Environ Monit Assess (2025) 197:474
https://doi.org/10.1007/s10661-025-13927-5

RESEARCH

Urban air pollutant mapping and tracing using multi-points in situ measurements combined with clustering and trajectory analysis

Muhammad Rizky Mulyana · Yosi Aristiawan · Chairil Linggabinangkit · Rudi Anggoro Samodro · Hafiizh Prasetia · Nidaa Fauziyyah · Nikolas Jalu Padma Iswara · Adindra Vickar Ega · Yonan Prihhapso

Received: 14 October 2024 / Accepted: 17 March 2025 / Published online: 26 March 2025

Abstract Air pollution poses significant risks, particularly in developing countries where rapid urbanization exacerbates pollutant emissions. These pollutants impact local populations and contribute to global air quality challenges through long-range transport. Despite numerous studies, comprehensive data on pollutant characteristic in urban areas remain limited by the availability of air quality monitoring stations in emerging urban regions, especially at developing countries. This study addresses these gaps by employing a novel approach that combines multi-points in situ air quality measurements with clustering and back trajectory analysis to map and trace pollution sources across diverse urban environments. The use of low-cost and mid-cost portable instruments allows for resource-efficient data collection, enhancing the ability to identify pollution hotspots without requiring extensive infrastructure. The analysis revealed two distinct pollutant clusters: aerosol pollutants dominated in residential areas, while gaseous pollutants were more prevalent near traffic-heavy and construction areas. Although low-cost sensors have limited capabilities and should not be used for regulatory purposes, this methodology provides a scalable complementary addition to regular air quality monitoring and offers valuable insights into pollution source attribution, particularly in developing countries where resources for environmental monitoring are limited.

M. R. Mulyana (✉) · R. A. Samodro · A. V. Ega · Y. Prihhapso
Research Centre for Testing Technology and Standards, National Research and Innovation Agency (BRIN), South Tangerang 15314, Indonesia
e-mail: mriz002@brin.go.id

R. A. Samodro
e-mail: rrud001@brin.go.id

A. V. Ega
e-mail: adin009@brin.go.id

Y. Prihhapso
e-mail: yona001@brin.go.id

Y. Aristiawan · H. Prasetia
Research Centre for Chemistry, National Research and Innovation Agency (BRIN), South Tangerang 15314, Indonesia
e-mail: yosi002@brin.go.id

H. Prasetia
e-mail: hafi001@brin.go.id

C. Linggabinangkit
Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
e-mail: chairil.linggabinangkit@gmail.com

N. Fauziyyah · N. J. P. Iswara
PT. Nafas Aplikasi Indonesia, Infiniti Office, Jakarta 12210, Indonesia
e-mail: nidaa@nafas.co.id

N. J. P. Iswara
e-mail: nikolas@nafas.co.id

Keywords Pollutant mapping · Pollution source tracing · Low-cost and mid-cost sensors

Introduction

Air pollution has become a significant issue in many regions across the globe, posing serious risks to human health and the environment alike. Rapid industrialization, urbanization, and the increasing reliance on fossil fuels have contributed to the alarming rise in air pollutants, particularly in highly populated cities (Usman & Balsalobre-Lorente, 2022). As populations grow and urban areas expand, the demand for energy, transportation, and industrial activities increases, leading to the release of various pollutants into the atmosphere. These pollutants, originating from both natural and human-made sources, contribute to the degradation of air quality and have widespread consequences (Chandra Voumik et al., 2023). According to Health Effects Institute new State of Global Air 2024 Report, air pollution now ranks as the second leading risk factor for death, accounting for 8.1 million deaths globally in 2021 (HEI, 2024).

The health risks of air pollution are well-documented, with pollutants in both aerosol and gaseous phases linked to respiratory and cardiovascular diseases. Particulate matter (PM), particularly fine particles (<2.5 μm, $PM_{2.5}$), can penetrate deep into the lungs, causing inflammation, reduced lung function, and long-term conditions like asthma, bronchitis, lung cancer, and heart disease (Arias-Pérez et al., 2020; Karimi & Samadi, 2024; Thangavel et al., 2022). Gaseous pollutants also have severe effects. Ozone (O_3) irritates the respiratory system and worsens chronic conditions like asthma, while carbon monoxide (CO) impairs oxygen transport in the blood, leading to headaches, dizziness, and fatalities in extreme cases (Donzelli & Suarez-Varela, 2024). (Chen et al., 2021; Cheng et al., 2019). Sulfur dioxide (SO_2) and nitrogen oxides (NO_x) can exacerbate respiratory diseases and contribute to smog (Peel et al., 2013; Persinger et al., 2002; Zhou et al., 2022). These pollutants not only harm human health but also accelerate climate change. Carbon dioxide (CO_2), the main driver of global warming, reached record levels in 2023, directly correlating with rising temperatures (NOAA, 2024). Urban environments, where pollutants and CO_2 are co-emitted from sources like transportation and industry, highlight the intertwined nature of air pollution and climate change (IPCC, 2021; Tollefson, 2021). Efforts to reduce air pollution can thus also mitigate climate change impacts (WHO, 2024).

Due to their severely harmful effects, numerous studies have attempted to characterize urban air quality; however, these efforts are often limited by the availability of stationary monitoring systems, which provide insufficient spatial coverage and are resource-intensive to operate. Previous study has characterized the urban air quality in in Uganda, although limited to $PM_{2.5}$, NO_2, and O_3 (Okure et al., 2022) More studies about air pollutants characteristic were also conducted in India but also limited to some pollutants such as particulate matter, CO_2, and formaldehyde (Gautam et al., 2021), or CO_2, CH_4, and CO (Komal et al., 2024). The studied locations were also limited to traffic interjunctions, or a campus in rural area that could have different results in comparison to urban cities. Other studies characterized ambient air quality on two South African urban sites; however, there was a limitation in which data for NO_2 and O_3 were unavailable in one site thus causing difficulty for a comparison between the two study sites (Matandirotya et al., 2023). Most of these studies utilize conventional air quality monitoring systems, which require significant resources and stationary equipment to cover multiple areas. Important findings have been yielded but were often constrained by a limited focus on specific pollutants or locations. Comprehensive profiling of both gaseous and aerosol pollutants across multiple urban environments remains scarce, especially in developing nations.

This study tackles gaps in urban air quality research by introducing a practical and adaptable approach that combines portable in situ air quality measurements with clustering and trajectory analysis. Unlike traditional monitoring systems, which rely on stationary, resource-intensive equipment, this method allows for quick, cost-effective assessments across different urban locations. However, it is important to recognize that this approach provides only a snapshot of air pollution levels at specific times and places. Given the high variability of urban pollution, long-term monitoring remains essential for capturing trends and accurately linking emissions to sources. Rather than replacing conventional monitoring networks, this "snapshot" method can serve as a preliminary tool to identify pollution hotspots and guide

the placement of permanent air quality stations, particularly in resource-constrained areas such as South Tangerang of Indonesia. By mapping pollution from both local and transported sources, this study offers valuable insights for early-stage decision-making, helping policymakers and urban planners prioritize areas that require more in-depth monitoring and intervention.

