**PARAMETER**
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA

e-ISSN. 2829-6303
Vol 04, No 3, December 2025
Page 381-394

# Time Series Clustering of Rice Productivity Using Trimming Gaussian Mixture Models

**Sarah Fadhlia[1]*, Eko Primadi Hendri[2]**

[1]Study Program of Guidance and Counseling, Faculty of Education and Social Sciences, Universitas Indraprasta PGRI
Nangka Raya St, No. 58 C, South Jakarta, 12530, DKI. Jakarta, Indonesia
[2]Study Program of Road Transport Management, Politeknik Transportasi Darat Indonesia- STTD
Setu Street, No. 89, Bekasi, 17320, West Java, Indonesia

**E-mail Correspondence Author:** *sarah.m.amin@gmail.com*

***Abstract***
*This study investigates the application of the Trimming Gaussian Mixture Model (TGMM) for clustering monthly rice productivity time series data in West Java from 2018 to 2023. TGMM is a robust clustering approach that reduces the influence of outliers by trimming a specified portion of the data prior to parameter estimation. The dataset, sourced from Open Data Jabar, was analyzed to identify the most representative number of clusters using the Silhouette Score. The optimal clustering solution was achieved with two main clusters (k = 2) and a trimming proportion of 15%. The results revealed three distinct regional groups: two dominant clusters characterized by moderate-stable and high-consistent productivity patterns, and a separate group of outliers marked by low and highly fluctuating productivity. Cluster stability was assessed using the Adjusted Rand Index (ARI), yielding values of 0.41 (bootstrap) and 0.545 (subsampling), which indicate a reasonably consistent clustering structure. These findings demonstrate the effectiveness of TGMM in capturing underlying productivity patterns while accounting for noise and outliers, suggesting its potential as a robust decision-support tool for data-driven agricultural planning and policy formulation.*

***Keywords:*** *Adjusted Rand Index, rice productivity, Silhouette Score, time series clustering, trimming gaussian mixture model.*

: https://doi.org/10.30598/parameterv4i1pp381-394





*Submitted: Juni 2025*
*Accepted: December 2025*

## 1. INTRODUCTION

In the era of modern agriculture, optimal data utilization is key to enhancing food security and production efficiency. Statistics play a vital role in modeling and analyzing complex phenomena across various fields, including the agricultural sector. One increasingly adopted approach is time series clustering, which enables the grouping of objects based on similarities in temporal patterns. This technique is relevant for supporting data-driven decision-making, including understanding the dynamics of agricultural productivity [1], [2].

West Java Province is one of Indonesia's main rice-producing regions, experiencing significant monthly fluctuations in productivity. These variations are influenced by multiple factors such as weather conditions, cultivation technologies, and dynamic government policies. Such uncertainty directly impacts farmers' income and regional food security. Therefore, comprehensively understanding rice productivity patterns is crucial. Clustering regions based on time series patterns of rice productivity has the potential to assist governments and policymakers in formulating more effective, targeted, and responsive intervention strategies [3].

Previous research has widely explored time series clustering techniques, for instance, Dynamic Time Warping (DTW) and k-medoids, to analyze agricultural data. However, these approaches generally lack robustness when handling data with outliers or highly complex patterns, which may lead to biased clustering results [4], [5]. The Gaussian Mixture Model (GMM) is a widely adopted probabilistic clustering technique that represents the underlying structure of data as a mixture of multiple Gaussian distributions. This model offers high flexibility in capturing the latent structure of data [6], and its parameters are typically estimated using the Expectation-Maximization (EM) algorithm [2]. Nevertheless, GMM has a major drawback—its sensitivity to outliers, which can reduce the reliability of the clustering results [7], [8].

As a solution, the Trimming Gaussian Mixture Model (TGMM) was developed by integrating a trimming technique, which removes a portion of the data deemed as outliers before estimation. This technique maximizes the likelihood only on a subset of representative data, producing more stable and accurate estimates [8]. The concept of trimming was formally introduced in the context of clustering by Cuesta-Albertos et al. [9] through the trimmed k-means method, which aims to enhance the robustness of the algorithm against outliers by excluding a small proportion of the most extreme data points. Compared to other robust methods such as M-estimators or mixtures of t-distributions, TGMM has the advantage of explicitly identifying outliers without modifying the underlying distributional assumptions. This makes it more interpretable in practical contexts, including time series data on agricultural productivity [7].

In addition to previous studies on clustering agricultural time series data [10], [11], [12]. Recent research has emphasized the importance of robust techniques that can withstand noisy or nonstationary data [13], [14]. TGMM represents a compelling alternative due to its balance of flexibility and robustness [7], [8], [15]. Moreover, studies have demonstrated the usefulness of TGMM in other domains where time series behavior is irregular or subject to disruption [8]. These developments suggest a promising opportunity to bring these advances into the agricultural productivity domain.

These developments suggest a promising opportunity to bring these advances into the agricultural productivity domain. To date, however, the specific application of TGMM to rice productivity time series in Indonesia, particularly in West Java, has never been conducted. This represents a significant research gap, as agricultural data in tropical

regions are inherently prone to extreme shocks—such as climate anomalies or pest outbreaks—that act as natural outliers

To date, the specific application of TGMM to rice productivity time series in Indonesia, particularly in West Java, has never been conducted. This represents a significant research gap, as agricultural data in tropical regions are inherently prone to extreme shocks—such as climate anomalies or pest outbreaks—that act as natural outliers [16], [17], [18]. Most existing studies in this region still rely on conventional approaches [5], [19] that are less resilient to such disturbances, often resulting in biased clustering results [20], [21]. The novelty of this study lies in being the first to explicitly address this limitation by deploying TGMM as a methodological solution in this context. As the pioneering implementation of this approach in the region, this research demonstrates how the trimming mechanism effectively isolates agricultural-specific outliers, thereby recovering the true underlying productivity structure that conventional methods fail to capture [8], [9].

