Universiti Teknologi MARA

I-URIID 2025

INTERNATIONAL UNDERGRADUATE RESEARCH INNOVATION, INVENTION AND DESIGN

UNDERGRADUATES PIONEERING TOMORROW'S BREAKTHROUGH

E-BOOK OF EXTENDED ABSTRACTS

MULTIVARIATE ANALYSIS OF AIR POLLUTION IN SOUTHERN PENINSULAR MALAYSIA : PRINCIPAL COMPONENT ANALYSIS AND FACTOR ANALYSIS APPROACH

Sofea Nor Shamsul Harisah[1], Nur Hanis Najwa Mohd Helmi[1], Az'lina Abdul Hadi[1*], Nornadiah Mohd Razali[1], and Nur Niswah Naslina Mohd Azid@Maarof[2]

[1] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan Branch Seremban Campus, Persiaran Seremban Tiga 1, Seremban 3, 70300 Seremban, Negeri Sembilan, Malaysia
[2] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA(UiTM) Cawangan Kelantan Kampus Kota Bharu, Lembah Sireh, 15050 Kota Bharu Kelantan Darul Naim, Malaysia

*azlinahadi@uitm.edu.my

ABSTRACT

Air pollution in Southern Peninsular Malaysia has become a critical environmental issue, exacerbated by rapid urbanization and industrial activities. This study aims to identify the main contributors of air pollution and construct underlying factors of air pollution in Southern Peninsular Malaysia using Principal Component Analysis (PCA) and Factor Analysis (FA), respectively. The data used in this study were a daily dataset that covered the period from 1 January 2023 to 31 December 2023. Eleven air monitoring stations from three region (Negeri Sembilan, Melaka and Johor) were used to record concentrations of six air pollutants which were CO, O_3, SO_2, NO_2, PM_{10}, and $PM_{2.5}$ for this study. Two principal components were obtained after performing PCA, with the cumulative percentage of variance for both principal components being 64.274%. The first principal components consisted of PM_{10}, $PM_{2.5}$, NO_2, and CO, while the second principal component consisted of SO_2 and O_3. All the assumptions related to FA were conducted and found to be complied. Two factors were extracted using PCA, and after varimax rotation, the variances of factor 1 and 2 accounted for 43.422% and 20.852% respectively, with the cumulative 64.274%. Factor 1 consisted of PM_{10}, $PM_{2.5}$, NO_2, and CO, whereas factor 2 consisted of SO_2 and O_3. It can be concluded that factor 1 loaded together since the pollutants in factor 1 shared the same source which is vehicle emissions. Since they are linked and involved in the photochemical process, SO_2 and O_3 are loaded together in factor 2. In conclusion, the main contributors to PC1 are PM_{10}, $PM_{2.5}$, NO_2, and CO, while PC2 is SO_2 and O_3. The grouping of the pollutants was the same in FA, where factor 1, the air pollutants shared by the same sources which was vehicle emissions. While factor 2, the pollutants were linked in the photochemical process.

Keywords: Principal Component Analysis (PCA), Factor Analysis (FA), air pollution, principal component, factor

INTRODUCTION

Air pollution can be described as the contamination of either indoor or outdoor air with various gases and particles that include sulfur dioxide (SO2), nitrogen dioxide (NO2), carbon monoxide (CO), ozone (O3), and particulate matter (PM2.5 and PM10) change the natural makeup of the air (WHO, 2019). Sulfur dioxide (SO_2) is a colourless air pollutant with a sharp, irritating odour, primarily emitted from industrial sources such as paper manufacture, petroleum and metal refining (DCCEEW, 2022). Nitrogen dioxide (NO2) is a significant air pollutant generated during high-temperature combustion in the air reacts with oxygen. Particulate matter is a combination of solid particles and liquid droplets in the air. According to the California Air Resources Board (2015), PM10 and PM2.5 from different emissions sources for PM2.5 omitted from the combustion of gasoline, oil, diesel fuel or wood. Meanwhile, PM10 is omitted from construction sites, agriculture, landfills, and waste burning. Carbon monoxide (CO) is a highly toxic, colourless, odourless, and flammable gas that results primarily from the incomplete combustion of carbon-containing fuels. Ozone (O3) is an air pollutant that not directly emitted into the air. It is formed through chemical reactions in atmosphere. The southern region of Peninsular Malaysia, which includes states such as Johor, Melaka, and Negeri Sembilan, plays a vital role in the nation's economic growth, hosting numerous industrial zones, urban centers, and busy transportation hubs. However, these developments have also contributed to significant air pollution challenges. The primary aim of this study is to identify the main contributors to air pollution and construct underlying pollution factors in Southern Peninsular Malaysia using Principal Component Analysis (PCA) and Factor Analysis (FA).

