Ground-Level Ozone Pollution in Malaysia: Descriptive Analysis and Classification Using Hierarchical Clustering

Hazrul Abdul Hamid

School of Distance Education Universiti Sains Malaysia, 11800, USM, Penang, Malaysia

DOI: https://dx.doi.org/10.47772/IJRISS.2025.909000426

Received: 07 September 2025; Accepted: 13 September 2025; Published: 14 October 2025

ABSTRACT

Tropospheric ozone, a critical air pollutant, plays a significant role in determining ambient air quality. It forms through a photochemical reaction between sunlight, volatile organic compounds (VOCs), and nitrogen oxides, earning it the label of a secondary pollutant. Beyond its impact on human health, ozone also poses a threat to the environment and vegetation. This study hourly average of ozone (O_3) concentrations in 2022 across 46 air quality monitoring stations in Malaysia. Mean, standard deviation, skewness, and kurtosis were employed to understand the characteristics of ground-level ozone concentrations. Using Agglomerative Hierarchical Cluster analysis, the stations were grouped into three clusters based on the mean hourly average O_3 concentrations: High Pollution Region (HPR), Medium Pollution Region (MPR), and Low Pollution Region (LPR). Interestingly, some suburban stations, typically expected to have lower pollution levels, recorded higher O_3 readings and were categorized in the high pollution cluster. By clustering the monitoring stations into these three groups, future research can benefit from a more focused assessment of air quality, reducing the complexity of monitoring and enhancing the effectiveness of pollution management strategies.

Keywords: Air pollution, cluster analysis, Ground-level ozone

INTRODUCTION

In today's fast-paced world, air quality is a major concern for both developed and developing countries. According to the World Health Organization (WHO), air pollution is a serious health and environmental issue, leading to millions of deaths and loss of healthy life every year due to exposure [1]. In Malaysia, environmental issues weren't taken seriously until the haze problem in Southeast Asia in the 1980s. Recognizing the impact on human health and the environment, the Malaysian Department of Environment (DOE) introduced the Malaysia Ambient Air Quality Guidelines (MAAQG) in 1989 to define harmful levels of air pollutants [2]. In 2013, the MAAQG was replaced by the New Ambient Air Quality Standard, which updated the concentration limits for pollutants and added a new parameter. The six parameters now include ozone (O_3), nitrogen dioxide (NO2), carbon monoxide (CO), sulphur dioxide (SO_2), particulate matter less than 10 microns (PM10), and the newly added particulate matter less than 2.5 microns ($PM_{2.5}$). The Malaysia Air Quality Index (MAQI) was the first system established by the Department of Environment (DOE) in 1993 to report air quality status to authorities and the public. In 1996, the DOE updated this system and introduced the Air Pollutant Index (API), which is based on the United States' Pollutant Standard Index (PSI). The API reflects the air quality status and is displayed on the Air Pollution Index Management System (APIMS) and the DOE website.

The API is calculated using the concentrations of six main pollutants: ground-level ozone (O_3), nitrogen dioxide (NO_2), carbon monoxide (CO), sulfur dioxide (SO_2), particulate matter less than 10 microns (PM_{10}), and particulate matter less than 2.5 microns ($PM_{2.5}$) [3]. The air quality is then classified as good, moderate, unhealthy, very unhealthy, or hazardous.

Each air pollutant affects human health differently, with impacts ranging from mild to severe depending on the length of exposure [4]. Short-term exposure can cause mild symptoms like eye irritation or worsen pre-existing conditions such as asthma and chronic obstructive pulmonary disease (COPD). However, long-term exposure can lead to more serious health issues, including death. In 2019, the Malaysia Ministry of Health reported that respiratory illnesses caused by poor air quality were responsible for about 10,600 deaths, or 14.8% of all deaths, making it the second leading cause of death after cardiovascular diseases [5].

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume IX Issue IX September 2025

According to the United States Environmental Protection Agency (USEPA), air pollution sources can be natural or human-made, emitting gases and particles into the air. Human activities, known as anthropogenic sources, include industrial activities, vehicle emissions, construction, use of chemical products like paints, power plants, and open burning. Natural sources include biogenic emissions from plants and microbial decomposition, as well as geogenic emissions from volcanoes, wildfires, and lightning, which contribute to the formation of pollutants like carbon dioxide (CO_2), sulfur dioxide (SO_2), carbon monoxide (CO), and ozone (O_3).