Materials and methods

This study utilizes portable low-cost and mid-cost air quality measuring instruments, offering a flexible and resource-efficient alternative to traditional stationary monitoring methods. The portability of these instruments allows for complementary analysis alongside fixed monitoring stations across multiple urban locations. By utilizing proper quality control and frequent validation, this approach can cover spatial gaps in urban area outside the coverage of regular air monitoring stations. The mobility of the instruments enables real-time data acquisition across varied locations, greatly improving the ability to map and trace pollution sources in complex urban settings.

More details on the portable measuring instruments transported to the target locations can be viewed in Table 1. Aerosol pollutants in the form of particulate, namely, PM2.5 and PM10, were measured using Air Quality All in One Sensor Type RD-ACCOH-L01 (Honde Technology, China) and RS485 PM2.5/PM10 Particle Detection Sensor (Air Quality Transmitter, China). Both sensors used light scattering to detect the pollutants.

On the other side, gaseous pollutants such as SO_2, NO_2, NO, CO, and CO_2 were measured using Graywolf Model DS II Probe (Graywolf Sensing Solutions, USA), also utilizing electrochemical sensor and NDIR for CO_2. In addition, the Air Quality All in One Sensor was also used to measure other gaseous pollutants like O_3, H_2S, and NMHC. For consistency in comparing aerosol and gaseous pollutants, all measurement readings were converted to $\mu g/m^3$, addressing the challenge of unit conversion between particulate and gaseous matter.

Aerosol pollutant measurements were ensured using factory-calibrated sensors, deployed in a co-location setup for consistency. For gaseous pollutants, the Graywolf DS II was calibrated at an Indonesian-accredited laboratory in the scope of ambient air testing. Lastly, the Air Quality All-in-One Sensor relied on its factory-issued calibration, covering O_3, H_2S, and NMHC.

Figure 1 summarizes the location used in this study. South Tangerang, part of the Greater Jakarta metropolitan area, is one of the most densely populated regions in Indonesia. This area was selected as the study site due to its reputation as one of the most polluted urban cities in Southeast Asian region, representing highly polluted city of a developing country. Several urban locations in South Tangerang were chosen for this study, including traffic intersections, main roadsides, residential areas near and away from roads, and construction sites. Each location is expected to

Table 1 Air pollutant measuring instrument specifications

No	Name/brand/type	Measurand	Measurement range	Measuring principle	Uncertainty
1	RS485 $PM_{2.5}$/PM_{10} Modbus Particle Detection Sensor	$PM_{2.5}$	0–1000 $\mu g/m^3$	Light scattering	±1 $\mu g/m^3$
		PM_{10}	0–1000 $\mu g/m^3$	Light scattering	±1 $\mu g/m^3$
2	Graywolf Sensing Solutions DS II Probe	SO_2	0–50 ppm	Electrochemical	±0.3 ppm
		NO_2	0–20 ppm	Electrochemical	±0.5 ppm
		NO	0–250 ppm	Electrochemical	±5% relative
		CO	0–850 ppm	Electrochemical	±0.6 ppm
		CO_2	0–10000 ppm	NDIR	±46 ppm
		H_2S	0–50 ppm	Electrochemical	±0.5 ppm
3	Honde Air Quality All in One Sensor Type RD-ACCOH-L01	PM_{10}	0–1000 $\mu g/m^3$	Light scattering	±10% relative
		PM_{10}	0–500 $\mu g/m^3$	Light scattering	±10% relative
		O_3	0–5 ppm	Electrochemical	±10% relative
		NMHC	0–20 $\mu g/m^3$	Electrochemical	±10% relative

Fig. 1 Map of measurement locations (taken from customized Google™ Earth)

be affected by different sources of pollution, such as vehicle emissions, local household activities (like open waste burning), and construction dust.

A snapshot of peak-time concentration, containing 20 sets of measurement, were taken in one day at each location; all of them are during working days. With temporal resolution of 1 data/minute, a total of 20 min were required for one experiment in one location, if no errors can be found within the datasets. To minimize the effect of extreme weather events to the measurement, experiment was conducted during sunny days on September 2024. All measurements were conducted on ground level with instruments setup on top of car roof.

Other details of the measurement locations are shown in Table 2. Air pollutant measurements were conducted during peak activity periods at each typical location. For traffic interconnection and main roadside, measurements were taken during mid-daytime, typically between 11 a.m. and 1 p.m., coinciding with

Table 2 Details of measurement locations

No	Typical area	Peak activity period	GPS coordinates	Temperature range (°C)*	Humidity range (%RH)*
1	Traffic interconnection	Noon	6°14′41.82″S 106°38′57.6594″E	32.2–33.9	40.32–51.53
2	Main roadside	Noon	6°19′5.9874″S 106°40′49.4754″E	30.4–31.9	54.61–62.71
3	Roadside residential	Afternoon	6°15′47.4834″S 106°41′18.78″E	30.0–31.4	55.41–60.07
4	Offroad residential	Afternoon	6°19′30.4674″S 106°45′14.58″E	29.8–33.3	55.2–69.28
5	Construction site	Afternoon	6°13′16.68″S 106°39′31.572″E	31.2–32.9	47.47–57.42

*Taken from OpenWeather™ API

peak traffic flow due to lunch breaks and prayer times. For roadside residential and off-road residential areas, measurements were conducted in the afternoon, during peak hours between 3 to 5 p.m. when most people return home from work. Lastly, measurements at the construction site were performed during regular working hours around 3 p.m. when traffic was minimal, and the majority of construction activities were ongoing.

Two primary data analysis methods were used for air pollutant mapping and tracing in this study. The first method, namely, clustering analysis, was performed using the Orange™ Data Mining software (Demšar et al., 2013). Initially, to ensure comparability across variables with different units and scales, all numerical data were normalized to a (0,1) range using min–max normalization in the Orange™ software. This process adjusts each value based on the minimum and maximum values in the dataset into the scale of zero to one, allowing features with larger numerical ranges to be scaled proportionally. Normalization prevents any single variable from dominating clustering and classification analyses, ensuring a more balanced interpretation of pollutant patterns across different locations.

The *K*-means clustering algorithm was then applied to group the locations based on their pollutant characteristics, resulting in clusters that represented different urban environments with similar pollution characteristics. The number of clusters was selected based on the most optimal silhouette score (*k*). In addition, clustering box plots were utilized to observe the distinction between clusters.

Afterwards, the geographical locations of the clustered areas were manually overlaid onto Google™ Maps by matching the latitude and longitude data of each location. This allowed for a clear visualization of how pollution sources are distributed across the city. This manual overlay on Google™ Maps helped in the spatial representation of cluster groups, providing insights into how different urban areas experience distinct pollution patterns.

The second method, which is back trajectory analysis, was conducted using Python™. This model traced the origins of air masses arriving at each monitoring site by calculating backward trajectories during measurement time. Meteorological data from the OpenWeather™ API, including parameters such as wind speed, wind direction, temperature, humidity, and pressure, were converted into radians and used to compute the displacements of air masses.

For each location, multiple trajectories were generated to account for variability in meteorological conditions. The displacements in both the latitude and longitude directions were computed iteratively for each time step, tracing the movement of the air masses backward in time. The calculated trajectory points were then used to determine the most likely origin of the pollutants affecting each site. The density of the trajectory paths was visualized using Gaussian kernel density estimation (KDE), providing a clear indication of the most probable origins of the pollutants affecting each location.