Therefore, this study aims to apply and evaluate the performance of TGMM in clustering monthly rice productivity time series in West Java over the 2018–2023 period. In addition to comparing it with conventional methods, this study seeks to assess TGMM's ability to detect more accurate and outlier-resilient clusters. Thus, the resulting clusters are expected to provide more representative and useful insights for planning and decision-making in the agricultural sector.

Several recent publications have also explored high-resolution time series clustering for agricultural policy support using advanced models, such as deep learning or entropy-based clustering [22], [23]. Although powerful, such methods may lack interpretability compared to probabilistic models like TGMM [6], [8]. The current study builds on this foundation by combining statistical robustness with domain relevance, making it well-suited for application in regional agricultural analysis where interpretability is key. Furthermore, statistical techniques such as principal component analysis (PCA) and silhouette coefficients have been used to validate cluster structure in high-dimensional time series settings [24], [25], [26]. Incorporating these validation approaches strengthens the analytical reliability of the proposed TGMM-based clustering.

## 2. RESEARCH METHODS

### 2.1. Data and Sources

This study uses rice productivity data based on the Area Sampling Frame obtained from the official Open Data Jabar portal, managed by the West Java Provincial Government. The dataset was compiled by the Department of Food Crops and Horticulture and includes monthly rice productivity information from 2018 to 2023 for each regency and city in West Java. The dataset is openly available and can be accessed through the Open Data Jabar portal [27].

### 2.2. Research Procedure

All data processing, statistical modeling, and clustering procedures in this study were carried out using the R programming language, which provides a comprehensive platform for advanced statistical computing, time series analysis, and model-based clustering. R offers a wide range of specialized packages such as mclust, cluster, and tseries, which were used to implement the Gaussian Mixture Model (GMM) and the Trimming Gaussian Mixture Model (TGMM), as well as to perform data preprocessing and cluster evaluation [28]. The flexibility of R in handling various data types and applying robust algorithms makes it particularly suitable for time series

clustering in complex agricultural datasets. In addition, the visualization process was supported by packages such as ggplot2 and factoextra, which facilitated the interpretation and presentation of clustering results in an intuitive manner [29].
The analysis procedures were carried out in the following steps:

1. Data Exploration
   The analysis begins with an initial exploration of the rice productivity data to identify general characteristics, detect outliers, and understand the data distribution for each region using boxplots. Boxplots are standard visualization tools used to depict five-number summary statistics and outliers [30], [31].
2. Determining the Optimal Number of Clusters
   The optimal number of clusters is determined using the Silhouette Score, a measure that evaluates how well an object fits within its assigned cluster compared to its distance from other clusters. Higher values reflect stronger internal cohesion and greater separation between clusters [26]. The Silhouette Score is defined by the following formula:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{1}$$

   with:
   - $\boldsymbol{a(i)}$ represents the mean distance between point *i* and all other points within the same cluster.
   - $\boldsymbol{b(i)}$ denotes the smallest average distance from point $i$ to all points in the nearest neighboring cluster.

   A value of $\boldsymbol{s(i)}$ close to 1 indicates a well-clustered point.
3. Determining the Optimal Trimming Proportion
   The selection of the data proportion to be trimmed is based on the average Silhouette Score computed for the optimal number of clusters. This approach ensures that the chosen trimming proportion enhances clustering quality by maximizing both cluster cohesion and separation.
4. Construction of the Trimming Gaussian Mixture Model
   The Trimming Gaussian Mixture Model (TGMM) constitutes a robust refinement of the traditional Gaussian Mixture Model (GMM), designed to mitigate the influence of outliers by excluding a subset of data points during parameter estimation via the Expectation-Maximization (EM) algorithm. This enhancement results in more stable and representative model estimations [7], [8]. Formally, the GMM posits that the observed data $\boldsymbol{x}$ are generated from a finite mixture of $\boldsymbol{K}$ multivariate Gaussian distributions, expressed as.

$$p(x_k|\Theta) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(x|\mu_k, \Sigma_k) \tag{2}$$

   with:

   - $\pi_k$ indicates the mixing weight (prior probability) associated with the $k$-th component, subject to the constraints $\sum_{k=1}^{K} \pi_k = 1$ dan $\pi_k \geq 0$,
   - $\mathcal{N}(x|\mu_k, \Sigma_k)$ describes the multivariate Gaussian distribution characterized by the mean vector $\mu_i$ and and the covariance matrix $\Sigma_i$:

The objective of TGMM is to maximize the trimmed log-likelihood function on a subset of data that excludes the $\alpha$proportion of observations with the lowest likelihoods. The objective function is defined as:

$$\mathcal{L}_{\text{trim}}(\Theta) = \sum_{i \in H_\alpha} \log\left(\sum_{k=1}^{K} \pi_k \, \mathcal{N}(x_i \mid \mu_k, \Sigma_k)\right) \tag{3}$$

with $H_\alpha$ is the subset of indices containing the $\lceil n(1-\alpha) \rceil$ observations with the highest likelihood values. The model parameters $\Theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^{K}$ are obtained by maximizing the likelihood function, a process carried out using the Expectation-Maximization (EM) algorithm as described below.

