E-ISSN 2586-9027
Homepage : *https://tci-thaijo.org/index.php/SciTechAsia* Science & Technology Asia
Vol.30 No.4 October - December 2025 Page: [126-146]

Original research article

Analyzing $PM_{2.5}$ Levels in Indonesia Using Dynamic Time Warping and Fuzzy Clustering: A Time-Series Ecological Study

Novi Reandy Sasmita[1,*], Mhd Khairul[1], Nurhasanah Nurhasanah[1], Feby Apriliansyah[2], Teuku Rizky Noviandy[3], Lyra Yulianti[4], Syaharuddin Syaharuddin[5]

[1]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh City 23111, Indonesia*
[2]*Department of Regional and Urban Planning, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh City 23111, Indonesia*
[3]*Department of Information Systems, Faculty of Engineering, Universitas Abulyatama, Aceh Besar City 23372, Indonesia*
[4]*Department of Mathematics and Data Science, Faculty of Mathematics and Natural Sciences, Universitas Andalas, Padang City 25163, Indonesia*
[5]*Department of Mathematics Education, Universitas Muhammadiyah Mataram, Mataram City 83115, Indonesia*

Received 20 July 2025; Received in revised form 12 October 2025
Accepted 27 October 2025; Available online 17 December 2025

ABSTRACT

Air pollution from particulate matter (PM2.5) poses serious global health risks. This study evaluated five fuzzy clustering methods—Fuzzy C-Means (FCM), Fuzzy Possibilistic C-Means (FPCM), Possibilistic Fuzzy C-Means (PFCM), Fuzzy Gustafson-Kessel (FGK), and Fuzzy C-Shells (FCS)—integrated with Dynamic Time Warping (DTW) to cluster PM2.5 levels across 33 Indonesian capital cities. Using an ecological time series design, daily $PM_{2.5}$ data (March 6, 2023–March 5, 2024; 12,045 data points) from PlumeLabs were analyzed in R 4.4.2 with descriptive and inferential statistics, including the Kolmogorov-Smirnov, Kruskal-Wallis, and Dunn tests. Three clusters—high, moderate, and low pollution—were identified. The DTW+FCS method showed the best performance (PCI: 0.798, MPCI: 0.697, PEI: 0.357, XBI: 0.197) with significant differences ($p < 0.000$). These findings highlight DTW+FCS as the optimal approach and emphasize targeted air quality strategies for Indonesia's high-pollution areas.

Keywords: About PM2.5 pollution; Fuzzy clustering; Dynamic time warping; Air quality management; Indonesia

***Corresponding author:** novireandys@usk.ac.id **doi:** 10.14456/scitechasia.2025.75

1. Introduction

Air pollution has become a critical environmental and public health concern globally, with particulate matter (PM) posing significant risks to human well-being. Among these pollutants, $PM_{2.5}$ (particulate matter with an aerodynamic diameter of ≤ 2.5 μm) is particularly hazardous due to its ability to penetrate deep into the respiratory system, leading to various health complications, including cardiovascular and respiratory diseases [1]. In urban environments, $PM_{2.5}$ originates from vehicular emissions, industrial activities, and biomass burning [2]. Effective monitoring and clustering of $PM_{2.5}$ pollution levels are essential for developing mitigation strategies and improving air quality management. However, traditional clustering techniques often fail to account for air pollution data's dynamic and uncertain nature [3].

Clustering is a widely employed data mining technique used to classify datasets into distinct groups based on shared characteristics. While hard clustering methods such as K-means assign each data point to a single cluster, they lack the flexibility to capture uncertainty inherent in environmental data [4,5]. To address this limitation, fuzzy clustering has emerged as an effective approach, allowing data points to belong to multiple clusters with varying degrees of membership. Several fuzzy clustering methods have been proposed, including Fuzzy C-Means (FCM), Fuzzy Possibilistic C-Means (FPCM), Possibilistic Fuzzy C-Means (PFCM), Fuzzy Gustafson-Kessel (FGK), and Fuzzy C-Shells (FCS), each demonstrating different strengths in handling complex datasets [6].

Recent studies have explored the use of fuzzy clustering in diverse domains. For instance, [7] applied FCM for tumour region classification, highlighting its efficacy in distinguishing overlapping features. Similarly, [8] demonstrated the superior performance of PFCM in classifying agricultural datasets with high ambiguity. [9] Implemented FGK for districts/cities based on poverty issues factors in Kalimantan Island, Indonesia, while [10] applied FCS to stock market clustering. These studies emphasize the growing relevance of fuzzy clustering in handling complex, ambiguous data structures.

The increasing complexity of air pollution data necessitates more advanced analytical techniques. While traditional clustering methods have been applied to $PM_{2.5}$ classification [11], their ability to manage time series data remains limited. Given that $PM_{2.5}$ concentrations exhibit temporal fluctuations influenced by meteorological conditions and human activities, an advanced technique is required to capture these variations. Dynamic Time Warping (DTW) has emerged as a powerful method for analyzing time series data, enabling the comparison of sequences with varying temporal alignments [12, 13]. Integrating DTW with fuzzy clustering methods offers a novel approach to classifying $PM_{2.5}$ pollution levels dynamically.

Despite the advancements in fuzzy clustering and DTW, there remains a gap in research regarding the optimal fuzzy clustering method for $PM_{2.5}$ time series data. Previous studies primarily focused on regional air pollution clustering without comprehensively comparing multiple fuzzy clustering techniques [13]. Moreover, existing research has largely overlooked the integration of DTW in $PM_{2.5}$ classification, leaving a significant gap in understanding the most effective approach for analyzing dynamic pollution data.

To address this gap, this study aims to evaluate and compare five fuzzy clustering methods: FCM, FPCM, PFCM, FGK, and FCS integrated with DTW for classifying $PM_{2.5}$ time series data from Indonesia. The research seeks to determine the most effective method for identifying high, moderate, and low pollution clusters.

The integration of DTW with fuzzy clustering methods is a crucial aspect of this study, designed to enhance the analysis of time-series data on $PM_{2.5}$ concentrations. The procedure begins with the application of DTW as a pre-processing step to measure the similarity between the $PM_{2.5}$ time series of different cities. Unlike traditional distance metrics such as Euclidean distance, which compares time series on a point-by-point basis, DTW finds the optimal alignment between two time series by ”warping” the time axis, thus accommodating variations in timing and speed [15]. This process results in a distance matrix where each entry represents the DTW distance between the $PM_{2.5}$ profiles of two cities. This matrix, which captures the shape-based similarity of the pollution patterns, is then used as the primary input for the various fuzzy clustering algorithms. By feeding this DTW-derived distance matrix into the clustering algorithms, we can group cities based on the similarity of their pollution pattern trajectories over the entire study period, rather than merely their average pollution levels.

The novelty of this study lies in its comprehensive comparison of multiple fuzzy clustering techniques applied to $PM_{2.5}$ time series data using DTW. While previous research has explored fuzzy clustering in various domains, this study uniquely integrates DTW to enhance the accuracy of air pollution classification. The findings will contribute to developing data-driven air quality management strategies, supporting policymakers in mitigating $PM_{2.5}$ related health risks.

2. Materials and Methods

2.1 Study design, setting, and study size

This study employed a time series ecological study to systematically classify $PM_{2.5}$ pollution levels in Indonesia. Its study is particularly effective for identifying and analyzing temporal trends and associations within historical data sets [14], making it ideal for assessing $PM_{2.5}$ pollutant levels. The study included 33 cities selected to represent major urban centres across Indonesia. The final dataset comprised 365 daily observations per city, totalling 12,045 data points, ensuring a geographically diverse representation of Indonesia's air pollution landscape.

2.2 Data source, data preprocessing and parameter specification

All analyses are based on a time-series dataset of daily mean $PM_{2.5}$ concentrations (μg/m^3) for 33 provincial capitals in Indonesia. The data were obtained from Plume Labs Air Quality and cover one year from March 6, 2023, to March 5, 2024. The data is publicly available under a creative commons license, from historical archives available at https://plumelabs.com/. Prior to analysis, the dataset underwent a rigorous preprocessing protocol. Time series for each city were inspected for missing values. In all analyses of this study, basic data in the form of daily $PM_{2.5}$ averages for each city were used.