In Malaysia, industrial activities are the main contributors to air pollution, especially in urban areas. In 2021, there were 13,989 industrial sources of air pollution in the country [6]. The aims of this research are to compare the characteristics of ground-level ozone (O_3) concentrations in Malaysia using descriptive statistics based on the locations of monitoring stations and to classify the stations according to these characteristics.

LITERATURE REVIEW

Air pollution is an escalating global concern that demands urgent and comprehensive action from human society. The World Health Organization (WHO) has revealed alarming statistics indicating that the air inhaled by 99% of the global population exceeds safe guideline limits. This widespread issue is largely attributed to industrialization and urbanization, as evidenced by recent studies. The situation is particularly dire in low- and middle-income countries, where rapid economic growth and urban expansion have severely compromised air quality [1].

In the pursuit of economic development, many nations have experienced a significant decline in air quality, driven by the emission of various pollutants. Malaysia, a developing country undergoing rapid industrialization and urbanization, mirrors this trend. The country's growing population has led to an increased demand for industrial production, transportation, and energy generation, all of which contribute to air pollution [7]. The pollutants released from these activities, particularly from industrial processes and vehicle emissions, directly enter the atmosphere, exacerbating pollution levels.

Air pollution is typically defined as the presence of pollutants in the atmosphere at concentrations that pose a threat to human health. The gravity of this issue is underscored by the fact that air pollution is responsible for approximately seven million premature deaths worldwide each year. Research has shown that 11.65% of global deaths, particularly in low-income countries, are linked to air pollution, highlighting the disproportionate burden borne by these regions due to limited resources to combat pollution [8].

Air quality is primarily determined by the concentrations of six key pollutants: particulate matter (PM_{10} and $PM_{2.5}$), ground-level ozone (O_3), carbon monoxide (CO), sulfur dioxide (SO_2), and nitrogen dioxide (NO2) [6]. Among these, the concentration of ground-level ozone has become increasingly important due to its significant health risks compared to other pollutants [9]. Ozone, a secondary pollutant, is formed through the reaction of volatile organic compounds (VOC) and nitrogen oxides (NOx) with solar radiation. The harmful effects of ozone are well-documented, as seen in China, where 89,391 premature deaths were reported in 2014 due to chronic obstructive pulmonary disease (COPD) caused by ozone exposure [10]. Ozone exposure not only poses a threat to human health but also damages materials, plants, and trees [11]. Furthermore, research indicates that ozone levels and related mortality rates are highest in Asian and African cities, although even high-income regions like North America and the Asia Pacific experience significant ozone concentrations and associated health impacts [12].

The factors contributing to air pollution can be broadly categorized into socioeconomic and natural meteorological factors. Socioeconomic factors include government policies on pollution control, economic scale, city size, industrial organization, and traffic intensity. Natural meteorological factors encompass temperature, humidity, air pressure, wind speed, and precipitation. For instance, changes in meteorological conditions can significantly influence ozone formation. High temperatures, often associated with intense sunlight, promote the chemical reactions that produce ozone, leading to increased concentrations [13]. Conversely, lower temperatures can inhibit these reactions, reducing ozone levels. Wind patterns also play a crucial role in the dispersion of ozone and its precursors, affecting pollution levels across different regions.

In Malaysia, the increasing concentrations of ozone are particularly concerning. Data from the Department of

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume IX Issue IX September 2025

Environment (DOE) in 2021 revealed a 28.15% increase in the number of vehicles compared to 2021, while emissions from power plants, a significant source of NO2, recorded the highest readings at 67%. Although there was a slight decrease in the annual average concentration of ozone compared to 2021, the levels remain concerning when compared to the Malaysian Ambient Air Quality Standards (MAAQS) [6].

The characteristics of ground-level ozone are heavily influenced by meteorological factors. For example, temperature plays a significant role in ozone dispersion. Low temperatures reduce the chemical reactions necessary for ozone formation, while high temperatures, coupled with increased sunlight, significantly boost ozone production. Studies have shown that ozone levels peak around 2:00 PM when human activities, motor vehicle emissions, and sunlight are at their highest, and then gradually decrease by 9:00 PM as solar radiation diminishes [14]. Wind also acts as a transporter of ozone and its precursors, further affecting pollution levels depending on wind speed and direction. The DOE's 2022 report noted that ozone concentrations were particularly high in suburban areas due to the downwind effects of pollutants from industrial and vehicular emissions.

