk factor for non-communicable disease. Several studies have reported adverse effects of air pollution on allergic and lung inflammation reactions (Isa et al. 2020) and diabetes (Wong et al. 2020). Exposure to air pollutants may cause bronchitis and respiratory complications (Arifuddin, 2019; Kamaruddin et al. 2019; Tellez-Rojo et al. 2020), asthma (Zainal-Abidin et al. 2014; Zakaria et al. 2012) related to genotoxicity due to air pollutants (Hisamuddin et al. 2022). In recent years there has been an emerging body of evidence of epidemiological studies linking air pollution exposure to neurological disorders (Shehab and Pope, 2019) and cognitive functions in children and the elderly (Chen et al. 2021; Cheng et al. 2014). Studies on air pollution exposure to cardiovascular and cerebrovascular diseases suggest a harmful impact on the brain and cognitive processes through vascular and inflammatory mechanisms (Peters et al. 2015). A previous study (Bell et al. 2014) suggested that controlling some of the sources of the pollutant could protect public health more efficiently than the regulation of particle concentration. Thus, the possible reduction in health risks from the predominant sources of air pollutants is desired as part of the mitigation strategy. Besides, this study engages and strategizes environmentally sustainable policy making.

As a developing country, Malaysia must have a good air quality monitoring system as DOE has organized programs on air pollution in Malaysia. However, statistical tools such as multivariate techniques have not been properly applied for pattern recognition in the extensive analyses of air quality data. The multivariate technique also known as chemometric analysis is one of the most recent and reliable statistical tools used by researchers to analyse enormous amounts of data. It is based on the statistical principle, which entails simultaneously observing and analysing multiple variables while keeping the process to a manageable level. These techniques are the best approaches to apply to a large amount of complex environmental monitoring data because they can avoid misinterpretation of results produced during data analysis (Shafii et al. 2019). It has been proven to be a better tool for analysing air quality than traditional statistical methods such as spatial variations which provides an understanding of the important patterns and underlying relationships in data, for contamination sources identification, data reduction and interpretation (Azid et al. 2015). These strategies aid in the reduction of database complexity, allowing for a better understanding and interpretation of air quality data, as well as effective management of air quality monitoring programmes. Many scientific research (Rahman et al. 2022; Dominick et al. 2012; Zakri et al. 2018) have used various environmental approaches such as principal component analysis (PCA), and agglomerative hierarchical cluster analysis (AHC) and discriminant analysis (DA), particularly in air quality monitoring. The use of these methods for interpreting complicated databases allows for a better knowledge of the air quality in the research region, as well as the development of appropriate plans for the effective administration of air quality monitoring programmes (Azid et al. 2014).

In this study, the new approach clustering method (PCA factor score, AHC) were applied to group sites across the central region in Peninsular Malaysia based on air quality pollutants using data collected between 2018 and 2019. The objectives of the study are to find out whether this hybrid cluster technique will allow for a better understanding of the heterogeneity in air quality pollutants. This demonstrates that the analysis technique utilised in this study is capable of producing more precise pollutant distributions that are useful in air pollution studies. In this study, multivariate analysis was performed to classify and interpret large data sets. We hope that the identified clusters can be used to further investigate the heterogeneity in the relationship between air pollutants concentration of the sampling sites and morbidity across Peninsular Malaysia.

## MATERIALS AND METHODS

### Study area

All 11 air quality monitoring sites located in the central region of Peninsular Malaysia were selected to give a general representation of the air quality status in Malaysia. These monitoring stations are under the supervision and control of a private company (Pakar Scieno TW Sdn. Bhd.) on behalf of the Department of Environment Malaysia (DOE). Furthermore, the location of the study area which involved four states based on their latitudes and longitudes are shown in Table 1.