Statistical analysis is applied in this study. Outliers were identified using the interquartile range (IQR) method. The clustering analysis was performed in R using the dtwclust package for DTW analysis, ppclust package for FCM, FPCM, PFCM, FGK analysis, $e1071$ package for FCS analysis, $fclust$ package for cluster evaluation, and $ggplot2$ package for visualisation and different tests.

In this study, the number of clusters was set to $k = 3$, corresponding to low, moderate, and high levels of $PM_{2.5}$ pollution. This pre-specification is based on the practical objective of generating policy-relevant classifications that are both interpretable and actionable for environmental management and public health interventions. Creating a small number of distinct categories aligns with the structure of most air quality indices used globally and simplifies the formulation of targeted mitigation strategies for distinct pollution tiers. This approach of defining a realistic number of clusters based on practical goals has been utilised in previous environmental studies analysing $PM_{2.5}$ data, which similarly aim to provide clear outputs for decision-makers like in some countries [11, 14, 17]. By defining the clusters a priori, we ensure the results are directly translatable to established air quality management frameworks.

2.3 Statistical analysis

This study employs statistical methods to derive generalizations from the findings [15, 16]. Data analysis was performed using R 4.4.2 for statistical computations and visualization. The analysis consisted of two main stages: descriptive statistics and inferential statistics. Descriptive statistics included central tendency and dispersion measures such as minimum, mean, maximum, standard deviation and interquartile range to summarize variable distributions [17, 18]. Furthermore, DTW is used to measure the similarity between two-time series. Fuzzy clustering as a soft computing method is also implemented to find the optimal cluster based on the degree of membership [19]. The clustering results using DTW integration and each fuzzy clustering method used will be validated to determine the best integration in clustering cities based on pollutant levels. This study divides pollutant levels into three categories: low, moderate and high.

Next up, inferential statistical analysis began with normality testing using the Kolmogorov-Smirnov test to determine whether the data followed a normal distribution [20]. This guided the selection of either parametric or nonparametric statistical tests. If data were normally distributed, Analysis of Variance (ANOVA) was applied as a parametric test, followed by the Bonferroni test for post hoc comparisons [21, 22]. If the data were not normally distributed, the Kruskal-Wallis was used as a nonparametric alternative, with the Dunn test employed for post hoc analysis [23]. This test is an alternative when the data does not meet the normality assumption required by the ANOVA test and can be applied to more than three groups of dependent variables [24, 25]. The data compared by cluster is the result of the best method. The difference test in this study was conducted to confirm that the three clusters formed were distinct and to ensure the accuracy of the formed clusters.

2.4 Dynamic Time Warping (DTW)

This stage calculates a distance matrix that measures the difference between each pair of time series data. DTW helps identify the similarity of $PM_{2.5}$ pollutant concentrations. Each pair of cities is compared against the concentration time series data to generate the DTW distance. The main characteristic of DTW is its ability to compare time series with different lengths. DTW performs dynamically within the time series to achieve an optimal fit, thus making a more flexible and accurate comparison between two-time series [26]. The result of this process is the dynamic distance, a measure of similarity between the two matched time series [27]. The DTW method involves the following steps:

Step 1-Calculating the local cost matrix (D). This step calculates the distance between each element of order a and each element of order b. The distance (d_{ij}) is the absolute value of the element difference, where d_{ij} is the distance between elements a_i and b_j.

$$d_{ij} = |a_i - b_j|, \tag{2.1}$$

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1j} \\ d_{21} & d_{22} & \cdots & d_{2j} \\ \vdots & \vdots & \vdots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nm} \end{bmatrix}. \tag{2.2}$$

Step 2: Construct the global cost matrix (G). This step is used to determine the optimal warping path. The matrix G is the same size as the matrix D. g_{ij} is the element of the global cost matrix. The element of the G matrix G (g_{ij}) can be calculated with the following equation.

$$g_{ij} = d_{ij} + \min(g_{(i-1),j}, g_{i,(j-1)}, g_{(i-1),(j-1)}), \tag{2.3}$$

$$G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1j} \\ g_{21} & g_{22} & \cdots & g_{2j} \\ \vdots & \vdots & \vdots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nm} \end{bmatrix}. \tag{2.4}$$

Step 3-Determine the cost function that becomes the starting point of the warping path. The starting point of the warping path is the point with the lowest cumulative cost in the global cost matrix. The starting point of the warping path is g_{nm}. Step 4-Determine the optimal warping path. The warping path is a path through matrix G that starts from the starting point and ends at $g_{1,1}$, with the lowest cumulative cost. Starting from g_{nm}, select neighbouring cells from the three options $g_{(i-1),j}, g_{i,(j-1)}, g_{(i-1),(j-1)}$ until reaching $g_{1,1}$.

Step 5-Calculating DTW distance (d_{DTW}). The DTW distance is calculated as the cumulative value of the optimal warping path. Calculate the cumulative value by taking the total cost along the warping path.

$$d_{DTW} = \sum_{(i,j)\in path} (g_{ij}). \tag{2.5}$$

2.5 Fuzzy clustering

The fuzzy clustering methods used in this study are FCM, which focuses on cluster centering; FPCM, which integrates possibilistic principles into fuzzy clustering; PFCM, which is a variation of FPCM with a different emphasis; FGK, which takes into account the shape of the cluster; and FCS, which is used to identify clusters with non-linear shapes. Each of these methods uses a different analysis to analyze the data.

FCM is a data clustering method of the k-means method, which incorporates fuzzy principles into its membership function [28]. In FCM, the objective function P_t uses μ_{ik} as a random number of membership elements, was the rank of fuzzy weights, v_{kj} as the cluster centre, and d_{ik} as the distance of sample x_{ij} to cluster centre v_{kj}. The FCM objective function at the t-th iteration (P_t) is as follows [29].

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}((\mu_{ik})^w(\sum_{j=1}^{m}(x_{ij} - v_{kj})^2)). \tag{2.6}$$

FPCM is the development of FCM and Possibilistic C-Means. FPCM relies on two parameters, namely fuzziness and possibilistic parameters [30]. In FPCM, the objective function P_t uses μ_{ik} and τ_{ik} as random numbers of membership elements and typicality, w and v as the rank of fuzzy and possibilistic weights, and d_{ik} as the distance of sample x_{ij} with cluster center v_{kj}. The FPCM objective function at the t-th iteration (P_t) is as follows [31].

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} (((\mu_{ik})^w + (\tau_{ik})^v)(\sum_{j=1}^{m} (x_{ij} - v_{kj})^2)). \quad (2.7)$$

Furthermore, the method used is PFCM. The difference between PFCM and FPCM is the order of application, PFCM first uses a fuzzy approach to determine the cluster center and calculate the membership degree [32]. After that, PFCM applies the possibilistic principle. In PFCM, the objective function P_t uses μ_{ik} and τ_{ik} as random numbers of membership elements and typicality, w and v as fuzzy and possibilistic weighting ranks, d_{ik} as the distance of sample x_{ij} with cluster center v_{kj}, Ω_k as a random number of reference distance. The PFCM objective function at the t-th iteration (P_t) is as follows [31].

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} (((\mu_{ik})^w + (\tau_{ik})^v)(\sum_{j=1}^{m} (x_{ij} - v_{kj})^2)) + \sum_{i=1}^{n} (1 - \tau_{ik})^n \sum_{k=1}^{c} \Omega_k. \quad (2.8)$$

FGK is a method that classifies data based on the membership value of each data in a cluster [33]. In FGK, the objective function P_t uses μ_{ik} as a random number of membership elements, w as the rank of fuzzy weights, A_k as the cluster norm inducer, d_{ikA_k} as the distance of sample x_{ij} and cluster center v_{kj} with cluster norm inducer, F_k as the k-th adaptive covariance matrix element, and v_{kj} as the cluster center. The FGK objective function at the t-th iteration (P_t) is as follows [33].