Effective air quality monitoring is crucial to addressing the health impacts of air pollution. In Malaysia, there are 65 air quality monitoring stations and three mobile monitoring stations [6]. However, the vast amount of data generated can make it challenging to accurately interpret the current air quality status. To improve monitoring efficiency and reduce costs, air quality monitoring stations can be grouped into clusters based on similar characteristics. This approach not only enhances the management of public health risks but also optimizes the use of resources.

Clustering analysis is a valuable technique for organizing monitoring stations into smaller groups based on shared characteristics. This method allows for a more efficient analysis of air quality data, making it easier to identify patterns and assess the status of air pollution [15].

In conclusion, air pollution remains a critical global challenge, with ozone pollution posing significant risks to both human health and the environment. Understanding the factors that contribute to ozone formation and dispersion, as well as employing effective monitoring and clustering techniques, is essential for mitigating these risks and improving air quality management.

METHODS AND STUDY AREA

A. Data and Study Areas

Malaysia's air quality is meticulously monitored by 65 stations distributed across the country. These stations, managed by the Department of Environment (DOE), are strategically classified according to land use categories, including industrial, rural, urban, and suburban areas. The classification helps to provide a comprehensive overview of air quality across diverse environments, reflecting the different sources and levels of pollution.

For the analysis, secondary data was obtained from the Department of Environment Malaysia. However, it is important to note that data from 19 of these stations were excluded from the analysis due to incomplete records. As a result, the study focused on the hourly average concentrations of ground-level ozone (O_3) from the remaining 46 monitoring stations only for 2022. Data before 2022 was ignored due to COVID19 and the data might be invalid. This data provides valuable insights into the spatial and temporal variations of ozone levels across different regions in Malaysia, helping to understand the impact of various factors such as land use, traffic, industrial activity, and meteorological conditions on air quality.

B. Methodology

In this study, descriptive statistics are employed to summarize and provide insights into the dataset's key characteristics, focusing on measures such as mean, standard deviation, skewness, and kurtosis. The mean serves as an indicator of central tendency, reflecting the average value around which data points cluster. The standard deviation measures the degree of dispersion or variability in the dataset, with larger values indicating greater spread. Skewness examines the asymmetry of the data distribution, where values near zero denote a symmetrical distribution, and positive or negative skewness suggests longer tails on the right or left side, respectively. Kurtosis assesses the "tailedness" or peakedness of the distribution, with higher kurtosis indicating sharper peaks and heavier tails, and lower kurtosis suggesting flatter distributions. These descriptive statistics provide an essential understanding of the data's characteristics, guiding further analytical processes [16].

In air pollution research, cluster analysis is a widely used technique for grouping data points with similar characteristics [15]. Among the various methods available, k-means and agglomerative hierarchical cluster analysis (AHC) are particularly popular. In this study, AHC was employed to classify 65 air quality monitoring stations across Malaysia into smaller, more manageable groups based on their ground-level ozone (O3) characteristics.

Agglomerative hierarchical cluster analysis begins by treating each monitoring station as its own cluster. The AHC analysis in this study was conducted using Ward's Method, a technique that measures the dissimilarity, or distance, between clusters using Euclidean distance. Ward's Method operates by initially considering each monitoring station as an individual cluster. The Euclidean distance between each pair of clusters is then calculated, and the two clusters with the smallest distance (i.e., the most similar clusters) are merged into a single cluster. This process of calculating distances and merging clusters is repeated iteratively until all stations are combined into a single cluster.

The process and results of Ward's Method can be visually represented in a dendrogram, a tree-like diagram that illustrates the sequence in which clusters are merged. The Euclidean distance, which quantifies the similarity between clusters, is a critical component of this method and is calculated using the following equation:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

This equation represents the distance d(x,y) between two points x and y across n dimensions (in this case, the characteristics of ground-level ozone). The smaller the Euclidean distance, the more similar the clusters, guiding the merging process in Ward's Method. Through this approach, the study aims to group the monitoring stations in a way that reflects their similarities in air quality characteristics, allowing for more efficient and targeted analysis of ozone pollution patterns across Malaysia.