### Frame of data

The central region of Malaysia's air quality was measured at 11 stations located throughout the country to continuously monitor (Table1) and detect any significant change in air quality that could harm human health and the environment. Two years of data from January 1st, 2018 to December 31st, 2019 for air pollutants ($PM_{10}$, $PM_{2.5}$, $SO_2$, $NO_2$, $O_3$ and CO) and meteorological parameters (wind direction, wind speed, temperature, humidity, solar radiation) were obtained from the Air Quality Division, Department of Environment (DOE), monitored and collected by the DOE authorised agency named Pakar Scieno TW Sdn. Bhd. All data were obtained from a monthly average that was established from the hourly monitoring sites. Total data sets were 6780 at 11 air monitoring stations in four states.

**Table 1. Location of the study area (central region) based on latitude and longitude**

| States | Station ID | Air monitoring stations | Longitude and latitude |
|---|---|---|---|
| Kuala Lumpur | CA0015 | Sek. Keb. Batu Muda, Batu Muda | 03° 12' 44.78" N 101° 40' 56.02" E |
| | CA0016 | Sek. Men. Keb. Seri Permaisuri, Cheras | 03° 06' 22.44" N 101° 43' 04.50" E |
| Putrajaya | CA0017 | Sek. Keb. Presint 18, Putrajaya | 02° 54' 53.33" N 101° 41' 24.17" E |
| Selangor | CA0018 | Sek. Men. Sains Selangor, Kuala Selangor | 03° 19' 16.70" N 101° 15' 22.47" E |
| | CA0019 | Sek. Keb. Bandar Utama, Petaling Jaya | 03° 07' 59.40" N 101° 36' 28.83" E |
| | CA0020 | Sek. Keb. TTDI Jaya, Shah Alam | 03° 06' 16.98" N 101° 33' 22.39" E |
| | CA0021 | Klinik Kesihatan Pandamaran, Klang | 03° 00' 53.60" N 101° 24' 47.19" E |
| | CA0022 | Kolej Mara Banting, Banting | 02° 49' 00.08" N 101° 37' 23.36" E |
| Negeri Sembilan | CA0023 | Sek. Keb. Taman Semarak (Fasa 2), Nilai | 02° 49' 18.09" N 101° 48' 41.34" E |
| | CA0024 | Sek. Men. Teknik Tuanku Jaafar, Seremban | 02° 43' 24.17" N 101° 58' 06.58" |
| | CA0025 | Pusat Sumber Pendidikan N. Sembilan, Port Dickson | 02° 26' 28.97" N 101° 52' 00.68" E |

Initially, there were eight parameters, which consisted of air pollutant variables that were collected to be used in this study. However, only six out of eight parameters were selected for further analysis due to the high percentage of missing data for two parameters, namely NO and NOx. The data produced by Pakar Scieno TW Sdn. Bhd will be examined and validated by the DOE as an authority before giving it to the stakeholders.

### Multivariate analysis method

This study identified three methods using multivariate analysis for achieving the objective. There were principal component analysis (PCA), agglomerative hierarchical cluster analysis (AHC) and discriminant analysis (DA) via XLSTAT software (XLSTAT, 2019, Addinsoft, New York, NY, USA). Using this multivariate approach, we were able to determine the selected stations and find a pattern.

PCA is a statistical approach that allows to summarise the content of big data tables using a smaller number of "summary indices" that can be displayed and analysed more readily (Azid et al. 2015). The links between observations and variables, as well as among the variables, may be revealed by this overview (Shafii et al. 2019). The most common application of PCA is to represent a multivariate data table as a smaller number of variables (summary indices). PCA is used in this study before clustering the data to get a better picture of the variables. In practice, it is thought to improve clustering results (noise reduction). This strategy was used to explain the variation of a large number of connected variables by translating them into a smaller number of uncorrelated (independent) variables known as principal components (PCs) (Dominick et al. 2012).

