UITMCNS SEREMBAN

# J E M U R

JOURNAL OF EXPLORATORY MATHEMATICAL UNDERGRADUATE RESEARCH

RECENT **MATHEMATICAL PROJECTS** AT UITM, *SEREMBAN CAMPUS*

Vol. 002 | PREPARED BY: KPPIM, UiTM N. SEMBILAN. | DATE: 05 OCT 2024


*Photography courtesy of Porapak Apichodilok*

# CLASSIFICATION OF AIR QUALITY IN THE KLANG VALLEY USING K-MEANS CLUSTERING

**Nur Aliah Amira Abdul Hamid[1], Siti Khairun Najwa Kamil[2] and Noorezatty Mohd Yusop[3,*]**

[1,2,3]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Cawangan Negeri Sembilan Kampus Seremban, 70300 Negeri Sembilan

*noorezatty@uitm.edu.my

**Abstract**

Air pollution is a major global concern, and its adverse impacts are most likely to affect urban areas. The Klang Valley, located on the southwest coast of Peninsular Malaysia, is also affected by this global issue. As one of the most industrialized and populated areas, controlling air pollution is a challenge. Thus, this paper aims to classify air monitoring stations in the Klang Valley into distinct clusters based on six major air pollutants. Data on pollutant levels collected by the Department of Environment from air monitoring stations throughout the region was utilised. Fourteen stations representing seven areas in Klang Valley were assessed in two weather conditions: the dry season (May-September) and the rainy season (November-March), typically known as monsoon seasons in Malaysia. Two clusters of air monitoring stations were identified using K-Means clustering. The first cluster, comprising four stations, showed better air quality with lower pollutant levels. In contrast, the second cluster, which includes ten stations, showed higher pollutant levels. However, the pollutant levels in both clusters were within the permissible limits according to the Malaysian Ambient Air Quality Guidelines. Furthermore, the location of the monitoring station can influence pollutant levels, whereas seasonal variations (dry or rainy) have a lesser impact. Consistent monitoring is crucial for tracking air pollution changes and adjusting policies accordingly.

***Keywords:*** Air pollution, Clustering Algorithm, K-Means Cluster Analysis

## 1. Introduction

Pollution issues, particularly air pollution pose major global challenges with urban areas particularly vulnerable to its negative consequences. There are a variety of air pollutants and the release into the air can cause serious health concerns and can even be deadly in small concentrations. In 2021, air pollution caused 8.1 million fatalities worldwide, making it the second most significant cause of death, including children under the age of five (Health Effect Institute, 2024).

Monitoring air quality is essential to protect public health, and ensuring compliance with regulations. It aligns with the World Health Organization's Sustainable Development Goals (SDGs), particularly Goal 3 (Good Health and Well-being) and Goal 11 (Sustainable Cities and Communities), by emphasizing the importance of monitoring air quality to safeguard public health and promote sustainable urban development. Consistent monitoring is crucial for tracking air pollution changes and adjusting policies accordingly.

Currently, the Air Pollution Index (API) is used as an indicator to monitor air quality status. The API value in Table 1 is calculated from six main air pollutants; Particulate Matter of less than 10 Microns ($PM_{10}$), Particulate Matter of less than 2.5 Microns ($PM_{2.5}$), Ozone ($O_3$), Carbon Monoxide ($CO$), Nitrogen Dioxide ($NO_2$) and Sulphur Dioxide ($SO_2$).

Table 1: Air pollution index

| API Range | Air Pollution Status |
|---|---|
| 0-50 | Good |
| 51-100 | Moderate |
| 101-200 | Unhealthy |
| 201-300 | Very Unhealthy |
| >300 | Hazardous |

Being one of the most densely populated and industrialized regions in Malaysia, the Klang Valley is increasingly faced with problems relating to air pollution. Klang Valley measures 2911.5 square kilometres in area and is located over several districts in Selangor and the Federal Territory of Kuala Lumpur. Potential sources of air pollution in Klang Valley include biomass burning/smoke (31.6%), soil dust (10.3%), industries (8.4%), motor vehicles (7.1%), sea spray (6.3%) and secondary sulphate (5.5%) (Elias et al., 2023) and the sources continuously contribute to the deterioration of air quality in the Klang Valley. The air monitoring stations measuring pollution levels in the region are located at Batu Muda, Petaling Jaya, Cheras, Shah Alam, Klang, Banting, Kuala Selangor, and Putrajaya.

