12 (1) 2026 45-57 ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
Register: *Jurnal Ilmiah Teknologi Sistem Informasi*

Contents lists available at www.journal.unipdu.ac.id

Register

Journal Page is available to www.journal.unipdu.ac.id/index.php/register

Research article

Agroecological Zoning of Bangkalan Regency Using K-Means and HDBSCAN Based on Integrated Soil Fertility and Climate Features

Wahyudi Agustiono[a], Giraldo Stevanus[b], Yoga Dwitya Pramudita[c], Wahyudi Setiawan[d,], Deshinta Arrova Dewi[e]*

[a,b,d] *Department of Information System, University of Trunodjoyo Madura, Bangkalan, Indonesia*
[c] *Department of Informatics, University of Trunodjoyo Madura, Bangkalan, Indonesia*
[e] *Center for Data Science and Sustainable Technologies, INTI International University, Malaysia. Jalan Persiaran Perdana Bandar Baru Nilai, Malaysia*
email: [a] wahyudi.agustiono@trunojoyo.ac.id, [b] 220441100064@student.trunojoyo.ac.id, [c] yoga@trunojoyo.ac.id, [d,*] wsetiawan@trunojoyo.ac.id, [e] deshinta.ad@newinti.edu.my * *Correspondence*

ARTICLE INFO

Article history:
Received September 28th, 2025
Revised February 15th, 2026
Accepted June 18th, 2026
Available online June 23th, 2026

Keywords:
Agroecological Zoning
K-Means Clustering
HDBSCAN
Soil–Climate Integration
Crop Suitability Mapping

Please cite this article in IEEE style as:
W. Agustiono, G. Stevanus, Y.D. Pramudita, W. Setiawan, D.A. Dewi, "Agroecological Zoning of Bangkalan Regency Using K-Means and HDBSCAN Based on Integrated Soil Fertility and Climate Features", : Jurnal Ilmiah Teknologi Sistem Informasi, vol. 12, no. 1, pp. 45-57, 2026.

ABSTRACT

Agroecological heterogeneity poses challenges for agricultural planning in Bangkalan Regency, Indonesia. This study aimed to delineate agroecological zones by integrating soil fertility, climate, and topographic variables using K-Means clustering and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). A total of 11,000 geospatial observations obtained from Google Earth Engine were aggregated into 277 village-level units. The dataset included soil nutrients (nitrogen, phosphorus, and potassium), the Soil Quality Index, temperature, rainfall, humidity, elevation, and slope. Data preparation, modeling, and evaluation were performed as the primary methodological steps. Min-Max Scaling was applied to normalize the data. The optimal K-Means configuration (K = 3) achieved a Silhouette Score of 0.2668, an Inertia value of 294.5529, and a Calinski-Harabasz Index (CHI) of 75.8821. The resulting clusters were classified as High-Potential (52 villages), Moderate-Potential (142 villages), and Environmental-Constraint (83 villages) zones. HDBSCAN was used to validate clustering patterns and detect environmental anomalies. The optimal HDBSCAN configuration identified two density-based clusters and five noise villages. These villages showed exceptionally high nitrogen, phosphorus, and Soil Quality Index values, indicating localized agroecological hotspots. The integration of K-Means and HDBSCAN offers a comprehensive framework for agricultural planning, resource allocation, and sustainable land management.

1. Introduction

Agriculture plays a central role in food security and economic development, particularly in agrarian economies. Agricultural productivity is increasingly affected by soil nutrient imbalances and climate variability [1]. Essential nutrients, such as nitrogen (N), phosphorus (P), potassium (K), and soil acidity (measured by pH) are critical determinants of crop growth and yield [2]. Nitrogen deficiency has been reported to reduce rice yields by up to 50%, while potassium deficiency may cause losses of up to 70% [3]. Extreme temperatures exceeding 35 °C and irregular rainfall can disrupt crop production and, in severe cases, lead to total crop failure [4].

These challenges are also evident in Bangkalan Regency, East Java, which has predominantly dryland agricultural systems and considerable environmental heterogeneity. The regency experiences low and variable rainfall, prolonged dry seasons, and unevenly soil fertility distribution across villages. Topographic conditions range from low-lying coastal areas to inland regions with higher elevations and steeper slopes [5], [6]. Such spatial variability leads to unequal agricultural productivity and complicates the formulation of site-specific development strategies. Therefore, a systematic agroecological zoning framework is needed to identify areas with similar soil and climate characteristics [7], [8]

Machine learning approaches have been applied in agriculture for crop yield prediction, soil quality assessment, land suitability analysis, and pest monitoring. Several studies have shown the effectiveness of clustering techniques, including K-Means, for identifying spatial patterns in agricultural and environmental datasets [9]. However, most previous studies have focused on either soil properties [10], [11]. Village-level agroecological zoning that integrates soil fertility, climate, and topographic variables within a single framework remains limited in the literature. The potential influence of extreme observations or agroecological hotspots is also often overlooked.

Although previous studies have applied clustering techniques to agricultural zoning, most have focused on either soil properties or climatic variables in isolation, while only a few have integrated soil fertility, climate, and topographic factors into a unified agroecological framework. Furthermore, existing studies generally employ partition-based clustering methods that assign all observations to predefined groups, potentially overlooking localized agroecological hotspots and anomalous environmental conditions. Consequently, the identification of villages with exceptionally distinct soil fertility characteristics remains limited.

To address this gap, this study proposes an integrated Agroecological Zoning (AEZ) framework using K-Means clustering and HDBSCAN. The framework integrates soil nutrient variables (N, P, and K), soil quality indicators, climate parameters, and topographic characteristics to delineate agroecological zones across the 277 villages in Bangkalan Regency. HDBSCAN is used as a complementary density-based method to identify agroecological hotspots and validate the robustness of the clustering structure. Table 1 shows a comparison with previous studies.

