Published online on the journal's webpage: **http://journals.telkomuniversity.ac.id/index.php/dinda**

Journal of Dinda

Data Science, Information Technology, and Data Analytics

Vol. 6 No. 1 (2026) 9 - 17 E-ISSN: 2809-8064

Clustering of Provinces in Indonesia Based on Land Cover and Environmental Indicators using PAM Clustering

Retno Mayapada[1*]

[1*]Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sulawesi Barat

[1*]retnomayapada@unsulbar.ac.id

Abstract

Land cover and environmental conditions in Indonesia play a strategic role in maintaining ecosystem sustainability, biodiversity, and supporting sustainable development. However, pressures arising from economic development, population growth, land conversion, and ecosystem degradation have resulted in significant environmental disparities across provinces. These variations necessitate the mapping and clustering of regions based on environmental indicators so that the characteristics, levels of pressure, and management needs of each province can be understood in a more systematic and structured manner. This study aims to classify 33 provinces in Indonesia based on land cover and environmental indicators, including the percentage of protected forest, mangrove realization, land cover quality index, conservation land area, forest rehabilitation, and hotspot density as an indicator of environmental pressure. The Partitioning Around Medoids (PAM) method was applied to standardized data due to its ability to produce clusters that are more robust to the presence of outliers, which are commonly found in environmental data. The optimal number of clusters was determined using using two validation approaches, namely internal cluster validation and cluster stability validation. The results indicate that Indonesian provinces can be grouped into six clusters with distinct environmental characteristics, ranging from provinces with relatively good land cover and conservation conditions to those experiencing high environmental pressure. Overall, this clustering provides a more comprehensive representation of the patterns and heterogeneity of environmental conditions across provinces and may serve as a basis for formulating more specific, targeted, and regionally characteristic-based environmental management policies.

Keywords: *cluster analysis; environmental indicators; land cover; Partitioning Around Medoids (PAM)*

1. Introduction

The environment constitutes a fundamental foundation for sustaining life and supporting sustainable development. Indonesia, as one of the countries with the highest levels of biodiversity in the world, bears a significant responsibility in preserving environmental sustainability, not only for national interests but also for global concerns [1]. Nevertheless, pressures arising from economic development, population growth, and land expansion for various purposes present challenges in balancing the utilization of natural resources with conservation efforts [2].

One of the key indicators used to assess environmental conditions and quality is land cover [3][4]. Land cover reflects how the Earth's surface is utilized and managed, including forest areas, agricultural land, settlements, and conservation zones. This indicator not only represents the physical extent of a region but also signifies ecosystem quality, conservation levels, and the pressures experienced by the environment. In Indonesia, forests constitute the most dominant component of land cover, covering approximately 120.5 million hectares or about 63% of the country's total land area [5][6]. Indonesian forests function not only as the lungs of the world but also as habitats for thousands of endemic flora and fauna species, as well as sources of livelihood for millions of people living in and around forest areas.

However, land cover conditions in Indonesia continue to face various pressures that persist over time. Based on data from the Ministry of Forestry [7], deforestation trends in recent years indicate a slight increase compared to previous years, although the figures remain lower than the average deforestation rate over the past decade. The loss of forest cover is still influenced by land conversion, illegal logging activities, and forest fires [8]. In addition, forest degradation and damage to mangrove ecosystems

Received: 12-01-2026 | Revised: 11-02-2026 | Published: 24-02-2026

remain significant issues, particularly in coastal areas that experience pressure from aquaculture activities and infrastructure development [9]. Furthermore, land cover conditions across provinces in Indonesia exhibit considerable variation, both in terms of the extent of protected areas, land cover quality, and the level of threats faced, such as forest fires reflected by hotspot density.

Indonesia consists of provinces with diverse geographical, ecological, and socio-economic characteristics; therefore, environmental management requires a more targeted approach. Clustering provinces based on land cover conditions and conservation efforts is necessary to identify similarities in patterns and characteristics across regions. Consequently, a statistical approach capable of grouping provinces objectively and systematically based on available indicator data is required.

One method that can be used to group provinces with similar characteristics is cluster analysis. Cluster analysis is an unsupervised learning method that aims to group objects into clusters based on the similarity of their characteristics, such that objects within the same cluster exhibit a high degree of similarity, while objects across different clusters show significant differences [10].

Several clustering methods are commonly used in data analysis, including Partitioning Around Medoids (PAM), K-Means, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). While these methods share the same objective of grouping objects based on similarity, they differ in their clustering mechanisms and sensitivity to outliers. K-Means is known to be more sensitive to outliers because it relies on centroids derived from mean values, whereas PAM and DBSCAN are relatively more robust in handling data with extreme characteristics.

A number of previous studies have compared the performance of these methods across different application contexts.[11] and [12] reported that PAM produced better clustering quality than K-Means. In addition, [13] found that PAM outperformed DBSCAN in clustering regions based on stunting prevalence in Indonesia, as indicated by higher silhouette coefficient values. Collectively, these findings suggest that PAM consistently provides more stable and representative clustering results, particularly for datasets characterized by substantial variability or extreme values.