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} ((\mu_{ik})^w (\sum_{j=1}^{m} (x_{ij} - v_{kj})^2) A_k^2). \quad (2.9)$$

The FCS method is a development method of FCM. Both FCS and FCM share similar characteristics in terms of having cluster centres. However, FCS adds a unique dimension in its clustering process by including radius as an additional parameter. In FCS, the objective function P_t uses μ_{ik} as the random number of membership elements, w as the rank of fuzzy weights, v_{kj} as the cluster center, and d_{ik} as the sample distance between sample x_{ij} and prototype (v_{kj}, r_k). The FCS objective function at the the t-th iteration (P_t) is as follows [34].

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} (\mu_{ik})^w (d_{ik})^2. \quad (2.10)$$

2.6 Index of cluster validation

Cluster validation is a step to evaluate cluster analysis results based on quantitative criteria and objective reality. Cluster validation not only aims to assess the results of cluster analysis but is also a method to evaluate the performance of the clustering method that has been implemented [35]. There are four cluster validations used in this study, namely Partition Coefficient Index (PCI), Modified Partition Coefficient Index (MPCI), Partition Entropy Index (PEI), and Xie-Benni Index (XBI).

The Partition Coefficient Index PCI is used to evaluate the compactness of fuzzy clusters by measuring how strongly data points belong to specific clusters. The formula gives it:

$$PCI = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik})^2, \qquad (2.11)$$

where n represents the total number of data points, c is the number of clusters, and μ_{ik} denotes the membership value of the i-th data point in the k-th cluster, which ranges between 0 and 1. PEI values close to 0 indicate that the clustering results are more optimal [36]. A higher PCI value indicates well-separated clusters, but this index tends to favour a higher number of clusters. The higher the PCI value, the more optimal the clustering result is in distinguishing and grouping the data into specific clusters [37]. Further, MPCI is an improved version of PCI that normalizes the coefficient to avoid overestimating the number of clusters. It is defined as:

$$MPCI = \frac{PCI}{\max(PCI)}, \qquad (2.12)$$

where $\max(PCI)$ is the maximum PCI value obtained across different cluster numbers. By normalizing the PCI value, MPCI provides a better estimate of the correct number of clusters without biasing toward a larger number of clusters. A high MPCI value indicates a good level of separation between clusters, as well as an indication that the data in each cluster has high homogeneity. The higher the index value, the better the quality of separation and clarity between the clusters formed [38].

PEI measures the fuzziness in a clustering solution, with lower values indicating well-defined clusters. It is calculated as:

$$PEI = \frac{1}{n}(\sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik}\log(\mu_k))), \qquad (2.13)$$

where n is the number of data points, c is the number of clusters, and μ_{ik} is the membership value of the i-th data point in the k-th cluster. Since the logarithmic term amplifies lower membership values, PEI effectively captures the degree of fuzziness in the clustering structure, with smaller values indicating more distinct clusters. Lastly, XBI evaluates both the compactness and separation of clusters, helping to identify the optimal clustering structure. It is expressed as:

$$XBI = \frac{\sum_{i=1}^{n}\sum_{k=1}^{c} u_{ik}^2 \|x_i - v_j\|^2}{n \times \min_{ik}\|v_j - v_k\|^2}, \qquad (2.14)$$

where x_i represents the i-th data point, v_j is the centre of cluster j, and $\|x_i - v_j\|$ is the Euclidean distance between a data point and its assigned cluster centre. The denominator, $\min_{ik}\|v_j - v_k\|^2$, represents the minimum squared Euclidean distance between any two cluster centres. A lower XBI value indicates well-separated and compact clusters, making it an effective measure for assessing clustering quality. XBI is a validation to assess both the compactness of a cluster and the separation between clusters, thus providing a balanced way to evaluate cluster structure. A smaller XBI value indicates a more optimal cluster [39].

2.7 Different test and post hoc test

The ANOVA test is a parametric statistical analysis that evaluates the mean difference between three or more groups. The assumption in applying the ANOVA test is that the variance between groups must be normal [22]. The Kruskal-Wallis test is a nonparametric statistical analysis utilized to identify whether there are statistically significant differences between three or more groups. This test is an alternative when the data does not meet the normality assumption required by the ANOVA test and can be applied to more than three groups of dependent variables [24].

Furthermore, in the Post Hoc test stage, the Bonferroni Test is a post-hoc correction method used after ANOVA if the null hypothesis is rejected [40]. The Bonferroni test is a widely used method in statistical analysis to address the problem of type I error rates that can arise from multiple comparisons between groups [41]. Dunn's test was used to identify different groups after Kruskal-Wallis showed an overall median difference. This test is particularly useful for comparing rankings between groups and correcting for multiple comparisons, generally using methods such as Bonferroni correction to preserve type I error rates [23].

3. Result and Discussion

Table 1 provides an overview of $PM_{2.5}$ pollutant conditions and environmental challenges faced by various cities in Indonesia. Jakarta stands out with an average $PM_{2.5}$ pollutant value of 70.84 $\mu g/m^3$, far exceeding the average values of other cities, and recording a maximum recorded pollutant value of 173 $\mu g/m^3$. This shows that Jakarta frequently experiences high levels of air pollution and faces extreme pollution peaks. In contrast, Palangka Raya showed a different phenomenon with the maximum value of $PM_{2.5}$ pollutant reaching 203 $\mu g/m^3$, the highest value among all cities studied. However, the average value was not as high as Jakarta. This result shows that Palangka Raya experiences extreme air pollution events even though the average pollution is lower than Jakarta. A visualization of the average $PM_{2.5}$ levels for each city in Indonesia during the specified study period is illustrated in Fig. 1.

A comparison between Indonesian cities shows significant variations in pollution levels, which can be attributed to various factors, such as geographical conditions, industrial activity, and traffic congestion. Ambon, with a mean value of 0.17 $\mu g/m^3$, a maximum of 4 $\mu g/m^3$, and a standard deviation of 0.554 $\mu g/m^3$, shows consistent pollutant levels with less fluctuation than Jakarta and Palangka Raya. This indicates that $PM_{2.5}$ pollutants in Ambon are only sourced from consistent daily activities, without many extreme events that increase pollutant levels drastically.

Furthermore, while the majority of cities have lower and relatively uniform levels, a few cities stand out with higher levels. Jakarta faces great challenges in managing air quality. This comparison is important as it shows areas that need urgent attention and reveals successes in lowering air pollution. Cities with lower levels may indicate better environmental management practices or geographical luck that results in less intense dispersion of pollutants. Then, figure 2 displays the median marked by the center line, showing the middle $PM_{2.5}$ concentration of each city. The length of the box representing the IQR shows the variation of concentration in each city within the middle 50% of the data, and the points outside the whisker represent outlier values.

Fig.2 shows significant variation between cities. Some cities, such as Jakarta and Palangka Raya show a wide range, indicated by the length of the whiskers and the presence of outliers. In contrast, cities such as Ambon, Gorontalo, and Ternate have very small boxes and short whiskers, indicating small variations in $PM_{2.5}$ concentrations. The standard deviation and IQR values of $PM_{2.5}$ for each city influence the condition. Each city has outlier data above Q3. These outliers in the $PM_{2.5}$ data indicate that there are events with pollutant concentrations that are very different from normal. Palangka Raya has high outlier data compared to other cities.

Based on these outlier data and highly diverse data, a fuzzy clustering method was used to group the 33 cities.

Table 1. Distribution of $PM_{2.5}$ (μg/m^3) for 33 provincial capitals in Indonesia.