In contrast to the manual overlay used in the clustering analysis, the back trajectory results were overlaid onto customized Google™ Maps directly within the Python environment, using the Cartopy library to plot the trajectories and the pollutant density on a geographical map (Elson et al., 2024). This automatic overlay provided a precise spatial representation of the pollution sources, allowing for a more detailed analysis of external pollutant contributions.

By incorporating these two methods onto multipoint in situ air quality measurements, this study introduces a novel approach to urban air quality analysis, addressing the limitations of conventional stationary monitoring. This methodology allows for a more dynamic and flexible assessment of pollution sources and their effects on various urban environments.

Results and Discussion

Clustering analysis was applied to the measurement results from all locations using the Orange™ Data Mining software. For this analysis, all measurement data was normalized with the Preprocessing widget in the software prior to the clustering. Using *k*-means clustering algorithm, optimal silhouette score was achieved with $k = 2$. This value indicates that air pollutant characteristic in each target location can be distinctively clustered in two groups (Gustriansyah et al., 2020; Zhang et al., 2021).

Results of cluster analysis application to all measured pollutants can be seen in Table 3, showing the normalized concentration of pollutants in each cluster. The order of the mentioned pollutant was written following the classification by Orange™ Data

Table 3 Cluster analysis results for all measured pollutants

No*	Pollutant	Cluster 1	Cluster 2	Student's *t*
1	NO_2	0.043 ± 0.043	0.827 ± 0.230	5.751
2	$PM_{2.5}$	0.191 ± 0.230	0.948 ± 0.052	5.490
3	PM_{10}	0.208 ± 0.259	0.959 ± 0.041	4.939
4	NMHC	0.030 ± 0.030	0.572 ± 0.377	2.477
5	NO	0.394 ± 0.278	0.800 ± 0.200	1.900
6	SO_2	0.000 ± 0.000	0.500 ± 0.500	1.414
7	O_3	0.000 ± 0.000	0.456 ± 0.413	1.915
8	CO	0.115 ± 0.115	0.588 ± 0.427	1.821
9	CO_2	0.398 ± 0.398	0.669 ± 0.385	0.755
10	H_2S	0.244 ± 0.244	0.438 ± 0.410	0.666

*Number based on order of relevance to subgroups following Orange™ Data Mining

Mining, which starts from the most relevant pollutant down to the least ones. NO_2 was found to be the most distinctive pollutant which separates Cluster 1 from Cluster 2, followed by $PM_{2.5}$, indicating the existence of different air pollution sources on each cluster. On the other hand, CO_2 and H_2S were among the least relevant pollutant to each cluster. This can be interpreted as the possibility of similar sources of the lower relevance pollutants on both clusters.

Figure 2 represents the distribution of two most relevant pollutant concentrations (NO_2, and PM_2.5) across Clusters 1 and 2. The blue bars represent the normalized data range, with the blue line indicating the mean concentration within each cluster. The error bars correspond to the standard deviation of the data. N represents the number of samples included in the statistical analysis. The horizontal axis denotes the normalized concentration (0–1), while the vertical labels indicate the clusters

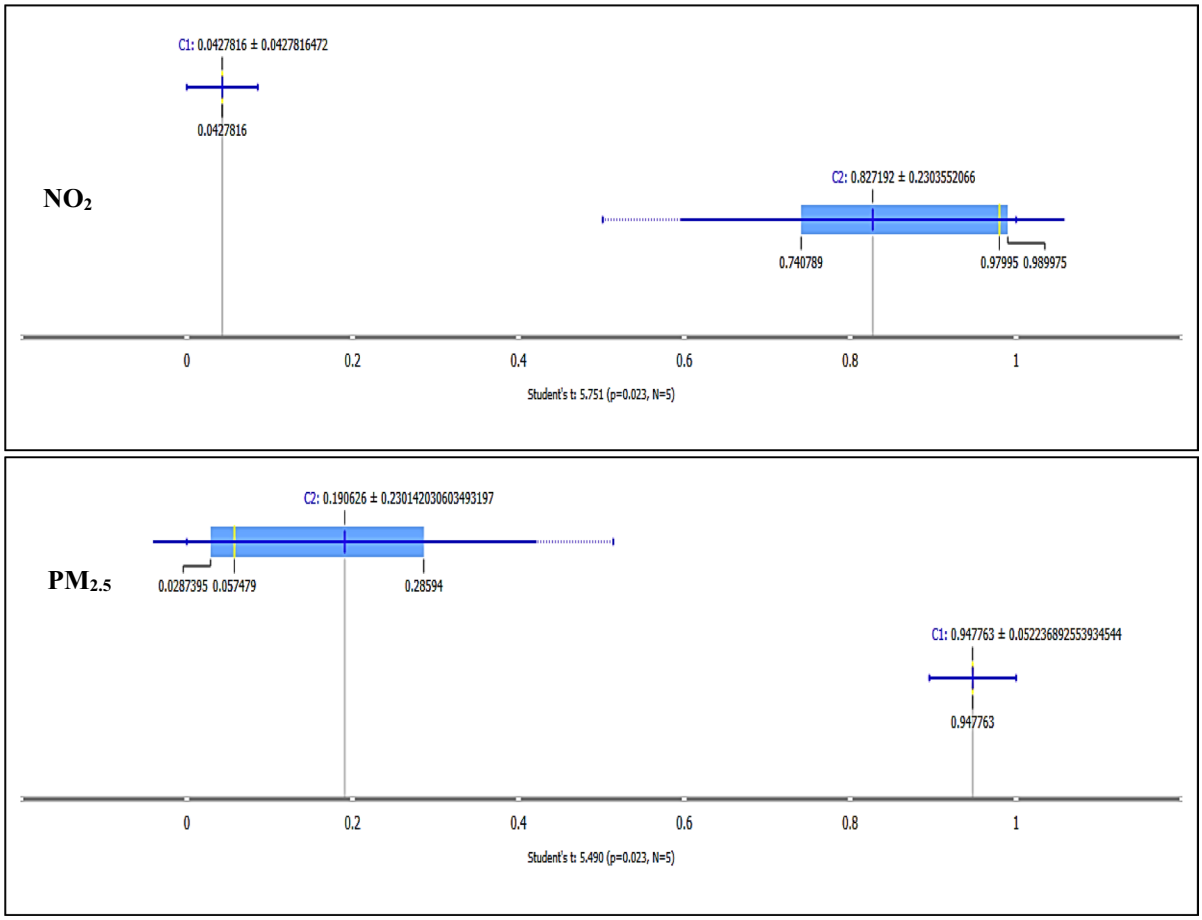

Fig. 2 Cluster box plot for two most relevant pollutants taken directly from Orange™ (more details can be observed on Table 3)

assigned based on the analysis. Statistical significance is tested using the Student's *t*-test, with the corresponding p-values displayed in each plot. The result of the clustering can be plotted with latitude and longitude as *y* and *x* axis, then overlaid to Google™ Maps as seen in Fig. 3 to observe the clustering result in actual geographical position.