Based on these considerations, this study employs the PAM clustering method to group Indonesian provinces according to land cover and environmental conservation indicators. This approach is expected to reveal patterns in environmental conditions across provinces, identify regions with contrasting environmental characteristics, and provide a foundation for more targeted and sustainable environmental policy recommendations.

The novelty of this study lies not only in the application of PAM with comprehensive internal and stability validation, but also in explicitly translating the resulting cluster structures into policy-oriented typologies, an aspect that is rarely emphasized in provincial-scale environmental clustering studies in Indonesia.

2. Research Methods

This study is a quantitative study using secondary data from 2024 covering 33 provinces in Indonesia. The research data were obtained from Statistics Indonesia [14] and the Ministry of Forestry [5]. This study employs six variables representing land cover and environmental conditions in each province, namely the Land Cover Quality Index, which describes the quality and condition of land cover based on relevant institutional standards; the Percentage of Protected Forest, which indicates the proportion of protected forest area relative to the total regional area; the Mangrove Realization Ratio, which represents the proportion of existing mangroves relative to total mangrove potential, including both existing mangroves and unrealized potential; the Percentage of Conservation Land Area, which indicates the proportion of terrestrial conservation areas relative to the total regional area; the Percentage of Forest Rehabilitation Area, which reflects the extent of forest rehabilitation efforts relative to the regional area; and Forest Hotspot Density, which indicates the intensity of hotspot occurrences in forest areas as a proxy for the level of environmental pressure.

The steps undertaken in the implementation of this study are as follows. In the first stage, data collection was conducted. Secondary data for the year 2024 were collected from Statistics Indonesia and the Ministry of Forestry, covering 33 provinces in Indonesia.

In the next stage, data exploration was conducted to obtain an initial overview of data characteristics and to identify the possible presence of extreme values (outliers).

Subsequently, data standardization was performed to equalize the scale across variables so that each variable contributes proportionally to the clustering analysis.

Following the preprocessing stages, clustering analysis using the PAM algorithm was conducted. Clustering analysis was carried out using the Partitioning Around Medoids (PAM) algorithm to group provinces in Indonesia based on six land cover and environmental indicator variables. The steps of the PAM algorithm refer to Almaza and Yel [15] and Bhat [16], and are described as follows.

To begin with, determination of the number of clusters (k) was conducted. The number of clusters was determined based on the objectives of the analysis and the characteristics of the data.

After determining the number of clusters, selection of initial medoids was performed. A total of k objects were randomly selected from the original dataset as initial medoids representing each cluster.

Next, distance calculation and initial clustering were conducted. The distance between each non-medoid data object and all medoids was calculated using the Euclidean distance, after which each object was assigned to the cluster with the nearest medoid, formulated as follows:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2} \quad (1)$$

Subsequently, selection of candidate medoids and calculation of total distance difference were carried out. The difference in total distance (S) was calculated by comparing the total distance before and after exchanging a medoid m_k with a non-medoid object x_i. The value of S indicates the change in total distance resulting from the medoid exchange.

Following this step, medoid exchange was evaluated. If $S < 0$, the candidate medoid produces a smaller total distance, and the old medoid is replaced. Conversely, if $S > 0$, the medoid is not replaced.

Thereafter, reassignment of data objects was conducted. Distances between each object and the updated medoids were recalculated, and objects were reassigned to the cluster with the nearest medoid.

Furthermore, algorithm iteration was performed. The process of calculating the total distance difference and reassigning clusters was repeated iteratively.

Finally, iteration stopping criterion was applied. The iterative process was terminated when no medoid exchange resulted in a decrease in total distance or when the medoid positions remained unchanged, indicating that optimal clusters had been obtained.

In the final stage, interpretation and conclusion were conducted. The clustering results were interpreted by examining the dominant characteristics of each cluster across the selected environmental indicators. Based on this interpretation, clusters were further translated into policy-oriented typologies reflecting different levels of environmental conditions and management priorities, and conclusions were drawn accordingly.

3. Results and Discussion

3.1. *Exploratory Data Analysis*

The initial step in this study involved conducting data exploration by examining descriptive statistics to provide a general overview of the characteristics of the land cover and environmental indicators used in this research. This descriptive analysis aims to identify the range of values, central tendencies, and levels of variability for each variable, thereby helping to understand the fundamental patterns of the data prior to further analysis. The descriptive statistics presented include the minimum, maximum, mean, and variance of each variable. A summary of the descriptive analysis results for the land cover and environmental indicators is presented in Table 1.