No	Cities	Summary Statistics				
		Minimum	**Mean**	**Maximum**	**Standard Deviation**	**IQR**
1	Ambon	0	0.17	4	0.55	0.00
2	Banda Aceh	0	1.77	15	2.81	2.00
3	Bandar Lampung	0	13.30	55	10.33	13.00
4	Bandung	6	34.69	87	12.81	16.00
5	Banjarmasin	0	7.07	88	8.08	9.00
6	Bengkulu	0	5.93	31	5.21	6.00
7	Denpasar	0	1.79	26	4.45	1.00
8	Gorontalo	0	0.41	5	0.89	0.00
9	Jakarta	1	70.84	173	30.21	41.75
10	Jambi	0	10.45	36	7.21	10.00
11	Jayapura	0	0.92	4	0.79	1.00
12	Kendari	0	3.20	18	4.42	6.00
13	Kupang	0	0.83	12	1.79	1.00
14	Makassar	0	4.73	22	5.35	8.00
15	Manado	0	1.46	16	2.65	2.00
16	Mataram	0	2.31	24	3.79	3.00
17	Medan	0	29.19	74	12.98	15.00
18	Padang	0	8.52	29	5.65	6.00
19	Palangkaraya	0	13.49	203	26.45	6.00
20	Palembang	0	15.59	99	16.51	17.00
21	Palu	0	6.83	18	3.50	5.00
22	Pangkalpinang	0	1.91	35	4.14	2.00
23	Pekanbaru	0	16.09	59	10.23	13.00
24	Pontianak	0	12.39	71	13.15	14.00
25	Samarinda	0	5.22	26	5.17	9.00
26	Semarang	0	23.51	72	13.51	16.75
27	Serang	0	38.93	125	26.83	40.00
28	Sorong	0	0.95	11	1.83	1.00
29	Surabaya	0	35.31	118	20.96	28.00
30	Tanjung Pinang	0	1.74	30	3.91	2.00
31	Tarakan	0	3.52	21	3.70	5.75
32	Ternate	0	0.52	12	1.42	0.00
33	Yogyakarta	0	14.99	81	17.67	20.00

Clustering Indonesian cities based on $PM_{2.5}$ pollutants, it is important to identify patterns and trends that are not immediately apparent through descriptive statistical analysis. Fuzzy clustering provides flexibility in clustering, allowing cities to be grouped based on non-strict membership levels and accommodating ambiguity in time series data. Focusing on cities that belong to clusters with both the highest and most variable pollutant levels can prevent or reduce negative impacts on public health.

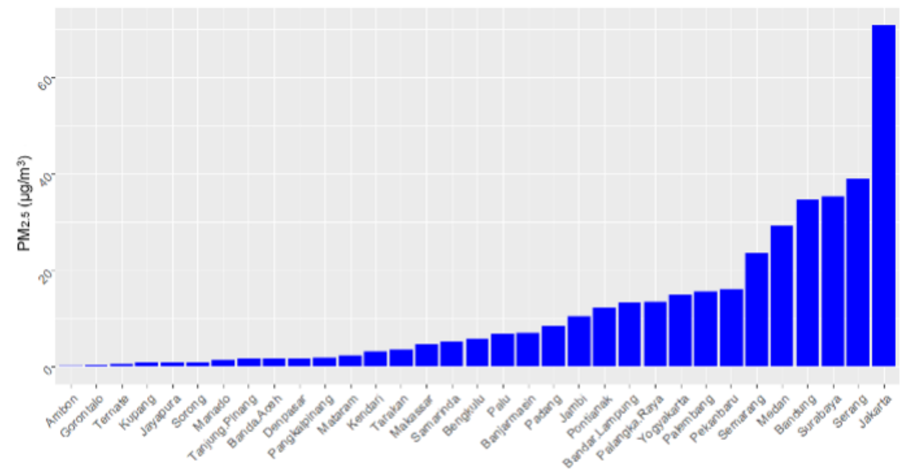

Fig. 1. Average $PM_{2.5}$ pollutants from 33 cities in Indonesia

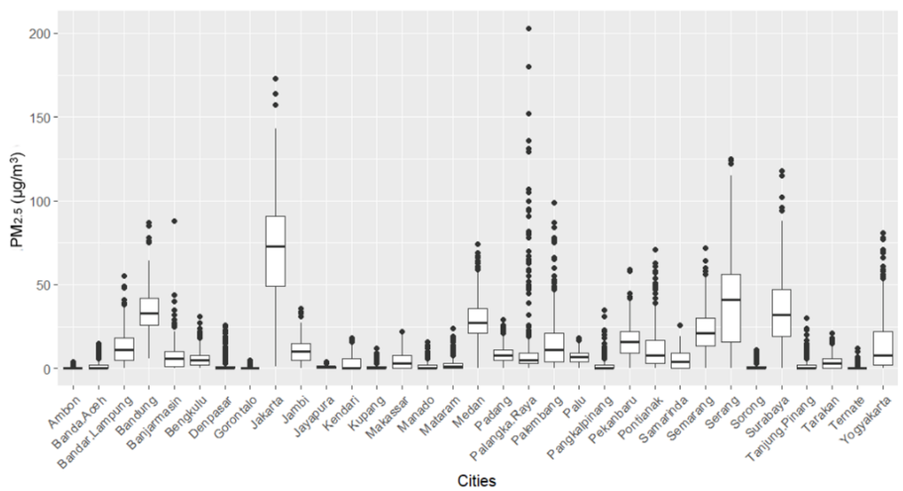

Fig. 2. Distribution of $PM_{2.5}$ from 33 cities in Indonesia.

3.1 Distribution of cities based on cluster

Table 2 presents the distribution of the 33 cities in Indonesia based on clustering results using the integration of DTW and each fuzzy clustering method. Through the application of DTW matrix, FCM method clusters 9 cities as cluster 1, 20 cities as cluster 2, and 4 cities as cluster 3. FPCM method clusters 10 cities as cluster 1, 17 cities as cluster 2, and 6 cities as cluster 3. Then, the PFCM method groups 4 cities as cluster 1, 3 cities as cluster 2, and 26 cities as cluster 3. Furthermore, the FGK method groups 9 cities as cluster 1, 19 cities as cluster 2, and 5 cities as cluster 3. Meanwhile, the FCS method groups 14 cities as cluster 1, 6 cities as cluster 2, and 13 cities as cluster 3.

3.2 Evaluation of cluster results

The selection of the best clustering method is based on maximizing PCI and MPCI while minimizing PEI and XBI. Table 3 indicates that DTW+FCS is the most effective for clustering cities in Indonesia based on time-series $PM_{2.5}$ pollutant data among the five tested methods. The FCS method achieves the highest PCI and MPCI values, 0.798 and 0.697, respectively. A maximum PCI and MPCI value in fuzzy clustering indicates well-defined and compact clusters with minimal overlap, suggesting that the chosen clustering structure effectively partitions the data. Higher values imply that data points have strong membership to specific clusters, reducing fuzziness and enhancing interpretability, making it the optimal choice for this study. The DTW+FCS method also achieves the lowest PEI and XBI values, 0.357 and 0.197, respectively, indicating well-separated and compact clusters. The minimal PI value suggests a lower degree of fuzziness, while the low XBI value reflects high intra-cluster similarity and clear cluster distinction, further confirming the effectiveness of DTW+FCS in producing optimal clustering results.

The superior performance of the FCS method, when integrated with DTW, can be attributed to its unique clustering mechanism. Unlike point-based algorithms like FCM that identify spherical clusters based on prototype centers, FCS identifies clusters based on hyperspherical shells, making it adept at detecting contours and boundaries in the data [45,46]. When analyzing time-series data processed by DTW, the resulting distance matrix reflects complex, shape-based similarities between pollution patterns that may not form simple spherical groups. The FCS algorithm's ability to define clusters by a prototype shell defined by its center and radius allows it to capture these more complex, non-spherical relationships more effectively. Therefore, the combination of DTW's robust time-series similarity measurement and FCS's flexible, shell-based clustering geometry provides a more accurate and nuanced classification of cities with similar $PM_{2.5}$ temporal dynamics, which explains its higher validation scores in our study.

3.3 Distribution of DTW+FCS cluster result

The results of normality testing with a 5% significance level using the Kolmogorov-Smirnov test showed that each cluster formed from DTW+FCS did not follow a normal distribution (p-value <0.000). Therefore, the Kruskal-Wallis test based on the median was applied to see the difference in statistical distribution.