RESULTS AND DISCUSSION

The following results were based on daily air quality data that have been retrieved from Department of Environment between 1 January 2023 and 31 December 2023 from 11 monitoring stations across Johor, Melaka, and Negeri Sembilan. The dataset included concentrations of six pollutants: SO_2, NO_2, CO, O_3, PM_{10}, and $PM_{2.5}$. This study introduces a dual analytical approach, integrating PCA and FA to simultaneously identify dominant air pollutants in a high-growth industrial and urban region. In additional, this study provide a statistical basis for developing real-time monitoring dashboards and mobile applications that can predict and communicate pollution risks in future. By combining advanced multivariate analysis with practical application potential, the study bridges the gap between academic research and real-world environmental management tools, a perspective not commonly applied in previous regional studies. Table 1 shows the list of eigenvalues associated with each linear component before and after extraction. Principal components 1 and 2 should be retained since the eigenvalues of PC1 are 2.631 and PC2 are 1.225, with both eigenvalues greater than one. The results of cumulative variance showed that two principals components should be retained since the cumulative variance of PC2 is 64.274%, which is more than 60%.

Table 1: Total Varianced for Principal Component Analysis

Component	Initial Eigenvalues			Extraction Sum of Squared Loadings		
	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative (%)
1	2.631	43.857	43.857	2.631	43.857	43.857
2	1.225	20.417	64.274	1.225	20.417	64.274
3	0.937	15.619	79.893			
4	0.724	12.066	91.959			
5	0.436	7.272	99.232			
6	0.046	0.768	100.00			

Table 2 represents the factor loadings and correlation that consisted of 2 components. The first component consisted of PM_{10}, $PM_{2.5}$, NO_2 and CO. PM_{10} had the highest contribution to the first component with strong negative loading with -0.940, followed by $PM_{2.5}$ with -0.923. PM_{10} and $PM_{2.5}$ were known as pollutants with similar sources from vehicle emissions, and various industrial processes and construction sites. Among these, vehicle emissions were the most contributor for PM_{10} and $PM_{2.5}$ (Azliyana, 2020). The increased trend in new vehicle numbers across three regions (Negeri Sembilan, Melaka and Johor) showed the fact that PM_{10} and $PM_{2.5}$ were the main contributor of air pollution in southern peninsular Malaysia due to vehicle emissions. The second principal component consisted of only 2 pollutants, SO_2 and O_3. SO_2, with a moderate value of -0.416, separates the pollutant from the first principal component. This could be because SO_2 was lower or less dominant in its sources, such as in the combustion of fossil fuels and industrial processes, compared to other pollutants. O_3 was separated from other pollutants because O_3 was a secondary pollutant unlike other pollutants that directly emitted from their sources.

Table 2: Factor Loading and Factor Analysis

Variables	Factor Loading Before Rotation		Factor Loading After Rotation	
	Component		Component	
	1	2	1	2
PM_{10}	0.940	0.126	**0.948**	0.003
$PM_{2.5}$	0.923	0.214	**0.944**	-0.086
SO_2	0.209	-0.416	0.150	**0.440**
NO_2	0.666	-0.493	**0.593**	0.579
O_3	0.170	0.861	0.285	**-0.830**
CO	0.617	-0.77	**0.600**	0.160

For factor analysis (FA) the results of the KMO test and Barlett′s test. The value of KMO was 0.628, which is in the mediocre category since the value was between 0.60 to 0.69 and was considered acceptable. The p-value of Barlett′s test was 0.000. Hence, adequate to conduct FA. Principal component analysis (PCA) was used as an extraction method in this analysis. After rotation, factor 1 accounted for 43.422% and for factor 2 it was 20.852%. Hence, the cumulative variance of the 2 factors was 64.274%.