Table 1. Descriptive statistics of land cover and environmental indicators

Variable	Min	Max	Mean	Variance
Protected Forest (%)	0.0068	9.6031	1.6446	2.7923
Mangrove Realization (%)	7.9787	99.9260	72.590	673.8185
Land Cover Quality Index	27.49	100.00	62.353	337.9848
Conservation Land (%)	0.0287	8.2772	1.1767	2.0700
Percentage of Forest Rehabilitation Area (%)	0.0007	0.1436	0.0384	0.0017
Hotspot Density per Forest Area (points/1,000 km²)	0.0000	2.1033	0.1085	0.1321

Based on Table 1, the land cover and environmental indicators exhibit considerable variability across provinces. This is reflected in the substantial differences between the minimum and maximum values for nearly all variables, indicating the presence of provinces with relatively poor environmental conditions, while others demonstrate significantly better indicator performance. Overall, the descriptive statistical results indicate that, on average, land cover and environmental conditions in Indonesia remain relatively low and uneven across provinces. This finding highlights the importance of further analysis to cluster provinces based on similarities in indicator characteristics, thereby enabling the formulation of more targeted and region-based environmental management policies.

In Figure 1, the boxplots present the distribution of six land cover and environmental indicators across 33 provinces in Indonesia. Overall, the plots indicate that most variables exhibit asymmetric distributions and the presence of outliers in several indicators, reflecting disparities in environmental conditions among provinces. The presence of outliers and asymmetric data distributions in several variables constitutes an important consideration in selecting the clustering method used in this study. This study employs the Partitioning Around Medoids (PAM) algorithm, as it is more robust to outliers because cluster centers are determined based on medoids [17]. Therefore, based on the data characteristics illustrated by the boxplots, the PAM algorithm is applied in this study to cluster provinces based on similarities in land cover and environmental indicators in a more stable and representative manner.

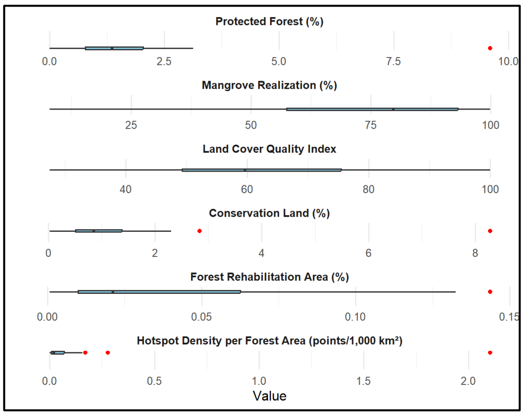

Figure 1. Boxplots of Land Cover and Environmental Indicator Data

3.2. *Determination of the Optimal Number of Clusters*

After conducting data exploration through descriptive statistical analysis and boxplot visualization, the next step in this study is to apply the Partitioning Around Medoids (PAM) algorithm to cluster provinces based on similarities in land cover and environmental indicators. One of the critical stages in implementing a clustering method is determining the optimal number of clusters (k). In this study, the optimal number of clusters is determined using two validation approaches, namely internal cluster validation and cluster stability validation, each assessed using several evaluation indices.

The number of clusters examined in this study ranges from $k = 2$ to $k = 8$. The evaluation indices used for internal cluster validation at each value of $k$include the Silhouette, Davies–Bouldin (DB), Calinski–Harabasz (CH), Dunn, and Connectivity indices [18]. The optimal number of clusters is selected based on clusters that exhibit the highest values of the Silhouette, Calinski–Harabasz, and Dunn indices, as well as the lowest values of the Davies–Bouldin and Connectivity indices. The results of the internal cluster validation for each number of clusters are presented in Table 2.

Table 2 shows that the optimal number of clusters based on the Silhouette and Davies–Bouldin indices is six clusters, whereas the Calinski–Harabasz and Dunn indices indicate that eight clusters represent the optimal solution. In contrast to these four indices, the Connectivity index suggests that two clusters constitute the most optimal number of clusters.

Table 2. Internal cluster validation index values for different numbers of clusters (k)

Cluster	Silhouette	DB	CH	Dunn	Connectivity
2	0.252	1.927	8.344	0.092	**12.624**
3	0.254	1.246	12.667	0.132	15.008
4	0.287	1.155	12.210	0.135	21.137
5	0.307	0.853	18.555	0.160	23.955
6	**0.340**	**0.741**	21.448	0.222	25.950
7	0.198	0.890	19.229	0.201	35.508
8	0.254	0.912	**22.650**	**0.282**	38.629

Nevertheless, although the eight-cluster solution yields the highest values for the Calinski–Harabasz and Dunn indices, the differences are relatively small, and the corresponding values are nearly comparable to those obtained for the six-cluster solution. Furthermore, when comparing the six- and eight-cluster solutions, the Connectivity value for eight clusters is substantially higher, indicating poorer clustering quality. Therefore, it can be concluded that six clusters represent the most optimal number of clusters based on the results of internal cluster validation.