Fig. 3. illustrates the distribution of $PM_{2.5}$ concentrations across three clusters identified using the DTW+FCS method. The Kruskal-Wallis test (χ^2=4,621.62, p-value < 0.000) indicates a statistically significant difference in $PM_{2.5}$ levels among the clusters, further confirmed by the Dunn pairwise test, which shows significant differences between all pairs with p-value < 0.000. Cluster 1, with a median $PM_{2.5}$ of 4 μg/m^3, represents areas with moderate pollution levels, likely indicating relatively clean air conditions. Cluster 2, with a median $PM_{2.5}$ of 33 μg/m^3, signifies highly polluted regions, possibly urban or industrial zones with higher pollution exposure. Meanwhile, Cluster 3, with a median $PM_{2.5}$ of 1.00 μg/m^3, represents areas with very low pollution levels, potentially rural or less industrialized regions. The location of each city based on its cluster can be seen in Fig. 4.

Cluster 2 exhibits the widest variability in $PM_{2.5}$ values, suggesting a diverse pollution range within this group. In contrast, Cluster 3 has the lowest $PM_{2.5}$ concentrations with minimal variation, indicating

Table 2. Distribution of 33 cities in Indonesia based on cluster results.

No	Cities	DTW+FCM	DTW+FPCM	DTW+PFCM	DTW+FGK	DTW+FCS
1	Ambon	2	2	3	3	3
2	Banda Aceh	2	2	2	1	1
3	Bandar Lampung	1	1	3	2	1
4	Bandung	3	3	1	1	2
5	Banjarmasin	1	1	3	1	3
6	Bengkulu	2	1	3	3	3
7	Denpasar	2	2	3	2	3
8	Gorontalo	2	2	3	2	3
9	Jakarta	3	3	3	3	2
10	Jambi	2	1	3	2	3
11	Jayapura	2	2	3	1	3
12	Kendari	2	2	3	2	1
13	Kupang	2	2	3	1	3
14	Makassar	2	2	3	2	1
15	Manado	2	2	2	2	1
16	Mataram	2	2	3	2	1
17	Medan	1	3	1	2	2
18	Padang	2	1	3	2	3
19	Palangka Raya	1	1	3	1	1
20	Palembang	1	1	3	1	1
21	Palu	2	2	3	2	1
22	Pangkalpinang	2	2	3	2	3
23	Pekanbaru	1	1	3	2	1
24	Pontianak	1	1	3	1	1
25	Samarinda	2	2	3	2	3
26	Semarang	1	3	3	2	2
27	Serang	3	3	1	2	2
28	Sorong	2	2	3	3	3
29	Surabaya	3	3	1	1	2
30	Tanjung Pinang	2	2	3	3	3
31	Tarakan	2	2	2	2	1
32	Ternate	2	2	3	2	1
33	Yogyakarta	1	1	3	2	1

Table 3. Cluster evaluation results of $PM_{2.5}$ Pollutant data.

No	Methods	Cluster Validity Indices			
		PCI	MPCI	PEI	XBI
1	DTW+FCM	0.789	0.684	0.392	0.238
2	DTW+FPCM	0.790	0.685	0.390	0.239
3	DTW+PFCM	0.655	0.482	0.580	2.025
4	DTW+FGK	0.439	0.158	0.925	1.036
5	DTW+FCS	0.798	0.697	0.357	0.197

consistently clean air quality. The significant differences between clusters confirm that the DTW+FCS clustering method effectively distinguishes regions based on air pollution levels, making it a valuable tool for air quality monitoring and targeted environmental policy-making.

Furthermore, the spatial distribution of $PM_{2.5}$ pollution across Indonesia, classified using the DTW+FCS clustering method, provides a comprehensive insight into regional air quality variations. By clustering cities into three distinct clusters, the analysis highlights significant disparities in pollution levels influenced by factors such

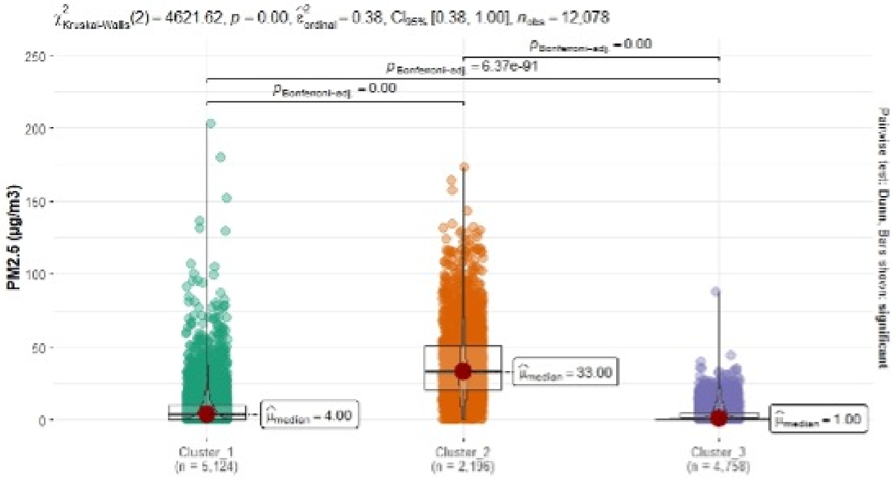

Fig. 3. Comparison cluster-based distribution of $PM_{2.5}$ using DTW+FCS method.

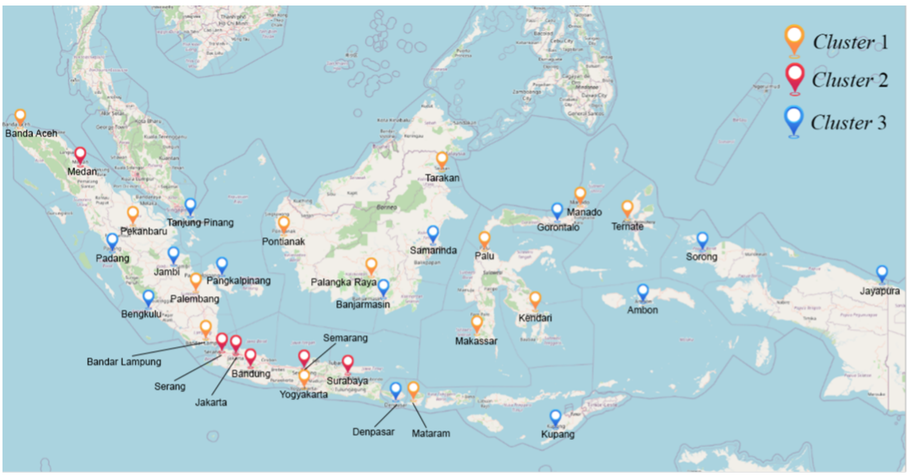

Fig. 4. Map of $PM_{2.5}$ pollutant clusters in Indonesia using DTW+FCS..

as urbanization, industrial activities, and geographical characteristics. The clustering results reveal a clear pattern where densely populated and industrialized regions exhibit the highest pollution levels, whereas less urbanized and ecologically preserved areas maintain better air quality.

Cluster 2 (High $PM_{2.5}$ Pollution) is a major urban and Industrial Hotspot. Cluster 2, marked in orange, represents the highest $PM_{2.5}$ concentrations and is predominantly located in Java and parts of Sumatra. This cluster includes Bandung, Jakarta, Medan, Semarang, Serang, and Surabaya, which are known for their dense population, heavy traffic emissions, and industrial activities. The high concentration of $PM_{2.5}$ in these areas can be attributed to several sources, including vehicle emissions, coal-fired power plants, and industrial waste.

The analysis results of this study are in line with several previous studies [42], which stated that large cities such as Jakarta, Medan, and Surabaya, which are industrial centers or have high population levels, contribute significantly to pollution. Fossil fuel combustion, including household heating and industrial activities, contributes significantly to high $PM_{2.5}$ concentrations in urban areas [2].

The severe air quality in these cities poses significant health risks, increasing the prevalence of respiratory diseases, cardiovascular issues, and reduced life expectancy. Immediate and stringent air pollution control measures, including vehicle emission regulations, industrial pollution monitoring, and reforestation programs, are necessary to mitigate the environmental impact in these regions [43].