- E-Step (Expectation): Calculate the posterior probability (soft assignment) that data point $x_i$ belongs to cluster $k$:

$$\gamma_{ik}^{(t)} = \frac{\pi_k^{(t-1)} \mathcal{N}\left(x_i | \mu_k^{(t-1)}, \Sigma_k^{(t-1)}\right)}{\sum_{j=1}^{K} \pi_j^{(t-1)} \mathcal{N}\left(x_i | \mu_j^{(t-1)}, \Sigma_j^{(t-1)}\right)} \tag{4}$$

  Compute the individual likelihood of each data point:

$$L_k^{(t)} = \sum_{k=1}^{K} \pi_k^{(t-1)} \, \mathcal{N}(x | \mu_k^{(t-1)}, \Sigma_k^{(t-1)}) \tag{5}$$

- Trimming step:
  Order the data instances from lowest to highest according to their corresponding likelihood values $L_i^{(t)}$. Select a subset $X_{subset}$ consisting of the top $n(1-\alpha)$ data points with the highest likelihoods, where $\alpha$ is the trimming proportion.
- M-Step (Maximization) : Update the parameters using only the trimmed subset $X_{subset}^{(t)}$ :

$$\pi_k^{(t)} = \frac{1}{\left|X_{subset}^{(t)}\right|} \sum_{i \in X_{subset}^{(t)}} \gamma_{ik}^{(t)} \tag{6}$$

$$\mu_k^{(t)} = \frac{\sum_{i \in X_{subset}^{(t)}} \gamma_{ik}^{(t)} \, x_i}{\sum_{i \in X_{subset}^{(t)}} \gamma_{ik}^{(t)}} \tag{7}$$

$$\Sigma_k^{(t+1)} = \frac{\sum_{i \in X_{subset}^{(t)}} \gamma_{ik}^{(t)} \left(x_i - \mu_k^{(t)}\right)\left(x_i - \mu_k^{(t+1)}\right)^T}{\sum_{i \in X_{subset}^{(t)}} \gamma_{ik}^{(t)}} \tag{8}$$

Repeat the E-step, trimming, and M-step iteratively until convergence, indicated by changes in the parameters $\Theta$ or the trimmed log-likelihood falling below a predefined threshold.

5. Cluster Visualization
   After completing the clustering process using the Trimming Gaussian Mixture Model (TGMM), scatter plots are employed to visualize the cluster distribution in a low-dimensional space. Since the rice productivity time series data are high-dimensional, dimensionality reduction techniques such as Principal Component Analysis (PCA) **[25]** are first applied. The resulting reduced data are then visualized in a two-dimensional scatter plot, where each point represents a single

regency or city unit, and its color indicates the cluster assignment determined by TGMM.

6. Exploration of Each Cluster
   Descriptive statistics and boxplots are utilized to compare characteristics across clusters, such as average productivity or variability among regions, thereby enriching the interpretation of clustering results.
7. Cluster Stability Evaluation
   To ensure that the clustering results are stable and not overly sensitive to data variation, cluster stability is assessed using the Adjusted Rand Index (ARI). ARI quantifies the similarity between two clustering outcomes—specifically, the original clustering and a replicated clustering derived from modified versions of the dataset. ARI serves as a measure of agreement between two partitions, correcting for chance. Given a set of $\boldsymbol{n}$ elements and two partitions $\boldsymbol{U} = \{\boldsymbol{u_1}, \dots, \boldsymbol{u_R}\}$ and $\boldsymbol{V} = \{\boldsymbol{v_1}, \dots, \boldsymbol{v_C}\}$, the ARI is computed as follows:

$$\text{ARI} = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]/\binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]/\binom{n}{2}} \tag{9}$$

   where $\boldsymbol{n_{ij}}$represents the number of objects common to clusters $\boldsymbol{u_i}$and $\boldsymbol{v_j}$, while $\boldsymbol{a_i}$and $\boldsymbol{b_j}$denote the number of objects in clusters $\boldsymbol{u_i}$and $\boldsymbol{v_j}$, respectively. An ARI value close to 1 implies high stability, whereas a value near 0 indicates random labeling. Two approaches are employed: bootstrap sampling, which involves generating multiple resampled datasets with replacement from the original data [7], and subsampling, where a portion of the data is selected without replacement [24]. This process provides insight into how consistent the resulting clusters are under data variation. A high ARI value indicates that the clustering is stable and reliable, and not significantly affected by data fluctuations.

## 3. RESULT AND DISCUSSION

### 3.1. Data Exploration

Based on Figure 1, the boxplot of rice productivity by regency/city in West Java Province from January 2018 to December 2023 provides a comprehensive overview of the distribution and variability patterns across regions. Each boxplot represents the monthly productivity distribution for a given regency/city, with the central line indicating the median. The box illustrates the interquartile range (IQR), while the points outside the whiskers indicate the presence of outliers. It is evident that Depok City exhibits the widest spread, with extreme value ranges and a relatively high number of outliers, indicating significant fluctuations in rice productivity. Bekasi City and Bogor City also display similar characteristics, though to a lesser extent than Depok. In contrast, areas such as Bandung City, Karawang Regency, Kuningan Regency, and Indramayu Regency show narrower spreads with higher medians, indicating more stable and consistently strong productivity performance over time.