RESULTS AND DISCUSSION

In this study, the agglomerative hierarchical cluster (AHC) method was applied to analyze the hourly average concentrations of ground-level ozone (O3) collected from 46 air quality monitoring stations across Malaysia for the years of 2022. The results of the AHC are represented in the form of a dendrogram, which illustrates the hierarchical relationships and dissimilarities between the clusters formed. The dendrogram groups of monitoring stations that share similar characteristics, clustering them together based on the ozone concentration levels observed at each station.

From the analysis of the data collected, the monitoring stations were classified into three main clusters, which were labeled as High Pollution Region (HPR), Medium Pollution Region (MPR), and Low Pollution Region (LPR). This classification helps to categorize the stations according to the severity of ozone pollution they recorded, providing a clear understanding of the regional distribution of air quality issues. Figure 1 and Table 1 provide a detailed classification of monitoring stations based on the hourly average ozone concentrations using the AHC method for the year 2022. According to the analysis, the stations were categorized into three clusters: High Pollution Region (HPR), Medium Pollution Region (MPR), and Low Pollution Region (LPR). In 2022, nine stations were classified under HPR, 20 under MPR, and 16 under LPR.

The geographical distribution of these stations reveals that most of the HPR stations are located in the central region of Peninsular Malaysia, with a few in the northern region. The stations in the MPR are more widely spread, covering all regions of Peninsular Malaysia. In contrast, the LPR stations are primarily located in Sabah, Sarawak, and the eastern region of Peninsular Malaysia, indicating relatively better air quality in these areas. Table 2 shows the descriptive statistics for ozone concentrations in these regions for 2022. The mean hourly average ozone concentration was recorded at 0.01244 ppm for LPR, 0.015773 ppm for MPR, and 0.018442 ppm for HPR. These statistics indicate that the HPR stations, although fewer in number, recorded the highest average ozone concentrations, reflecting more severe pollution levels.

The land use classification of these stations further provides insights into the pollution levels. In the HPR, five stations are located in suburban areas, and four in urban areas, indicating that both suburban and urban regions in these clusters are experiencing high pollution levels. In the MPR, a majority of the stations (17) are situated in suburban areas, with five in urban areas, making it the region with the greatest number of stations. Lastly, in the LPR, 13 stations are in suburban areas, two in urban areas, and one in a background area, emphasizing that suburban regions play a significant role in the overall air quality assessment.

This detailed classification and analysis help in understanding the distribution of ozone pollution across different regions and land use categories, providing a basis for targeted air quality management strategies in Malaysia.