Cluster 1 (Moderate $PM_{2.5}$ Pollution) is a transitional region with emerging air quality concerns. Cluster 1, marked in green, represents moderate $PM_{2.5}$ pollution levels and includes cities distributed across Sumatra, Kalimantan, Sulawesi, and parts of Java such as Banda Aceh, Bandar Lampung, Kendari, Makassar, Manado, Mataram, Palangka Raya, Palembang, Palu, Pekanbaru, Pontianak, Tarakan, Ternate, dan Yogyakarta. These regions experience pollution levels that are neither as severe as Cluster 2 nor as clean as Cluster 3, indicating transitional zones where a mix of natural and anthropogenic factors influences air quality. The presence of mining, palm oil plantations, forest fires, and industrial expansion in Sumatra and Kalimantan contributes to fluctuating pollution levels. Additionally, cities such as Banda Aceh, Yogyakarta and East Kalimantan suggest that urbanization gradually impacts air quality. This is supported by the study of [44] who attributed the increase in $PM_{2.5}$ pollutants in Palangka Raya to forest fires.

Meanwhile, a study by [45, 46] found that land clearing and forest fires are one of the causes of the spread of $PM_{2.5}$ pollutants in Pekanbaru, and a study by Kemala et al. [47] found that population growth and increased transportation contributed to $PM_{2.5}$ pollutant pollution [48]. If proactive air quality management strategies are not implemented, these areas risk transitioning into high-pollution zones (Cluster 2) over time. Therefore, early intervention policies, such as sustainable urban planning, industrial emission controls, and stricter environmental regulations, are crucial to prevent further deterioration.

Cluster 3 (Low $PM_{2.5}$ Pollution) is Indonesia's least polluted and ecologically preserved area. Cluster 3, marked in purple, consists of cities with the lowest $PM_{2.5}$ concentrations, mainly found in eastern Indonesia, including Papua, Sulawesi, and parts of Kalimantan. This cluster includes Ambon, Banjarmasin, Bengkulu, Denpasar, Gorontalo, Jambi, Jayapura, Kupang, Padang, Pangkalpinang, Samarinda, Sorong dan Tanjung Pinang. These areas are characterized by low population density, minimal industrialization, and abundant natural forests, contributing to better air quality [49]. The lack of heavy traffic and large-scale industrial emissions helps maintain relatively pristine atmospheric conditions [50].

In recent assessments of air pollution, cities in eastern Indonesia, particularly in regions such as Papua, Sulawesi, and parts of Kalimantan, have been recognized as exhibiting some of the lowest $PM_{2.5}$ concentrations. This notable air quality can be attributed to several factors, including the abundance of natural vegetation, which helps to filter pollutants, and lower levels of industrialization compared to urban areas in other regions.

Cities such as Ambon and Jayapura stand out in this context. Extensive forest cover plays a critical role in maintaining air quality, as forests are known to absorb pollutants and produce oxygen. These cities generally experience less vehicular and industrial pollution, leading to lower $PM_{2.5}$ levels. Previous research has highlighted the correlation between urbanization and air quality, indicating that regions with limited urban expansion and industrial activities tend to have cleaner air [51, 52].

Moreover, the geographic and meteorological conditions unique to eastern Indonesia, including strong winds and humidity levels, contribute to the dispersion and dilution of air pollutants that may arise from natural and anthropogenic sources [53]. This combination of natural environmental factors and limited human impact ensures that cities in this region enjoy a relatively clean atmosphere, reinforcing their status as locations with some of the lowest $PM_{2.5}$ concentrations globally.

However, while these areas enjoy clean air, future infrastructure development and economic expansion could threaten their air quality. To ensure sustainable development, environmental conservation policies, green urban initiatives, and air pollution monitoring systems should be proactively implemented to maintain these regions' air quality. The clustering results reveal distinct regional pollution trends, highlighting disparities between western and eastern Indonesia.

Java, Sumatra, and parts of Kalimantan exhibit significantly higher pollution levels than Sulawesi, Papua, and eastern Kalimantan, which remain relatively unpolluted. The primary factors contributing to these differences include population density, industrial activity, and deforestation. Java, as the economic and industrial hub of Indonesia, experiences severe air pollution due to high energy consumption, traffic congestion, and manufacturing industries. Sumatra and Kalimantan, while less urbanized, face seasonal air quality degradation from agricultural burning, mining, and logging activities. In contrast, Papua and eastern Indonesia benefit from low industrialization and extensive forest coverage, acting as a natural buffer against air pollution [54].

Another critical observation is the geographical clustering of pollution sources. Coastal and inland cities in Java and Sumatra are disproportionately affected by dense transportation networks, port activities, and industrial zones. On the other hand, mountainous and forested regions in Sulawesi and Papua maintain cleaner air due to minimal human-induced pollution. These variations emphasize the need for region-specific air quality management strategies, ensuring each region receives tailored environmental policies based on its pollution profile.

Regarding implications for environmental policy and air quality management, this clustering analysis provides valuable insights for environmental policymakers to design targeted pollution control strategies. The categorization into three clusters allows for prioritized air quality interventions, ensuring that high-risk regions receive urgent mitigation measures, while cleaner areas are safeguarded against future degradation [55].

Immediate air pollution control measures must be implemented to address the severe pollution in Cluster 2 regions. One of the most critical steps is enforcing stricter vehicle emission standards and enhancing public transportation systems to reduce traffic-related pollution, a major contributor in densely populated cities. Additionally, industrial emission regulations should be strengthened, particularly in manufacturing hubs across Java and Sumatra, to limit the release of harmful pollutants. Another key intervention is the implementation of peatland fire prevention programs, which are essential for mitigating seasonal haze that severely impacts air quality in western Indonesia. These strategies are crucial in reducing $PM_{2.5}$ levels and improving public health in highly polluted areas.

For Cluster 1 regions, where pollution levels are moderate but rising, sustainable development planning is essential to prevent further environmental degradation. Encouraging eco-friendly industrial expansion is necessary to ensure economic growth does not lead to higher pollution levels. In addition, green urban planning should be strengthened to integrate sustainable transportation, green spaces, and smart city initiatives, which can help control air pollution in rapidly growing cities. Another critical aspect is promoting alternative energy sources, such as solar and wind power, to reduce reliance on coal-based industries, which are significant contributors to air pollution. These measures will help maintain environmental quality while supporting economic progress in these transitional regions.

In Cluster 3 regions, where air quality remains relatively clean, proactive environmental conservation strategies should be implemented to preserve the existing natural balance. Establishing air quality monitoring systems is vital for detecting early signs of pollution increases, allowing for preventive actions before conditions deteriorate. Additionally, sustainable infrastructure planning must be prioritized to ensure that future economic development does not come at the cost of environmental degradation. A crucial component of this strategy is the enforcement of forest conservation policies, which help maintain natural air purification processes and protect biodiversity. These regions can sustain their clean air status by implementing these measures while promoting responsible development.

For future research, several extensions could build upon our findings to provide a more comprehensive understanding of air pollution in Indonesia. A primary direction would be to conduct source apportionment studies within the high-pollution clusters identified by our analysis. This would help pinpoint the specific origins of $PM_{2.5}$, such as vehicular emissions, industrial activities, or biomass burning, allowing for more targeted and effective mitigation policies [61,62]. Additionally, incorporating spatial analysis could reveal patterns of transboundary pollution and the spatial dependency of $PM_{2.5}$ levels between adjacent regions, offering a more holistic view of air pollution dynamics across the archipelago [63,64]. Finally, linking our pollution clusters to public health data could quantify the specific health burdens, such as respiratory and cardiovascular morbidity, associated with different levels of $PM_{2.5}$ exposure in Indonesia. Such an analysis would strengthen the case for urgent policy intervention by highlighting the direct human cost of air pollution.

This study has several limitations. Firstly, it covers only 33 out of 38 provincial capitals in Indonesia, which may limit the representation of the country's diverse geographic distribution. The reason is because the other five cities are newly established provincial capitals, resulting in the unavailability of data from the sources used in this study. Given that Indonesia is a vast archipelagic nation with significant environmental variations, the findings may not fully capture the conditions in smaller cities or remote areas that were not included in the sample. Additionally, the study relies on data from a single year, which is insufficient for identifying long-term trends or variations in $PM_{2.5}$ pollution, so it does not account for the effects of seasonal variations and long-term climate changes on $PM_{2.5}$ levels, which could influence the overall findings and interpretations.