Cluster 1 shows higher concentration range of $PM_{2.5}$ and PM_{10} compared to Cluster 2, as can be observed on Table 4. This could be attributed to the fact that residential areas may have more localized sources of aerosol pollutants such as open burning, cooking emissions, or other domestic activities (Aslam et al., 2020; Muyemeki et al., 2021; Yan et al., 2016). Dust emission from unpaved roads in

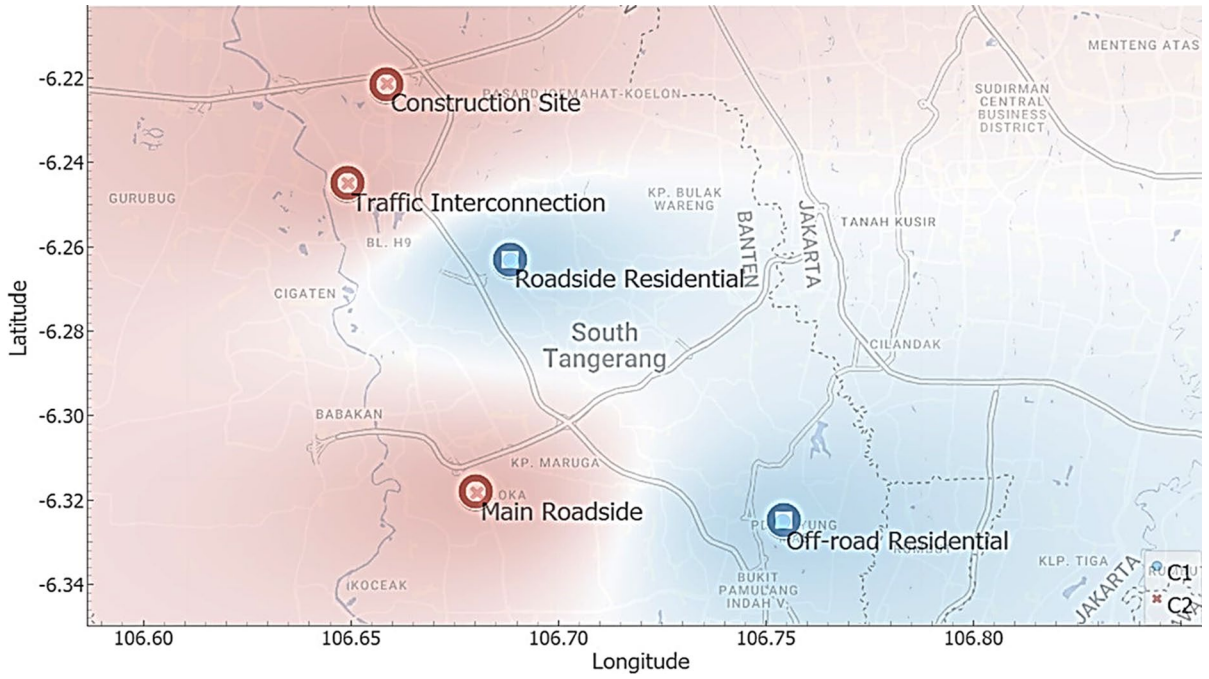

Fig. 3 Geographical map of air pollutant clusters (Cluster 1 in blue, Cluster 2 in red; background map taken from Google™ Maps)

Table 4 Concentration range of pollutants in each location

No*	Pollutant	Cluster 1 Concentration range (μg/m^3)		Cluster 2 Concentration range (μg/m^3)		
		Roadside residential	Offroad residential	Traffic interconnection	Main roadside	Construction site
1	$PM_{2.5}$	38.5–71.5	50–59.5	16–69.5	33–50	15–28
2	PM_{10}	45–78.5	56.5–63.5	16–71.5	38.5–36.5	15.5–31.5
3	NO_2	78.06–90.95	40.07–87.92	127.17–222.88	91.15–159.74	41.61–218.39
4	NMHC	0–76	0–136	0–482	0–744	0–89
5	O_3	n.d*	n.d*	0–47040	0–17640	n.d*
6	CO	15.7–3383.9	n.d*	1674.3–9905	619.7–8026.9	n.d*
7	NO	7727–7743.2	4700.9–4751.5	172.6–173.7	4569.7–4613.8	4671.5–4687.3
8	SO_2	n.d*	0–31.55	n.d*	n.d*	n.d*
9	H_2S	0–73.49	0–97.17	0–85.41	0–342.62	0–82.42
10	CO_2	762699–905613	853204 – 2162020	1,053,510–1,264,143	957,703–1,188,825	800,220–974,911

*not detected by the instruments

the offroad residential area also has the potential to contribute to higher particulate emission, as reported by previous studies (Cédric et al., 2023; Mri, 2005). The confidence intervals for $PM_{2.5}$ and PM_{10} in Cluster 1 indicate a consistent concentration level across residential locations, whereas in Cluster 2, PM levels are more variable which could be affected by several factors.

In Cluster 2, the levels of gaseous pollutant such as NO_2 are generally higher compared to Cluster 1, also confirmed by Table 4. This is likely due to emissions from vehicle exhaust and construction machinery, which are more prevalent in traffic heavy areas and construction sites. The confidence intervals for NO_2 in C2, as indicated by the blue bars, show a wider range compared to C1, which implies greater variability in NO_2 levels in traffic and construction areas. This could also be associated with the variability of the number of vehicles passing along the time. Other gaseous pollutant level like NMHC were also higher in Cluster 2, consistent with the contribution of vehicle emissions and construction activities, which release hydrocarbons into the atmosphere (Filonchyk & Peterson, 2024; Reşitoğlu et al., 2015; Wang et al., 2015).

As shown in Fig. 3, Cluster 1 (C1) are geographically located further from major traffic routes and construction activities. The highway on the west of roadside residential and offroad residential was newly developed and has not experienced as much traffic volume as the arterial road on the left of traffic interconnection location. The clustering results suggest that these areas experience different pollution characteristic, particularly characterized by higher concentrations of aerosol pollutants like $PM_{2.5}$ and PM_{10}, as seen in earlier analyses. The shading on the map indicates a predominantly blue color in these areas, reflecting their association with Cluster 1, where aerosol pollutants are more prominent due to domestic residential activities.

Cluster 2 (C2) includes the traffic interconnection, construction site, and main roadside locations, represented by red shading. These locations are situated along arterial road networks and construction zones, reflecting their association with sources of gaseous pollutants such as NO_2 and NMHC, which are typically emitted from vehicles and construction activities. The proximity of these locations to major traffic routes likely explains their elevated concentrations of gaseous pollutants shown in Table 4, as previously indicated by the cluster analysis.

Investigation for the potential sources and pathways of pollutants in both Cluster was performed using back trajectory analysis by utilizing the Python™-based model. The back trajectory analysis allows for an understanding of how far-traveled air masses, influenced by regional meteorological conditions, might contribute to the observed pollutant levels in the study area (Guan et al., 2024; Van Pinxteren et al., 2010). For this approach, meteorological data recorded from an hour before and during measurement time were taken from the OpenWeather™ API including wind speed, wind direction, temperature, and pressure. Furthermore, they were converted into radians to compute the backward movement of air masses.