Table 1. Comparison of Previous Studies and the Present Study

Study	Main Variables	Method	Agroecological Zoning	Soil–Climate Integration	Topographic Factors
[9]	Land suitability variables	PCA + K-Means	Partial	No	Limited
[7]	Soil fertility indicators	Statistical analysis	Yes	Partial	No
[8]	Soil fertility and agroforestry variables	Field-based analysis	Partial	Partial	No
[12]	Soil, climate, and slope	AEZ framework	Yes	Yes	Yes
[13]	Soil or climate variables	K-Means or related methods	Partial	Limited	Limited
Present Study	Soil fertility, climate, and topography	K-Means + HDBSCAN	Yes	Yes	Yes

Table 1 shows that previous studies generally focused on either soil properties or climatic factors. Density-based clustering for anomaly detection has rarely been incorporated. The present study combines K-Means and HDBSCAN to simultaneously delineate agroecological zones and identify localized agroecological hotspots at the village scale.

2. Materials and Methods

2.1. Materials

The materials used in this study consist of publicly available geospatial datasets representing the soil, climate, and topographic characteristics of Bangkalan Regency, East Java, Indonesia. Soil parameters, including N, P, potassium (cation exchange capacity/CEC), pH, and soil texture, were obtained from the SoilGrids v2 database [14] . Climate data, covering temperature, humidity, rainfall, wind speed, and sunshine duration, were collected from ERA5-Land [15] and CHIRPS [16] Topographic variables, such as elevation and slope, were extracted from the Shuttle Radar Topography Mission (SRTM DEM) [17]. To ensure adequate spatial representation, 11,000 random sampling points were selected within the administrative boundaries of Bangkalan Regency. Table 2 presents the dataset specifications used as input for data collection and processing.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 45-57

Table 2.Specification and Description of Data

Specification	Description
Subject area	Agriculture, Agroecology, Environmental Science
Type of data	Tabular data, geospatial points (CSV, raster)
Data format	Raw, processed
Number of records	11,000 sampling points
Parameters	Soil nutrients (N, P, K, pH, texture), climate (temperature, humidity, rainfall, sunshine, wind speed), topography (elevation, slope), geographic coordinates (Latitude and longitude for each sampling point)
Data source	Google Earth Engine (SoilGrids v2, ERA5-Land, CHIRPS, SRTM DEM)
Data period	2021–2025

Although the data sources covered the period from 2021to 2025, only the most recent available observations (2025) were used for the analysis. Therefore, the dataset represents a temporal snapshot of agroecological conditions rather than a long-term multi-year average.

2.2. Methods

To ensure a systematic and reproducible research process, this study employed a structured methodology organized into a series of sequential stages. The framework was designed to transform raw agroecological datasets into meaningful, actionable insights through data preprocessing, feature selection, dimensionality reduction, clustering, and interpretation. Each stage was carefully executed to maintain data quality, ensure analytical rigor, and enhance the validity of the resulting recommendations. The following subsections provide a detailed description of each step in the methodological process.

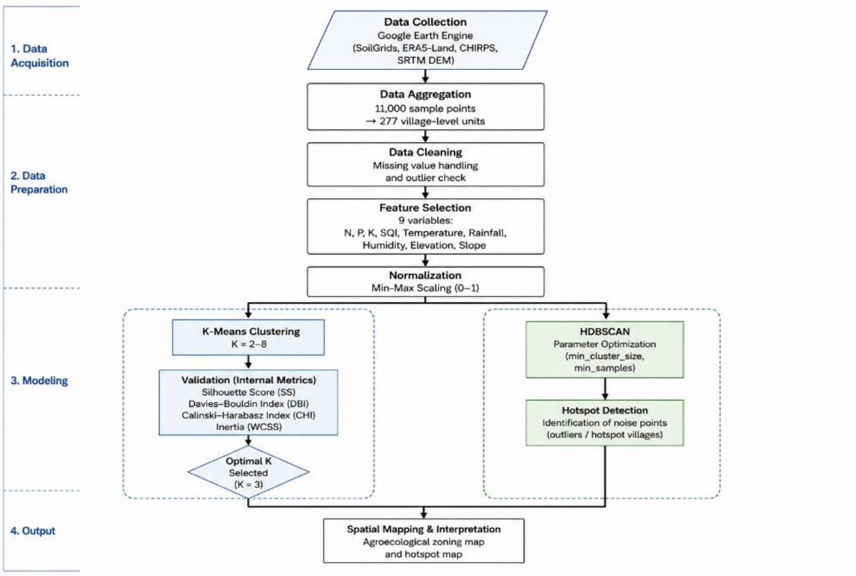

Fig. 1. Methodology research of the study

The overall methodological framework of this study is illustrated in Figure 1. The framework summarizes the sequential stages of the research, beginning with data collection and preprocessing, followed by feature selection and dimensionality reduction, clustering, and ending with interpretation and validation. This visual representation emphasizes the logical flow of the study and demonstrates how each analytical stage is systematically connected to subsequent stages.

2.2.1. Data Collection

The sampled point data were aggregated according to administrative boundaries of the 277 villages in Bangkalan Regency. Zonal statistics were calculated by averaging the values of all sampling points within each village boundary polygon. This aggregation transformed the initial dataset of 11,000 point observations into a village-level dataset comprising 277 records. Each record represents the mean soil, climate, and topographic characteristics of a village unit. The resulting dataset served as the primary input for subsequent clustering analysis.

48
ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 45-57

A total of 11,000 sampling points were generated using a uniform random sampling approach within the administrative boundary of Bangkalan Regency in Google Earth Engine. Under this approach, each location within the study area had an equal probability of being selected, minimizing spatial selection bias. Environmental attributes from SoilGrids, ERA5-Land, CHIRPS, and SRTM datasets were extracted at each sampling point and used for subsequent analysis.

