Environmental Science

AIMS Environmental Science, 9(6): 754–770.
DOI: 10.3934/environsci.2022043
Received: 16 July 2022
Revised: 01 September 2022
Accepted: 27 September 2022
Published: 02 November 2022

http://www.aimspress.com/journal/environmental

Research article

Clustering analysis of PM2.5 concentrations in the South Sumatra Province, Indonesia, using the Merra-2 Satellite Application and Hierarchical Cluster Method

Muhammad Rendana[1,*], Wan Mohd Razi Idris[2] and Sahibin Abdul Rahim[3]

[1] Department of Chemical Engineering, Faculty of Engineering, Universitas Sriwijaya, Indralaya 30662 South Sumatra, Indonesia
[2] Department of Earth Science and Environment, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
[3] Department of Environmental Science, Faculty of Science and Natural Resources, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia

* **Correspondence:** Email: muhrendana@ft.unsri.ac.id; Tel: +62 81379895949.

Abstract: The air quality monitoring system is the most prominent tool for monitoring air pollution levels, especially in areas where forest fires often occur. The South Sumatra Province of Indonesia is one of the greatest contributors to haze events in Indonesia due to peatlands fires. It does not sufficiently possess a ground monitoring system to cover rural areas, and thus, delayed actions can result in severe air pollution within this region. Therefore, the aim of this current study is to analyze the distribution and classification of PM2.5 observed from 2019 to 2021 within the South Sumatra Province, Indonesia. The acquisition of PM2.5 data was from the Merra-2 Satellite with a spatial resolution of 0.5° × 0.625° and an hourly interval. The hierarchical cluster analysis (HCA) was applied in this study for the clustering method. The result of the study revealed that the daily mean of PM2.5 levels varied from 5.9±0.01 to 21.3±0.03 μg/m^3. The study area was classified into three classes: high pollution areas (HPA), moderate pollution areas (MPA) and low pollution areas (LPA), based on the HCA method. The average level of PM2.5 observed in HPA was notably higher, at 16.8±0.02 μg/m^3, followed by MPA and LPA. Furthermore, this study indicated that the highest level of PM2.5 was found during 2019, with a severe haze event in the study area due to the intensive burning of forests, bush and peatlands. As a whole, the output of this study can be used by authorities for air quality management due to forest fire events in a certain area.

Keywords: PM2.5; cluster analysis; forest fire; haze event

1. Introduction

The severe air pollution that occurs in a polluted area can threaten human health due to respiratory diseases such as acute respiratory infection, asthma and lung cancer [1]. Recently, the COVID-19 outbreak has also been associated with respiratory diseases, which then could worsen human health when occurring at the same time. Several studies have supported this notion [2]. Furthermore, a total of 4.2 million deaths per year due to air pollution worldwide have been declared by the World Health Organization (WHO). Aerosols like particulate matter 2.5 (PM2.5) have been more detrimental to human health than other air pollutants [3]. Therefore, numerous studies have been concerned with PM2.5 because it affects environmental factors like climate and visibility, and it has the capability to enter lung tissues and then threaten human health [4,5].

Intense expansion and people's mobility have led to air pollution and raised the interest of scientists to investigate the impacts and implications of PM2.5. Yang et al. [6] stated that the majority of PM2.5 comes from fossil fuel combustion, power stations and industrial activities, while Hao et al. [7] reported that motor vehicle exhausts and coal-fired power facilities were the majority sources of PM2.5 emissions in urban regions. PM2.5 can also be emitted from densely commercial and business places [8]. Goudarzi et al. [9] assumed that high PM2.5 levels in Middle East regions have resulted from dust storm occasions. High levels of PM2.5 in Indonesia have been related to the unsupervised burning of forests in some main areas of the country, such as the Sumatra and Kalimantan areas [10]. Peatland fires are known as a predominant cause of particulate matter emission in the South Sumatra region due to the forest fire events.

The high PM2.5 trend every year indicates that this pollutant needs quick actions to mitigate this issue. As the most populous country in the Southeast Asian region, it is important for Indonesia to possess a good air pollution observation system to measure PM2.5 levels for the health of society. The major improvement to the Indonesian air pollution observation system has established real-time stations to measure PM2.5 in major cities in Indonesia, but it is still limited in covering many rural areas in the country. Recent remote sensing techniques can be used to measure air pollutants and give larger coverage by using satellite-based data. Many studies have validated that the PM2.5 acquired from the satellite-based data had a strong relationship with ground monitoring data [11–13].

The clustering method is a technique to explore and analyze the central structure of data. In general, hierarchical clustering and k-means are two popular approaches in clustering analysis. Both have been widely used in air pollution studies all over the world [14,15]. Previous studies on clustering analysis, such as Gönençgil [16] and Mahmud et al. [17], have largely been concerned with climate variables, with little concern for air pollution. Considering the significant effects of particulate matter on humans, a better notion of the spatiotemporal pathways and transports of particles is crucial at this moment. The HCA is a method for clustering data into several classes in which the data inside a class is the same as each other, and vice versa [18]. Thus, this current study aimed to elucidate the PM2.5 trend from 2019 to 2021 in the South Sumatra region, Indonesia, according to the classification of the hierarchical cluster analysis. The association between meteorological variables and PM2.5 concentration was also analyzed.

2. Materials and methods

2.1. Study area

The South Sumatra province is geographically located in the ranges of 1° 37’52’’N – 4° 57’45’’N and 102° 3’25’’E – 106° 5’25’’E in the south of the Southeast Asian region and in the West Indian Ocean (Figure 1). The province is situated on Sumatra Island. The total area of Sumatra Island is about 437,481 km^2 with the South Sumatra province contributing nearly 79% (91,592.43 km^2) of the total area. Peatland forests compose almost half of the South Sumatra region, with most of them located in the northern and eastern parts. Palembang city is situated in the center of South Sumatra province and is the most populous, developed and congested area. The study area had annual mean rainfall and temperature around 2,598 mm and 27.8 °C, respectively. It experiences two kinds of seasons, the wet and dry seasons. The wet season carries heavy rainfall, while the dry season has less rainfall and haze events. Therefore, in this region, there are two primary calamities, namely, floods and forest or land fires.

Figure 1. Location of the South Sumatra province, its districts and the observed stations that were used in this study.

2.2. Satellite-based data for PM2.5 concentration

The recent satellite-based remote sensing techniques have allowed us to detect spatial and temporal aerosols in areas with limited ground monitoring stations like the study area. The Merra-2 satellite data with spatial resolution of 0.5° × 0.625° utilized in this research were from 1 Jan 2019 to 31 Dec 2021. The data were obtained hourly at the study area and passed through an accuracy assessment to testify to their validity. The fine-resolution aerosol optical depth (AOD) product from the Merra-2 satellite was used to reconstruct ground PM2.5 concentrations and assess the spatial and temporal distribution data. We used a geographically and temporally weighted regression (GTWR) model to validate the PM2.5 concentrations based on satellite data with ground measurements. The accuracy results obtained a highly accurate value, with $R^2 = 0.85$.

Gueymard and Yang [13] stated during their study that the aerosol optical depth (AOD) data from the Merra-2 satellite showed an accurate prediction, with a lowest value of root mean square error of 0.017. The Merra-2 data were extracted from the GIOVANNI website (https://giovanni.gsfc.nasa.gov/). Figure 1 shows the location of the South Sumatra province and the 40 stations which were used for the PM2.5 acquisition using the Merra-2 satellite. A total of 40 stations have been distributed in the whole area and were chosen according to the location of each district within the South of Sumatra province, and we have labeled the stations as AQS (Air Quality Station). Because the study area was large, we classified these AQSs into five regions: Northern (N), Eastern (E), Southern (S), Central (C) and Western (W) regions (Figure 2). This classification would assist in determining the most polluted area within the South of the Sumatra region. In addition, we collected meteorological variables such as temperature, rainfall, humidity, wind speed and sunshine hour from the Meteorology, Climatology and Geophysical Agency. These variables were used to analyze the association between PM2.5 and meteorological variables.