These differences in distribution patterns reflect the heterogeneity of regional characteristics in terms of cultivation techniques, agroclimatic conditions, and socio-economic factors influencing agricultural output. Regions with numerous outliers and wide spreads are likely to face extreme seasonal disruptions, climate dependency, or even

structural issues within their production systems. Meanwhile, regions with more stable distributions tend to have well-established agricultural systems that are more resilient to seasonal variability.

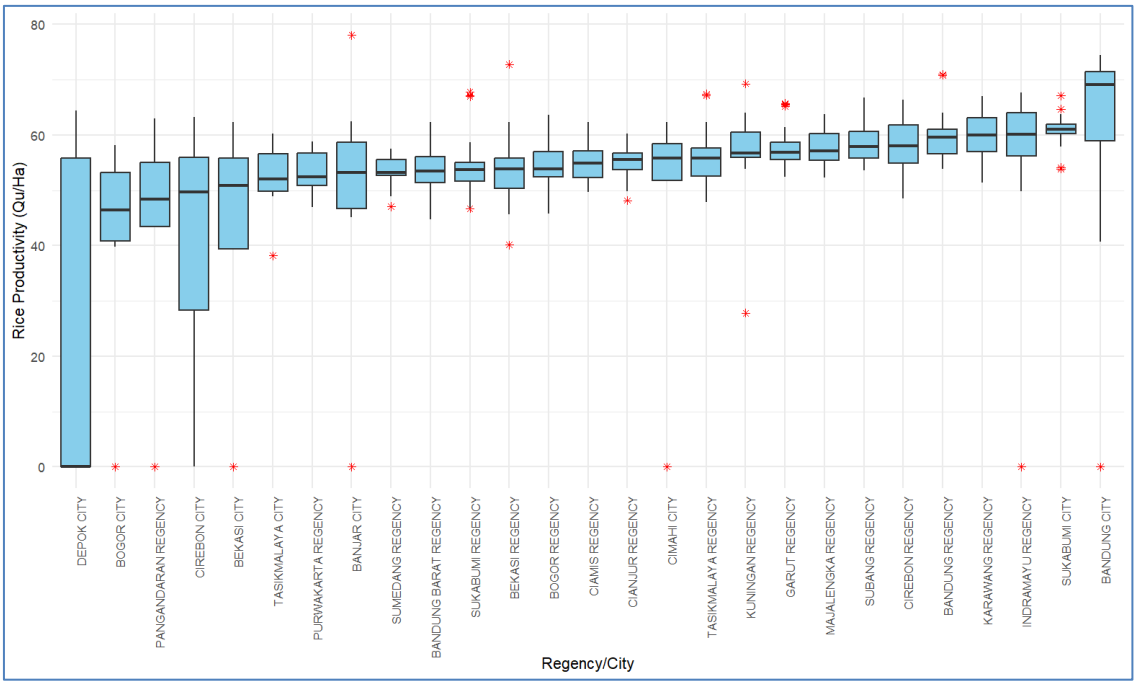


**Figure 1. Boxplot of Rice Productivity by Administrative Region**

### 3.2. Trimming Gaussian Mixture Model

Selecting the optimal cluster count is a fundamental aspect of the clustering procedure, especially when applying the Trimming Gaussian Mixture Model (TGMM). **Figure 2** displays the average Silhouette Scores for different numbers of clusters (k), evaluated over a range from 2 to 10. The Silhouette Score serves as an indicator of clustering quality, where higher values reflect clearer separation between clusters and stronger internal cohesion. Based on **Figure 2**, the highest Silhouette Score is achieved at two clusters ($\boldsymbol{k = 2}$), approximately 0.42. The score then declines significantly from $\boldsymbol{k = 3}$ to $\boldsymbol{k = 10}$, indicating that increasing the number of clusters does not necessarily improve data segmentation quality.

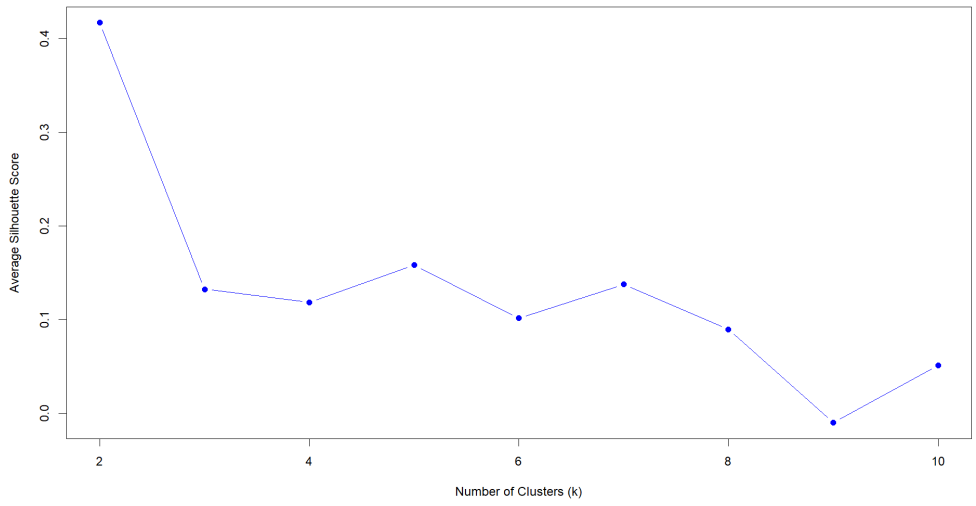


**Figure 2. Optimal Clustering Based on Silhouette Score**

These findings suggest that the natural structure of the monthly rice productivity data in West Java tends to form two statistically distinct groups. In the context of TGMM, selecting $\boldsymbol{k = 2}$ as the optimal number of clusters implies that the trimming process effectively excludes outliers and yields two dominant clusters that best represent the general productivity patterns.