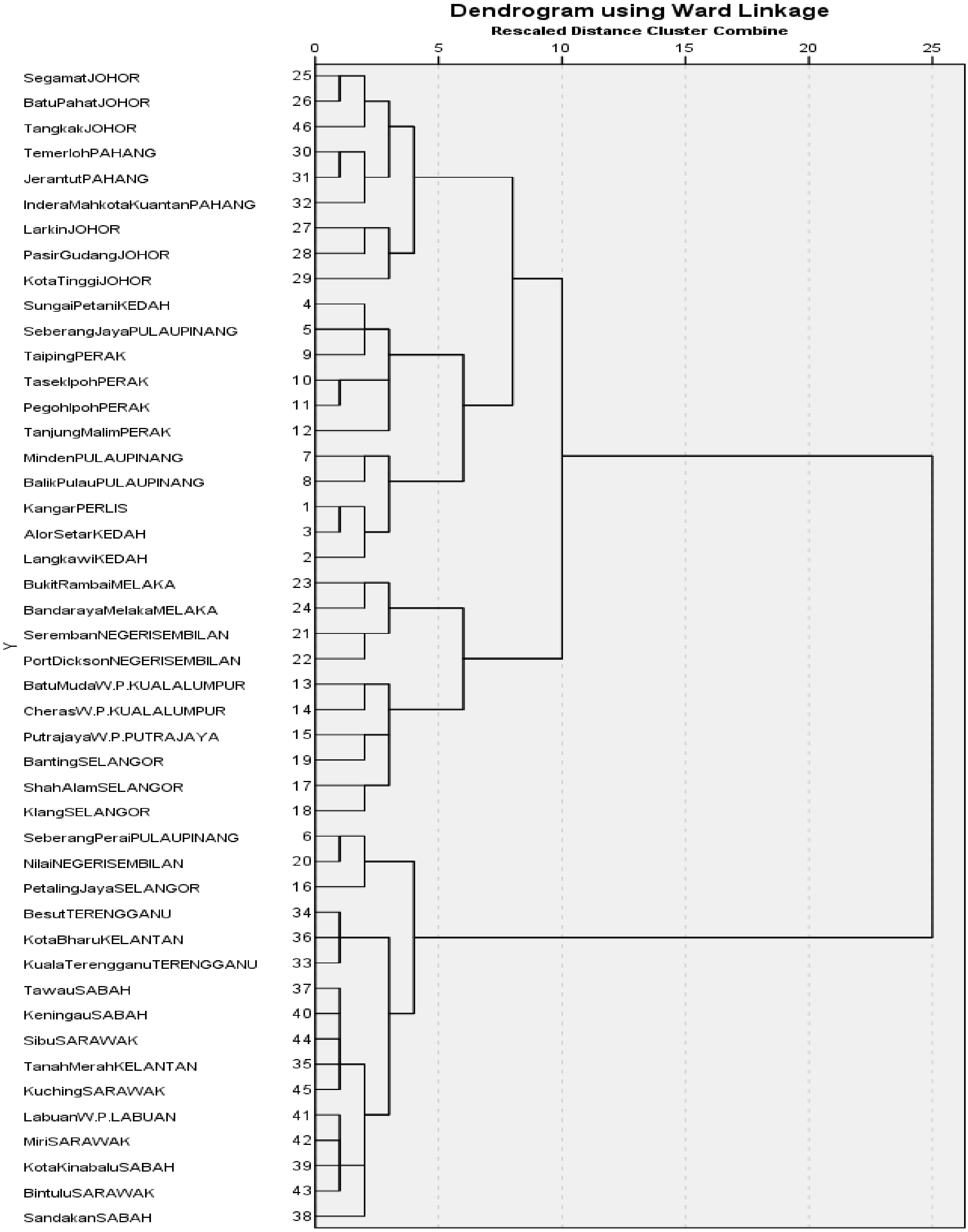

Figure 1. Stations cluster

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume IX Issue IX September 2025

Table I List Of Stations Based On High Pollution Regions (Hpr), Medium Pollution Regions (Mpr) And Low Pollution Regions (Lpr) 2022

CLUSTER 1 (HPR)	CLUSTER 2 (MPR)	CLUSTER 3 (LPR)
Bandaraya Melaka, Melaka	Alor Setar, Kedah	Seberang Perai, P. Pinang
Seremban, Negeri Sembilan	Balik Pulau, P.Pinang	Nilai, Negeri Sembilan
Port Dickson, Negeri Sembilan	Batu Pahat, Johor	Petaling Jaya, Selangor
Batu Muda, W.P. Kuala Lumpur	Bukit Rambai, Melaka	Besut, Terengganu
Cheras, W.P. Kuala Lumpur	Cheras, W.P. Kuala Lumpur	Kota Bharu, Kelantan
Putrajaya, W.P. Putrajaya	Indera Mahkota Kuantan, Pahang	Kuala Terengganu, Terengganu
Banting, Selangor	Jerantut, Pahang	Tawau, Sabah
Shah Alam, Selangor	Kangar, Perlis	Keningau, Sabah
Klang, Selangor	Kota Tinggi, Johor	Sibu, Sarawak
	Langkawi, Kedah	Tanah Merah, Kelantan
	Larkin, Johor	Kuching, Sarawak
	Minden, P.Pinang	Labuan, W.P. Labuan
	Pasir Gudang, Johor	Miri, Sarawak
	Pegoh Ipoh, Perak	Kota Kinabalu, Sabah
	Seberang Jaya, P.Pinang	Bintulu, Sarawak
	Segamat, Johor	Sandakan, Sabah
	Sungai Petani, Kedah	
	Taiping, Perak	
	Tangkak, Johor	
	Tanjung Malim, Perak	
	Tasek Ipoh, Perak	
	Temerloh, Pahang	

Table 2 Descriptive Statistics

	LPR	*MPR*	*HPR*
Mean	0.01244	0.015773	0.018442
Standard Deviation	0.002056	0.002864	0.003785
Kurtosis	-1.41271	-0.78895	2.07734
Skewness	-0.02127	-0.43766	0.331262
Minimum	0.00965	0.010089	0.011868
Maximum	0.015488	0.019997	0.025921