2.3. Multivariate and correlation analysis

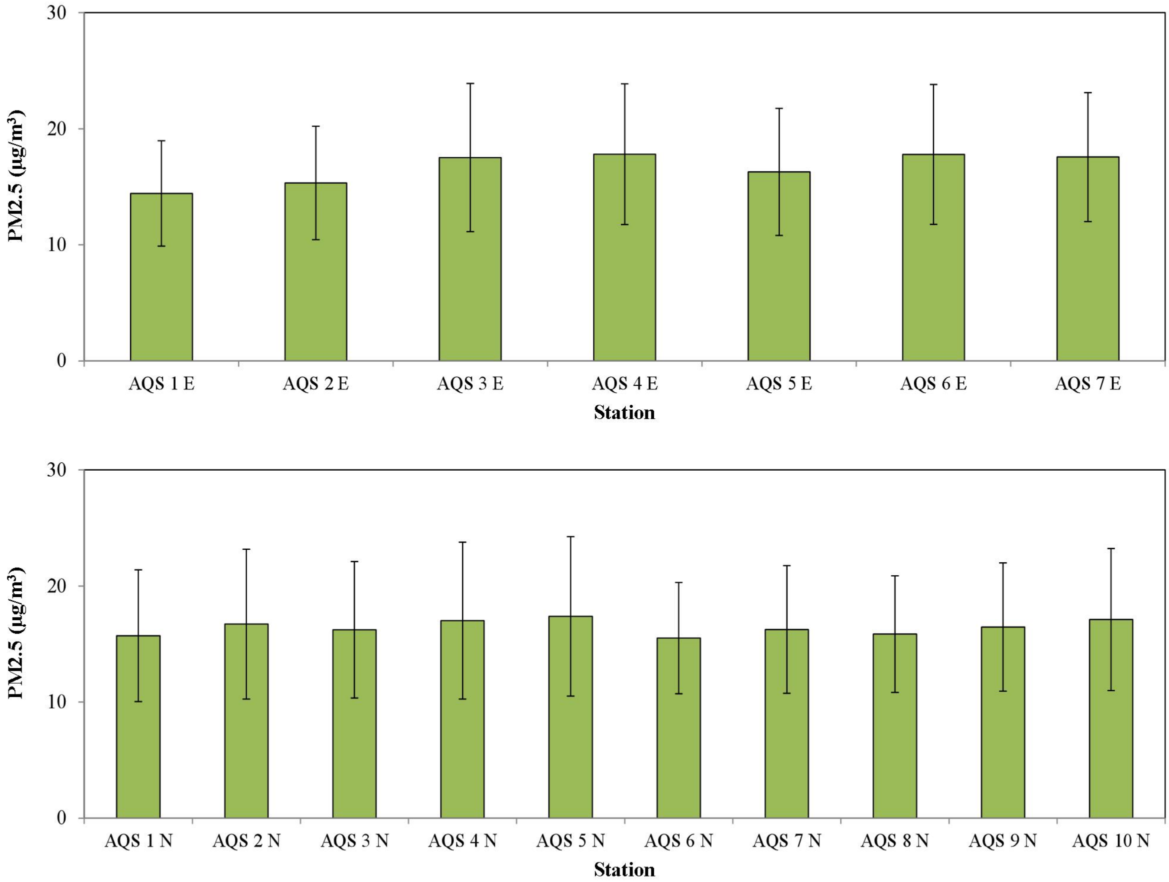
The daily mean was computed using hourly PM2.5 data. The daily mean PM2.5 data from the 40 observation stations were examined using the HCA method according to the PM2.5 characteristic through the South Sumatra region. Hierarchical Cluster Analysis (HCA) is one of the statistical multivariate methods applied in this study. The HCA analysis will merge pairs of clusters, generating smaller clusters with many associations [19]. The centroid clustering approach is one of the popular cluster methods and shows good clustering outputs for HCA analysis [20]. The clusterization of the stations is described in a dendrogram. The dendrogram indicated the rank of sameness between them that is calculated by the Euclidean distance using the centroid clustering approach [21]. The Euclidean distance is measured according to the closest neighbor method [22]. The threshold value for PM2.5 (15 $\mu g/m^3$) was obtained from the Ministry of Environment and Forestry of Indonesia. For clustering analysis, we only used daily mean PM2.5 data and did not include meteorological parameters. The relationship between meteorological variables and PM2.5 is determined by the Pearson correlation. The correlation analysis used data from 40 stations around the study area. The strong relationship is indicated with a correlation coefficient of ($r = 1$ or -1), while a weak relationship is indicated with a correlation coefficient of ($r = 0.1$–0.2). The variables have no relationship if the correlation coefficient value is 0 [23]. All statistical analyses were carried out using IBM SPSS Statistics 20.

3. Results and Discussion

3.1. Distribution of PM2.5 levels

Figure 2 present the daily mean of PM2.5 in five regions of the South Sumatra Province. We divided these regions to determine the highest PM2.5 concentrations within the South Sumatra

Province, and thus we could identify possible PM2.5 sources and their effects on the adjacent regions. The trend chart showed that PM2.5 levels were above a threshold value for almost all stations, except for some stations below Indonesia's standard of PM2.5 (15 μg/m^3). The average of PM2.5 obtained at all stations varied from 11.2 to 18.4 μg/m^3. The lowest level of mean PM2.5 was in Musi Rawas (AQS 1 W), while the highest concentrations were obtained in OKU Selatan (AQS 1 S) and Prabumulih (AQS 1 C). Most of the stations revealed an average of PM2.5 higher than the threshold value. This might have occurred because South Sumatra was struck by a heavy haze incident due to the forest fires locally within this region in 2019 (month of September), where the peak level of PM2.5 was obtained. The local open burning, especially in peatland areas, was the major reason for the high level of PM2.5. This is in line with Xu et al. [24], who found the increment of PM2.5 could contribute to particle transport in the air during the hazy condition. The high PM2.5 influenced rural and urban regions because of the open burning during dry weather [25]. Compared with the Southeast Asian countries, the PM2.5 concentration is relatively high in Indonesia. Ly et al. [26] found, during a haze episode in Hanoi city, Vietnam, the mean PM2.5 concentration was higher than 50 μg/ m^3 during the dry period. Hassan et al. [27] found the daily PM2.5 level during the non-hazy period ranged from 9.80 to 59.9 μg/m^3 in Petaling Jaya, Malaysia. Lung et al. [28] reported that Dhaka city revealed the highest PM2.5 concentration in South and Southeast Asian areas, with 77.1 μg/m^3 in 2020, compared with Jakarta (39.6 μg/m^3), Kuala Lumpur (16.5 μg/m^3), Metro Manila (13.1 μg/m^3), Taipei (12.6 μg/m^3), Bangkok (20.6 μg/m^3) and Hanoi (37.9 μg/m^3).