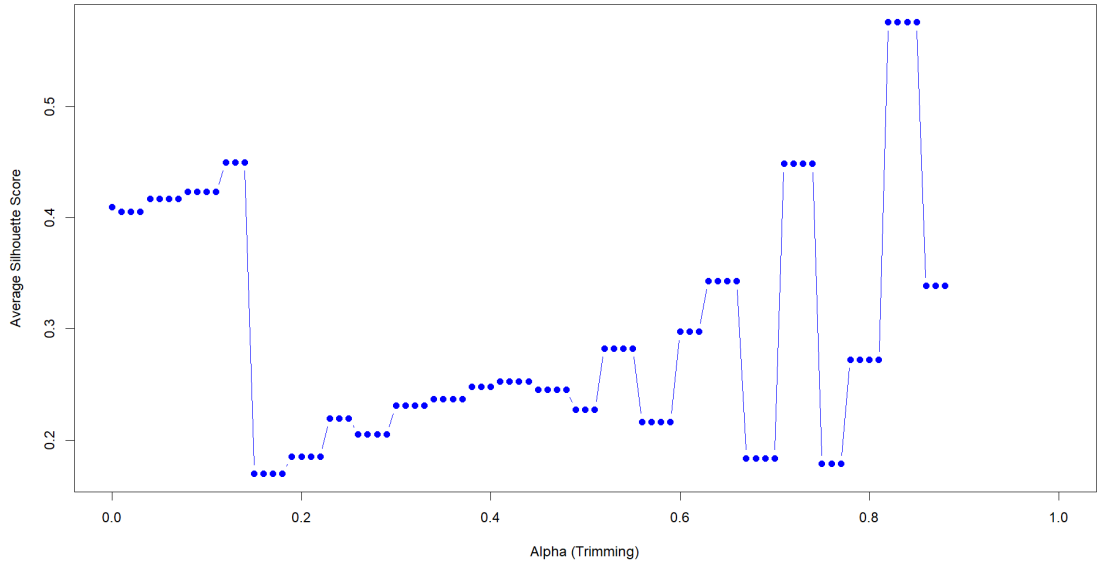


**Figure 3. Silhouette Score as a Function of Trimming Proportion in 2-Cluster Analysis**

In the parameter optimization stage for the Gaussian Mixture Model (GMM), a trimming analysis was conducted to identify the most appropriate trimming proportion, particularly when the number of clusters ($\boldsymbol{k}$) was fixed at 2. **Figure 3** presents a plot of the average Silhouette Score against varying values of the trimming parameter $\boldsymbol{\alpha}$ (Alpha). Higher Silhouette Scores indicate better clustering quality. From the graph, it can be observed that the Silhouette Score fluctuates significantly as $\alpha$ increases. The highest peak in the Silhouette Score occurs around $\boldsymbol{\alpha = 0.15}$, reaching an average value above 0.45. Although there are other peaks at higher $\alpha$ values, $\boldsymbol{\alpha = 0.15}$ was selected as the optimal trimming proportion. This choice is based on the consideration that excessively high trimming values may remove a large portion of relevant data, whereas 0.15 successfully maintains good clustering quality (indicated by the high Silhouette Score) without sacrificing too many observations. Thus, trimming at 0.15 is considered most effective in enhancing the robustness of the GMM against outliers while preserving a clear cluster structure for $\boldsymbol{k = 2}$.

After determining the optimal trimming parameter ($\boldsymbol{\alpha = 0.15}$), TGMM was applied to cluster the rice productivity data. **Figure 4** shows the clustering results with two clusters $k$ = 2 in the PC1 and PC2 space obtained through PCA. PCA reduces the data's dimensionality while preserving the largest variance. PC1 and PC2 capture the first and second highest variances, respectively, and are orthogonal. Each region's position in the plot reflects its rice productivity characteristics, making cluster patterns easier to interpret. The objective of this clustering is to categorize regions based on their rice productivity characteristics, while simultaneously excluding outlier observations identified through the trimming process. As illustrated in **Figure 4**, the data are partitioned into two primary clusters—depicted in green and blue—whereas several observations, identified as outliers and shown in red, are excluded from the main clusters due to their deviation from the overall data distribution.