Subsequently, the determination of the optimal number of clusters also considers cluster stability validation. A clustering solution is considered stable if it produces relatively consistent grouping structures when the clustering process is applied to multiple subsamples drawn from the original dataset (Leisch, 2016, as cited in [19]). To assess this stability, cluster instability is quantified by calculating pairwise distances between clustering results obtained from different subsamples. Accordingly, a good clustering solution is characterized by the lowest level of instability. In this study, cluster instability is measured using several indices, namely APN (Average Proportion of Non-overlap), AD (Average Distance), ADM (Average Distance between Means), and FOM (Figure of Merit) [20][21].

Table 3. Internal cluster validation index values for different numbers of clusters (k)

Cluster	APN	AD	ADM	FOM
2	**0.102**	2.586	**0.582**	0.983
3	0.222	2.303	0.689	0.990
4	0.144	2.022	0.665	0.957
5	0.169	1.651	0.598	**0.779**
6	0.253	1.510	0.610	**0.779**
7	0.263	1.432	0.671	0.780
8	0.180	**1.283**	0.594	0.785

Based on the results of cluster stability validation presented in Table 3, each index exhibits a different tendency with respect to the optimal number of clusters. The APN and ADM indices tend to yield the lowest values for relatively small numbers of clusters, particularly at $k = 2$, indicating that simpler cluster structures tend to be more stable. Meanwhile, the AD index shows a decreasing trend as the number of clusters increases, with the lowest value observed at $k = 8$. In contrast to the other indices, the FOM index reaches its minimum values at $k = 5$ and at $k = 6$, suggesting that these two clustering solutions exhibit relatively good stability.

Although not all stability indices point to the same optimal number of clusters, the stability values at $k = 6$ remain within an acceptable range and do not indicate extreme instability. Furthermore, the results of the internal cluster validation show that $k = 6$provides the best balance between cluster compactness and inter-cluster separation. By jointly considering the results of internal validation and cluster stability validation, as well as taking into account the presence of several outliers identified during the data exploration stage, the number of clusters $k = 6$ is determined to be the most optimal solution in this study.

3.3. *Clustering Results Using the PAM Algorithm*

Based on the results of the previous determination of the optimal number of clusters, the best clustering solution consists of six clusters. Subsequently, the Partitioning Around Medoids (PAM) algorithm was applied to cluster 33 provinces in Indonesia based on similarities in land cover and environmental indicators. The purpose of this clustering analysis is to identify patterns of similarity and differences in characteristics across provinces, thereby providing a clearer representation of regional group structures.

The visualization of the PAM clustering results is presented in a two-dimensional space constructed through dimensionality reduction of the data, as shown in Figure 2. Each point in the visualization represents a province, while colors and symbols indicate cluster membership. The numbers displayed on each point correspond to province codes listed in Table 4.

Figure 2 presents the results of PAM clustering applied to 33 provinces in Indonesia with an optimal number of six clusters. The visualization indicates a relatively clear separation among clusters, reflecting differences in land cover and environmental indicator characteristics across provinces. Some clusters consist of a relatively large number of member provinces, while others contain only one or two provinces, indicating the presence of regions with characteristics that are markedly different from the others.

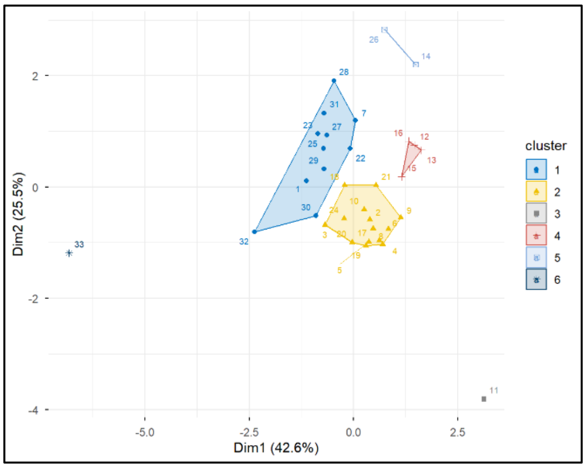

Figure 2. Visualization of Clustering Results Using the PAM Algorithm

Table 4. Codes of Indonesian Provinces

No	Province	No	Province
1	Aceh	18	Nusa Tenggara Barat
2	Sumatera Utara	19	Nusa Tenggara Timur
3	Sumatera Barat	20	Kalimantan Barat
4	Riau	21	Kalimantan Tengah
5	Jambi	22	Kalimantan Selatan
6	Sumatera Selatan	23	Kalimantan Timur
7	Bengkulu	24	Sulawesi Utara
8	Lampung	25	Sulawesi Tengah
9	Kep. Bangka Belitung	26	Sulawesi Selatan
10	Kep. Riau	27	Sulawesi Tenggara
11	DKI Jakarta	28	Gorontalo
12	Jawa Barat	29	Sulawesi Barat
13	Jawa Tengah	30	Maluku
14	DI Yogyakarta	31	Maluku Utara
15	Jawa Timur	32	Papua Barat
16	Banten	33	Papua
17	Bali		

Based on the clustering results, Cluster 1 consists of 11 provinces and Cluster 2 comprises 14 provinces. Meanwhile, Cluster 3 and Cluster 6 each contain only one province, indicating provinces with indicator characteristics that are highly distinct compared to other provinces (these single-member clusters are examined in greater detail in the subsequent discussion). In contrast, Cluster 4 consists of four provinces, and Cluster 5 consists of two provinces. Overall, these clustering results demonstrate that land cover and environmental conditions in Indonesia are heterogeneous, with pronounced differences across groups of provinces. These findings provide a basis for further analysis of the characteristics of each cluster.