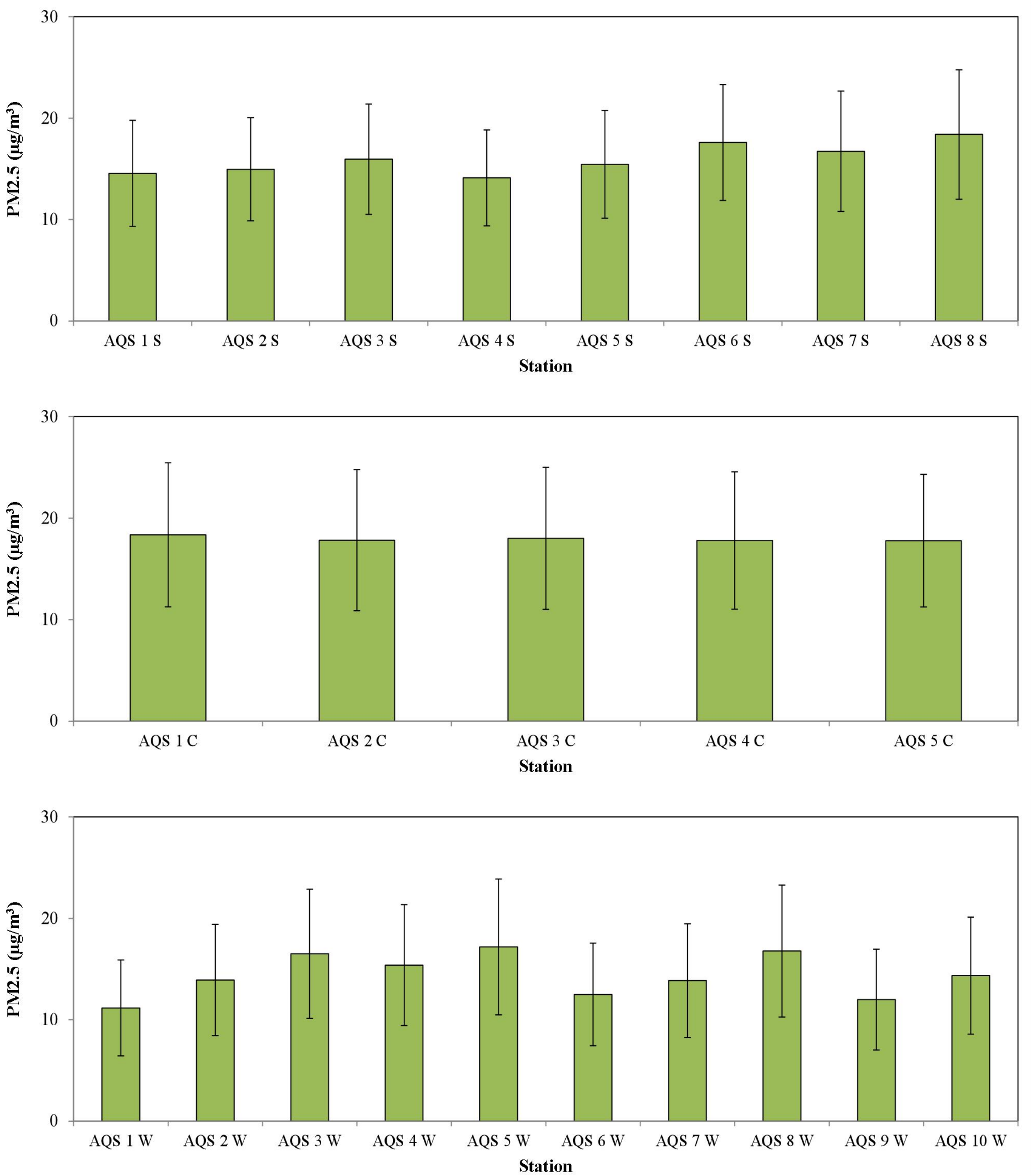


Figure 2. Daily mean of PM2.5 for 40 air quality stations based on the (a) Eastern, (b) Northern, (c) Southern, (d) Central and (e) Western regions.

3.2. The clustering analysis of PM2.5 level

The HCA analysis was applied to cluster daily mean levels of PM2.5 obtained from the 40 stations with distinct areas from 2019 to 2021. A total of three clusters that characterized the homogeneity are depicted in Figure 3. The HCA results showed different characteristics among the clusters. The three clusters were labeled as high pollution areas (HPA), moderate pollution areas (MPA) and low pollution areas (LPA). The classification of stations using HCA (HPA, MPA and LPA) according to PM2.5 levels in South Sumatra is exhibited in Figure 4. Overall, there were 31 stations in HPA, 6 stations in MPA and 3 stations in LPA. Most HPA stations were distributed almost in the whole area (northern to southern and central parts), but MPA and LPA stations were in the peripheral areas.

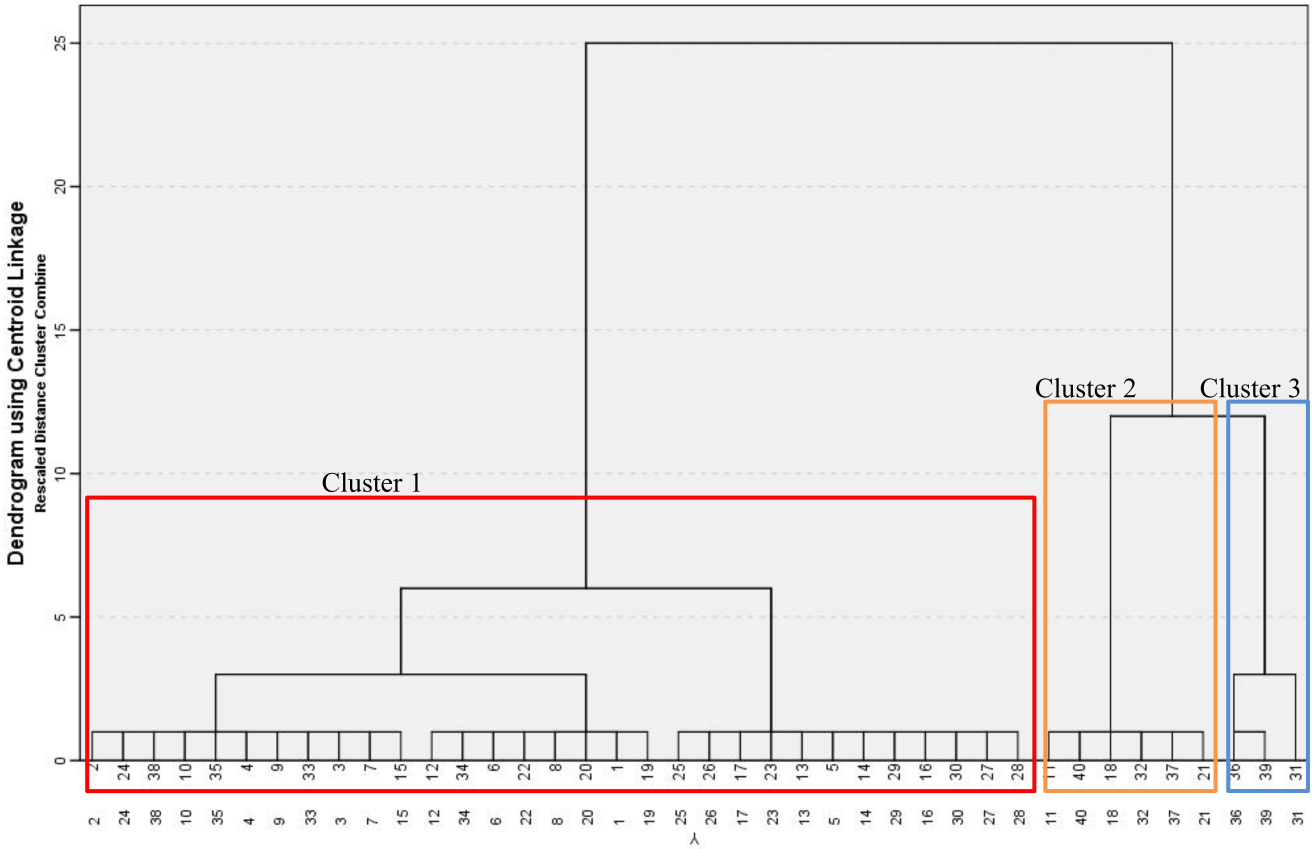


Figure 3. Clustering the 40 stations in South Sumatra using dendrogram (note: 1–10: stations in the northern area, 11–17: stations in the eastern area, 18-25: stations in the southern area, 26–30: stations in the central area, 31–40: stations in the western area).