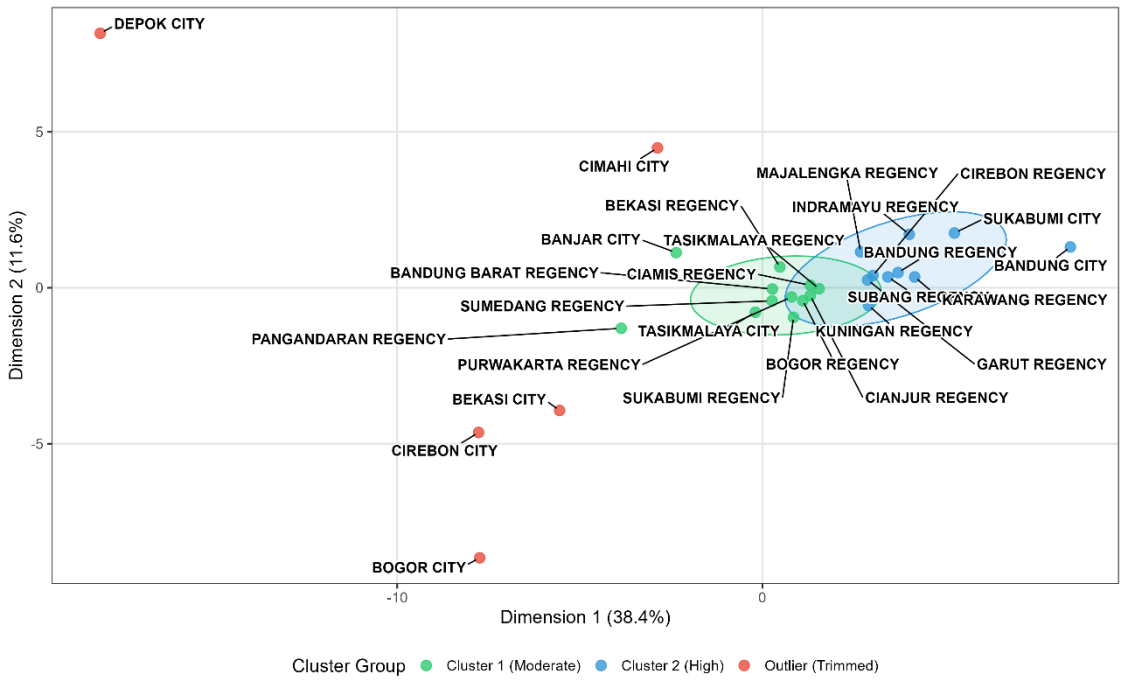


**Figure 4. Trimming GMM Clustering of Rice Productivity**

Cluster 0 (represented in red) comprises observations that were trimmed or identified as outliers, including regions such as Depok City, Cimahi City, Bekasi City, and Bogor City. Their considerable distance from the centers of the other clusters indicates that these regions exhibit markedly different or atypical rice productivity characteristics compared to the majority of other areas. This likely reflects specific conditions such as high levels of urbanization, extremely limited agricultural land, or an economic focus not centered on rice cultivation, resulting in rice productivity levels that fall outside the general pattern. These observations are effectively isolated through the trimming process to prevent bias in the formation of the main clusters.

Cluster 2 (depicted in blue), which comprises regions such as Kuningan Regency, Majalengka Regency, Indramayu Regency, Subang Regency, and Bandung Regency, exhibits a concentration that is somewhat distinct from Cluster 1. The regions in this cluster are likely characterized by high or optimal rice productivity. These areas may serve as key rice-producing centers, benefiting from favorable conditions such as fertile soil, suitable climate, and effective agricultural practices. Their relatively close proximity within the cluster indicates similar productivity patterns, suggesting a high degree of efficiency and strong production potential.

Based on **Figure 5**, clear differences in characteristics are observed among the clusters. Cluster 0 (red), previously identified as trimmed observations or outliers, exhibits a very wide and generally low range of rice productivity. Although the median productivity is 50.0 quintals per hectare, the mean is only 37.6 quintals per hectare, indicating the presence of several extremely low values that pull the average downward. The notably high standard deviation (24.3) further confirms the extreme variability within this cluster, with many data points falling below the first quartile and several extreme values even below zero. This supports the interpretation that Cluster 0 comprises regions with anomalous or inconsistent rice productivity, often substantially lower than those in the other clusters, which explains why these observations were trimmed by the GMM model.

Cluster 1 (green) exhibits a more consistent distribution of rice productivity compared to Cluster 0. With a mean of 52.6 quintals per hectare and a median of 53.9 quintals per hectare, this cluster represents regions with moderate to relatively high levels of rice productivity. The lower standard deviation (10.1) indicates more controlled variability within the data, suggesting that regions in this cluster tend to have similar productivity patterns that exceed the overall average but do not reach the highest productivity levels.