2.2.2. Data Preprocessing

Data preprocessing was performed to ensure the dataset was suitable for clustering using K-Means and HDBSCAN. Following aggregation, the dataset comprised 277 villages, each represented by aggregated soil, climate, and topographic attributes. The first stage involved data cleaning and handling missing values. Several records contained incomplete information, particularly for the soil type variable. A data quality assessment revealed that 17 out of 277 villages (6.14%) had missing information in the Soil Type attribute. Mode imputation was applied by assigning the most frequent soil type, Vertisol, to the missing observations. This approach was used because Vertisol was the most dominant soil type across the study area. Following imputation, no missing values remained in the dataset.

The second stage consisted of feature selection. Nine variables with high agricultural relevance were selected as input features: N, P, K, slope, elevation, rainfall, humidity, temperature, and the Soil Quality Index (SQI). Variables excluded were Soil Type (with near-zero variability), Soil pH (with a limited range of 54 to 63), and geographic coordinates (to avoid proximity-based bias).

The final preprocessing stage involved data normalization using Min-Max Scaling. This procedure transformed all variables into a common scale ranging from 0 to 1. Normalization was particularly important because nutrient variables such as Phosphorus had values in the hundreds of thousands, whereas slope had substantially smaller numerical ranges in comparison.

2.2.3. Clustering Method

This study used two unsupervised machine learning algorithms: K-Means and HDBSCAN, to delineate agroecological zones and identify environmental anomalies. The combined use of these algorithms provides complementary perspectives into agroecological heterogeneity.

K-Means clustering was used as the primary method. The algorithm partitions observations into K predefined clusters by minimizing the within-cluster sum of squared distances between observations and their respective cluster centroids. Villages with similar soil fertility, climate, and topographic characteristics are grouped into the same cluster based on the Euclidean distance metric. The clustering process used nine selected features: N, P, K, slope, elevation, rainfall, humidity, temperature, and Soil Quality Index.

To determine the optimal number of clusters, several internal validation metrics were evaluated, including the Silhouette Score, Davies-Bouldin Index (DBI), Within-Cluster Sum of Squares (WCSS/Inertia), and the Calinski-Harabasz Index (CHI). The final cluster configuration was selected based on the combined interpretation of these metrics and the practical interpretability of the resulting clustered zones.

HDBSCAN was applied as a complementary density-based clustering method. Unlike K-Means, HDBSCAN does not require the number of clusters to be specified in advance and can identify observations that do not belong to any dense cluster. This characteristic makes HDBSCAN suitable for detecting agroecological hotspots and identifying anomalous villages with extreme environmental characteristics. The algorithm was optimized by varying the parameters min_cluster_size (5, 10, 15, 20, 25, and 30) and min_samples (1, 3, 5, and 10). The optimal parameter combination was selected based on the highest Silhouette Score obtained.

In this study, HDBSCAN served two purposes. First, it was used to validate the robustness of the agroecological patterns identified by K-Means clustering. Second, it enabled the detection of villages with exceptionally distinct soil fertility characteristics. Villages classified as noise (label = -1) were interpreted as agroecological hotspots requiring special consideration in agricultural land-use planning.

2.2.4. Model Evaluation

The Silhouette Score (SS) measures the degree of cohesion within clusters and separation between clusters, with higher values indicating better clustering performance [18]. The Davies-Bouldin Index (DBI) evaluates the ratio of within-cluster dispersion to between-cluster separation, where lower values

 https://doi.org/10.26594/register.v12i1.5969

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 45-57

represent more distinct clusters [19]. Inertia (WCSS) measures cluster compactness by summing the squared distances between observations and their assigned centroids[18], [19], [20], CHI assesses the ratio of between-cluster dispersion to within-cluster dispersion, with higher values indicating better-defined cluster structures[21].

For K-Means, the number of clusters was varied from K = 2 to K = 8 clusters. Each configuration was evaluated using all four metrics. For HDBSCAN, the cluster_selection_method parameter was systematically evaluated. The SS metric was used throughout the analysis as the primary evaluation criterion.

3. Results and Discussion

3.1. Result

3.1.1 K-Means Cluster Optimization

The optimal number of clusters was determined by evaluating K-Means performance using four internal validation metrics: SS, DBI, WCSS/Inertia, and CHI indices. These metrics were evaluated for cluster sizes ranging from K = 2 to K = 8. The comparative results are presented in Table 4 and Figure 2.

Table 4. Internal Validation Metrics for Determining the Optimal Number of Clusters

Cluster-Size	SS	DBI	Inertia	CHI
K=2	0.2284	1.7552	374.6914	62.4542
K=3	**0.2668**	1.3708	**294.5529**	**75.8821**
K=4	0.2534	1.3323	227.193	73.1205
K=5	0.2458	1.3373	178.1324	70.5518
K=6	0.2498	1.3882	139.4003	68.2234
K=7	0.245	1.3305	116.6279	65.9012
K=8	0.2524	**1.3098**	96.6532	63.4429

Figure 2 shows that the Silhouette Score (SS) reached its highest value of 0.2668 at K = 3, indicating that the clusters at this configuration exhibit the best balance between cohesion and separation. A higher Silhouette Score suggests that observations are well matched to their own cluster and distinctly separated from neighboring clusters. Therefore, the SS metric supports K = 3 as the optimal number of clusters.