3.4. *Cluster Characteristics*

Based on the results of PAM clustering in the previous section, six clusters of provinces in Indonesia were obtained. The next step is to interpret the clustering results by analyzing the characteristics of each cluster. This analysis is conducted based on the mean values of land cover and environmental indicators for each cluster, as well as the composition of provinces included in each cluster. The mean values of each indicator are presented in Table 4, while the list of member provinces for each cluster is provided in Table 5.

Table 4a. Mean values of environmental indicators based on PAM clustering results

Cluster	Number of Provinces	Protected Forest (%)	Mangrove Realization (%)	Land Cover Quality Index
1	11	2.110	71.800	80.100
2	14	1.210	88.000	54.200
3	1	0.007	94.000	27.500
4	4	0.470	29.700	45.700
5	2	1.320	30.900	53.800
6	1	9.600	98.200	100.000

Table 4b. Mean values of environmental indicators based on PAM clustering results

Cluster	Conservation Land (%)	Forest Rehabilitation (%)	Hotspot Density (points/1,000 km²)
1	1.393	0.059	0.010
2	0.897	0.018	0.070
3	0.041	0.002	2.100
4	0.523	0.021	0.060
5	0.269	0.138	0.080
6	8.277	0.003	0.000

Table 5. Cluster membership of provinces

Cluster	Member Provinces
1	Aceh, Bengkulu, Gorontalo, Kalimantan Selatan, Kalimantan Timur, Maluku, Maluku Utara, Papua Barat, Sulawesi Barat, Sulawesi Tengah, Sulawesi Tenggara
2	Bali, Jambi, Kalimantan Barat, Kalimantan Tengah, Kep. Bangka Belitung, Kep. Riau, Lampung, Nusa Tenggara Barat, Nusa Tenggara Timur, Riau, Sulawesi Utara, Sumatera Barat, Sumatera Selatan, Sumatera Utara
3	DKI Jakarta
4	Banten, Jawa Barat, Jawa Tengah, Jawa Timur
5	DI Yogyakarta, Sulawesi Selatan
6	Papua

Based on Table 4 and Table 5, clear differences are observed in land cover and environmental indicator values across clusters. Papua Province, which is the sole member of Cluster 6, exhibits the highest values for nearly all positive indicators, namely Protected Forest, Mangrove Realization, Land Cover Quality Index, and Conservation Land, along with a very low hotspot density. This indicates very favorable environmental conditions. In contrast, DKI Jakarta Province, the only member of Cluster 3, has very low values for the Land Cover Quality Index and Conservation Land, accompanied by the highest hotspot density.

Cluster 1 is characterized by relatively high land cover quality with very low hotspot levels, whereas Cluster 2 stands out for its high mangrove realization but exhibits a moderate level of land cover quality. Meanwhile, Cluster 5 shows the highest forest rehabilitation values, although its conservation and land cover indicators remain at moderate levels. Cluster 4, which consists predominantly of provinces located on Java Island, tends to have lower and relatively uniform indicator values, with no single indicator being dominant.

The clustering results obtained using the PAM algorithm show that two provinces form single-member clusters, namely Papua (Cluster 6) and DKI Jakarta (Cluster 3). This condition requires further validation to determine whether these provinces represent extreme outliers that arise artificially due to methodological limitations, or whether they genuinely reflect a natural cluster structure. To assess the robustness of the clustering results, a comparison is conducted using the K-Means method with the same number of clusters ($k = 6$).

Table 6. Cluster membership of provinces using K-Means algorithm

Cluster	Member Provinces
1	Banten, Jawa Barat, Jawa Tengah, Jawa Timur
2	Aceh, Nusa Tenggara Barat, Kalimantan Selatan, Kalimantan Timur, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat
3	Bengkulu, Gorontalo, DI Yogyakarta, Sulawesi Selatan
4	DKI Jakarta
5	Papua
6	Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, Bali, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Sulawesi Utara

Based on Table 6, despite their methodological differences, particularly in terms of sensitivity to outliers, the K-Means method also assigns Papua and DKI Jakarta to single-member clusters, although with different cluster labels. This consistency indicates that the formation of single-member clusters reflects a natural data structure, in which these two provinces exhibit environmental profiles that are substantially different from those of other provinces.