In the year 2022, the API index for Klang Valley areas mostly fell within the moderate range compared to API readings in the good range as shown in Figure 1. Putrajaya recorded the highest number of good API readings at 98 days, while Klang recorded the highest number of moderate API readings at 360 days and Cheras recorded the highest unhealthy API reading at 5 days with most days falling within moderate status Deparment of Environment (2023). Relying solely on API assessments provides a narrow perspective on pollutant dynamics, overlooking their intricate patterns and interactions. Pollution levels can fluctuate across regions and weather conditions. Classifying monitoring stations based on pollution patterns helps target interventions to address specific pollution sources and patterns effectively.

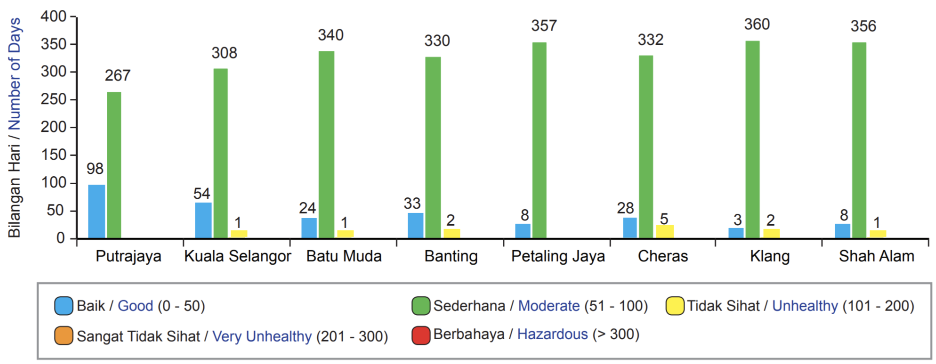


Figure 1: Air quality status in Klang Valley in 2022

Research interest towards air pollution analysis has increased due to its potential to protect public health, preserve environmental sustainability, and ensure compliance with air quality regulations. Previous studies utilizing air pollution data aim to understand pollutant sources (Richard et al., 2023), to predict air quality (Kaur et al., 2023; Vyas et al., 2023) or identify pollutant sites (Goyal et al., 2021). Statistical analysis is essential to simplify the complexities of air pollution data and is used in developing predictive models, identifying factors of air quality and describing pollutant levels. Additionally, advanced methods such as clustering algorithms are employed to reveal hidden patterns.

Clustering algorithms are used to group data points with common attributes, enabling the identification of distinct pollution patterns and sources. This technique assists in categorizing sites with similar pollution profiles, facilitating targeted interventions and policy implementations (Mohamad et al., 2015; Shafi and Waheed, 2020; Sunori et al., 2021). The use of clustering techniques enhances understanding of air quality dynamics. This information, combined with ongoing efforts to minimise air pollution, represents alternate solutions to provide clean air and a sustainable environment for current and future generations. Hence, this paper seeks to employ clustering algorithms to obtain valuable insights into air pollution in urban areas such as Klang Valley enabling effective and targeted environmental interventions.

## 2. Literature Review

Air pollution is a major environmental issue that has severe health implications. Previous research has been conducted to distinguish the levels of air pollution in different areas, as well as to investigate the cause of high pollutant concentrations. The study conducted by Nguyen and Kim (2006) assessed the regional and temporal distribution patterns of $SO_2$ in Korea. Data from four different groups of air quality monitoring stations (AQMS) was utilized, encompassing the period spanning from 1998 to 2003. The

AQMS covered densely populated areas in seven major cities (with urban traffic and urban environments), as well as less populated areas in nine major provinces (with suburban environments and rural environments). The result showed significant differences in the average levels of $SO_2$ among all four groups of stations. A comprehensive analysis comparing pairs of stations (urban traffic vs. urban environment and suburban environment vs. rural environment) revealed that urban traffic stations indicated significantly higher concentrations of $SO_2$ compared to urban environment stations.