The computed trajectory points were statistically analyzed using Gaussian kernel density estimation (KDE) to iteratively reconstruct air mass displacements rather than emphasizing single pathway lines. The results were automatically overlaid onto customized Google Earth maps using the Cartopy library to reflect actual conditions. Figure 4 illustrates the results of the back trajectory analysis for the traffic interconnection and main roadside measurement location, covering the study area of South Tangerang and its surroundings. The green highlights alongside the blue trajectory lines describe the density estimation result of air mass displacements during measurement time, while each point describes the computed positions of air masses at specific time intervals as they move backward in time.

Observation in Fig. 4 shows that, during the measurement time at traffic interconnection, the majority of air masses originate from the west side, passing over some factory complexes. Some of the notable factories in this area include a pulp and paper plant and a ceramic toilet manufacturing factory. These types of industries are typically associated with gaseous emissions such as NOx rather than aerosol pollutants (Badyda et al., 2022; NCASI, 2024). This observation further confirms the cluster analysis result, where traffic-heavy areas like the traffic interconnection were characterized by lower $PM_{2.5}$ and PM_{10} concentrations and higher levels of gaseous pollutants such as NO_2 compared to residential areas.

For the main roadside, the air masses were found to be revolved around the vicinity of the measurement

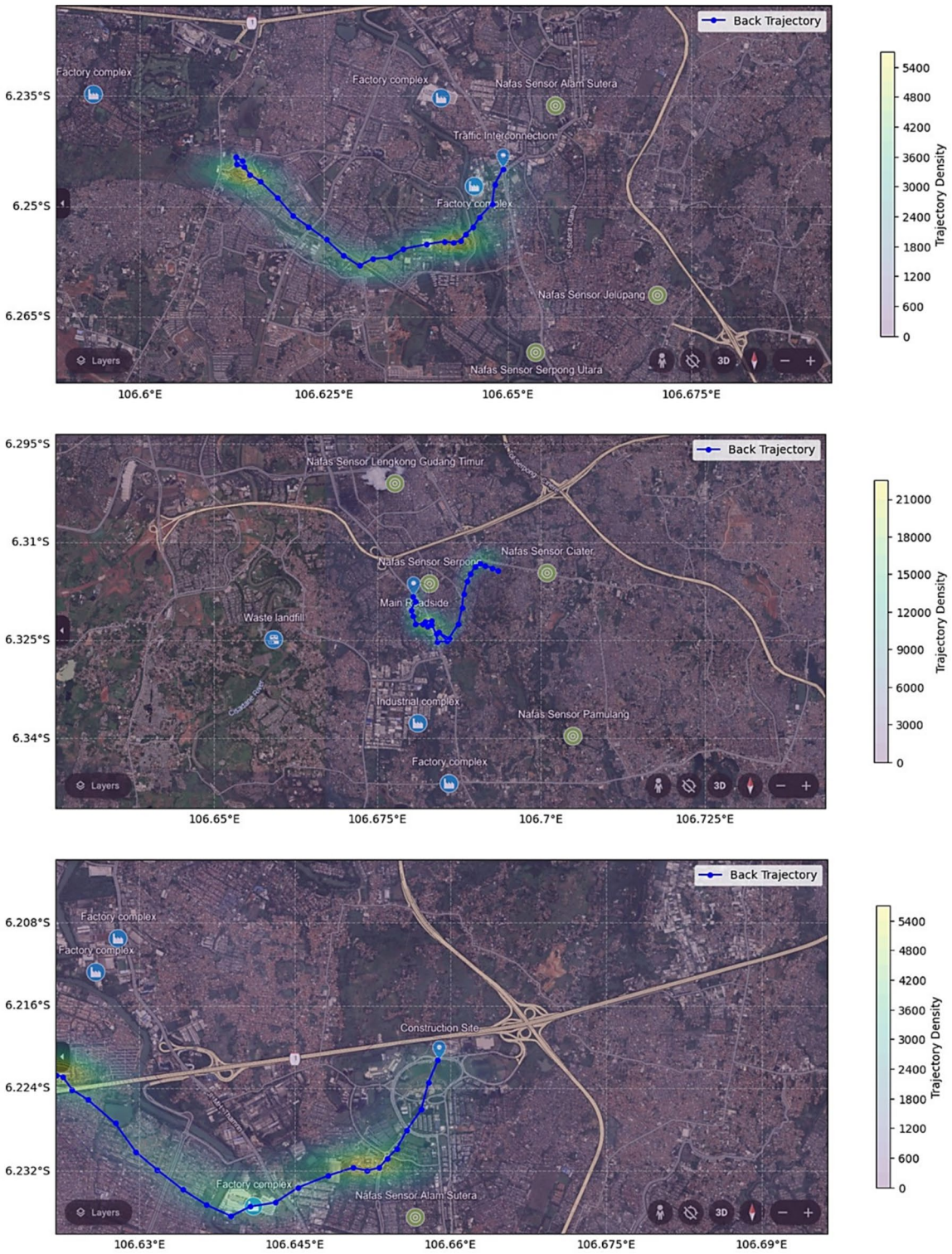

Fig. 4 Back trajectory map for areas in Cluster 2: traffic interconnection (above), main roadside (middle), and construction site (below)

location. The green trajectory density area was also smaller than the one found on traffic interconnection area, indicating lower possibility of transboundary pollutants impact from faraway sources. Therefore, the air profile in the area was mostly affected by the surrounding environments, where a waste landfill and some industrial complexes can be found. The waste landfill has been reported to be a significant source of gaseous pollutants like H_2S (Nunes et al., 2021; Panza & Belgiorno, 2010; Wang et al., 2023), which further confirms its high concentration measured in the main roadside location. The relatively lower level of aerosol pollutants at the main roadside location compared to residential areas could be explained by the lack of significant PM sources along the transport pathway, as the industries and landfill mainly emit gaseous pollutants along with the passing vehicles.

Lastly, for the construction site, the trajectories predominantly originate from the far western areas filled with some factory complexes, before reaching the construction site. High trajectory density illustrated in green-yellow color gradient can be observed over the factory complex to the southwest of the construction site, known to be ceramic toilet manufacturing plant as mentioned earlier. Elevated concentration of NO_2 observed at the construction site could be partly affected by transported emissions from this factory (Kerton, 1993). Other industrial emissions might also be contributed by other factories around the trajectory pathway. These factors are consistent with the pollutants detected in higher concentrations at this site, such as NMHC and NO_2 (Keawboonchu et al., 2023; Zhu & Xu, 2022).

Back trajectory analysis was also conducted for the areas in Cluster 1 as shown in Fig. 5. One key characteristic of Cluster 1 is that it comprises residential areas that are situated farther from major high-traffic roads compared to the locations in Cluster 2. This distance from heavy traffic significantly influences the pollutant characteristic observed in Cluster 1.

In the roadside residential area, the trajectory map indicates that air masses predominantly originate from nearby regions, and they do not pass through any notable industrial or traffic-heavy zones. Despite the presence of several factories to the west, the trajectory does not cross them, suggesting that these factories are not major contributors to the local pollution. Given the relatively low proximity to major traffic sources, gaseous pollutants such as NO_2 and CO, which are primarily emitted by vehicles (Bradley et al., 1999; Kurtenbach et al., 2012), are present at lower concentrations in Cluster 1 compared to Cluster 2. On the other hand, local activities, such as residential waste burning and smaller-scale emissions, are likely contributing to the elevated particulate matter levels observed in this area.