CONCLUSIONS

Based on the findings, it appears that the categories of monitoring stations set by the Department of Environment do not necessarily reflect the actual readings of O_3 concentrations. This discrepancy suggests that the categorization of these stations might have been determined based on factors beyond just ozone levels, such as land use patterns and the presence of other pollutants. These additional considerations could explain why the current classification system does not align perfectly with the observed ozone pollution data. The persistent high pollution levels in HPR areas could be attributed to ongoing urbanization, industrial activities, and vehicular emissions, which are significant contributors to ozone formation. This observation underscores the need for a more refined categorization approach that considers the specific pollutants of concern, such as O3, to better reflect the actual air quality conditions at each monitoring station. It also highlights the importance of focused air quality management and mitigation strategies in regions like Melaka, Negeri Sembilan, Kuala Lumpur and Selangor, where high ozone levels have been consistently recorded.

ACKNOWLEDGMENT

Special thanks to Universiti Sains Malaysia for Internationalisation Incentive Scheme provided (R502-KR-ARP004-00AUPRM003-K134)

REFERENCES

1. World Health Organization. (2023). Air Pollution. World Health Organization; World Health Organization. Retrieved from https://www.who.int/health-topics/air-pollution#tab=tab_1
2. Rani, N. L. A., Azid, A., Khalit, S. I., Gasim, M. B., & Juahir, H. (2018). Selected Malaysia air quality pollutants assessment using chemometrics techniques. Journal of Fundamental andApplied Sciences, 9(2S), 335. https://doi.org/10.4314/jfas.v9i2s.23
3. Mohd Halim, N.F., Mohd Zahid, A.Z., Salleh, M.Z.M., Abu Bakar, A.A. (2022). Air Quality Status During Pandemic Covid 19 in Urban and Sub-Urban Area in Malaysia, IOP Conf Series : Earth & Environment Science, 1019.
4. Chen, H., Oliver, B.G., Pant, A., Olivera, A., Poronnik, P., Pollock, C.A., Saad, S. (2022). Effects of air pollution on human health – Mechanistic evidence suggested by in vitro and in vivo modelling, Environmental Research, 212.
5. Centre for Research on Energy and Clean Air (2022) The Health & Economic Impact of Ambient Air Quality in Malaysia.
6. Department of Environment (2022). Environment Quality Report 2022, Ministry of NaturalResources & Environment http://www.doe.gov.my/
7. Jaffar, M.I., Hamid, H.A., Yunus, R., Raffee, A.F. (2018) Fitting statistical distribution on air pollution : An Overview, International Journal of Engineering and Technology, 7(23).
8. Babatola, S.S. (2021) Global burden of disease attributable to air pollution, J Public Health Afr, 9(3).
9. Sicard, P. (2021) Ground-level ozone over time : An observation-based global overview, Current Option in Environmental Science & Health, 29.
10. Xu, X. (2021) Recent advances in studies of ozone pollution and impacts in China : A short review, Current Opinion in Environmental Sciences & Health, 19.
11. Ramya, A., Dhevagi, P., Poornima, R., Avudainayagam, S., Watanabe, M., Agathokleous, E. (2023) Effect of ozone stress on crop productivity : A threat to food security, Environmental Research, 236(2).
12. Malashock, D., DeLang, M., Becker, J. S., Serre, M. L., J. Jason West, Chang, K., Cooper, O.R., & Anenberg, S. C. (2022). Estimates of ozone concentrations and attributable mortality inurban, peri-urban and rural areas worldwide in 2019. Environmental Research Letters, 17(5).
13. Li, K., Chen, L., Ying, F., White, S.J., Jang, C., Wu, X., Gao, X., Hong, S., Shen, J., Azzi, M., Chen, K. (2017) Meteorological and chemical impacts on ozone formation : A case study in Hang Zhou China, Atmospheric Research, 196.
14. Raffee, A. F., Hamid, H. A., Rahmat, S. N., & Jaffar, M. I. (2022). The cause-and-effect analysis of ground level ozone (O3), air pollutants and meteorological parameters using thecausal relationship approach. Journal of Engineering Research.
15. Govender, P., Sivakumar, V. (2020) Application of k-means and hierarchical clustering techniques for analysis of air pollution : A review (1980 – 2019), Atmospheric Pollution Research, 11(1).
16. Sahoo, J. (2022). Introduction to Statistical Data Analysis: Concepts and Applications. *Springer Nature*.