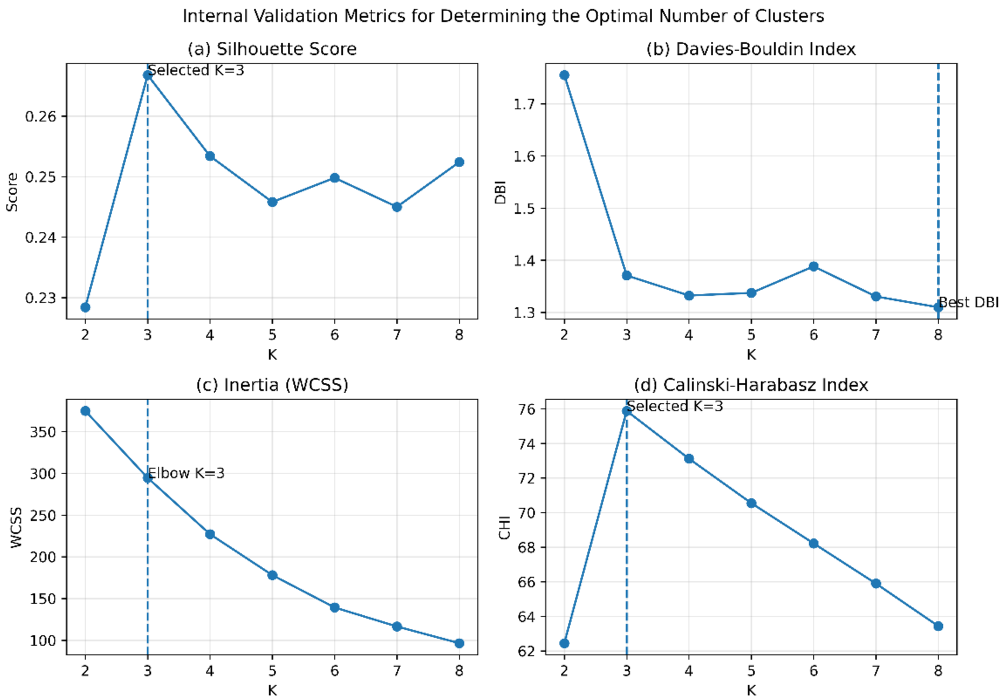

Fig. 2. Determination of the Optimal K-Means Cluster Number Based on Internal Validation Metrics

 https://doi.org/10.26594/register.v12i1.5969

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 45-57

The Davies–Bouldin Index (DBI) achieved its lowest value of 1.3098 at K = 8 clusters. Since lower DBI values indicate better clustering quality, this metric suggests that increasing the number of clusters improves the compactness and distinctiveness of the cluster boundaries. Consequently, DBI identifies K = 8 as the optimal solution based on this criterion.

The WCSS (Inertia) curve shows a substantial decrease from K = 2 to K = 3, followed by a more gradual decline for higher K values. This pattern forms an "elbow" around K=3, indicating that adding more clusters beyond this point yields diminishing improvements in within-cluster compactness. Thus, the elbow method based on inertia also favors K = 3.

The Calinski–Harabasz Index (CHI), also known as the Variance Ratio Criterion, evaluates clustering quality by measuring the ratio of between-cluster dispersion to within-cluster dispersion. Higher CHI values indicate that clusters are compact internally while remaining well separated from one another. In this study, the CHI reached its maximum value of 75.8821 at K = 3, suggesting that this configuration provides the most effective balance between cluster cohesion and separation among all tested configurations. The decline in CHI values for K > 3 clusters indicates that additional clusters increase model complexity without producing a proportional improvement in cluster distinctiveness. Therefore, the CHI results strongly support selecting K = 3 as the optimal clustering configuration for agroecological zoning in this study.

Overall, three of the four validation metrics (SS, Inertia, and CHI) consistently identify K = 3 as the optimal number of clusters, while only the Davies–Bouldin Index favors K = 8 clusters. Considering the agreement among most metrics and the principle of model parsimony, K = 3 was selected as the final clustering solution, balancing cluster quality, interpretability, and its practical usefulness for agroecological zone delineation.

Although the Davies–Bouldin Index achieved its optimum value at K = 8 clusters, the improvement compared with K = 3 was relatively modest. In contrast, the Silhouette Score and Calinski–Harabasz Index both reached their highest values at K = 3, while the Inertia curve exhibited a clear elbow point at the same cluster size. These findings indicate that K = 3 provides the best compromise between cluster quality, model simplicity, and interpretability in practice. Furthermore, the three resulting clusters can be meaningfully interpreted as High-Potential, Moderate-Potential, and Environmental-Constraint zones, making the solution more useful for agricultural planning than a more fragmented eight-cluster configuration. Figure 3 shows the AEZ Zoning Map of Bangkalan Regency.