AHC is an unsupervised statistical strategy for grouping or clustering observations based on their similarities or differences. A dendrogram that evaluates the degree of risk homogeneity using Ward's approach and Euclidean distance measurement can be used to demonstrate this spatial classification of air quality monitoring stations (Jamalani et al. 2016). The quotient between the linkage distance divided by the maximal distance [($D_{link}/D_{max}$)], which is multiplied by 100 to standardise the linkage distance indicated by the y-axis, is used to calculate Euclidean distance (Hamza Ahmad & Azman, 2015)

In this study, DA was used on data for spatial analysis in the two clusters developed by AHC using standard mode, backward stepwise mode and forward stepwise mode to see if the groups differed in terms of the variable's mean (Sahrir et al. 2019). DA also has been used to predict group membership based on another studies (Azid et al. 2017). Clusters 1 and 2 were chosen as independent variables, whereas air pollutants ($PM_{10}$, $PM_{2.5}$, $SO_2$, $NO_2$, $O_3$ and CO) were chosen as dependent variables. In the forward stepwise, variables were added one by one until no significant changes were noticed, but in the reverse stepwise mode, variables were removed one by one, starting with the least significant variable until no significant changes were observed. All analyses were summarized in Figure 1.

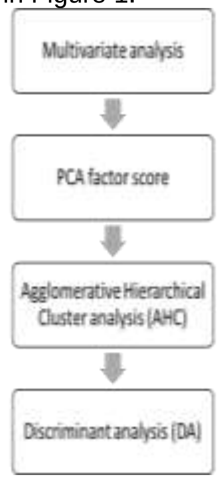


**Figure 1. Flowchart of the statistical and multivariate analysis**

## RESULTS AND DISCUSSION

### Air quality pattern using the hybrid clustering method (PCA factor score, AHC)

There are 11 air monitoring stations in the central region of Peninsular Malaysia and six variables are measured to be calculated as API. DOE is categorized according to urban, suburban, industrial and rural settings to easily monitor any implications for human health and the environment. In this study, six air pollutants have been used in PCA as independent variables and 11 air monitoring stations as dependent variables. The result showed that all six variables in Table 2 have a significant correlation toward air quality monitoring stations. $PM_{10}$ and $PM_{2.5}$ showed significantly positively correlation to each other as compared to other pollutants.

As all six variables were found to be significant in air quality, the PCA factor score was able to summarise all 11 stations with six variables into ten stations as the significant observation outcome. PCA was generally used to identify potential sources of variation (Hua, 2018; Zakri et al. 2018; Sahrir et al. 2019). The additional variables known as principal components (PCs) were subsequently investigated using PCA to identify emission sources. On the other hand, in this work, PCA was used to analyse data in which observations were described by several interconnected quantitative dependent variables. Through PCA analysis, the most important information is extracted and presented in a new space as a set of linear and orthogonal variables called principal components (sometimes called factors) (F1+F2....+Fn),

**Table 2. Correlation matrix (Pearson (n))**

| Variables | $PM_{10}$ MAX µg $m^{-3}$ | $PM_{2.5}$ MAX µg $m^{-3}$ | $SO_2$ MAX ppm | $NO_2$ MAX ppm | $O_3$ MAX ppm | CO MAX ppm |
|---|---|---|---|---|---|---|
| $PM_{10}$ MAX µg m-3 | **1** | **0.99** | 0.08 | **0.17** | -0.08 | **0.33** |
| $PM_{2.5}$ MAX µg m-3 | **0.99** | **1** | 0.08 | **0.16** | -0.07 | **0.34** |
| $SO_2$ MAX ppm | 0.08 | 0.08 | **1** | 0.04 | 0.02 | 0.10 |
| $NO_2$ MAX ppm | **0.17** | **0.16** | 0.04 | **1** | 0.10 | **0.66** |
| $O_3$ MAX ppm | -0.08 | -0.07 | 0.02 | 0.10 | **1** | -0.10 |
| CO MAX ppm | **0.33** | **0.34** | 0.10 | **0.66** | -0.01 | **1** |