Figure 5 presents the daily mean level of PM2.5 from 2019 to 2021 for HPA, MPA and LPA. The highest PM2.5 concentration was observed in HPA stations, while the lowest was observed in LPA stations. The status of air quality in each area was shown by the number of exceedances of stations in HPA, MPA, and LPA (Figure 6). The year 2019 had the highest number of exceedances in all three areas. The number of exceedances in MPA did not show significant changes in 2020 and 2021. There were zero exceedances recorded in LPA in 2020 and 2021. In 2019, the number of exceedances were higher, compared with 2020 and 2021. HPA, MPA and LPA obtained 31, 6 and 3 occurrences of unhealthy levels (exceeding Indonesia's policy, 15 μg/m^3). The high number of occurrences of unhealthy levels in 2019 was because of severe haze events from South Sumatra during that year. This haze event could have been caused by open burning from peatland fires and oil palm replanting. The haze events have significantly impacted rural areas in LPA, such as Pagar Alam, West Lahat and West Musi Rawas, resulting in the mean level of PM2.5 in 2019 being higher than in 2020 to 2021. This result was consistent with another report by Tarigan et al. [29], which found that those areas had high forest fire proclivity, as we can see in Figure 7. Generally, the HPA obtained the highest annual mean PM2.5 levels compared with MPA and LPA. The average level of PM2.5 was

16.8±0.02 μg/m³ in HPA, 14.2±0.03 μg/m³ in MPA and 11.9±0.01 μg/m³ in LPA. The variation of PM2.5 values had a positive skew, which indicated significant pollution values [30]. The LPA showed a high skewness value, which suggested the area was the most impacted during the period of study because of the level of PM2.5 in the ambient air.

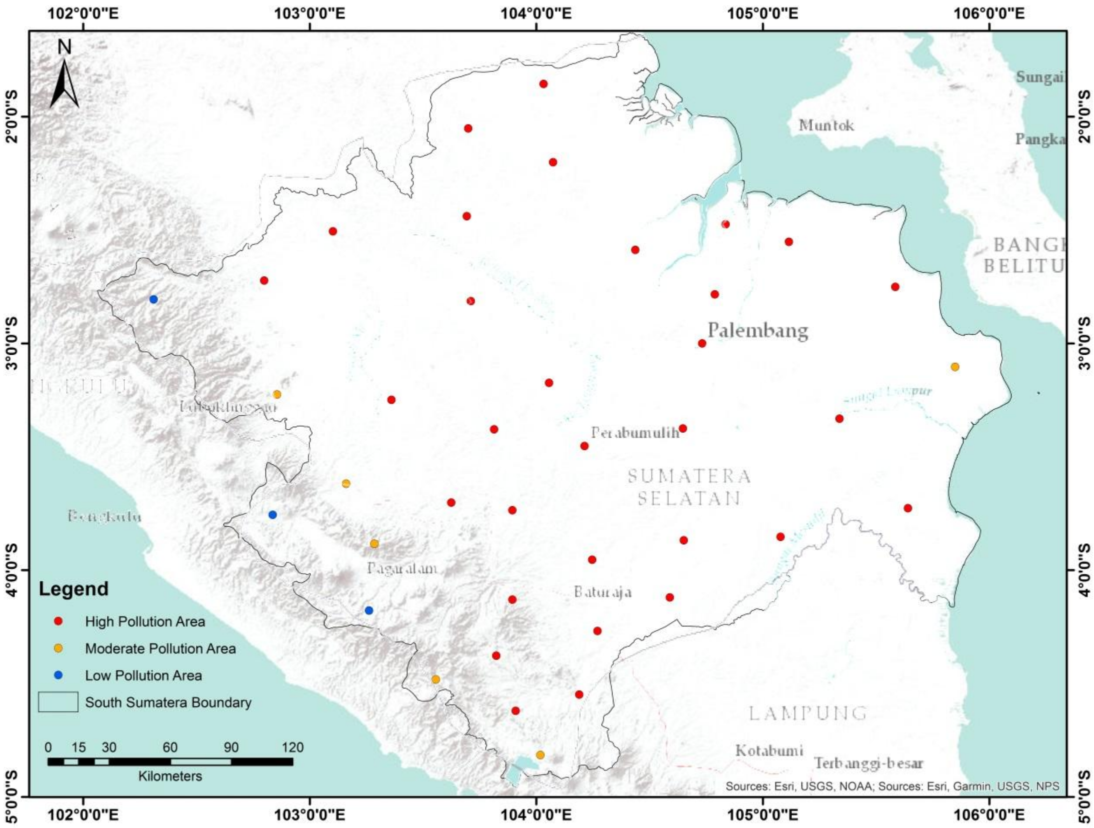


Figure 4. Classification of stations using hierarchical cluster analysis according to the daily mean of PM2.5 levels.

3.3. PM2.5 concentrations in three clusters (HPA, MPA and LPA)

The numbers of stations with distinct clusters according to HPA, MPA and LPA are presented in Table 1. Every cluster exhibited stations situated in urban (U), rural (R) and industrial (I) areas. The graph of PM2.5 according to I, U and R areas is depicted in Figure 8 for HPA, MPA and LPA. There were 3 stations located in industrial areas in HPA but none in industrial areas in MPA and LPA. Our study gave a different result as compared with other previous studies, where they generally assumed urban or industrial areas had higher pollution than rural areas. However, interestingly, in our study, we found most stations have been classified into high pollution areas (HPA), with twenty stations situated in rural areas, three in industrial areas and one in urban areas (Table 1). This could occur because forest fires are frequently located in rural areas. The distance between city and rural areas and meteorological factors (i.e., wind, rainfall) had significant roles in air pollutants' transport from the source to other locations. A relatively stable wind condition during the middle of the year in the study area has resulted in the concentrated air pollutants only at the source point.

The annual mean of PM2.5 in rural areas also almost had the same value as in urban and industrial areas. For example, in 2019, the mean PM2.5 of around 26.62 μg/m^3 (Figure 8) was observed during that period, and the value surpassed Indonesia's standard value (15 μg/m^3). Some stations situated mostly in the northern to eastern regions were categorized as HPA. The massive conversion of lands located in the northern and eastern parts from peatland forests into agricultural areas and settlements has led to air pollution from open burning, biomass and replanting processes. The serious haze event which occurred in 2019 was the effect of these activities. Furthermore, for the urban area, the air pollution was mostly emitted from motors, cars, trucks, industrial sectors and also migration activities. Particulate matter was the primary pollutant in urban regions [31]. In addition, cars and motor vehicles were the main active origins of air contamination in congested cities [32].

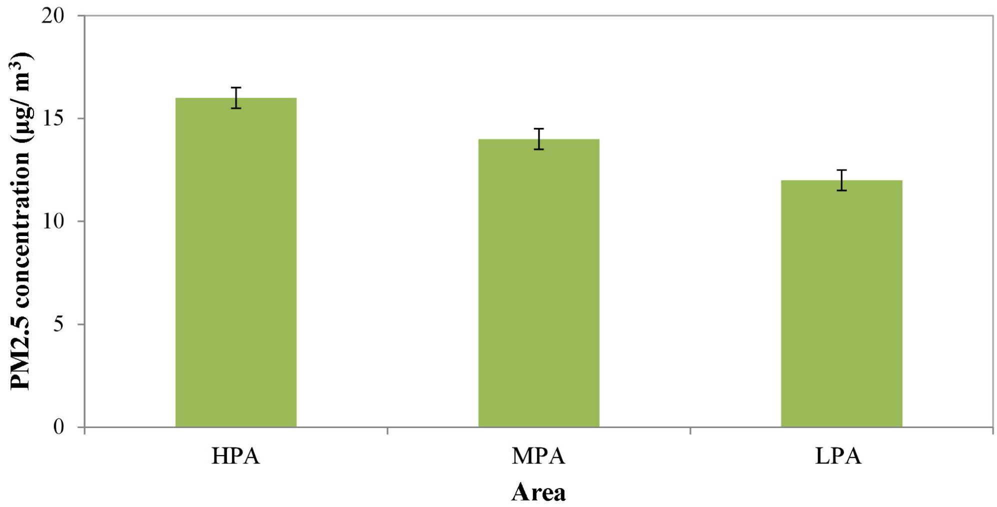


Figure 5. The daily mean level of PM2.5 from 2019 to 2021 for three cluster areas.