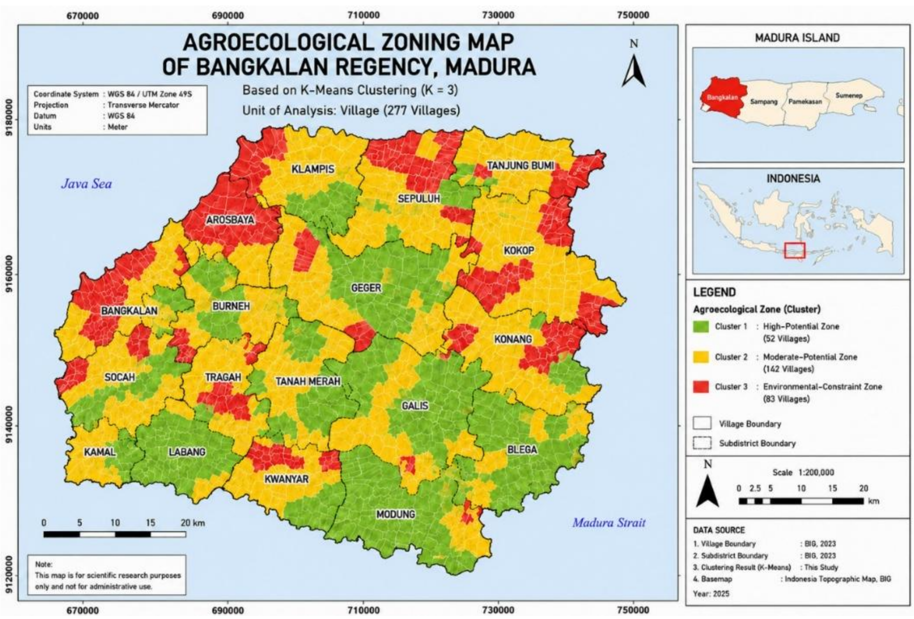

Figure 3. Spatial distribution of agroecological clusters in Bangkalan Regency. Cluster 1 (High-Potential Zone) is concentrated in the western and central regions, whereas Cluster 2 and Cluster 3 occupy transitional and environmentally constrained areas, respectively.

https://doi.org/10.26594/register.v12i1.5969

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 45-57

The spatial distribution of Cluster 1 reveals that villages with the highest agroecological potential are concentrated primarily in the western, central, and southern parts of Bangkalan Regency. Based on the village coordinate analysis, Cluster 1 villages form several contiguous spatial groups, particularly around Bangkalan, Arosbaya, Burneh, Kwanyar, Galis, and parts of Tragah Districts. This spatial concentration suggests that favorable agroecological conditions are not randomly distributed but occur in geographically connected areas characterized by relatively high soil fertility and overall environmental suitability.

The quantitative analysis further supports this spatial pattern. Cluster 1 exhibits the highest mean values of Nitrogen (2412.38), Phosphorus (324,840.61), Potassium (203.44), and Soil Quality Index (618,744.02) among all clusters identified. These conditions indicate that villages within this zone have superior nutrient availability and overall soil quality, both essential factors for agricultural productivity and sustainability. In addition, the average elevation (37.37 m) and slope (1.38) remain relatively moderate, reducing topographic constraints on agricultural activities in these areas.

The spatial clustering of high-potential villages suggests that agricultural intensification programs can be implemented more efficiently through area-based development strategies instead of village-by-village interventions. The concentration of Cluster 1 in geographically connected regions may facilitate infrastructure development, irrigation management, agricultural extension services, and coordinated land-use planning processes. Therefore, Cluster 1 represents the primary target area for productivity enhancement and sustainable agricultural development efforts in Bangkalan Regency.

3.1.2 Cluster-to-Crop Mapping Results

The K-Means algorithm partitioned the 277 villages into three agroecological zones. The mean values of the selected variables for each cluster are presented in Table 5.

Table 5. Average Soil Fertility, Climatic, and Topographic Characteristics of Each Agroecological Cluster

Feature	Cluster 1	Cluster 2	Cluster 3
Potassium (K)	203.44	192.54	184.08
Nitrogen (N)	2412.38	2223.80	2160.52
Phosphorus (P)	324,840.61	299,449.55	290,926.83
Elevation (m)	37.37	38.40	59.39
Humidity (%)	79.46	79.56	79.61
Slope	1.38	1.39	1.94
Soil Quality Index	618,744.02	570,380.09	554,146.35
Temperature (°C)	27.37	27.36	27.30
Rainfall	6.18	6.14	6.15

Cluster 1 represents the most favorable agroecological conditions and is classified as the High-Potential Zone. Villages in this cluster exhibit the highest average values of N (2412.38), P (324,840.61), and K (203.44). This cluster also recorded the highest Soil Quality Index (618,744.02) value. The average elevation (37.37 m) and slope (1.38) are relatively moderate, indicating minimal topographic constraints on agricultural activities. These characteristics suggest that Cluster 1 is the most suitable zone for intensive agricultural production systems.

Cluster 2 is characterized as the Moderate-Potential Zone, representing villages with intermediate agroecological suitability. The average values of N (2223.80), P (299,449.55), K (192.54), and Soil Quality Index (570,380.09) are lower than those of Cluster 1 but remain higher than those of Cluster 3. Topographic conditions are relatively favorable, with low elevation (38.40 m) and gentle slopes (1.39). Agricultural productivity in this cluster is primarily constrained by moderate levels of soil fertility rather than adverse environmental conditions.

Cluster 3 is classified as the Environmental-Constraint Zone and is the largest cluster in the study area. It exhibits the lowest K concentration (184.08) and the lowest Soil Quality Index (554,146.35). A distinguishing characteristic is its topographic condition, with the highest average elevation (59.39 m) and slope (1.94), indicating more challenging terrain. The combination of lower soil quality and more challenging topography may limit agricultural productivity. Villages in Cluster 3 would benefit from

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 45-57

conservation-oriented agricultural practices, erosion control measures, and crop varieties adapted to marginal environmental conditions.

The comparison among clusters shows that soil fertility variables, particularly N, P, K, and Soil Quality Index, play a dominant role in differentiating agroecological zones in the study area. Cluster 1 consistently exhibits the highest nutrient availability, making it the most suitable zone for agricultural intensification activities. Cluster 3 is distinguished primarily by higher elevation and slope, indicating that topographic constraints contribute to its lower agricultural potential. Cluster 2 occupies an intermediate position between the other two zones.

Table 6. HDBSCAN Parameter Optimization Results

min_cluster_size \ min_samples	1	3	5	10
5	**0,4651**	**0,4651**	0,3341	0,3820
10	0,3812	0,3812	0,3812	0,3812
15	0,3795	0,3795	0,3795	0,3795
20	0,3795	0,3795	0,3795	0,3795
25	0,3795	0,3795	0,3795	0,3795
30	0,3795	0,3795	0,3795	0,3795

The highest Silhouette Score (0.4651) was achieved with min_cluster_size = 5 and min_samples = 1 or 3. This result indicates that HDBSCAN effectively distinguished high-potential agroecological hotspots from surrounding villages in the study area. When min_cluster_size increased from 15 to 30, the Silhouette Score remained stable at approximately 0.3795. Therefore, min_cluster_size = 5 and min_samples = 1 were selected as the optimal parameters for the final model configuration.