Azmi et al. (2010) on the other hand, assessed air quality patterns in Klang Valley. Data on five major air pollutants ($PM_{10}$, $CO$, $SO_2$, $O_3$, and $NO_2$) was gathered from monitoring stations in Petaling Jaya, Shah Alam, and Gombak. These data were compared with benchmark stations in Jerantut, Pahang. Petaling Jaya has higher levels of $CO$, $NO_2$, and $SO_2$ as a consequence of high traffic volume and industrialization. Regional factors such as biomass burning and sunlight affect $PM_{10}$ and $O_3$ concentration in the area. Although pollutant levels are lower than the permissible limit, they are significantly higher compared to the benchmark station, indicating localized pollution sources.

Furthermore, Anand et al. (2019) examined the levels of particulate matter pollutants ($PM_{2.5}$ and $PM_{10}$) in four major Indian cities: Delhi, Mumbai, Pune and Ahmedabad, in the North-West region of India. Subsequently, employing data from SAFAR observational network to determine whether geographical location influences particulate matter variability. Besides, the study revealed that although Delhi has the highest concentration of particulate matter, Mumbai shows a higher proportion of $PM_{2.5}$ as a percentage of $PM_{10}$ (60%) than Delhi (50 %). Differences were caused by the fact that Delhi is on land whereas Mumbai is on the coast, which affects the dispersion of air pollution.

In contrast, Mohd Anuar et al. (2021) explored the concentrations of $SO_2$ in the Universiti Tun Hussein Onn Malaysia (UTHM) Campus Area in Batu Pahat. The study examined air pollution levels in thirteen different areas in different time frames and discovered significant variations in $SO_2$. The highest recorded levels were detected in the afternoon at TDI Residential College A, with a reading of 0.198$ppm$, whereas morning measurements, particularly low at University Health Centre with a value of 0.03$ppm$, indicated changes associated with the distance from pollution sources.

These studies highlight that the variation in air pollution levels across various geographic regions is driven by factors such as burning activities, manufacturing activities, and traffic density.

## 3. Methodology

### 3.1. *Source of Data*

This study used secondary data obtained from the Malaysian Department of Environment (DOE). Data on the concentrations of six major air pollutants were obtained during the dry season (May - September) and the rainy season (November - March) in 2022. A total of 2407 records were utilized to generate average concentration values for each pollutant in every monitoring station at the Klang Valley, Selangor, specifically in Cheras, Putrajaya, Shah Alam, Batu Muda, Banting, Klang and Petaling Jaya.

### 3.2. *Description of Variables*

In the context of assessing air quality, six indicators were used as depicted in Table 2. These indicators provided information about different aspects of air characteristics.

Table 2: Scale of measurements for variables

| Variable | Unit | Measurement |
|---|---|---|
| Air Monitoring Station | - | Nominal |
| Particulate Matter ($PM_{2.5}$) | $\mu$g/$m^3$ | Ratio |
| Particulate Matter ($PM_{10}$) | $\mu$g/$m^3$ | Ratio |
| Nitrogen Dioxide ($NO_2$) | ppm | Ratio |
| Sulphur Dioxide ($SO_2$) | ppm | Ratio |
| Carbon Dioxide ($CO$) | ppm | Ratio |
| Ozone ($O_3$) | ppm | Ratio |

Table 3 contains information on the cluster members, which are the air monitoring stations in Klang Valley. Fourteen monitoring stations were assessed, with data collected from selected areas in Klang Valley during two seasons: dry (May-September) and rainy (November-March). This setup allows for data comparisons between the two time periods.

Table 3: Air monitoring stations in Klang Valley

| Station ID | Station Name | Season |
|---|---|---|
| BM1 | Batu Muda | Dry |
| BM2 | Batu Muda | Rainy |
| C1 | Cheras | Dry |
| C2 | Cheras | Rainy |
| PW1 | Putrajaya | Dry |
| PW2 | Putrajaya | Rainy |
| PJ1 | Petaling Jaya | Dry |
| PJ2 | Petaling Jaya | Rainy |
| SA1 | Shah Alam | Dry |
| SA2 | Shah Alam | Rainy |
| K1 | Klang | Dry |
| K2 | Klang | Rainy |
| B1 | Banting | Dry |
| B2 | Banting | Rainy |