4. Conclusion

This study presents a comprehensive evaluation and comparison of five fuzzy clustering methods: FCM, FPCM, PFCM, FGK, and FCS integrated with Dynamic Time Warping (DTW) for classifying $PM_{2.5}$ time series data from 33 capital cities in Indonesia. The key findings reveal that the integration of DTW with Fuzzy C-Shells (DTW+FCS) emerged as the optimal method for clustering Indonesian cities based on $PM_{2.5}$ pollution levels. This approach demonstrated superior performance in cluster compactness and separation, as evidenced by the highest PCI and MPCI values and the lowest PEI and XBI values among the tested methods. The clustering analysis identified three distinct groups of cities: high pollution (Cluster 2), moderate pollution (Cluster 1), and low pollution (Cluster 3). The clusters formed confirm that there are statistically significant differences between clusters based on the results of different tests. Major urban and industrial centres, primarily located in Java and parts of Sumatra, were classified in the high pollution cluster, highlighting the significant environmental challenges faced by these regions. Conversely, cities in eastern Indonesia, particularly in Papua, Sulawesi, and parts of Kalimantan, exhibited the lowest $PM_{2.5}$ concentrations, emphasizing the role of lower population density, minimal industrialization, and abundant natural forests in maintaining better air quality. The findings align with the study's objectives by successfully identifying the most effective clustering method for $PM_{2.5}$ data and providing a comprehensive spatial distribution of air pollution across Indonesia. The findings underscore the critical need for targeted interventions in high-pollution areas and proactive measures to preserve air quality in less polluted regions, ultimately supporting the development of more effective strategies for improving public health and environmental sustainability across the Indonesian archipelago.

References

[1] Fuller R, Landrigan PJ, Balakrishnan K, Bathan G, Bose-O'Reilly S, Brauer M, et al. Pollution and Health: a Progress Update. Lancet Planet Heal. 2022;6:535–47.

[2] Khomenko S, Cirach M, Pereira-Barboza E, Mueller N, Barrera-Gómez J, Rojas-Rueda D, et al. Premature Mortality Due to Air Pollution in European Cities: A Health Impact Assessment. Lancet Planet Heal. 2021;5:121–34.

[3] Patil RM, Dinde HT. Status of Ambient Air Pollution in Different States of India during 1990–2015. Curr World Environ. 2023;18:245–64.

[4] Ahmed M, Seraj R, Islam SMS. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. Electronics. 2020;9:1295.

[5] Ikotun AM, Ezugwu AE, Abualigah L, Abuhaija B, Heming J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Inf Sci (Ny). 2023;622:178–210.

[6] Miyamoto S, Ichihashi H, Honda K. Algorithms for Fuzzy Clustering. Berlin: Springer; 2008.

[7] Krasnov D, Davis D, Malott K, Chen Y, Shi X, Wong A. Fuzzy C-Means Clustering: A Review of Applications in Breast

Cancer Detection. Entropy. 2023;25:1–14.

[8] Özdemir O, Kaya A. Comparison of FCM, PCM, FPCM and PFCM Algorithms in Clustering Methods. Afyon Kocatepe Univ J Sci Eng. 2019;19:92–102.

[9] Paradilla YS, Hayati MN, Sifriyani S. Implementation of the fuzzy gustafson-kessel method on grouping districts/cities in Kalimantan island based on poverty issues factors. BAREKENG J Ilmu Mat Dan Terap. 2023;17:0125–34.

[10] Pratiwi NBI, Saputro DRS, Widyaningsih P. Fuzzy C-Shells for Clustering of Data's LQ45 in Indonesia Based on Earnings per Share and Price Earning Ratio. J Phys Conf Ser. 2021;1776:1–9.

[11] Chuang HC, Shie RH, Chio CP, Yuan TH, Lee JH, Chan CC. Cluster Analysis of Fine Particulate Matter (PM2.5) Emissions and Its Bioreactivity in the Vicinity of A Petrochemical Complex. Environ Pollut. 2018;236:591–7.

[12] Suris FNA, Bakar MAA, Ariff NM, Mohd Nadzir MS, Ibrahim K. Malaysia PM10 Air Quality Time Series Clustering Based on Dynamic Time Warping. Atmosphere (Basel). 2022;13.

[13] Ulhaq MZ, Farid M, Aziza ZI, Nuzullah TMF, Syakir F, Sasmita NR. Integration of machine learning and time series analysis for upwelling prediction dashboard in lake Laut Tawar, Indonesia: A study based on climate forecasting. Theor Appl Climatol. 2025;156:1–12.

[14] Rahman EA, Hamzah FM, Latif MT, Dominick D. Assessment of PM2.5 Patterns in Malaysia Using the Clustering Method. Aerosol Air Qual Res. 2022;22:1–15.

[15] Liu Y, Zhang Y-A, Zeng M, Zhao J. A novel distance measure based on dynamic time warping to improve time series classification. Inf Sci (Ny). 2024;656:1–16.

[16] Carlos de Abreu L, Elmusharaf K, Eduardo Gomes Siqueira C. A time-series ecological study protocol to analyze trends of incidence, mortality, lethality of COVID-19 in Brazil. J Hum Growth Dev. 2021;31:491–5.

[17] Rendana M, Razi Idris WM, Abdul Rahim S. Clustering analysis of PM2.5 concentrations in the South Sumatra Province, Indonesia, using the Merra-2 Satellite Application and Hierarchical Cluster Method. AIMS Environ Sci. 2022;9:754–70.

[18] Azharuddin, Sasmita NR, Idroes GM, Andid R, Raihan, Fadlilah T, et al. Patient Satisfaction And Its Socio-Demographic Correlates In Zainoel Abidin Hospital, Indonesia: A Cross-Sectional Study. Unnes J Public Heal. 2023;12:57–67.

[19] Sasaki D, Sofyan H, Sasmita NR, Affan M, Nizamuddin N. Assessing the Intermediate Function of Local Academic Institutions During the Rehabilitation and Reconstruction of Aceh, Indonesia. J Disaster Res. 2021;16:1265–73.

[20] Earlia N, Bulqiah M, Muslem M, Karma T, Suhendra R, Maulana A, et al. Protective Effects of Acehnese Traditionally Fermented Coconut Oil (Pliek U Oil) and its Residue (Pliek U) in Ointment against UV Light Exposure: Studies on Male Wistar Rat Skin (Rattus novergicus). Sains Malaysiana. 2021;50:1285–95.

[21] Saputra A, Sofyan H, Kesuma ZM, Sasmita NR, Wichaidit W, Chongsuvivatwong V. Spatial patterns of tuberculosis in Aceh Province during the COVID-19 pandemic: a geospatial autocorrelation assessment. IOP Conf Ser Earth Environ Sci. 2024;1356:012099.

[22] Klawonn F, Hoffmann G. Using Fuzzy Cluster Analysis to Find Interesting Clusters. In: Build. Bridg. between Soft Stat. Methodol. Data Sci., Bern: Springer; 2023. p. 231–9.

[23] Rahayu L, Sasmita NR, Adila WF, Kesuma ZM, Kruba R. Spatial Estimation of Relative Risk for Dengue Fever in Aceh Province using Conditional Autoregressive Method. J Appl Data Sci. 2023;4:466–79.

[24] Donmez MB, Olcay EO, Demirel M. Load-to-Failure Resistance and Optical Characteristics of Nano-Lithium Disilicate Ceramic after Different Aging Processes. Materials (Basel). 2022;15:4011.

[25] Nguyen D, Kim E, Wang Y, Pham TV, Chen Y-H, Kromrey JD. Empirical Comparison of Tests for One-Factor ANOVA Under Heterogeneity and Non-Normality: A Monte Carlo Study. J Mod Appl Stat Methods. 2020;18:2–30.