Meanwhile, for offroad residential area, the air masses predominantly flow from the southeast, passing directly over a waste processing site and a waste burning site before reaching the monitoring point. These findings suggest that particulate matter emissions from waste burning and processing activities are the primary contributors to the elevated aerosol pollutant levels in Cluster 1 (Wang et al., 2023; Yan et al., 2016). Similar to the roadside residential area, the offroad residential area's relative distance from high-traffic roads means it is less affected by vehicle emissions, leading to lower concentrations of gaseous pollutants compared to Cluster 2. This difference between clusters further highlights the impact of proximity to major roads and localized pollutant sources on the overall pollutant characteristic.

The combination of local waste management activities, including processing and burning, likely explains the high levels of particulate matter in Cluster 1, while the lower gaseous pollutant concentrations reflect the reduced influence of traffic emissions in these areas, as opposed to the traffic-heavy zones of Cluster 2.

Conclusion

In this study, a novel approach combining in situ air quality measuring instruments with clustering and back trajectory analysis was successfully developed and employed in the mapping and tracing of air pollutants across various urban locations. This method provided a flexible and resource-efficient alternative to traditional stationary air monitoring systems, enabling spatially diverse data collection across multiple locations without the need for extensive infrastructure.

The method clearly revealed two distinct clusters of pollution sources: areas primarily impacted by gaseous pollutants linked to traffic and industrial emissions and others affected by aerosol pollutants predominantly originating from localized activities

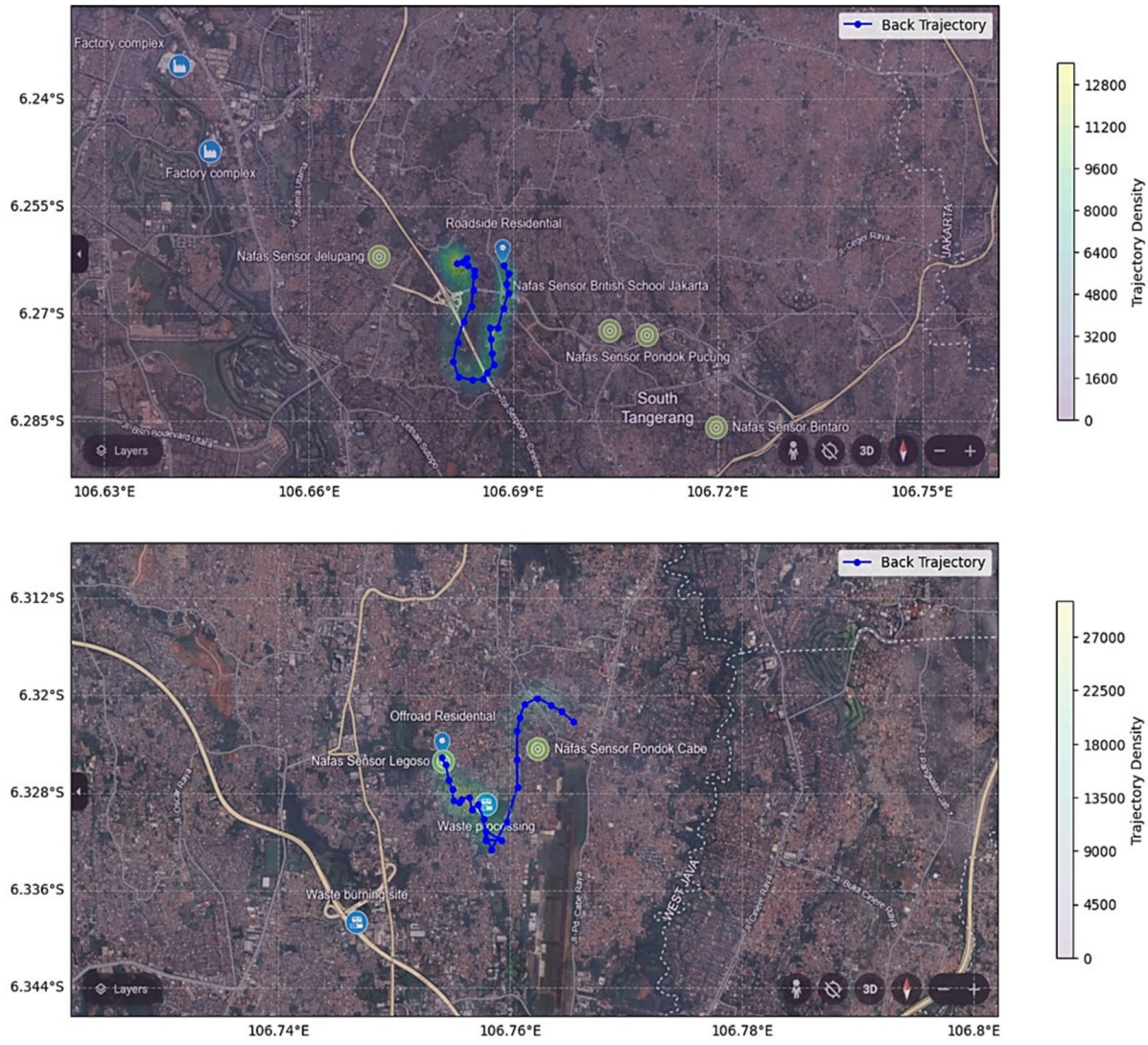

Fig. 5 Back trajectory map for areas in Cluster 1: roadside residential (above) and offroad residential (below)

such as waste processing and burning. These findings demonstrate the ability of the methodology to distinguish between pollution sources and provide a clearer understanding of how urban areas are affected by both local and transported pollutants.

It should be noted, however, that for the scope of this study, the aim was focused on pollution source investigation with close proximity to the five targeted locations. Due to the high spatial variability of some pollutants, such as $PM_{2.5}$ and PM_{10}, more spatial data will be needed to cover the vast area of South Tangerang. Therefore, a significantly more measurement locations are needed to accurately describe the air pollution condition of the big city. Longer measurement time will also be required to provide temporal profile of the air quality condition. As stated in the "Materials and methods" section, the measurement data in this study only represent one snapshot of the pollution characteristic during the peak time activity.

Starting from a study case at the highly polluted city of South Tangerang, this approach and its findings have broader implications, particularly for urban areas in developing countries facing similar pollution challenges. The combination of clustering and trajectory analysis with multi-point in situ measurements demonstrated here provides a scalable, cost-effective

tool for improving the understanding of air pollution dynamics in other cities, contributing to more effective pollution mitigation strategies on a global scale.

Acknowledgements The authors would like to express their acknowledgements for data and resources supported by Nafas Indonesia and for the facilities, scientific and technical support from Advanced Chemical Characterization Laboratory, National Research and Innovation Agency through E- Layanan Sains—BRIN.