More specifically, based on the characteristics of the resulting clusters obtained using the PAM algorithm, as presented in Table 4, Papua represents the province with the best environmental conditions in Indonesia.This is indicated by the highest values for nearly all positive environmental indicators, such as the extent of protected forests, the realization of mangrove rehabilitation, the Land Cover Quality Index, and conservation land, along with a very low density of forest fire hotspots. This exceptionally favorable environmental profile cannot be equated with that of any other province in Indonesia, and therefore the formation of a single-member cluster represents an accurate and ecologically meaningful characterization.

Under these circumstances, the single-member cluster represented by Papua can be categorized as a *priority conservation cluster*, referring to regions that should be prioritized for environmental protection rather than exploitation. Policies relevant to this cluster should not focus on rehabilitation or pressure reduction, but instead emphasize preventive strategies against environmental degradation, such as restricting the expansion of extractive industries, strengthening the protection of forest and mangrove areas, and enforcing strict monitoring of land cover changes. Accordingly, the clustering results provide a quantitative basis for designating Papua as a national conservation priority area, where policies are directed toward maintaining its highly favorable environmental conditions and preventing future degradation [22].

Meanwhile, DKI Jakarta represents the province experiencing the highest level of environmental pressure in Indonesia. This condition is reflected by a very low Land Cover Quality Index compared to other provinces, minimal conservation land area, and limited presence of protected forests and forest rehabilitation activities. In addition, the high density of hotspots indicates substantial environmental pressure in the region. These characteristics are closely linked to Jakarta's role as the national capital, which is associated with extremely intensive urbanization, resulting in an environmental profile that differs significantly from other provinces, including nearby urban regions such as West Java and Banten.

Based on these characteristics, DKI Jakarta can be categorized as a *high-pressure cluster,* referring to regions that simultaneously experience multiple environmental pressures due to dense urban activities. These pressures include declining environmental quality, habitat fragmentation, and increasing climate-related risks, which collectively threaten biodiversity and the quality of life of urban populations [23]. Therefore, policies relevant to high-pressure clusters such as Jakarta should focus on managing and reducing environmental pressures, including expanding green open spaces, enforcing stricter land-use controls, and developing more environmentally friendly transportation systems to reduce urban emissions.

Meanwhile, Aceh, Bengkulu, Gorontalo, South Kalimantan, East Kalimantan, Maluku, North Maluku, West Papua, West Sulawesi, Central Sulawesi, and Southeast Sulawesi, which are grouped into Cluster 1, are characterized by relatively high land cover quality and very low hotspot levels. This indicates that regions within this cluster remain in relatively stable environmental conditions and have not yet experienced significant degradation pressure. Therefore, this cluster can be categorized as a *conservation maintenance cluster.*

In regions with low levels of environmental disturbance and relatively stable conditions, policy interventions are more appropriately directed toward maintenance and degradation prevention efforts. Ecosystem management approaches emphasize the importance of maintaining ecosystem integrity and functionality, as well as applying the precautionary principle to reduce potential impacts before environmental damage occurs [24].

In contrast, Bali, Jambi, West Kalimantan, Central Kalimantan, Bangka Belitung Islands, Riau Islands, Lampung, West Nusa Tenggara, East Nusa Tenggara, Riau, North Sulawesi, West Sumatra, South Sumatra, and North Sumatra, which are grouped into Cluster 2, are characterized by high levels of mangrove realization, while overall land cover quality remains at a moderate level. This condition suggests that successful mangrove restoration has not yet been fully translated into improvements in overall landscape quality.

Previous research by [25] indicates that increased mangrove restoration efforts in Indonesia have been strongly driven by ecosystem-based climate change mitigation agendas. However, because mangrove management falls under both forestry and marine sectors, governance fragmentation may limit the contribution of mangrove restoration to broader improvements in land cover quality [25]. Therefore, policy interventions for Cluster 2 can be categorized as a *sectoral restoration cluster*, referring to regions that have demonstrated successful rehabilitation in specific ecosystems but still require cross-sectoral and cross-ecosystem integration to ensure that restoration efforts deliver broader impacts on landscape quality and overall environmental performance.

Furthermore, DI Yogyakarta and South Sulawesi, which belong to Cluster 5, exhibit relatively high levels of forest rehabilitation compared to other clusters, while conservation indicators and overall land cover quality remain at moderate levels. This condition indicates that although rehabilitation interventions have been implemented intensively, their outcomes have not yet been fully reflected in improvements in land cover quality and conservation performance. This finding is consistent with the global meta-analysis by [26], which shows that ecosystem restoration generally enhances biodiversity and ecosystem services, but its outcomes remain lower than those of undisturbed natural ecosystems. Accordingly, the high level of forest rehabilitation observed in Cluster 5 does not automatically translate into optimal conservation and land cover conditions. Therefore, policy attention should be directed toward improving the effectiveness and quality of forest rehabilitation implementation to ensure that ecological benefits can be achieved more optimally. Based on these characteristics, Cluster 5 can be categorized as a *rehabilitation-focused cluster.*

Finally, Cluster 4, which predominantly consists of provinces located on Java Island, is characterized by relatively low and uniform environmental indicator values, with no single indicator appearing dominant. This condition suggests that environmental performance in this cluster is not driven by strength in a particular sector, but rather may be influenced by structural constraints occurring simultaneously across multiple aspects of environmental management. Consequently, partial or sector-specific policy approaches are likely to be less effective [27], indicating the need for more comprehensive and integrated management strategies to gradually improve environmental performance. Therefore, Cluster 4 can be categorized as an *integrated management priority cluster*.