**Values in bold are different from 0 with a significance level alpha=0.05*

where n is the total number of variables. A geometric projection known as a factor score or factor loading is used to depict each variable or observation on each primary component. PCA associations can be visualised using a variety of tools, including correlation plots, scree plots, and bi-plots. Variables or observations with bigger squared cosines have a considerable impact on a component, as well as on total variability (Kamalha and Omollo, 2017). We have yet to come across any additional relevant material demonstrating the usage of these two multivariate methods in the relationship of air quality classification. PCA was used in our work to convert air quality observations into factorial axes and then analyse the observations-variables connections. We employed PCA to elucidate the most important contributing observations to air quality variability. Finally, our study was able to sort out the most significant observations with the most significant variables from all stations in Peninsular Malaysia.

AHC was then examined using the summary indices from the PCA factor score to discover trends in air quality in the central region of Peninsular Malaysia. It is based on all variables found in PCA results from 2018 to 2019, as well as a monthly fluctuation from January to December, which refers to the selected data for that year. The analysis was centred on classes, with a focus on the cluster that was created. AHC was utilised to investigate the pattern of air quality, which is depicted as a dendrogram in Figure 2. It represents how the algorithm works to group the observations, and then the subgroups of observations. All of the observations were successfully grouped by the algorithm. The monthly average air quality data from ten monitoring sites was analysed using AHC analysis based on factor scores from prior PCA analysis in this study. This section uses AHC to categorise the most important air quality stations based on their homogeneity level by observing the historical values of six air contaminants one at a time. The dendrogram's AHC results illustrate the dissimilarity between the clusters involved as High Pollution Regions (HPR) and Low Pollution Regions (LPR).

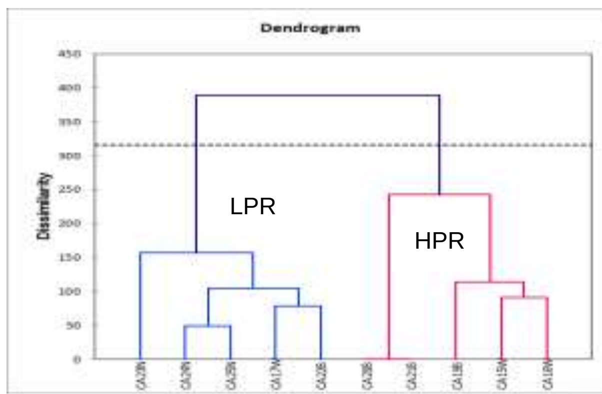


**Figure 2. Dendrogram of different clusters of air monitoring stations**

Figure 3 shows the classification of stations in Malaysia using AHC (HPR and LPR) based on six air pollutants concentrations. In total, HPR has five stations and LPR has five stations. The majority of HPR stations are located in Klang Valley and most of the LPR stations are located in Negeri Sembilan with two stations in Putrajaya and Banting respectively.

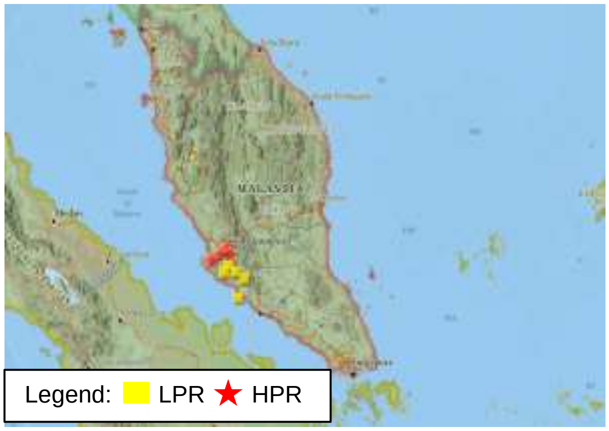


**Figure 3. Classification of stations using hybrid clustering method based on the monthly average of six air pollutants concentrations.**

**Differentiate between classes using discriminant analysis (DA)**

Further analysis by using DA, which exposed dissimilarities within the study sites was implemented given the clustering obtained from the AHC for HPR and LPR. Here, the class of clusters were considered as the independent parameters, while the air pollutants were considered as the dependent parameters.