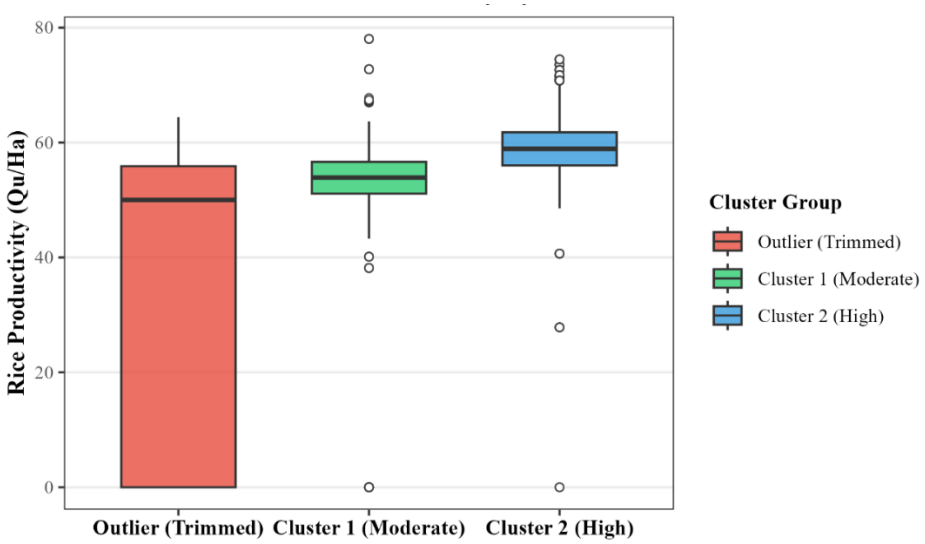


**Figure 5. Boxplot Distribution of Rice Productivity by Cluster**

Cluster 2 (blue) stands out with the highest and most consistent rice productivity levels among the three groups. With a mean of 59.2 quintals per hectare and a median of 58.9 quintals per hectare, this cluster represents regions characterized by optimal and stable rice productivity. The narrowest boxplot range and the lowest standard deviation (6.39) confirm that the regions within this cluster exhibit highly uniform and elevated productivity levels. This finding reinforces the assumption that Cluster 2 comprises key rice-producing areas that are highly efficient and contribute the most to overall rice productivity.

Overall, this distribution analysis confirms that the trimming Gaussian Mixture Model (TGMM) successfully clustered regions based on both the level and consistency of their rice productivity, effectively separating outliers (Cluster 0) from the moderate productivity group (Cluster 1) and the high productivity group (Cluster 2). This classification provides a deeper understanding of the characteristics of each cluster and their implications for rice productivity management.

### 3.3. Cluster Stability

Following clustering using the Gaussian Mixture Model (GMM) with trimming, evaluating cluster stability is essential to ensure consistent results. Stability was assessed using the Adjusted Rand Index (ARI) through two internal approaches: bootstrap and subsampling. The ARI varies between -1 and 1, with values approaching 1 signifying greater clustering stability. The bootstrap test yielded an average ARI of 0.41, indicating moderate stability and some variability in cluster assignments due to resampling with replacement. In contrast, the subsampling test produced a higher average ARI of 0.545, reflecting greater stability and consistency despite sampling without replacement.

Overall, the ARI results from both methods suggest that the obtained cluster partitions are reasonably stable and reliable for data interpretation, although there remains room for improvement, particularly in the bootstrap approach. These findings align with Hennig [24], which emphasized that trimming in GMM improves robustness by excluding extreme observations, thus enhancing stability under noisy conditions. Furthermore, Mouret et al. [32] demonstrated that robust GMM techniques could effectively reconstruct satellite-derived time series and detect agricultural anomalies even in contaminated datasets, reinforcing the suitability of trimming-based clustering for spatio-temporal agricultural analysis. In this study, the application of TGMM to rice productivity data confirms its potential to produce stable clusters that remain interpretable and resilient to variability, supporting informed decision-making in agricultural policy.

## 4. CONCLUSION

This study successfully applied the Trimming Gaussian Mixture Model (TGMM) to cluster monthly rice productivity time series data in West Java (2018–2023), identifying an optimal structure with two main clusters and a trimming proportion of 0.15. The results effectively distinguished significant productivity patterns, comprising a high-productivity group characterized by low variability, a moderate-stable group, and a distinct outlier group marked by low productivity and high fluctuations. The stability of these clusters was confirmed through the Adjusted Rand Index (ARI), yielding values of 0.41 (bootstrap) and 0.545 (subsampling), which demonstrates the consistency and reliability of TGMM as a robust method for agricultural time series analysis. While this study was limited to the geographic scope of West Java and a specific number of clusters, future research is recommended to extend this approach to other regions or commodities, explore a more diverse range of cluster numbers, and integrate TGMM with predictive models to provide a comprehensive tool for agricultural planning and policy formulation.


### Funding Information
The authors state no funding is involved.


### Author Contributions Statement
**Sarah Fadhlia:** Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing-original draft, writing-review & editing, supervision, project administration. **Eko Primadi Hendri:** Methodology, software, validation, formal analysis, investigation, data curation, visualization, writing-review & editing. All authors discussed the results and contributed to the final manuscript.

### Conflict Of Interest Statement
Authors state no conflict of interest.

### Data Availability
The data that support the findings of this study are openly available at the West Java Open Data portal (https://opendata.jabarprov.go.id) under the dataset titled *"Produktivitas Padi Kerangka Sampel Area (KSA) Berdasarkan Bulan di Jawa Barat"*, accessible at: https://opendata.jabarprov.go.id/id/dataset/produktivitas-padi-kerangka-sampel-area-ksa-berdasarkan-bulan-di-jawa-barat.

## REFERENCES


[1] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering – A decade review," *Inf Syst*, vol. 13, pp. 16–38, 2015.

[2] R. Umatani, T. Imai, K. Kawamoto, and S. Kunimasa, "Time series clustering with an EM algorithm for mixtures of linear Gaussian state space models," *Pattern Recognit*, vol. 138, no. 15, p. 109375, 2023.

[3] M. Ulinnuha, F. M. Afendi, and I. M. Sumertajaya, "Study of Clustering Time Series Forecasting Model for Provincial Grouping in Indonesia Based On Rice Price," *Indonesian Journal of Statistics and Its Applications*, vol. 6, no. 1, pp. 50–62, 2022.

[4] A. D. Munthe, "Penerapan Klastering Deret Waktu untuk Pengelompokan Provinsi di Indonesia Berdasarkan Nilai Produksi Padi," *Jurnal Litbang Sukowati*, vol. 2, no. 2, pp. 1–11, 2019.