Kalisa et al. [33] explained that air pollutant levels were higher in urban areas compared with in rural and native environments. Thus, the high number of stations located in urban regions have been classified as HPA clusters. Meanwhile, peatland fires in the study area were some of the events that promote haze incidence in Indonesia during the dry period and governed many rural areas clustered as HPA situated in the northern to eastern parts of the study area in the districts of Musi Banyuasin, Banyuasin, Ogan Komering Ilir, Ogan Komering Ulu Timur, Ogan Komering Ulu, Muara Enim, Panukal Abab Pematang Ilir and Ogan Ilir. A total of 27 rural areas categorized into the HPR clusters obtained a high level of PM2.5 caused by open burning activities and local biomass fires in peatlands. This was in line with Figure 9, which shows the highest forest fire proclivity was mainly found in the north to the eastern part of the South Sumatra region.

Peatland fires are difficult to mitigate because they can burn for a long time, and the thicker the peat layer is, the longer it will burn [34]. This condition would contribute to a hazy environment with high PM2.5 levels. Many stations situated from the southern to western areas were categorized into MPA. Most MPA clusters were rural areas that consisted predominantly of agricultural and land-use conversion activities, which were the same activities in the studies of Rondhi et al. [35], Zhou et al. [36] and Li et al. [37]. There was one station with the urban area located in the western part (Lubuk Linggau area) in MPA; all sources of air pollution in this area could be emitted from vehicles, power plants, urban development and construction and urbanization activities, as well as open burning.

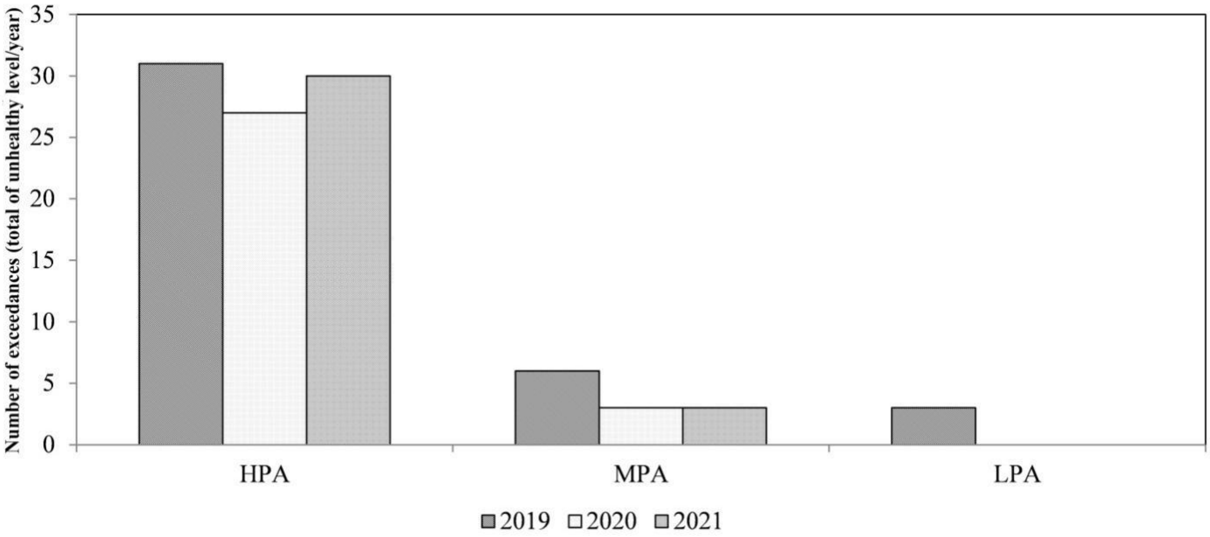


Figure 6. The number of exceedances of PM2.5 concentrations for three cluster areas.

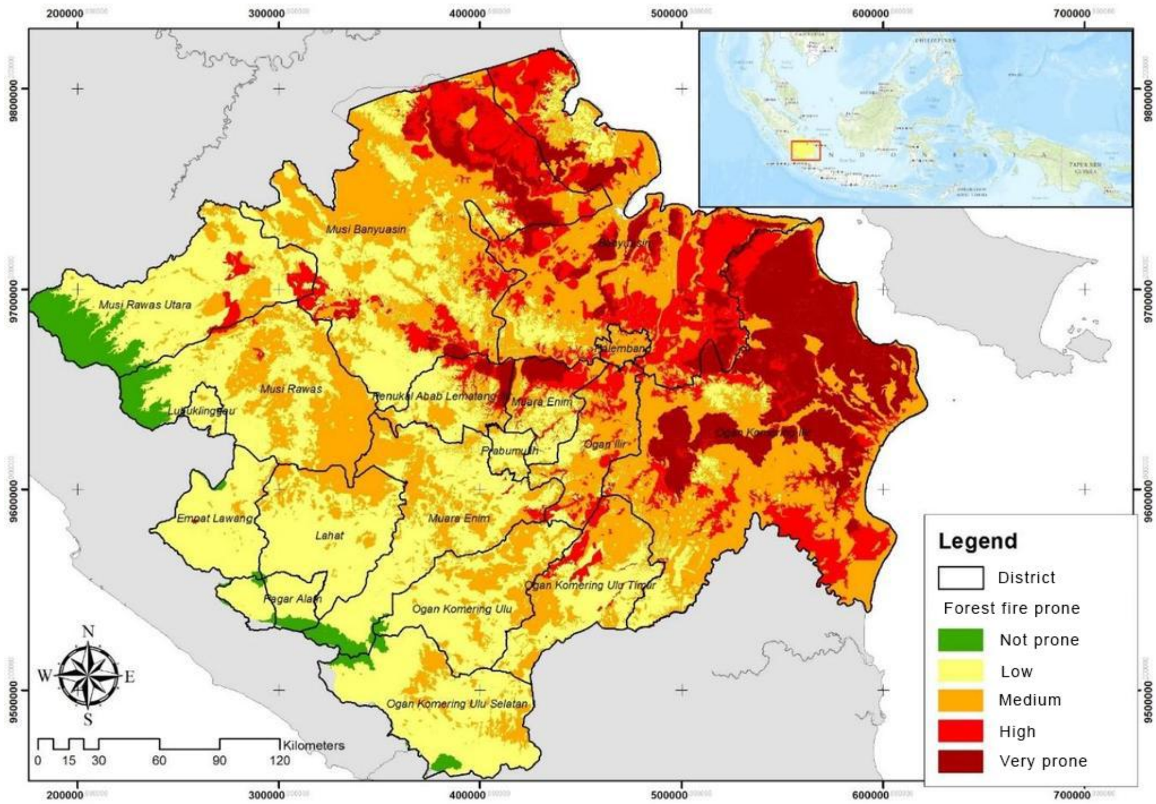


Figure 7. Forest fire proclivity map in the study area (source: Tarigan et al. [29]).