Author contribution M.R.M.: Conceptualization, Software, Validation, Formal analysis, Investigation, Data Curation, Writing Original Draft. Y.A.: Formal analysis, Investigation, Data Curation, Resources. C.L.: Data curation, Investigation, Resources, Software. R.A.S.: Funding acquisition, Supervision, Software, Resources. H.P.: Resources, Investigation, Data Curation. N.F. and N.J.P.U.: Resources, Investigation, Data Curation. A.V.E. and Y.P.: Resources and Investigation. All authors reviewed the manuscript.

Funding This study was funded by the collaborative financial support between Indonesia Endowment Fund for Education Agency (LPDP) and National Research and Innovation Agency (BRIN) under the funding scheme of the 4th Riset Inovasi untuk Indonesia Maju (RIIM 4).

Data availability Data is provided within the manuscript.

Declarations

Ethics approval All authors have read, understood, and com plied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors.

Competing interests The authors declare no competing interests.

References

Arias-Pérez, R. D., Taborda, N. A., Gómez, D. M., Narvaez, J. F., Porras, J., & Hernandez, J. C. (2020). Inflammatory effects of particulate matter air pollution. *Environmental Science and Pollution Research, 27*(34), 42390–42404. https://doi.org/10.1007/S11356-020-10574-W

Aslam, A., Ibrahim, M., Shahid, I., Mahmood, A., Irshad, M. K., Yamin, M., Ghazala, T. M., & Shamshiri, R. R. (2020). Pollution characteristics of particulate matter (PM2.5 and PM10) and constituent carbonaceous aerosols in a South Asian Future Megacity. *Applied Sciences, 10*(24), 8864. https://doi.org/10.3390/APP10248864

Badyda, A., Tzanis, C. G., Giacosa, G., Barnett, C., Rainham, D. G., & Walker, T. R. (2022). Characterization of annual air emissions reported by pulp and paper mills in Atlantic Canada. *Pollutants, 2*(2), 135–155. https://doi.org/10.3390/POLLUTANTS2020011

Bradley, K. S., Stedman, D. H., & Bishop, G. A. (1999). A global inventory of carbon monoxide emissions from motor vehicles. *Chemosphere - Global Change Science, 1*(1–3), 65–72. https://doi.org/10.1016/S1465-9972(99)00017-3

Cédric, N. Y., Cyrille, M. A., Charitos, M. S., & David, M. (2023). Using a Gaussian model to estimate the level of particle matter concentration on paved and unpaved roads in urban environment A case study. *Journal of Air Pollution and Health, 8*(2), 183–204. https://doi.org/10.18502/JAPH.V8I2.12917

Chandra, V. L., Mahinur, M. B., & Raihan, A. (2023). Nexus between urbanization, industrialization, natural resources rent, and anthropogenic carbon emissions in South Asia: CS-ARDL approach. *Anthropocene Science, 2*, 48–61. https://doi.org/10.1007/s44177-023-00047-3

Chen, R. J., Lee, Y. H., Chen, T. H., Chen, Y. Y., Yeh, Y. L., Chang, C. P., Huang, C. C., Guo, H. R., & Wang, Y. J. (2021). Carbon monoxide-triggered health effects: The important role of the inflammasome and its possible crosstalk with autophagy and exosomes. *Archives of Toxicology, 95*(4), 1141–1159. https://doi.org/10.1007/S00204-021-02976-7

Cheng, J., Xu, Z., Zhang, X., Zhao, H., & Hu, W. (2019). Estimating cardiovascular hospitalizations and associated expenses attributable to ambient carbon monoxide in Lanzhou, China: Scientific evidence for policy making. *Science of The Total Environment, 682*, 514–522. https://doi.org/10.1016/J.SCITOTENV.2019.05.110

Demšar, J., Erjavec, A., Hočevar, T., Milutinovič, M., Možina, M., Toplak, M., Umek, L., Zbontar, J., & Zupan, B. (2013). Orange: Data mining toolbox in Python Tomaž Curk Matija Polajnar Laň Zagar. *Journal of Machine Learning Research, 14*, 2349–2353.

Donzelli, G., & Suarez-Varela, M. M. (2024). Tropospheric ozone: A critical review of the literature on emissions, exposure, and health effects. *Atmosphere, 15*(7), 779. https://doi.org/10.3390/ATMOS15070779

Elson, P., Andrade, E. S. de, Lucas, G., May, R., Hattersley, R., Campbell, E., Comer, R., Dawson, A., Little, B., Raynaud, S., scmc72, Snow, A. D., lgolston, Blay, B., Killick, P., lbdreyer, Peglar, P., Wilson, N., Andrew, Kirkham, D. (2024). SciTools/cartopy: REL: v0.24.1. *Zenodo*. https://doi.org/10.5281/ZENODO.13905945

Filonchyk, M., & Peterson, M. P. (2024). Analysis of air pollution from vehicle emissions for the contiguous United States. *Journal of Geovisualization and Spatial Analysis, 8*(1), 1–11. https://doi.org/10.1007/S41651-024-00180-6/TABLES/2

Gautam, S., Sammuel, C., Bhardwaj, A., Shams Esfandabadi, Z., Santosh, M., Gautam, A. S., Joshi, A., Justin, A., John Wessley, G. J., & James, E. J. (2021). Vertical profiling of atmospheric air pollutants in rural India: A case study on particulate matter (PM10/PM2.5/PM1), carbon dioxide, and formaldehyde. *Measurement, 185*, 110061. https://doi.org/10.1016/J.MEASUREMENT.2021.110061

Guan, H., Chen, Z., Tian, J., & Xiao, H. (2024). Assessing PM2.5 dynamics and source contributions in Southwestern China: Insights from winter haze analysis. *Atmosphere, 15*(7), 855. https://doi.org/10.3390/ATMOS15070855

Gustriansyah, R., Suhandi, N., & Antony, F. (2020). Clustering optimization in RFM analysis based on k-means. *Indonesian Journal of Electrical Engineering and Computer Science, 18*(1), 470–477. https://doi.org/10.11591/ijeecs.v18.i1.pp470-477

Health Effects Institute. (2024). *State of global air report 2024*. https://www.stateofglobalair.org/resources/report/state-global-air-report-2024

IPCC. (2021). *Climate change widespread, rapid, and intensifying*. https://www.ipcc.ch/2021/08/09/ar6-wg1-20210809-pr/

Karimi, B., & Samadi, S. (2024). Long-term exposure to air pollution on cardio-respiratory, and lung cancer mortality: A systematic review and meta-analysis. *Journal of Environmental Health Science and Engineering, 22*(1), 75–95. https://doi.org/10.1007/S40201-024-00900-6

Keawboonchu, J., Thepanondh, S., Kultan, V., Pinthong, N., Malakan, W., & Robson, M. G. (2023). Integrated sustainable management of petrochemical industrial air pollution. *International Journal of Environmental Research and Public Health, 20*(3), 2280. https://doi.org/10.3390/IJERPH20032280/S1

Kerton, C. P. (1993). Behaviour of volatile materials in cement kiln systems. *Gas cleaning at high temperatures*, 589–603. https://www.academia.edu/24939106/BEHAVIOUR_OF_VOLATILE_MATERIALS_IN_CEMENT_KILN_SYSTEMS

Komal, S. D., Singh, K., & Aggarwal, S. G. (2024). Comparative measurement of CO_2, CH_4 and CO at two traffic interjunctions having inflated vehicular flow in Delhi. *Journal of Environmental Sciences, 141*, 314–329. https://doi.org/10.1016/J.JES.2023.06.023

Kurtenbach, R., Kleffmann, J., Niedojadlo, A., & Wiesen, P. (2012). Primary NO_2 emissions and their impact on air quality in traffic environments in Germany. *Environmental Sciences Europe, 24*(6), 1–8. https://doi.org/10.1186/2190-4715-24-21/FIGURES/7

Matandirotya, N. R., Dangare, T., Matandirotya, E., & Mahed, G. (2023). Characterisation of ambient air quality over two urban sites on the South African Highveld. *Scientific African, 19*, e01530. https://doi.org/10.1016/J.SCIAF.2022.E01530

Mri. (2005). Analysis of the fine fraction of particulate matter in fugitive dust final report for Western Governors' Association Western Regional Air Partnership (WRAP).