[5] A. M. Yolanda and H. Savira, "Segmentation of Provinces in Indonesia Using Time Series Data of Rice," *Jurnal Pangan*, vol. 33, no. 3, pp. 169–177, 2024.

[6] D. A. Reynolds, "Gaussian Mixture Models," in *Encyclopedia of Biometrics*, Springer, 2009, pp. 659–663.

[7] P. Coretto and C. Hennig, "Robust Improper Maximum Likelihood: Tuning, Computation, and a Comparison With Other Methods for Robust Gaussian Clustering," *J Am Stat Assoc*, vol. 111, no. 516, pp. 1648–1659, 2016.

[8] L. A. García-Escudero, A. Gordaliza, C. Matrán, and A. Mayo-Iscar, "A general trimming approach to robust cluster Analysis," *The Annals of Statistics*, vol. 36, no. 3, pp. 1324–1345, 2008.

[9] J. Cuesta-Albertos, A. Gordaliza, and C. Matrán, "Trimmed k-Means: An Attempt to Robustify Quantizers," *The Annals of Statistics*, vol. 25, no. 2, pp. 553–576, 1997.

[10] P. D'Urso, L. De Giovanni, and V. Vitale, "Robust DTW-based entropy fuzzy clustering of time series," *Ann Oper Res*, 2023.

[11] M. Ishanifa, "Use of Hierarchical Clustering Method with Complexity Invariant Distance on Provincial Rice Prices in Indonesia," *Journal of Applied Statistics and Data Science*, vol. 2, no. 1, pp. 45–57, 2025.

[12] I. P. Kurniawati, H. Pratiwi, and S. Sugiyanto, "Indonesian Territory Clustering Based On Harvested Area and Rice Productivity Using Clustering Algorithm," *Journal of Social Science*, vol. 4, no. 1, pp. 100–110, 2023.

[13] L. Gandharum, M. E. Mulyani, D. M. Hartono, A. Karsidi, and M. Ahmad, "Remote sensing versus the area sampling frame method in paddy rice acreage estimation in Indramayu regency, West Java province, Indonesia," *Int J Remote Sens*, vol. 42, no. 5, pp. 1738–1767, 2021.

[14] X. Song, Q. Wen, Y. Li, and L. Sun, "Robust Time Series Dissimilarity Measure for Outlier Detection and Periodicity Detection," in *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22)*, Atlanta, Georgia, 2022.

[15] A. Blázquez-García, U. M. A. Conde, and J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ACM Comput Surv*, vol. 54, no. 3, pp. 1–33, 2021.

[16] T. Iizumi and N. Ramankutty, "Changes In Yield Variability of Major Crops for 1981–2010 Explained by Climate Change," *Environmental Research Letters*, vol. 11, no. 3, p. 034003, Mar. 2016, doi: 10.1088/1748-9326/11/3/034003.

[17] C. Lesk, P. Rowhani, and N. Ramankutty, "Influence of Extreme Weather Disasters On Global Crop Production," *Nature*, vol. 529, pp. 84–87, Jan. 2016, doi: 10.1038/nature16467.

[18] D. K. Ray, J. S. Gerber, G. K. MacDonald, and P. C. West, "Climate Variation Explains A Third of Global Crop Yield Variability," *Nat Commun*, vol. 6, p. 5989, Jan. 2015, doi: 10.1038/ncomms6989.

[19] I. Mahmudiati and R. Fajriyah, "Grouping Indonesian Province Farmers' Term of Trade Using Dynamic Time Warping," *Indonesian Journal of Applied Statistics*, vol. 7, no. 2, pp. 112–120, 2024.

[20] E. Keogh and S. Kasetty, "On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration," *Data Min Knowl Discov*, vol. 7, no. 4, pp. 349–371, Oct. 2003, doi: 10.1023/A:1024988512476.
[21] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. Wiley, 1987. doi: 10.1002/0471725382.
[22] A. Javed, B. S. Lee, and D. M. Rizzo, "A Benchmark Study on Time Series Clustering," *Machine Learning with Applications*, vol. 1, 2020.
[23] P. Senin, "Dynamic Time Warping Algorithm Review," 2008, *Honolulu, USA*.
[24] C. Hennig, "Cluster-wise assessment of cluster stability," *Comput Stat Data Anal*, vol. 52, no. 1, pp. 258–271, 2007.
[25] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag, 2002. doi: 10.1007/b98835.
[26] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to The Interpretation and Validation of Cluster Analysis," *J Comput Appl Math*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.
[27] W. J. P. Government, "Rice productivity using the Area Sample Framework (ASF) by month in West (2018-2023)," 2024.
[28] C. Ritz and J. C. Streibig, *Nonlinear Regression with R*. New York: Springer, 2008.
[29] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag, 2016.
[30] J. W. Tukey, *Exploratory Data Analysis*. Massachusetts: Addison-Wesley, 1977.
[31] R. Mcgill, J. W. Tukey, and W. A. Larsen, "Variations of Box Plots," *Am Stat*, vol. 32, no. 1, pp. 12–16, 1977.
[32] F. Mouret, M. Albughdadi, S. Duthoit, D. Kouamé, G. Rieu, and J.-Y. Tourneret, "Reconstruction of Sentinel-2 derived time series using robust Gaussian mixture models — Application to the detection of anomalous crop development," *Comput Electron Agric*, vol. 198, pp. 148–163, 2022.