**Table 3 Discriminant analysis for all classes (standard)**

| Sampling stations | Cluster 1 HPR | Cluster 2 LPR | Total | % Correct |
|---|---|---|---|---|
| Cluster 1 HPR | 115 | 5 | 120 | 95.83% |
| Cluster 2 LPR | 8 | 112 | 120 | 93.33% |
| Total | 123 | 117 | 240 | 94.58% |

Table 3 shows the classification matrix of DA, which involves two types of clusters, namely Cluster 1-HPR and Cluster 2-LPR. Based on the table, there were 120 numbers of data from Cluster 1-HPR with a percentage accuracy of 95.83 % and 120 numbers of data from Cluster 2-LPR with a percentage accuracy of 93.33 %. In general, the regions discriminated well with an average of 94.58 % correct classification. Compared with another study (Ismail et al. 2017) which was using normal clustering analysis, the percentage correction for each cluster was much lower than the current study. In a previous study, accuracy was 63.94 % of the average value of percentage correction, however in this study, accuracy was 94.58 %. It shows that through the hybrid clustering method, the percentage of correction either in each cluster and total clusters could be improved and have a better percentage rather than using a routine method for clustering analysis. However, there were few studies done previously using the normal clustering method that were able to perform a good percentage of accuracy (Azid et al. 2017; Azid et al. 2015; Hua, 2018). Their findings showed more than 90 % of accuracy for spatial variation. However, there was a limitation from the previous study because only selected stations in small number were chosen to examine the spatial variation rather than the present study. This novel approach was able to promote higher numbers of observations rather than involve selected stations to reduce error or avoid any bias from the data selection. A previous study by Rahman et al. (2022) also performed AHC about determining the spatial variation on $PM_{2.5.}$ However, no validation of the classification of the groups was done to verify the correction of the classification.

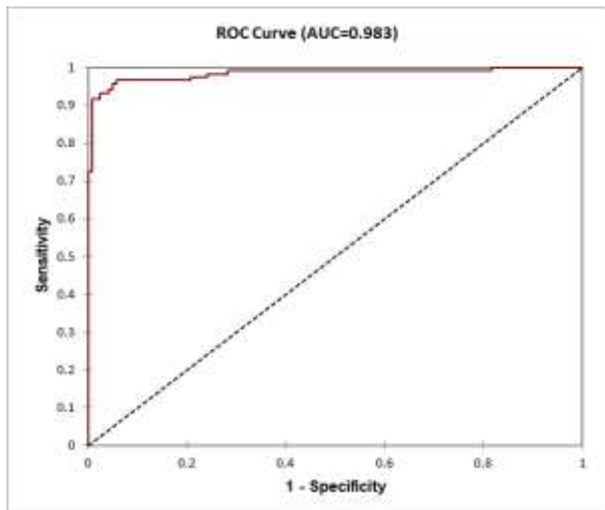


**Figure 4. Receiver Operating Characteristics (ROC) for the classification of regions**

Besides, Figure 4 also shows that these two classification regions were discriminating well based on the Receiver Operating Characteristics (ROC), which supported the accuracy of each cluster. According to a previous study (Hidalgo et al. 2018), the performance of a classifier can be assessed by the curve of ROC, whereby the accuracy of the classifier can be measured through the area under the curve (AUC). It is a perfect test if an area represents 1 while it is a worthless test if an area represents 0.5. The AUC curve for HPR and LPR was 0.983. The results obtained, it showed that the accuracy for each classification region was acceptable in the area represented above the worthless test value.

The test will become more accurate if the curve becomes closer to the left-hand border as well as to the top border space of ROC. In contrast, it becomes less accurate if the curve gets closer to the 45-degree diagonal space of ROC. Figure 5 shows that all of the classification regions following the left-hand border are closer to the top border of the ROC space.