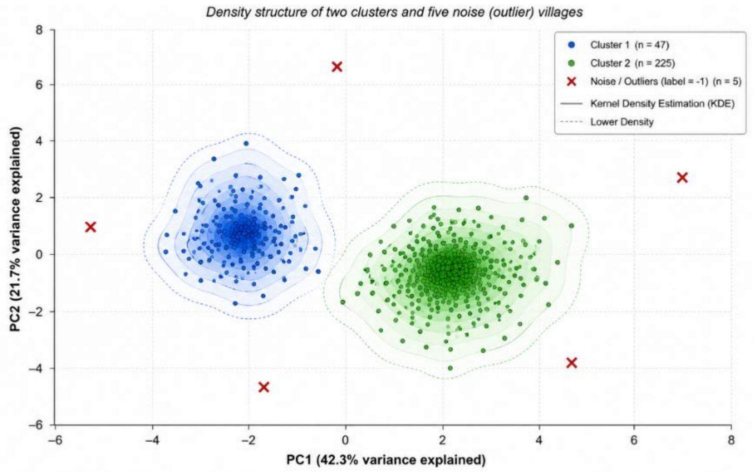

Figure 4. PCA Projection of HDBSCAN Clustering Results

Figure 4 presents the PCA projection of the HDBSCAN clustering results in this study. Two dominant density-based clusters were identified, representing the primary agroecological patterns across Bangkalan Regency. Cluster 1 forms a compact high-density region, indicating villages with relatively homogeneous environmental characteristics, whereas Cluster 2 encompasses the majority of villages with moderate-level agroecological profiles. The five villages classified as noise (label = −1) are located outside the principal density regions, suggesting distinctive combinations of soil fertility and environmental conditions that distinguish them from the dominant agroecological patterns observed.

3.1.3 HDBSCAN Clustering Results and Hotspot Detection

Based on parameter optimization, the best clustering performance was achieved with min_cluster_size = 5 and min_samples = 1, resulting in a Silhouette Score of 0.4651. The HDBSCAN analysis partitioned the dataset of 277 villages into two primary clusters and identified five villages (1.8%) as noise points (label = -1). The remaining 272 villages (98.2%) were assigned to one of the two clusters identified. Cluster 1 consisted of 47 villages, while Cluster 2 contained 225 villages. The distribution of villages across clusters and noise categories is presented in Table 7.

Table 7. Distribution of Villages Across HDBSCAN Clusters and Noise Category

https://doi.org/10.26594/register.v12i1.5969

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 45-57

Category	Number of Villages	Percentage (%)
Cluster 1	47	17.0
Cluster 2	225	81.2
Noise (Outliers)	5	1.8
Total	277	100.0

Figure 5 shows that all hotspot villages are concentrated in a relatively small area in the western part of Bangkalan Regency, suggesting that exceptional agroecological conditions are spatially clustered rather than randomly distributed across the region. These villages exhibit unusually high Nitrogen concentrations and Soil Quality Index values compared with the regional average, indicating distinct localized environmental advantages. The spatial concentration of hotspot villages highlights the potential influence of local soil formation processes, land management practices, and environmental conditions that distinguish them from surrounding villages in the study area. Consequently, these locations may serve as priority areas for precision agriculture, field validation, and site-specific agricultural development initiatives.

One of the key advantages of HDBSCAN is its ability to identify observations that do not belong to any dense cluster structure. In this study, five villages were classified as noise because their soil fertility characteristics were substantially different from those of other villages in Bangkalan Regency as a whole. These villages exhibited exceptionally high values of Nitrogen (N), Phosphorus (P), and Soil Quality Index, making them distinct agroecological hotspots within the region.

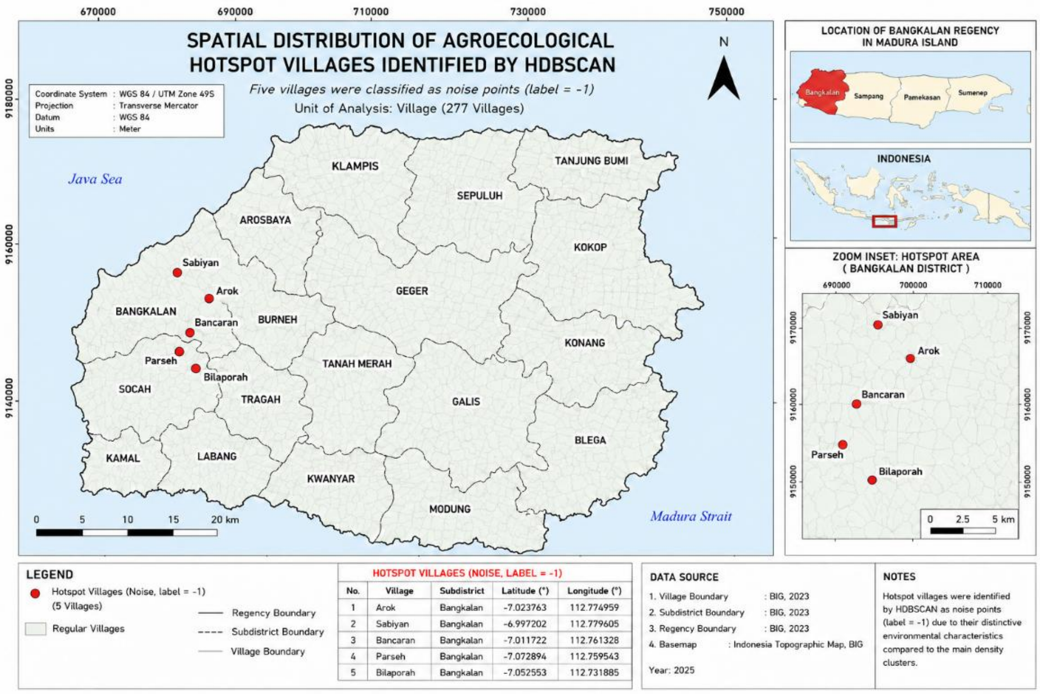

Fig. 5. Spatial Distribution of Agroecological Hotspot Villages Identified by HDBSCAN