4. Conclusion

This study identifies six optimal clusters of Indonesian provinces based on land cover and environmental indicators using the PAM algorithm, supported by internal and stability validation results. Among these, two clusters consist of a single province, namely Papua and DKI Jakarta. The analysis indicates that these single-member clusters are characterized by environmental profiles that are distinct from those of other provinces.

Each cluster corresponds to a specific policy orientation. Papua is classified as a *priority conservation cluster*, requiring preventive policies to maintain its highly favorable environmental conditions. In contrast, DKI Jakarta represents a *high-pressure cluster*, where policy efforts should focus on reducing environmental stress associated with intensive urbanization. Cluster 1 is categorized as a *conservation maintenance cluster*, emphasizing ecosystem preservation, while Cluster 2 represents a *sectoral restoration cluster*, highlighting the need for cross-sectoral integration. Cluster 5 is identified as a *rehabilitation-focused cluster*, requiring improvements in restoration effectiveness, and Cluster 4 is classified as an integrated *management priority cluster*, indicating the need for comprehensive environmental governance.

Overall, these findings demonstrate substantial environmental heterogeneity across Indonesia and provide a concise, evidence-based framework for cluster-specific environmental policy formulation.

References

[1] Gusmardi, I., Hidayat, F., Subrata, E., Fauzan, Heriyanto and Rivai, F.S.H., 2023. Keberadaan avifauna sebagai dampak pembangunan taman keanekaragaman hayati (KEHATI) PT Tirta Investama Pabrik Solok. *Menara Ilmu: Jurnal Penelitian dan Kajian Ilmiah*, 17(1), pp.61–70.

[2] Ramadhan, A.M., 2023. Dampak pertumbuhan ekonomi terhadap kualitas lingkungan hidup di Provinsi Pulau Sumatera. *Determinasi: Jurnal Penelitian Ekonomi Manajemen dan Akuntansi*, 1(2), pp.1–12. Available at: https://doi.org/10.23917/determinasi.v1i2.40.

[3] Andresi, B., Septyana, D., Mardin, R., Silvana and Firman, T., 2023. Pemantauan indeks kualitas tutupan lahan di Kota Palu tahun 2023. *Jurnal Peweka Tadulako*, 2(2), pp.236–243. Available at: https://doi.org/10.22487/peweka.v2i2.23.

[4] Masyruroh, A. and Binyati, 2021. Kajian indeks kualitas lingkungan hidup Kota Serang. *Jurnal Lingkungan dan Sumberdaya Alam (JURNALIS)*, 4(2), pp.151–162. Available at: https://doi.org/10.47080/jls.v4i2.1463.

[5] Kementerian Kehutanan Republik Indonesia, 2025a. *Statistik Kementerian Kehutanan 2024: Volume I.* Jakarta Pusat: Pusat Data dan Informasi Kementerian Kehutanan.

[6] Hasanah, F.T., 2020. Karakteristik wilayah daratan dan perairan di Indonesia. *Jurnal Geografi: Geografi dan Pengajarannya*, 20(1), pp.1–6.

[7] Kementerian Kehutanan Republik Indonesia, 2025b. Hutan dan deforestasi Indonesia tahun 2024. *Siaran pers*, 20 March. Available at: https://www.kehutanan.go.id/news/article-10.

[8] Yamani, S.A.Z., Amar, A. and Ria, D., 2025. Analisis laju deforestasi penutupan lahan hutan Kabupaten Bireuen tahun 2020–2024 menggunakan sistem informasi geografis. *Jurnal Agrium*, 22, pp.350–357.

[9] Hidayat, A. and Rachmawatie, D., 2021. Deforestasi ekosistem mangrove di Pulau Tanakeke, Sulawesi Selatan, Indonesia. *Jurnal Ilmu dan Teknologi Kelautan Tropis*, 13(3), pp.439–454. Available at: https://doi.org/10.29244/jitkt.v13i3.38502.

[10] Monalisa, S. and Kurnia, F., 2019. Analysis of DBSCAN and K-means algorithm for evaluating outlier on RFM model of customer behaviour. *TELKOMNIKA (Telecommunication Computing*

Electronics and Control), 17(1), pp.110–118. Available at: https://doi.org/10.12928/telkomnika.v17i1.9394.