## CONCLUSION

Based on the results, multivariate analysis for environmental measurement was able to assess air monitoring patterns. The framework provided here offers a novel tool for identifying and classifying stations based on air quality variables. The use of the PCA factor score in conjunction with the clustering technique is particularly important. AHC created two distinct air quality clusters. In light of the connection link that fixes an object to one class, this present study claims that hybrid clustering (PCA factor score, AHC) is a better strategy for similar iterations. Furthermore, discriminant analysis (DA) has presented the differences between the parameters and the clustering established significant value to validate the cluster classifications and the results showed that they were reliable variables in the central region of Peninsular Malaysia. This demonstrates that the hybrid cluster analysis technique utilized in this study is capable of producing more precise pollutant distributions that are useful in air pollution studies.

## CONFLICT OF INTEREST

The authors declared that the present study was performed in absence of any conflict of interest.

## ACKNOWLEDGEMENT

This work is part of a research project, UniSZA/2021/DPU1.0/20/R0324. NAA is supported by Skim Latihan Akademik IPTA-SLAI (Ministry of Higher Education, Malaysia) and Universiti Sultan Zainal Abidin. We also thank the Department of Environment of Malaysia for their support.

## AUTHOR CONTRIBUTIONS

AA: Suggested the work protocol, analyzed & interpreted the data, and revised the manuscript. AM, ASO and SNS: Conceived & designed the framework, contributed materials, analysis tools or data, and analyzed & interpreted the data. NAA: Performed the data gathering, analyzed, and wrote the paper. All authors read and approved the final version.



## REFERENCES

Afroz R, Hassan MN, Ibrahim NA, 2003. Review of air pollution and health impacts in Malaysia. Environ Res, 92(2): 71-77. doi:10.1016/s0013-9351(02)00059-2

Arifuddin AA, Jalaludin J, Hisamuddin NH, 2019. Air pollutants exposure with respiratory symptoms and lung function among primary school children nearby heavy traffic area in Kajang. Asian J. Atmos. Environ. 13: 21–29. doi:10.5572/ajae.2019.13.1.021

Azid A, Abd Rani N, Samsudin M, Khalit SI, Yunus K, Mohd Saudi AS, Gasim M, 2017. Air quality modelling using chemometric techniques. Journal of Fundamental and Applied Sciences, 9: 443-466. doi:10.4314/jfas.v9i2s.30

Azid A, Juahir H, Ezani E, Toriman ME, Endut A, Rahman MNA, . . . Umar R, 2015. Identification Source of Variation on Regional Impact of Air Quality Pattern Using Chemometric. Aerosol and Air Quality Research, 15(4): 1545-1558. doi:10.4209/aaqr.2014.04.0073

Azid A, Juahir H, Toriman ME, Kamarudin MKA, Saudi ASM, Hasnam CNC, . . . Yamin M, 2014. Prediction of the Level of Air Pollution Using Principal Component Analysis and Artificial Neural Network Techniques: a Case Study in Malaysia. Water, Air, & Soil Pollution, 225(8): 2063. doi:10.1007/s11270-014-2063-1

Azmi SZ, Latif MT, Ismail AS, Juneng L, Jemain AA, 2010. Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. Air Quality, Atmosphere & Health, 3(1): 53-64. doi:10.1007/s11869-009-0051-1

Bell ML, Ebisu K, Leaderer BP, Gent JF, Lee HJ, Koutrakis P, . . . Peng RD, 2014. Associations of $PM_{2.5}$ constituents and sources with hospital admissions: analysis of four counties in Connecticut and Massachusetts (USA) for persons >/= 65 years of age. Environmental health perspectives, 122(2): 138-144. doi:10.1289/ehp.1306656

Chen MC, Wang CF, Lai BC, Hsieh SW, Chen SC, Hung CH, Kuo CH, 2021. Air Pollution Is Associated with Poor Cognitive Function in Taiwanese Adults. International Journal of Environmental Research and Public Health, 18(1): 316. doi:10.3390/ijerph18010316