### 3.3. *K-Means Clustering Algorithm*

The K-Means clustering algorithm is employed to categorise monitoring stations into different groups based on similar properties. K-Means algorithm operates by taking an input of objects (in this study the monitoring stations), denoted as C and an integer value K, ultimately producing a partition of C into subsets $C_1, C_2, \ldots, C_k$. The approach is regarded as an optimization criterion aimed at minimizing the sum of squared distances of objects within their respective clusters from the cluster centres (Chen et al., 2002). The sum of squared criterion is defined by the cost function,

$$W(C_i) = \sum_{r=1}^{|C_i|} \sum_{s=1}^{|C_i|} (d(x_r^i, x_s^i))^2 \tag{1}$$

Where
$x_r^i$ = The $r^{th}$ element of $C_i$
$x_s^i$ = The $s^{th}$ element of $C_i$
$|C_i|$ = Number of elements in $C_i$
$d(x_r^i, x_s^i)$ = Distance between a data point $x_r^i$ and $x_s^i$

The K-Means algorithm calculates the centroid of each cluster $C_i$ (denoted by $x_i$) and optimizes the cost function that can be defined as;

$$W(C_i) = \sum_{r=1}^{|C_i|} (d(x^i x_r^i))^2 \tag{2}$$

The K-Means algorithm consists of several steps:

(1) Determine the K number of clusters in which objects are categorized, assigning them to one of these K groups.
(2) Allocate each object to the group that has the closest centroid (mean) with respect to Euclidean distance. The Euclidean distance can be defined as the distance between $x_i$ and $y_i$ in $n$-dimensional space,

$$d_E(x_i, y_i) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3}$$

Here, $n$ is the number of indicators.

(3) Once all the objects have been assigned, recalculate the centroid for the clustering gaining the new object and adjust the centroid for the cluster losing the object.

(4) Repeat steps 2 and 3 until some of the convergence is met. That is when centroids reach a state where they remain static. The convergence criteria either no or minimal reassignment of data points to new cluster centres and a minimal decrease in the sum of squared error.

### 3.4. *Cluster Validity Index*

Finding the optimal number of clusters in a dataset is a critical aspect of clustering analysis. Cluster validity indices play a significant role in this process by providing objective criteria to evaluate and compare clustering results for different numbers of clusters.

Krzanowski-Lai Index assists in finding the optimal number of clusters in the data Syahputri et al. (2024). This method is defined by the difference between cluster $k$ with $k-1$ using the equation in (4).

$$DIFF(k) = (k-1)^{\frac{2}{p}}W_{k-1} - k^{\frac{2}{p}}W_k \tag{4}$$

The cluster number, $k$ is selected by maximizing the quantity based on equation (5).

$$KL(k) = |\frac{DIFF(k)}{DIFF(k+1)}| \tag{5}$$

Where:
$p$ = Number of dimensions (indicator)
$k$ = Number of clusters

## 4. Result and Discussion

In this section, the clusters of air monitoring stations are identified. There are fourteen air monitoring stations in total, spanning seven areas across two seasons. The optimal number of clusters was determined using the Krzanowski-Lai Index, and K-Means analysis was employed to identify the cluster members. Additionally, a 2D clustering plot is used to visually represent the distinct clusters.

### 4.1. *Determining the Optimal Number of Cluster*

Krzanowski-Lai Index was utilized for this purpose. Table 4 shows the Krzanowski-Lai Index score for cluster number 2 to 5. The highest score 1.8513, was obtained at cluster number 2, affirming that the optimal number of clusters is 2.

Table 4: Krzanowski-Lai index score

| Cluster Number | Score |
|---|---|
| 2 | **1.8513** |
| 3 | 1.4078 |
| 4 | 0.4893 |
| 5 | 0.9142 |

Figure 2 displays the score in a plot with the x-axis denoting the cluster count and the y-axis indicating the corresponding Krzanowski-Lai Index score. The plot demonstrates that the highest point occurs at 2.