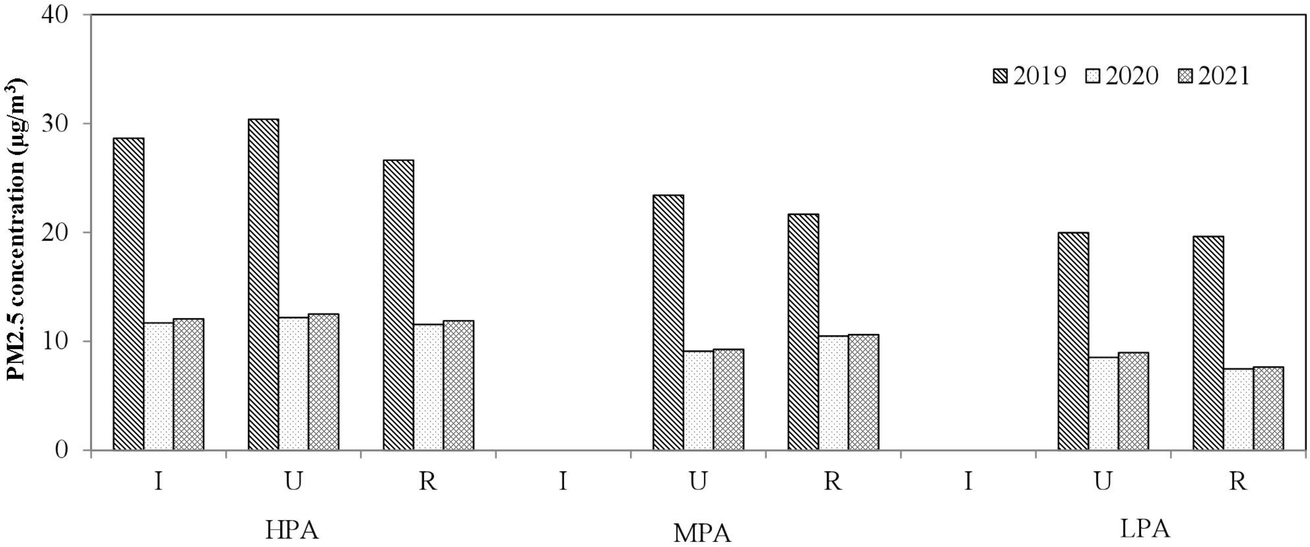


Figure 8. The annual mean level of PM2.5 for three cluster areas based on type of land use: industrial (I), urban (U) and rural (R).

Table 1. The number of stations that relate to land use was based on three cluster areas.

Area	Classification of stations			Total
	Industrial	Urban	Rural	
HPA	3	1	27	31
MPA	0	1	5	6
LPA	0	1	2	3

Meteorology is one of the prominent parameters affecting the particulate matter in distribution and transport in all three clusters. Although the effect of changes in sources was important to evaluated, the changes in climate and weather shift had significant impacts on particulate matter properties [38]. Xie et al. [39] estimated that wind speed was a prominent parameter in the incidence of particle transport from urban or hotspot zones to rustic areas. Many stations in LPR were categorized as low contamination because they were located farther from urbanized, industrial and peatland fire or hotspot areas. These stations were located in the southern to western parts, where they were not susceptible to land fires due to fewer peatlands in that location. The PM2.5 in short-range transport from hotspot areas, more located in the northern and eastern parts, made a significant impact on air pollution in the whole study area.

Tariq et al. [39] stated that forest fires were mostly caused by anthropogenic activities during dry periods. With recent developments in technology, forest or land fires could be easily located by using satellite data. Sari and Putra [40] found a total of 1,369 hotspots in South Sumatra province, which was the second-highest reported among Sumatra provinces in 2019. Most open fire events originated from peatland fires, biomass, agricultural waste burning and some household burning activities which took place during the dry period. Therefore, the local burning in rural areas was a primary contributor to the heavy pollution in the study area. Moreover, motor vehicles were the main precursors to bad air quality in most urban regions that contained particulate matter, lead and dust.

Several sources could result in air pollution in urban areas, such as vehicles and automobiles, power stations, industries, waste burning and also dusts from establishments and mining sites in urban areas. The escalated urbanization process has contributed to the increased pollution concentration in urban areas. Singh and Chauhan [41] assumed that poor air quality normally existed in densely populated areas. This linked with our study, where we found industrial areas were grouped into HPA according to PM2.5 level, whereas urban and rural areas noted high annual levels of PM2.5.

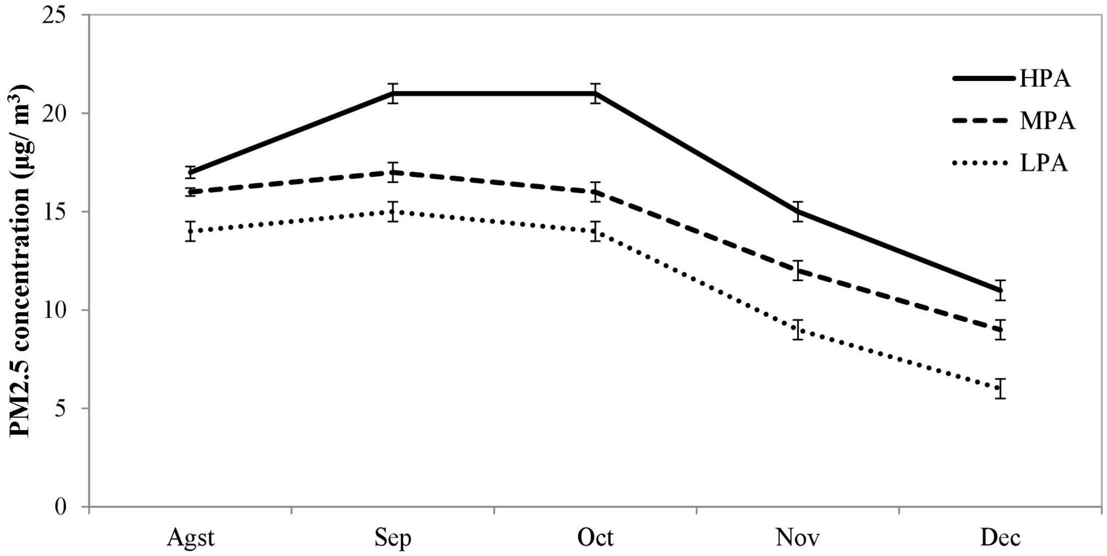


Figure 9. The monthly mean level of PM2.5 from 2019 to 2021 according to three cluster areas.

Dias and Tchepel [42] stated that the dissimilarity of air pollutant levels in urban, industrial and rural areas was affected by the spatial and temporal transport of pollutants. The diversity of station locations could also obtain various values for each area. Thus, the diverse pollutant sources in urban, industrial and rural areas drove an iniquitous level of pollutants. To study the trend of monthly PM2.5 in the study area (Figure 9), we analyzed five months of the study period (from August to Dec) because forest fire events frequently occurred during this dry period. A higher daily level of PM2.5 was observed in HPA than in MPA and LPA across the month. Meanwhile, the level of PM2.5 in LPA is quite lower than in MPA throughout the study period. September and October had the highest daily PM2.5 concentration, and December recorded the lowest daily PM2.5 concentrations in 2019.

The effect of the COVID-19 pandemic that occurred from 2020 to 2021 also caused a significant change in PM2.5 concentration. The implementation of the lockdown policy during the pandemic period restricted almost all human and social activities, with the closures of industries and commercial and public places, which could reduce the PM2.5 emissions. Many worldwide studies have found that PM2.5 also depends on meteorological factors. This was consistent with our study, where we found a highly significant correlation between PM2.5 and temperature ($r = 0.906$, $p<0.01$). Moreover, the dry season that occurred between August and December had the responsibility for the highest daily PM2.5 levels in September. PM2.5 levels were higher during the southwest monsoon from May to September in the Southeast Asian region [43]. The higher PM2.5 levels during this episode were generally because of hot weather status, local effects, steady air and particle transport from burning sources and hotspots. A little rainfall and stable meteorological circumstances contributed to the high PM2.5 levels, while a violent climate sped up the spread of particles.