[11] Wahyudi, M., Pujiastuti, L. and Solikhun, 2022. Komparasi K-means clustering dan K-medoids dalam mengelompokkan produksi susu segar di Indonesia. *Jurnal Bumigora Information Technology (BITe)*, 4(2), pp.243–254. Available at: https://doi.org/10.30812/bite.v4i2.2104.

[12] Marlina, D., Lina, N., Fernando, A. and Ramadhan, A., 2018. Implementasi algoritma K-medoids dan K-means untuk pengelompokkan wilayah sebaran cacat pada anak. *Jurnal CoreIT*, 4(2), pp.64–71.

[13] Sulano, T., 2024. *Perbandingan algoritma Partitioning Around Medoid (PAM) dan Density-Based Spatial Clustering of Applications with Noise (DBSCAN) dalam klasterisasi kasus balita stunting di Indonesia*. Skripsi. Palembang: Universitas Sriwijaya.

[14] Almazar, N. and Yel, M.B., 2024. Application of the K-medoids algorithm in clustering PAM customers based on provinces in Indonesia. *International Journal of Software Engineering and Computer Science (IJSECS)*, 4(1), pp.210–224.

[15] Badan Pusat Statistik, 2025. *Statistik Lingkungan Hidup Indonesia 2025 (Environment Statistics of Indonesia 2025)*. Vol. 44. Jakarta: Badan Pusat Statistik.

[16] Bhat, A., 2014. K-medoids clustering using partitioning around medoids for performing face recognition. *International Journal of Soft Computing, Mathematics and Control (IJSCMC)*, 3(3), pp.1–12.

[17] Nahdliyah, M.A., Widiharih, T. and Prahutama, A., 2019. Metode K-medoids clustering dengan validasi silhouette index dan C-index (studi kasus jumlah kriminalitas kabupaten/kota di Jawa Tengah tahun 2018). *Jurnal Gaussian*, 8(2), pp.161–170.

[18] Rendon, N., Giraldo, J.H., Bouwmans, T., Rodríguez-Buritica, S., Ramirez, E. and Isaza, C., 2023. Uncertainty clustering internal validity assessment using Fréchet distance for unsupervised learning. *Engineering Applications of Artificial Intelligence*, 124, 106635. Available at: https://doi.org/10.1016/j.engappai.2023.106635.

[19] Gere, A. (2023). Recommendations for validating hierarchical clustering in consumer sensory projects. Current Research in Food Science, 6, 100522. Available at: https://doi.org/10.1016/j.crfs.2023.100522

[20] Brock, G., Pihur, V., Datta, S. and Datta, S., 2008. clValid: An R package for cluster validation. Journal of Statistical Software, 25(4). Available at: https://doi.org/10.18637/jss.v025.i04.

[21] Anand, S.K. and Kumar, S., 2022. Experimental comparisons of clustering approaches for data representation. ACM Computing Surveys (CSUR), 55(3), pp.1–33. Available at: https://doi.org/10.1145/3490384.

[22] Chen, Y., Yang, J., Zhang, W., Guan, X., Pan, L., Liu, M. and Xiao, N., 2025. Identifying conservation priority areas through the integration of biodiversity, ecosystem services and landscape patterns in the Wujiang River Basin. Land, 14(12), 2335. Available at: https://doi.org/10.3390/land14122335 Available at: https://doi.org/10.1145/3490384.

[23] Mori, E., Di Lorenzo, T., Viviano, A., Jakovljević, T., Marra, E., Baesso Moura, B., Garosi, C., Manzini, J., Ancillotto, L., Hoshika, Y. and Paoletti, E., 2025. Under pressure: Environmental stressors in urban ecosystems and their ecological and social consequences on biodiversity and human well-being. *Stresses*, 5(4), 66. Available at: https://doi.org/10.3390/stresses5040066.

[24] Leech, S., Wiensczyk, A. and Turner, J., 2009. Ecosystem management: A practitioners' guide. BC Journal of Ecosystems and Management, 10(2), pp.1–12. Available at: https://www.forrex.org/publications/jem/ISS51/vol10_no2_art1.pdf.

[25] Sidik, F., Lawrence, A., Wagey, T., Zamzani, F. and Lovelock, C.E., 2023. Blue carbon: A new paradigm of mangrove conservation and management in Indonesia. *Marine Policy*, 147, 105388. Available at: https://doi.org/10.1016/j.marpol.2022.105388.

[26] Rey Benayas, J.M., Newton, A.C., Díaz, A. and Bullock, J.M., 2009. Enhancement of biodiversity and ecosystem services by ecological restoration: A meta-analysis. Science, 325(5944), pp.1121–1124. Available at: https://doi.org/10.1126/science.1172460.

[27] Nilsson, M., Zamparutti, T., Petersen, J.E., Nykvist, B., Rudberg, P. and McGuinn, J., 2012. Understanding policy coherence: Analytical framework and examples of sector–environment policy interactions in the EU. Environmental Policy and Governance, 22, pp.395–423. Available at: https://doi.org/10.1002/eet.1589