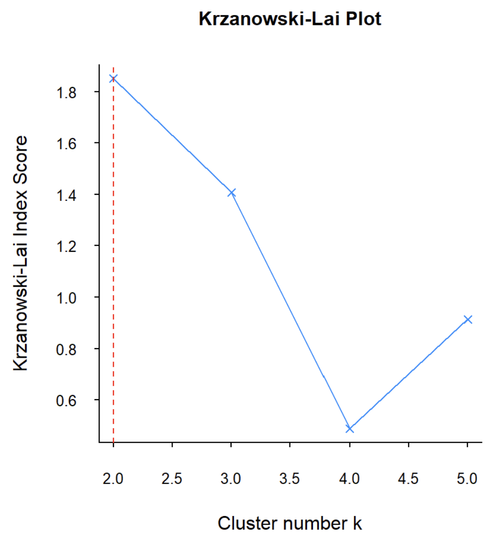


Figure 2: Krzanowski-Lai index plot

### 4.2. *Identifying Monitoring Stations by Cluster*

The 2D clustering plot in Figure 3 visually presents the outcome of the K-Means analysis. By employing K-Means analysis, two unique clusters of air monitoring stations were identified. Each point on the plot corresponds to a monitoring station at a specific location, with the numbers 1 or 2 representing seasonal points, dry or rainy season. Points of the same color are grouped to form a cluster, positioned close to their cluster centroid. The cluster numbers (e.g., 1, 2, etc.) are assigned according to the sequence in which the centroids are initialized and subsequently updated during the clustering algorithm's iterations. Consequently, these numbers are somewhat arbitrary and do not inherently correspond to any specific criteria

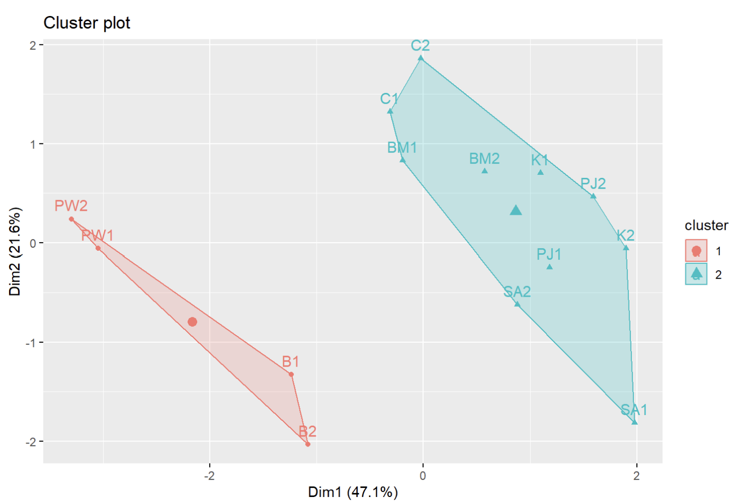


Figure 3: Clusters of air monitoring stations in Klang Valley

The results of the clustering analysis are summarized in Table 5, which illustrates the distribution of stations across different seasons for each obtained cluster. The monitoring stations are divided into two clusters. Cluster 1 contains four stations: Putrajaya and Banting, during both the dry and rainy seasons. The remaining ten stations fall into Cluster 2, which includes Batu Muda, Cheras, Petaling Jaya, Shah Alam, and Klang, also encompassing both the dry and rainy seasons.

Table 5: Cluster members

| Cluster | Station ID | Station Name | Season |
|---|---|---|---|
| 1 | PW1 | Putrajaya | Dry |
| | PW2 | Putrajaya | Rainy |
| | B1 | Banting | Dry |
| | B2 | Banting | Rainy |
| 2 | BM1 | Batu Muda | Dry |
| | BM2 | Batu Muda | Rainy |
| | C1 | Cheras | Dry |
| | C2 | Cheras | Rainy |
| | PJ1 | Petaling Jaya | Dry |
| | PJ2 | Petaling Jaya | Rainy |
| | SA1 | Shah Alam | Dry |
| | SA2 | Shah Alam | Rainy |
| | K1 | Klang | Dry |
| | K2 | Klang | Rainy |

Table 6 presents summary statistics for six air pollutants across Clusters 1 and 2, focusing on average concentrations. The table indicates that air pollutant levels in both clusters are within permissible limits. Cluster 2 shows higher average concentrations for four pollutants ($PM_{10}$, $PM_{2.5}$, $CO$, and $NO_2$), whereas Cluster 1 exhibits higher pollutant levels for two types of pollutants ($O_3$ and $SO_2$).