[26] Zarauz C, Pradíes GJ, Chebib N, Dönmez MB, Karasan D, Sailer I. Influence of Age, Training, Intraoral Scanner, and Software Version on the Scan Accuracy of Inexperienced Operators. J Prosthodont. 2023;32:135–41.

[27] Nwobi F, Akanno F. Power Comparison of ANOVA and Kruskal–Wallis Tests when Error Assumptions are Violated. Adv Methodol Stat. 2021;18.

[28] Sofyan H, Diba F, Susanti SS, Marthoenis M, Ichsan I, Sasmita NR, et al. The state of diabetes care and obstacles to better care in Aceh, Indonesia: a mixed-methods study. BMC Health Serv Res. 2023;23:1–7.

[29] Wang W, Lyu G, Shi Y, Liang X. Time Series Clustering Based on Dynamic Time Warping. In: Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS, IEEE; 2018. p. 487–90.

[30] Dai J, He Y, Wang Z, Wang M. An Unsupervised Grey Incidence Clustering Method Based on Multi-Dimensional Dynamic Time Warping Distance. In: 2018 IEEE 3rd Int. Conf. Cloud Comput. Big Data Anal., IEEE; 2018. p. 88–92.

[31] Pedrycz W. An Introduction to Computing with Fuzzy Sets. vol. 190. Cham: Springer International Publishing; 2021.

[32] Rahhal T Al, Rençber ÖF. Comparisons of K-Means and Fuzzy C-Means Algorithm by Clustering Countries. Ankara: Iksad Publications House; 2022

[33] Lahjouji M, Alami J el, Hlyal M. Advanced Intelligent Systems for Sustainable Development (AI2SD'2019). vol. 1104. Cham: Springer International Publishing; 2020.

[34] Giordani P, Ferraro MB, Martella F. An Introduction to Clustering with R. vol. 1. Singapore: Springer Singapore; 2020.

[35] Maji KJ, Dikshit AKu, Deshpande AW. Recent Developments and the New Direction in Soft-Computing Foundations and Applications. vol. 393. Cham: Springer International Publishing; 2021.

[36] Scitovski R, Sabo K, Martínez-Álvarez F, Ungar Š. Cluster Analysis and Applications. Cham: Springer International Publishing; 2021.

[37] Dave RN. Fuzzy shell-clustering and applications to circle detection in digital images. Int J Gen Syst. 1990;16:343–55.

[38] Escudero LAG, Alfonso G, Agustín M, Lubiano GMA, Angeles GM, Przemyslaw G, et al. Building Bridges between Soft and Statistical Methodologies for Data Science. vol. 1433. Cham: Springer International Publishing; 2023.

[39] Bowden R. The Information Theory of Comparisons. Singapore: Springer Singapore; 2018.

[40] Bezdek JC. Pattern Recognition with Fuzzy Objective Function Algorithms. Boston, MA: Springer US; 1981.

[41] Heumann M, Kraschewski T, Werth O, Breitner MH. Reassessing taxonomy-based data clustering: Unveiling insights

and guidelines for application. Decis Support Syst. 2024;187:114344.

[42] Xie XL, Beni G. A validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell. 1991;13:841–7.

[43] Lee S, Lee DK. What is the proper way to apply the multiple comparison test? Korean J Anesthesiol. 2018;71:353–60.

[44] Soriano MA, Jiménez-Ormeño E, Haff GG, Comfort P, Giráldez-Costas V, Ruiz-Moreno C, et al. A comparison of three load-velocity based methods to estimate maximum overhead press performance in weightlifters. Int J Sports Sci Coach. 2023;18:1585–93.

[45] Mittal H, Pandey AC, Saraswat M, Kumar S, Pal R, Modwel G. A comprehensive survey of image segmentation: clustering methods, performance parameters, and benchmark datasets. Multimed Tools Appl. 2022;81:35001–26.

[46] Runkler TA. Relation Fuzzy Clustering. In: Oliveira JV de, Pedrycz W, editors. Adv. Fuzzy Clust., West Sussex: John Wiley & Sons, Ltd; 2007. p. 1–457

[47] Auwalin IL, Rumayya, Sukartini NM. Environmental impacts of green open space in urban Indonesia: a difference-in-differences analysis. Asian Dev Rev. 2024;1:1–32.

[48] Jiang B, Li Y, Yang W. Evaluation and Treatment Analysis of Air Quality Including Particulate Pollutants: A Case Study of Shandong Province, China. Int J Environ Res Public Health.

[49] Santoso M, Hopke PK, Damastuti E, Lestiani DD, Kurniawati S, Kusmartini I, et al. The air quality of Palangka Raya, Central Kalimantan, Indonesia: The impacts of forest fires on visibility. J Air Waste Manage Assoc. 2022;72:1191–200.

[50] Santoso M, Lestiani DD, Kurniawati S, Damastuti E, Kusmartini I, Atmodjo DPD, et al. Assessment of Urban Air Quality in Indonesia. Aerosol Air Qual Res. 2020;20:2142–58.

[51] Hanami ZA, Amin M, Hustim M, Putri RM, Torabi SE, Ramadhani AAT, et al. Spatial–Temporal Changes in Air Pollutants in Four Provinces of Sumatra Island, Indonesia: Insights from Sentinel-5P Satellite Imagery. Urban Sci. 2025;9:42.

[52] Kemala N, Gani A, Mahidin. Evaluation of the Effect of Motorized Vehicles on Ambient Air Quality on Various Types of Road Sections in Banda Aceh City. J Penelit Transp Darat. 2020;21:131–44.

[53] Zhan C, Xie M, Lu H, Liu B, Wu Z, Wang T, et al. Impacts of urbanization on air quality and the related health risks in a city with complex terrain. Atmos Chem Phys. 2023;23:771–88.

[54] Qi G, Che J, Wang Z. Differential effects of urbanization on air pollution: Evidences from six air pollutants in mainland China. Ecol Indic. 2023;146:109924.

[55] Borck R, Schrauth P. Population density and urban air quality. Reg Sci Urban Econ. 2021;86:103596.

[56] Zhang L, You S, Zhang M, Zhang S, Yi S, Zhou B. The effects of urbanization on air pollution based on a spatial perspective: Evidence from China. Front Environ Sci. 2022;10:1–16.

[57] Bikis A. Urban Air Pollution and Greenness in Relation to Public Health. J Environ Public Health. 2023;2023:1–18.

[58] Nakyai T, Santasnachok M, Thetkathuek A, Phatrabuddha N. Influence of meteorological factors on air pollution and health risks: A comparative analysis of industrial and urban areas in Chonburi Province, Thailand. Environ Adv. 2025;19:1–9.

[59] Gunawan H, Setyawati T, Atmoko T, Subarudi, Kwatrina RT, Yeny I, et al. A

review of forest fragmentation in Indonesia under the DPSIR framework for biodiversity conservation strategies. Glob Ecol Conserv. 2024;51:1–21.

[60] Rendana M, Razi Idris WM, Rahim SA. Clustering Analysis of PM2.5 Concentrations in The South Sumatra Province, Indonesia, Using the Merra-2 Satellite Application and Hierarchical Cluster Method. AIMS Environ Sci. 2022;9:754–70.

[61] Mohyuddin S, Alam K, Zeb B, Khokhar MF, Mir KA, Wexler AS, et al. Characterization and source identification of PM2.5 during intense haze episodes in an urban environment of Lahore. Atmos Environ X. 2024;23:1–12.

[62] Zhu Q, Liu Y, Hasheminassab S. Long-term source apportionment of PM2.5 across the contiguous United States (2000–2019) using a multilinear engine model. J Hazard Mater. 2024;472:1–27.

[63] Srisaringkarn T, Aruga K. The Spatial Impact of PM2.5 Pollution on Economic Growth from 2012 to 2022: Evidence from Satellite and Provincial-Level Data in Thailand. Urban Sci. 2025;9:1–22.

[64] Jin Z, Velásquez Angel MA, Mura I, Franco JF. Enriched spatial analysis of air pollution: Application to the city of Bogotá, Colombia. Front Environ Sci. 2022;10:1–15.