3.4. The association between meteorological variables and PM2.5 level

The correlation analysis is employed to obtain the relationship between the two parameters. The correlation between PM2.5 concentration and meteorological parameters in the study area is presented in Table 2. Previous studies have found a strong correlation between PM2.5 and CO, PM10 and O_3 [44,45]. The strong correlation showed that PM2.5 was significantly related to air pollutants. CO was associated with PM2.5 levels due to the combustion process from vehicle emissions. Other pollutants were associated with the particulate matter due to the same sources. There was a strong significant association between meteorological factors and PM2.5, including temperature ($r = 0.906$, $p<0.01$) and humidity ($r = -0.748$, $p<0.05$).

Table 2. Correlation analysis between meteorological factors and PM2.5 concentration.

Meteorological variable	PM2.5
Temperature (˚C)	0.906**
Humidity (%)	−0.748*
Rainfall (mm)	−0.415
Sunshine hours (hour)	0.205
Wind speed (m/s)	−0.184

Note:**: Correlation is significant at the 0.01 level (2-tailed); *: Correlation is significant at the 0.05 level (2-tailed).

The non-significant correlations indicated that other meteorological parameters also had indirect impacts on PM2.5 levels. Moreover, the increasing air temperatures could accelerate the splitting of particulate matter through chemical processes, such that the levels of PM2.5 were raised circumstantially [46,47]. Although our study obtained a low connection between wind speed and PM2.5 ($r=-0.184$, $p>0.05$), the wind conditions were the main meteorological properties that spread pollutants. The Pearson correlation was successfully applied to determine the association between the meteorological factors and PM2.5 levels.

4. Conclusions

The annual PM2.5 trend all over the world shows a gradual increase, and thus quick action is needed to solve this problem. Indonesia is one of the most populous countries in the world, and it should possess a good air quality monitoring system. Although some substantial improvements have been carried out by the government until this moment, it is still limited in monitoring air quality in rural areas. Therefore, the use of remote sensing and the HCA method can be used as an alternative way to monitor air quality because they can help to cover rural areas such as those in the South Sumatra Province region. From our study, the total of 40 stations around the study area have been classified into HPA, MPA and LPA, with 31, 6 and 3 total stations, respectively. Rural areas in HPA were the sources of pollutants, since the burning sites in these areas have contributed to the haze events in other neighboring areas. The annual mean of PM2.5 was above Indonesia's and WHO's standards (15 and 10 μg/m^3, respectively) in HPA in 2019. The temperature and humidity showed highly significant correlations with PM2.5 concentrations in the study area. As a whole, this study is the first study with clustering of PM2.5 concentrations in this region. For future studies, other air pollutants like CO, NO_2, O_3 and SO_2 could be analyzed comprehensively based on distinct land use areas. Furthermore, the advanced regression analysis can be conducted to provide a deeper notion of the associations between particulate matter and other air pollutants and also the related meteorological parameters.

Acknowledgments

We would like to thank the Meteorology, Climatology and Geophysical Agency for providing the meteorological data to ensure the success of this research.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Xue Y, Chu J, Li Y, et al. (2020) The influence of air pollution on respiratory microbiome: A link to respiratory disease. *Toxicology Letters* 334: 14–20. https://doi.org/10.1016/j.toxlet.2020.09.007
2. Copat C, Cristaldi A, Fiore M, et al. (2020) The role of air pollution (PM and NO2) in COVID-19 spread and lethality: a systematic review. *Environmental research* 191: 110129. https://doi.org/10.1016/j.envres.2020.110129
3. Harishkumar KS, Yogesh KM, Gad I (2020) Forecasting air pollution particulate matter (PM2. 5) using machine learning regression models. *Procedia Computer Science* 171, 2057–2066. https://doi.org/10.1016/j.procs.2020.04.221
4. Polezer G, Tadano YS, Siqueira HV, et al. (2018) Assessing the impact of PM2. 5 on respiratory disease using artificial neural networks. *Environmental pollution* 235: 394–403. https://doi.org/10.1016/j.envpol.2017.12.111
5. Sun X, Zhao T, Liu D, et al. (2020) Quantifying the influences of PM2. 5 and relative humidity on change of atmospheric visibility over recent winters in an urban area of East China. *Atmosphere* 11: 461. https://doi.org/10.3390/atmos11050461
6. Yang HH, Arafath SM, Lee KT, et al. (2018) Chemical characteristics of filterable and condensable PM2. 5 emissions from industrial boilers with five different fuels. *Fuel* 232: 415–422. https://doi.org/10.1016/j.fuel.2018.05.080
7. Hao Y, Gao C, Deng S, et al. (2019) Chemical characterisation of PM2. 5 emitted from motor vehicles powered by diesel, gasoline, natural gas and methanol fuel. *Science of the total Environment* 674, 128–139. https://doi.org/10.1016/j.scitotenv.2019.03.410

8. Xu H, Ta W, Yang L, et al. (2020) Characterizations of PM2. 5-bound organic compounds and associated potential cancer risks on cooking emissions from dominated types of commercial restaurants in northwestern China. *Chemosphere* 261: 127758. https://doi.org/10.1016/j.chemosphere.2020.127758
9. Goudarzi G, Shirmardi M, Naimabadi A, et al. (2019) Chemical and organic characteristics of PM2. 5 particles and their in-vitro cytotoxic effects on lung cells: The Middle East dust storms in Ahvaz, Iran. *Science of The Total Environment* 655: 434–445. https://doi.org/10.1016/j.scitotenv.2018.11.153
10. Fujii Y, Tohno S, Kurita H, et al. (2021) Characteristics of organic components in PM2. 5 emitted from peatland fires on Sumatra in 2015: Significance of humic-like substances. *Atmospheric Environment: X* 11, 100116. https://doi.org/10.1016/j.aeaoa.2021.100116
11. Song Z, Fu D, Zhang X, et al. (2018) Diurnal and seasonal variability of PM2. 5 and AOD in North China plain: Comparison of MERRA-2 products and ground measurements. *Atmospheric Environment* 191: 70–78. https://doi.org/10.1016/j.atmosenv.2018.08.012
12. He L, Lin A, Chen X, et al. (2019) Assessment of MERRA-2 surface PM2. 5 over the Yangtze River Basin: Ground-based verification, spatiotemporal distribution and meteorological dependence. *Remote Sensing* 11: 460. https://doi.org/10.3390/rs11040460
13. Gueymard CA, Yang D (2020) Worldwide validation of CAMS and MERRA-2 reanalysis aerosol optical depth products using 15 years of AERONET observations. *Atmospheric Environment* 225: 117216. https://doi.org/10.1016/j.atmosenv.2019.117216
14. Franceschi F, Cobo M, Figueredo M (2018) Discovering relationships and forecasting PM10 and PM2. 5 concentrations in Bogotá, Colombia, using artificial neural networks, principal component analysis, and k-means clustering. *Atmospheric Pollution Research* 9: 912–922. https://doi.org/10.1016/j.apr.2018.02.006
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: 40-56. https://doi.org/10.1016/j.apr.2019.09.009
16. Gönençgil B (2020) *Evaluate Turkey's climate classification by clustering analysis method.* In Smart geography (pp. 41–53). Springer, Cham. https://doi.org/10.1007/978-3-030-28191-5_4
17. Mahmud S, Sumana FM, Mohsin M, et al. (2021) Redefining homogeneous climate regions in Bangladesh using multivariate clustering approaches. *Natural Hazards* 1: 1–22. https://doi.org/10.21203/rs.3.rs-633865/v1
18. Granato D, Santos JS, Escher GB, et al. (2018) Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends in Food Science & Technology* 72: 83–90. https://doi.org/10.1016/j.tifs.2017.12.006
19. Charikar M, Chatziafratis V, Niazadeh R, et al. (2019) Hierarchical clustering for euclidean data. In *The 22nd International Conference on Artificial Intelligence and Statistics* (pp. 2721–2730). PMLR.
20. Bu J, Liu W, Pan Z, et al. (2020) Comparative study of hydrochemical classification based on different hierarchical cluster analysis methods. *International journal of environmental research and public health* 17: 9515. https://doi.org/10.3390/ijerph17249515