Table 6: Average pollutant levels across two clusters

| Pollutants | Average exposure limit | Cluster | |
|---|---|---|---|
| | | 1 | 2 |
| Particulate Matter ($PM_{2.5}$) | $260\mu g/m^3$ | 24.0097 | **28.5323** |
| Particulate Matter ($PM_{10}$) | $260\mu g/m^3$ | 15.8125 | **19.3600** |
| Sulfur Dioxide ($SO_2$) | $0.04ppm$ | **0.0015** | 0.0013 |
| Nitrogen Dioxide ($NO_2$) | $0.04ppm$ | 0.0088 | **0.0135** |
| Ozone ($O_3$) | $0.06ppm$ | **0.0197** | 0.0154 |
| Carbon Monoxide ($CO$) | $9ppm$ | 0.5076 | **0.7015** |

Statistics in Table 6 indicate that the air monitoring stations in Cluster 1, which exhibit lower concentration values, are in Putrajaya and Banting for both the dry and rainy seasons. It shows that changes in the seasons do not affect the pollutant levels. Although it is anticipated that rainfall could lower the pollutant levels, however shorter exposure to rainfall which is less than an hour, keeps some pollutant level to remain constant (Hilario et al., 2022). Generally, Klang Valley develop short-duration rainfall between 15 to 45 minutes (Mamun et al., 2018).

Putrajaya is primarily designed as an administrative capital with extensive parks, green spaces, and water features. It aims to become a "green city" by implementing eco-friendly structures, promoting public transportation, and using renewable energy sources, all contributing to pollution reduction (Abd. Razak, 2014). In contrast, Banting, a suburban village nestled near hills, forests, and farmland, hosts industrial zones that contribute to air pollution, particularly during El Niño periods. However, it achieved a good Air Pollution Index (API) status in 2022, possibly due to pollutant dispersion in agricultural and forested areas (Shaadan et al., 2018).

In Klang Valley, stations like Cheras, Petaling Jaya, Batu Muda, Shah Alam, and Klang form a different cluster, showing higher pollutant concentrations and poorer air quality attributed to urban growth, industrial activities, and heavy traffic. Cheras had the highest unhealthy API, with significant carbon monoxide emissions from motor vehicles, while Klang reported the highest moderate API in 2022 (Deparment of Environment, 2023). Despite seasonal haze from fires in Sumatra and Borneo affecting the region, pollutant concentrations in Klang Valley generally remain within acceptable limits due to dispersion and absorption in natural environments (Deparment of Environment, 2023).

## 5. Conclusion

In conclusion, the analysis of air quality across different areas within the Klang Valley reveals diverse environmental dynamics and their impact on pollutant concentrations. The monitoring stations were successfully divided into two distinct clusters, with each member within a cluster exhibiting similar air pollution characteristics. The first cluster includes stations in Putrajaya and Banting, which are recognized for their comparatively better air quality and lower average pollution levels. In contrast, the second cluster comprises stations in more urbanized areas like Cheras, Petaling Jaya, Batu Muda, Shah Alam, and Klang, where higher average pollution levels are observed due to industrial emissions and vehicular pollution. Based on the pollutant levels, Cluster 1 has higher levels of $SO_2$ and $O_3$ which is indicating to higher concentrations of certain gases. Encouraging people to walk, bicycle, or use public transport can help reduce car emissions. On the other hand, Cluster 2 has higher level of $PM_{10}$, $PM_{2.5}$, $NO_2$, and $CO$. Strict regulations can help control these pollutants. Construction companies, for example, can take steps to reduce pollution by installing dust control measures on building sites and upgrading industrial equipment. Moving forward, it is essential to strengthen urban planning strategies and implement stricter regulatory measures to manage air quality effectively across Klang Valley. Continuously monitoring air quality across all clusters and enhancing public awareness about the impact of pollution on health and the environment are crucial steps in achieving sustainable development and improving quality of life in the region.

## Acknowledgment

The authors would like to thank Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban for the facilities involved in making this research a success.

## References

Abd. Razak, M. R. (2014). Putrajaya green cities: constraints and mitigation strategies towards low carbon cities [unpublished thesis]. Universiti Teknologi Malaysia.

Anand, V., Korhale, N., Rathod, A., and Beig, G. (2019). On processes controlling fine particulate matters in four indian megacities. *Environmental Pollution*, 254:113026.