Cheng D, Hao Y, Zhou W, Ma Y, 2014. The relationship between interleukin-18 polymorphisms and allergic disease: a meta-analysis. Biomed Res Int: 290687. doi:10.1155/2014/290687

Department of Environment, 2020. New Malaysia ambient air quality standard Department of Environment.http://www.doe.gov.my/portalv1/en/category/info-umum/indeks-pencemaran-udara Accessed March 2022

Dominick D, Juahir H, Latif MT, Zain SM, Aris AZ, 2012. Spatial assessment of air quality patterns in Malaysia using multivariate analysis. Atmospheric Environment, 60: 172-181. doi:https://doi.org/10.1016/j.atmosenv.2012.06.021

Hamza Ahmad I, Azman A, 2015. Air Quality Pattern Assessment in Malaysia Using Multivariate Techniques. Malaysian Journal of Analytical Sciences, 19(5): 966-978

Hidalgo MJ, Pozzi MT, Furlong OJ, Marchevsky EJ, and Pellerano RG, 2018. Classification of organic olives based on chemometric analysis of elemental data. *Microchemical Journal, 142*, 30-35. doi:https://doi.org/10.1016/j.microc.2018.06.002

Hisamuddin NH, Jalaludin J, Abu Bakar S, Latif MT, 2022. The Influence of Environmental Polycyclic Aromatic Hydrocarbons (PAHs) Exposure on DNA Damage among School Children in Urban Traffic Area, Malaysia. International Journal of Environmental Research and Public Health, *19*(4): doi:10.3390/ijerph19042193

Hua A, 2018. Applied Chemometric Approach in Identification Sources of Air Quality Pattern in Selangor, Malaysia. Sains Malaysiana, *47*: 471-479. doi:10.17576/jsm-2018-4703-06

Isa KNM, Hashim Z, Jalaludin J, Norbäck D, Jabbar MA & Hashim JH, 2020. The Impact of Exposure to Indoor Pollutants on Allergy and Lung Inflammation among School Children in Selangor, Malaysia: An Evaluation Using Factor Analysis. Aerosol and Air Quality Research, *20*(11): 2371-2383. doi:10.4209/aaqr.2020.03.0128

Ismail A, Abdullah A, Samah M, 2017. Environmetric Study on Air Quality Pattern for Assessment in Northern Region of Peninsular Malaysia. Journal of Environmental Science and Technology, 10: 186-196. doi:10.3923/jest.2017.186.196

Jamalani M, Abdullah A, Azid A, Ramli M, Baharudin MR, Bose M, . . . Yusuf Gumel D, 2016. Monthly analysis of $PM_{10}$ in Ambient Air of Klang Valley, Malaysia. Malaysian Journal of Analytical Sciences. doi:10.17576/mjas

Jamhari AA, Sahani M, Latif MT, Chan KM, Tan HS, Khan MF, Mohd Tahir N, 2014. Concentration and source identification of polycyclic aromatic hydrocarbons (PAHs) in $PM_{10}$ of urban, industrial and semi-urban areas in Malaysia. Atmospheric Environment, 86: 16-27. doi:https://doi.org/10.1016/j.atmosenv.2013.12.019

Kamalha E, Omollo E, 2017. Clustering and Classification of Cotton Lint Using Principle Component Analysis, Agglomerative Hierarchical Clustering, and K-Means Clustering. Journal of Natural Fibers, 15. doi:10.1080/15440478.2017.1340220

Kamaruddin AS, Jalaludin J, Hamedon TR, Hisamuddin N, 2019. FeNO as a biomarker for airway inflammation due to exposure to air pollutants among school children nearby industrial areas in Terengganu. Pertanika Journal of Science and Technology, *27*: 589-600.