Table 3: Factor Loading Before Rotation and After Rotation

	Factor Loading		Correlation	
	1	2	1	2
PM_{10}	-0.579	0.114	**-0.940**	0.126
$PM_{2.5}$	-0.569	0.193	**-0.923**	0.214
SO_2	-0.129	-0.376	-0.209	**-0.416**
NO_2	-0.411	-0.445	**-0.666**	-0.493
O_3	-0.105	0.778	-0.170	**0.861**
CO	-0.380	-0.070	**-0.617**	-0.077

Table 3 shows the factor loading after extraction in first factor contained PM_{10}, $PM_{2.5}$, NO_2 and CO since the loadings of variables exceed 0.4. The first components had a strong correlation of air pollutants (PM_{10}, $PM_{2.5}$, NO_2 and CO) due to a similar source which was vehicle emission. The second component was made up of O_3 and SO_2. SO_2 and O_3 loading together in second factor indicated there was significant correlation between them that most probably shared atmospheric processes that influence their concentration (Xiofan et al., 2024).

CONCLUSION

In conclusion, principal component analysis (PCA) and factor analysis (FA) were applied in this study to achieve their objectives. Two principal components were obtained from those methods, with 64.274% of the cumulative percentage of variance. The first component comprised PM10, PM2.5, NO2 and CO, while the second comprised SO2 and O3. The main contributors were PM10, PM2.5, NO2 and CO for the first principal component and SO2 and O3 for the second principal component. FA also conducted this study to construct underlying air pollution factors in Southern Peninsular Malaysia. The value of KMO was 0.628, which was acceptable, and the p-value of Barlett′s test was 0.000, which was significant, hence adequate to conduct FA. The first factor consisted of PM10, PM2.5, NO2 and CO, they loaded together because they were from the same sources, which were vehicle emissions. The second factor, SO2 and O3, are both pollutants in the photochemical process. The outcomes of this study are highly relevant to policymakers, public health agencies, and local communities. This study also brings benefits to both users and society by identifying the dominant pollutants, which helps authorities in making decisions for preventive health measures. These findings raise community awareness of the health risk associated with exposure to high level of pollutants. The study also supports Malaysia′s efforts to meet United Nations Sustainable Development Goals (SDGs), particularly Goal 3: Good Health and Well-Being and Goal 11: Sustainable Cities and Communities. The results of this study may also guide the Department of Environment (DOE) and government in enforcing stricter emission control on industries and vehicles. In terms of commercialization, the PCA and FA models developed in this study can be transformed into practical tools by integrating into real time air quality monitoring dashboards or mobile applications that alert the public about pollution risks in specific areas. This study not only contributes to academic knowledge but also provides benefits to society and shows potential for commercialization.

ACKNOWLEDGEMENT

We wish to express our profound appreciation to the Department of Environment for generously providing the dataset utilized in this study. Furthermore, we extend our sincere gratitude to the KPP Faculty of Computer and Mathematical Sciences for their continuous support, motivation, and encouragement, which greatly facilitated our participation in this innovation competition. (Project ID:ST-15)

REFERENCES

Azliyana, A., Abdul Halim, N. D., Othman, M., Latif, M. T., Juneng, L., Sofwan, N. M., Stocker, J., & Johnson, K. (2020). Highly spatially resolved emission inventory of selected air pollutants in Kuala Lumpur's urban environment. *Atmospheric Pollution Research*, *12*(2), 12–22. https://doi.org/10.1016/j.apr.2020.10.004

California Air Resources Board. (2022). *Combustion Pollutants & Indoor Air Quality | California Air Resources Board*. Ww2.Arb.ca.gov. htt ps://ww2.arb.ca.gov/resources/documents/combustion-pollutants-indoor-air-quality

DCCEEW.(2022).Dcceew.gov.au.https://www.dcceew.gov.au/environment/protection/npi/substances/fact-sheets/sulfur-dioxide.

WHO. (2019). What is Air Pollution. *World HealthOrganization*. https://cdn.who.int/media/docs/default-source/searo/wsh-och-searo/what-is-air-pollution-2019.pdf?sfvrsn=6dcc13ee_2

Xiaofan Lv, Lily, M., Tasheh, S. N., Ghogomu, J. N., Du, L., & Narcisse Tsona Tchinda. (2024). Enhanced Sulfate Formation from Gas-Phase SO2 Oxidation in Non−•OH−Radical Environments. *Atmosphere*, *15*(1), 64–64. https://doi.org/10.3390/atmos15010064