The identified outlier villages were Arok (Burneh District), Sabiyan (Bangkalan District), Bancaran (Bangkalan District), Parseh (Socah District), and Bilaporah (Socah District). Among these villages, Arok exhibited the highest nutrient concentrations, with Nitrogen (N) reaching 3.618, Phosphorus (P) 487.168, and a Soil Quality Index of 927.939. Similarly, elevated values were observed in the other four villages, indicating significantly higher and exceptionally fertile soil conditions compared conditions with the regional average.

The five villages classified as noise were identified as agroecological hotspots due to their distinct environmental characteristics. They exhibited exceptionally high Nitrogen concentrations and Soil Quality Index values compared with the regional average of the study area. The hotspot villages were Arok (Burneh District), Sabiyan (Bangkalan District), Bancaran (Bangkalan District), Parseh (Socah

 https://doi.org/10.26594/register.v12i1.5969

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 45-57

District), and Bilaporah (Socah District). Table 8 summarizes the characteristics of the detected outlier villages in detail.

Table 8. Summarizes the characteristics of the detected outlier villages.

Village	District	Nitrogen (N)	Phosphorus (P)	Soil Quality
Arok	Burneh	3.618	487.168	927.939
Sabiyan	Bangkalan	3.496	470.725	896.620
Bancaran	Bangkalan	3.464	466.427	888.432
Parseh	Socah	3.263	439.379	836.912
Bilaporah	Socah	3.121	420.258	800.493

All five villages are characterized by Vertisol soils, but their nutrient concentrations are at the extreme upper end of the observed distribution range. HDBSCAN classified them as outliers rather than assigning them to a standard conventional cluster. This finding indicates the presence of localized agroecological hotspots with exceptionally high soil fertility. All hotspot villages are located in the western part of Bangkalan Regency, suggesting that exceptional agroecological conditions are spatially clustered rather than randomly distributed across space.

The 47 villages in Cluster 1 represent areas with relatively high agricultural potential, with favorable nutrient conditions and soil quality indicators. Cluster 2, containing 225 villages, represents the dominant agroecological pattern across the regency. HDBSCAN merged villages that K-Means had previously separated into the moderate-potential and environmental-constraint groups. This result indicates that, from a density-based perspective, these villages share broadly similar agroecological characteristics.

3.2 Discussion

3.2.1 Comparison of K-Means and HDBSCAN

K-Means partitioned all 277 villages into three clusters representing High-Potential, Moderate-Potential, and Environmental-Constraint zones. This approach provides a comprehensive classification framework suitable for agricultural planning purposes. HDBSCAN identified two principal density-based clusters and classified five villages as noise points. Unlike K-Means, HDBSCAN does not force all observations into predefined groups, thereby enabling the detection of villages with distinct or exceptional environmental characteristics. The comparison between both methods is presented in Table 9.

Table 9. Comparative Analysis of K-Means and HDBSCAN Clustering Results

Aspect	K-Means	HDBSCAN
Clustering Principle	Centroid-based partitioning	Density-based clustering
Number of Clusters	Fixed (K = 3)	Automatically determined
Optimal Configuration	K = 3	min_cluster_size = 5, min_samples = 1
Number of Clusters Produced	3	2
Number of Noise Villages	0	5
Ability to Detect Outliers	No	Yes
Main Purpose	Agroecological zoning	Hotspot and anomaly detection
Strengths	Clear and interpretable zoning structure	Identifies unique agroecological conditions
Limitation	Forces all villages into clusters	Produces fewer interpretable clusters

The results show that K-Means is more suitable for regional agroecological zoning, whereas HDBSCAN is effective for identifying localized environmental anomalies and hotspots. The combined use of both methods provides a more comprehensive understanding of agroecological heterogeneity across Bangkalan Regency as a whole.

3.2.2 Comparison with Previous Studies

Previous research has consistently emphasized the influence of temperature and soil pH on agroecological differentiation processes. Table 10 summarizes these studies and their relevance to the present study context.

Table 4. Comparison with Previous Studies on Key Agroecological Drivers

https://doi.org/10.26594/register.v12i1.5969

Reference	Key Variables	Main Findings	Relevance to Present Study
[22]	Temperature	Heat stress above 35 °C disrupts pollen viability, reduces photosynthetic efficiency, and leads to significant maize yield decline in tropical/subtropical regions.	Strongly supports the finding that temperature is the primary driver of cluster separation in agroecological analysis.
[23]	Soil pH	Low soil pH significantly reduces phosphorus availability and fertilizer efficiency, limiting crop productivity.	Aligns with results showing soil pH as the second most influential factor influencing clustering.
[24]	Soil pH	Soil pH regulates phosphorus fractions, nutrient cycling, and microbial community activity, thereby strongly influencing soil fertility.	Reinforces the role of soil pH as a dominant variable in agroecological zoning.
[23]	Soil temperature, climate variability	Soil–climate interactions explain yield anomalies better than air temperature alone, emphasizing variability over absolute values.	Strengthens observation that temperature variability is the most influential factor in clustering.
[12]	Soil type, slope, climate	Agroecological zoning shows that soil physical properties, slope, and climate are crucial determinants of sugarcane suitability in East Java.	Provides direct regional evidence supporting the integration of soil and climate variables in agroecological clustering.
[25]	Soil pH, nutrients	Precise soil pH and nutrient availability determine phosphorus uptake, crop yield, and nutritional quality in tropical systems.	Validates the practical significance of using soil pH and nutrient data as actionable clustering features.