Azmi, S. Z., Latif, M. T., Ismail, A. S., Juneng, L., and Jemain, A. A. (2010). Trend and status of air quality at three different monitoring stations in the klang valley, malaysia. *Air Quality, Atmosphere & Health*, 3:53–64.

Chen, G., Jaradat, S., Banerjee, N., Tanaka, T., Ko, M. S., and Zhang, M. Q. (2002). Evaluation and comparison of clustering algorithms in analyzing ES cell gene expression data. *Statistica Sinica*, 12(1):241–262.

Deparment of Environment (2023). *Laporan Kualiti Alam Sekeliling (EQR) 2022*. Ministry Of Natural Resources And Environmental Sustainability, Petaling Jaya.

Elias, M. S., Hashim, A., Bahrudin, N. F. D., Sapiee, N. A., Paulus, W., Azman, M. A., Raduian, N. J., and Abdullah, I. M. (2023). Assessment and sources identification of air quality pollution in Klang Valley, Kuala Lumpur, Malaysia. *IOP Conference Series Materials Science and Engineering*, 1285(1):012017.

Goyal, P., Gulia, S., and Goyal, S. (2021). Identification of air pollution hotspots in urban areas-an innovative approach using monitored concentrations data. *Science of The Total Environment*, 798:149143.

Health Effect Institute (2024). *State of Global Air 2024*. Health Effects Institute, Boston, MA.

Hilario, M. R. A., Bañaga, P. A., Betito, G., Braun, R. A., Cambaliza, M. O., Cruz, M. T., Lorenzo, G. R., MacDonald, A. B., Pabroa, P. C., Simpas, J. B., et al. (2022). Stubborn aerosol: why particulate mass concentrations do not drop during the wet season in metro manila, philippines. *Environmental Science: Atmospheres*, 2(6):1428–1437.

Kaur, J., Parmar, K. S., and Singh, S. (2023). Autoregressive models in environmental forecasting time series: a theoretical and application review. *Environmental Science and Pollution Research*, 30(8):19617–19641.

Mamun, A. A., Salleh, M. N. B., and Noor, H. M. (2018). Estimation of short-duration rainfall intensity from daily rainfall values in Klang Valley, Malaysia. *Applied Water Science*, 8(7).

Mohamad, N. D., Ash'aari, Z. H., and Othman, M. (2015). Preliminary assessment of air pollutant sources identification at selected monitoring stations in Klang Valley, Malaysia. *Procedia Environmental Sciences*, 30:121–126.

Mohd Anuar, F. N., Othman, N., and Hamood Altowayti, W. A. (2021). Assessments of concentrations of sulfur dioxide (so2) in universiti tun hussein onn malaysia (uthm) campus area, batu pahat. In *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, pages 1–6.

Nguyen, H. T. and Kim, K. (2006). Evaluation of SO2 pollution levels between four different types of air quality monitoring stations. *Atmospheric Environment*, 40(36):7066–7081.

Richard, G., Izah, S. C., and Ibrahim, M. (2023). Air pollution in the niger delta region of nigeria: sources, health effects, and strategies for mitigation. *Journal of Environmental Studies*, 29(1):1–15.

Shaadan, N., Nazeri, L., Jalani, M., Rahman, N., and Roslan, R. (2018). Data visualization of temporal ozone pollution between urban and sub-urban locations in selangor malaysia. *Journal of Fundamental and Applied Sciences*, 9:490.

Shafi, J. and Waheed, A. (2020). K-means clustering analysing abrupt changes in air quality. *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 3:26–30.

Sunori, S. K., Negi, P. B., Maurya, S., Juneja, P., Rana, A., and Bhawana (2021). K-means clustering of ambient air quality data of uttarakhand, india during lockdown period of covid-19 pandemic. In *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, pages 1254–1259.

Syahputri, Z., Siregar, M. A. P., et al. (2024). Determining the optimal number of k-means clusters using the calinski harabasz index and krzanowski and lai index methods for groupsing flood prone areas in north sumatra. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 9(1):571–580.

Vyas, V., Mehta, K., and Sharma, R. (2023). The nexus between toxic-air pollution, health expenditure, and economic growth: An empirical study using ardl. *International Review of Economics & Finance*, 84:154–166.