Mohamad ND, Ashaari Z, Othman M, 2015. Preliminary Assessment of Air Pollutant Sources Identification at Selected Monitoring Stations in Klang Valley, Malaysia. Procedia Environmental Sciences, 30: 121-126. doi:10.1016/j.proenv.2015.10.021

Peters R, Peters J, Booth A, Mudway I, 2015. Is air pollution associated with increased risk of cognitive decline? A systematic review. Age Ageing, 44(5): 755-760. doi:10.1093/ageing/afv087

Rahman EA, Hamzah FM, Latif MT, Dominick D, 2022. Assessment of $PM_{2.5}$ Patterns in Malaysia Using the Clustering Method. Aerosol and Air Quality Research, 22(1): 210161. doi:10.4209/aaqr.210161

Sahrir S, Abdullah A, Ponrahono Z, Sharaai A. H, 2019. Environmetric Study on Air Quality Pattern for Assessment in Klang Valley, Malaysia. International Journal of Recent

Technology and Engineering, 8: 17-24.

Shafii NZ, Saudi ASM, Pang JC, Abu IF, Sapawe N, Kamarudin MKA, Saudi HFM, 2019. Application of chemometrics techniques to solve environmental issues in Malaysia. Heliyon, 5(10): e02534. doi:10.1016/j.heliyon.2019.e02534

Shehab MA, Pope FD, 2019. Effects of short-term exposure to particulate matter air pollution on cognitive performance. Sci Rep, 9(1): 8237. doi:10.1038/s41598-019-44561-0

Tellez-Rojo MM, Rothenberg SJ, Texcalac-Sangrador JL, Just AC, Kloog I, Rojas-Saunero LP, Gutierrez-Avila I, Bautista-Arredondo LF, Tamayo-Ortiz M, Romero M, Hurtado-Diaz M, Schwartz JD, Wright R, Riojas-Rodriguez H, 2020. Children's acute respiratory symptoms associated with $PM_{2.5}$ estimated in two sequential representative surveys from the Mexico City Metropolitan Area. Environ Res, 180, 108868. doi:10.1016/j.envres.2019.108868

Wong SF, Yap PS, Mak JW, Chan WLE, Khor GL, Ambu S, Chu WL, Mohamad MS, Ibrahim Wong N, Ab Majid NL, Abd Hamid HA, Rodzlan Hasani WS, Mohd Yussoff MFB, Aris HTB, Rahman EA, M Rashid ZB, 2020. Association between long-term exposure to ambient air pollution and prevalence of diabetes mellitus among Malaysian adults. Environ Health, 19(1): 37. doi:10.1186/s12940-020-00579-w

World Health Organisation, 2018. The global health observatory; Explore a world of health data https://www.who.int/data/gho/data/themes/theme-details/GHO/air-pollution Accessed May 2022.

World Health Organisation, 2021. Global air quality guidelines. Particulate matter ($PM_{2.5}$ and $PM_{10}$), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. https://www.who.int/publications/i/item/9789240034228 Accessed 20 May 2022.

Zainal-Abidin E, Semple S, Rasdi I, Ismail SNS, Ayres JG, 2014. The relationship between air pollution and asthma in Malaysian schoolchildren. Air Quality, Atmosphere & Health, 7(4): 421-432. doi:10.1007/s11869-014-0252-0

Zakaria J, Lye MS, Hashim Z, 2012. Asthma Severity and Environmental Health Risk Factor among Asthmatic Primary School Children in the Selected Areas. American Journal of Applied Sciences, 9: 1553-1560.

Zakaria UA, Saudi ASM, Abu IF, Azid A, Balakrishnan A, Amin NA & Rizman ZI, 2017. The assessment of ambient air pollution pattern in Shah Alam, Selangor, Malaysia. Journal of Fundamental and Applied Sciences, 9(4S): 772-788. doi:10.4314/jfas.v9i4s.43

Zakri NL, Saudi ASM, Juahir H, Toriman ME, Abu IF, Mahmud MM & Khan MF, 2018. Identification source of variation on regional impact of air quality pattern using chemometric techniques in Kuching, Sarawak. Int J Eng Technol, 7, 49. doi:10.14419/ijet.v7i3.14.16861