Muyemeki, L., Burger, R., Piketh, S. J., Language, B., Beukes, J. P., & van Zyl, P. G. (2021). Source apportionment of ambient PM10-2.5 and PM25 for the Vaal Triangle South Africa. *South African Journal of Science*, *117*(5–6), 1–11. https://doi.org/10.17159/SAJS.2021/8617

National Council for Air and Stream Improvement. (2024). *Technical Bulletin No. 0884: Compilation of criteria air pollutant emissions data for sources at pulp and paper mills including boilers*. https://www.ncasi.org/resource/technical-bulletin-no-0884-compilation-of-criteria-air-pollutant-emissions-data-for-sources-at-pulp-and-paper-mills-including-boilers/

National Oceanic and Atmospheric Administration. (2024). *No sign of greenhouse gases increases slowing in 2023 - NOAA Research.* Retrieved February 23, 2025, from https://research.noaa.gov/no-sign-of-greenhouse-gases-increases-slowing-in-2023/

Nunes, M. I., Kalinowski, C., Godoi, A. F. L., Gomes, A. P., & Cerqueira, M. (2021). Hydrogen sulfide levels in the ambient air of municipal solid waste management facilities: A case study in Portugal. *Case Studies in Chemical and Environmental Engineering, 4*, 100152. https://doi.org/10.1016/J.CSCEE.2021.100152

Okure, D., Ssematimba, J., Sserunjogi, R., Gracia, N. L., Soppelsa, M. E., & Bainomugisha, E. (2022). Characterization of ambient air quality in selected urban areas in Uganda using low-cost sensing and measurement technologies. *Environmental Science and Technology, 56*(6), 3324–3339. https://doi.org/10.1021/ACS.EST.1C01443

Panza, D., & Belgiorno, V. (2010). Hydrogen sulphide removal from landfill gas. *Process Safety and Environmental Protection, 88*(6), 420–424. https://doi.org/10.1016/J.PSEP.2010.07.003

Peel, J. L., Haeuber, R., Garcia, V., Russell, A. G., & Neas, L. (2013). Impact of nitrogen and climate change interactions on ambient air pollution and human health. *Biogeochemistry, 114*(1–3), 121–134. https://doi.org/10.1007/S10533-012-9782-4/TABLES/2

Persinger, R. L., Poynter, M. E., Ckless, K., & Janssen-Heininger, Y. M. W. (2002). Molecular mechanisms of nitrogen dioxide induced epithelial injury in the lung. *Molecular and Cellular Biochemistry, 234–235*(1), 71–80. https://doi.org/10.1023/A:1015973530559/METRICS

Reşitoɔlu, I. A., Altinişik, K., & Keskin, A. (2015). The pollutant emissions from diesel-engine vehicles and exhaust aftertreatment systems. *Clean Technologies and Environmental Policy, 17*(1), 15–27. https://doi.org/10.1007/S10098-014-0793-9/FIGURES/4

Thangavel, P., Park, D., & Lee, Y. C. (2022). Recent insights into particulate matter (PM2.5)-mediated toxicity in humans: An overview. *International Journal of Environmental Research and Public Health, 19*(12), 7511. https://doi.org/10.3390/IJERPH19127511

Tollefson, J. (2021). IPCC climate report: Earth is warmer than it's been in 125,000 years. *Nature, 596*(7871), 171–172. https://doi.org/10.1038/D41586-021-02179-1

Usman, M., & Balsalobre-Lorente, D. (2022). *Environmental concern in the era of industrialization: Can financial development, renewable energy and natural resources alleviate some load?* https://doi.org/10.1016/j.enpol.2022.112780

Van Pinxteren, D., Brüggemann, E., Gnauk, T., Müller, K., Thiel, C., & Herrmann, H. (2010). A GIS based approach to back trajectory analysis for the source apportionment of aerosol constituents and its first application. *Journal of Atmospheric Chemistry, 67*(1), 1–28. https://doi.org/10.1007/S10874-011-9199-9/TABLES/5

Wang, M., Shao, M., Chen, W., Lu, S., Liu, Y., Yuan, B., Zhang, Q., Zhang, Q., Chang, C. C., Wang, B., Zeng, L., Hu, M., Yang, Y., & Li, Y. (2015). Trends of non-methane hydrocarbons (NMHC) emissions in Beijing during 2002–2013. *Atmospheric Chemistry and Physics, 15*(3), 1489–1502. https://doi.org/10.5194/ACP-15-1489-2015

Wang, X., Firouzkouhi, H., Chow, J. C., Watson, J. G., Carter, W., & De Vos, A. S. M. (2023). Characterization of gas and particle emissions from open burning of

household solid waste from South Africa. *Atmospheric Chemistry and Physics, 23*(15), 8921–8937. https://doi.org/10.5194/ACP-23-8921-2023

WHO. (2024). *Air quality, energy and health*. https://www.who.int/teams/environment-climate-change-and-health/air-quality-energy-and-health/health-impacts/climate-impacts-of-air-pollution

Yan, F., Zhu, F., Wang, Q., & Xiong, Y. (2016). Preliminary study of PM2.5 formation during municipal solid waste incineration. *Procedia Environmental Sciences, 31*, 475–481. https://doi.org/10.1016/J.PROENV.2016.02.054

Zhang, Y., Zuo, Y., Li, X., Shouyu, L., Zhang, K., Fang, W., Zheng, S., Zhang, M., Zhou, H., Wang, W., Ma, Q., Liu, Y., Yao, N., Liu, J., Wang, Z., & Li, H. (2021). Clustering analysis method of power grid company based on K-means. *Journal of Physics: Conference Series, 1883*(1), 012072. https://doi.org/10.1088/1742-6596/1883/1/012072

Zhou, X., Gao, Y., Wang, D., Chen, W., & Zhang, X. (2022). Association between sulfur dioxide and daily inpatient visits with respiratory diseases in Ganzhou, China: A time series study based on hospital data. *Frontiers in Public Health, 10*, 854922. https://doi.org/10.3389/FPUBH.2022.854922/BIBTEX

Zhu, Z., & Xu, B. (2022). Purification technologies for NOx removal from flue gas: A review. *Separations, 9*(10), 307. https://doi.org/10.3390/SEPARATIONS9100307