21. Ciaramella A, Nardone D, Staiano A (2020) Data integration by fuzzy similarity-based hierarchical clustering. *BMC bioinformatics* 21: 1–15. https://doi.org/10.1186/s12859-020-03567-6
22. Lubis AR, Lubis M (2020) Optimization of distance formula in K-Nearest Neighbor method. *Bulletin of Electrical Engineering and Informatics* 9: 326–338. https://doi.org/10.11591/eei.v9i1.1464
23. Schober P, Boer C, Schwarte LA (2018) Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia* 126: 1763–1768. https://doi.org/10.1213/ANE.0000000000002864
24. Xu Q, Wang S, Jiang J, et al. (2019) Nitrate dominates the chemical composition of PM2. 5 during haze event in Beijing, China. *Science of the Total Environment* 689: 1293–1303. https://doi.org/10.1016/j.scitotenv.2019.06.294
25. Yabueng N, Wiriya W, Chantara S (2020) Influence of zero-burning policy and climate phenomena on ambient PM2. 5 patterns and PAHs inhalation cancer risk during episodes of smoke haze in Northern Thailand. *Atmospheric Environment* 232: 117485. https://doi.org/10.1016/j.atmosenv.2020.117485
26. Ly BT, Matsumi Y, Nakayama T, et al. (2018) Characterizing PM2. 5 in Hanoi with new high temporal resolution sensor. *Aerosol and air quality research* 18: 2487–2497. https://doi.org/10.4209/aaqr.2017.10.0435
27. Hassan H, Latif MT, Juneng L, et al. (2021) Chemical characterization and sources identification of PM2. 5 in a tropical urban city during non-hazy conditions. *Urban Climate* 39: 100953. https://doi.org/10.1016/j.uclim.2021.100953
28. Lung SCC, Thi Hien T, Cambaliza MOL, et al. (2022) Research Priorities of Applying Low-Cost PM2. 5 Sensors in Southeast Asian Countries. *International Journal of Environmental Research and Public Health* 19: 1522. https://doi.org/10.3390/ijerph19031522
29. Tarigan ML, Nugroho D, Firman B, et al. (2015) *Pemutakhiran Peta Rawan Kebakaran Hutan dan Lahan di Provinsi Sumatera Selatan. Sumatera Selatan*: Dinas Kehutan Provinsi Sumatera Selatan.
30. Warsono W, Antonio Y, Yuwono S (2021) Modeling generalized statistical distributions of PM2. 5 concentrations during the COVID-19 pandemic in Jakarta, Indonesia. *Decision Science Letters* 10: 393–400. https://doi.org/10.5267/j.dsl.2021.1.005
31. Wu J, Zheng H, Zhe F, et al. (2018) Study on the relationship between urbanization and fine particulate matter (PM2. 5) concentration and its implication in China. *Journal of cleaner production* 182: 872–882. https://doi.org/10.1016/j.jclepro.2018.02.060
32. Zhou Z, Tan Q, Liu H, et al. (2019) Emission characteristics and high-resolution spatial and temporal distribution of pollutants from motor vehicles in Chengdu, China. *Atmospheric Pollution Research* 10: 749-758. https://doi.org/10.1016/j.apr.2018.12.002
33. Kalisa E, Fadlallah S, Amani M, et al. (2018) Temperature and air pollution relationship during heatwaves in Birmingham, UK. *Sustainable cities and society* 43: 111–120. https://doi.org/10.1016/j.scs.2018.08.033
34. Taufik M, Widyastuti MT, Sulaiman A, et al. (2022) An improved drought-fire assessment for managing fire risks in tropical peatlands. *Agricultural and Forest Meteorology* 312: 108738. https://doi.org/10.1016/j.agrformet.2021.108738

35. Rondhi M, Pratiwi PA, Handini VT, et al. (2018) Agricultural land conversion, land economic value, and sustainable agriculture: A case study in East Java, Indonesia. *Land* 7: 148. https://doi.org/10.3390/land7040148
36. Zhou Y, Li X, Liu Y (2020) Land use change and driving factors in rural China during the period 1995-2015. *Land Use Policy* 99: 105048. https://doi.org/10.1016/j.landusepol.2020.105048
37. Li J, Bai Y, Alatalo JM. (2020). Impacts of rural tourism-driven land use change on ecosystems services provision in Erhai Lake Basin, China. *Ecosystem Services* 42: 101081. https://doi.org/10.1016/j.ecoser.2020.101081
38. Campbell-Lendrum D, Prüss-Ustün A (2019) Climate change, air pollution and noncommunicable diseases. *Bulletin of the World Health Organization*, 97: 160. https://doi.org/10.2471/BLT.18.224295
39. Xie J, Liao Z, Fang X, et al. (2019) The characteristics of hourly wind field and its impacts on air quality in the Pearl River Delta region during 2013–2017. *Atmospheric Research* 227: 112–124. https://doi.org/10.1016/j.atmosres.2019.04.023
40. Sari NA, Putra RA (2020) Analisis Statistik Deskriptif Titik Panas di Wilayah Sumatera Selatan. In *Prosiding Seminar Nasional Sains dan Teknologi Terapan* (Vol. 3, No. 1, pp. 51-57).
41. Singh RP, Chauhan A (2020) Impact of lockdown on air quality in India during COVID-19 pandemic. *Air Quality, Atmosphere & Health* 13: 921–928. https://doi.org/10.1007/s11869-020-00863-1
42. Dias D, Tchepel O (2018) Spatial and temporal dynamics in air pollution exposure assessment. *International journal of environmental research and public health* 15: 558. https://doi.org/10.3390/ijerph15030558
43. Suroto A, Shith S, Yusof NM, et al. (2020) Impact of high particulate event on the indoor and outdoor fine particulate matter concentrations during the Southwest monsoon season. In *IOP Conference Series: Materials Science and Engineering* (Vol. 920, No. 1, p. 012007). IOP Publishing. https://doi.org/10.1088/1757-899X/920/1/012007
44. Rendana M (2021) Air pollutant levels during the large-scale social restriction period and its association with case fatality rate of COVID-19. *Aerosol and Air Quality Research* 21: 200630. https://doi.org/10.4209/aaqr.200630
45. Rendana M, Idris WMR, Rahim SA (2021) Spatial distribution of COVID-19 cases, epidemic spread rate, spatial pattern, and its correlation with meteorological factors during the first to the second waves. *Journal of infection and public health*, 14: 1340–1348. https://doi.org/10.1016/j.jiph.2021.07.010
46. Rendana M, Idris WMR, Rahim SA, et al. (2019) Effects of Organic Amendment on Soil Organic Carbon in Treated Soft Clay in Paddy Cultivation Area. *Sains Malaysiana* 48: 61–68. https://doi.org/10.17576/jsm-2019-4801-07
47. Rendana, M, Idris, WMR (2021) New COVID-19 variant (B. 1.1. 7): forecasting the occasion of virus and the related meteorological factors. *Journal of infection and public health*, 14: 1320–1327. https://doi.org/10.1016/j.jiph.2021.05.019