Temperature is identified as the most influential factor. Djalovic et al. [22] showed that heat stress above 35 °C impairs pollen viability and grain filling, reducing yields in tropical regions. This aligns with the present study, where temperature was the primary driver of cluster separation in the dataset. Soil pH is consistently highlighted as a decisive determinant of nutrient dynamics and soil fertility [23], [24]. Purnamasari [12] demonstrated that soil physical characteristics, slope, and climate collectively determine land suitability in East Java, providing regional evidence that supports the present approach used in this study. The integration of K-Means and HDBSCAN extends previous AEZ research by enabling both regional classification and hotspot detection within a single analytical framework for agroecological analysis.

3.2.4 Limitation

Several limitations should be considered when interpreting the results. The analysis was based on agroecological variables aggregated from 11,000 observation points into 277 village-level units for analysis. This aggregation may mask within-village variability in soil fertility, topography, and climate conditions. The spatial resolution is limited to the village level, and higher-resolution analyses using field-level observations could provide more detailed and nuanced insights. The study used a selected set of variables, and other relevant factors such as land use, irrigation infrastructure, and socioeconomic conditions were not included in the analysis. Future research should integrate higher-resolution datasets, remote sensing indicators, and temporal climate variability to improve zoning accuracy and robustness.

4. Conclusion

This study analyzed agroecological zones in Bangkalan Regency using K-Means clustering and HDBSCAN based on integrated soil fertility, climate, and topographic variables datasets. A total of 11,000 observation points were aggregated into 277 village-level units for analysis purposes. Nine variables were used for clustering analysis, including N, P, K, elevation, slope, rainfall, humidity, temperature, and Soil Quality Index. The best K-Means configuration was K = 3, supported by the Silhouette Score, CHI, and the Elbow Method results. The three clusters represent High-Potential, Moderate-Potential, and Environmental-Constraint zones. HDBSCAN identified five hotspot villages (Arok, Sabiyan, Bancaran, Parseh, and Bilaporah) with exceptionally high N concentrations and Soil

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 45-57

Quality Index values compared to the regional average. Furthermore, future research should incorporate field-based soil measurements, crop yield observations, and higher-resolution environmental datasets to improve zoning accuracy and practical applicability in agricultural planning.

Author Contributions

Wahyudi Agustiono: Conceptualization, methodology design, data preprocessing, formal analysis, writing – original draft. Giraldo Stevanus: Data collection, coding implementation, software development, visualization, validation. Yoga Dwitya Pramudita: Methodology refinement, result interpretation, writing – review & editing, visualization support. Wahyudi Setiawan: Supervision, project administration, resources, writing – review & editing, correspondence. Deshinta Arrova Dewi: Funding Acquisition, Resources, Writing – Review & Editing, Project Coordination).

Acknowledgment

This work was supported by the Directorate of Research and Community Service (DRPM) under the Fundamental Research Regular Scheme, Grant No. B/031/UN46.1/PT.01.03/BIMA/PL/2025

Declaration of Competing Interest

We declare that we have no conflict of interest.

References

[1] R. Virtriana *et al.*, “Effects of extreme climate on agriculture in paddy field area in West Java Province (Indonesia) using multitemporal scenarios, GIS, and remote sensing,” *Geomatics, Natural Hazards and Risk*, vol. 16, no. 1, p. 2455487, Dec. 2025, doi: 10.1080/19475705.2025.2455487.

[2] M. S. Farooq *et al.*, “Partial replacement of inorganic fertilizer with organic inputs for enhanced nitrogen use efficiency, grain yield, and decreased nitrogen losses under rice-based systems of mid-latitudes,” *BMC Plant Biol.*, vol. 24, no. 1, p. 919, 2024, doi: 10.1186/s12870-024-05629-w.

[3] W. J. Brownlie, P. Alexander, M. Maslin, M. Cañedo-Argüelles, M. A. Sutton, and B. M. Spears, “Global food security threatened by potassium neglect,” *Nat. Food*, vol. 5, no. 2, pp. 111–115, 2024, doi: 10.1038/s43016-024-00929-8.

[4] X. Fan *et al.*, “Impacts of Extreme Temperature and Precipitation on Crops during the Growing Season in South Asia,” *Remote Sens. (Basel).*, vol. 14, no. 23, Dec. 2022, doi: 10.3390/rs14236093.

[5] W. Utama *et al.*, “Application of Flow Coefficients to Support High Economical Plant Cultivation (Case Study: Kwanyar Bangkalan, Indonesia),” *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 9, pp. 6828–6833, Sep. 2023, doi: 10.29303/jppipa.v9i9.3307.

[6] E. Fauziyah, I. Maflahah, and D. R. Hidayati, “Policy Impacts on Farm Efficiency: Fertilizer Subsidies in Corn Production in Madura Island, Indonesia,” *Research on World Agricultural Economy*, Sep. 2025, doi: 10.36956/rwae.v6i4.2264.

[7] W. I. Susanti, S. N. Cholidah, and F. Agus, “Agroecological Nutrient Management Strategy for Attaining Sustainable Rice Self-Sufficiency in Indonesia,” Jan. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/su16020845.

[8] J. Chavez, V. Nijman, D. K. T. Sukmadewi, M. D. Sadnyana, S. Manson, and M. Campera, “Impact of Farm Management on Soil Fertility in Agroforestry Systems in Bali, Indonesia,” *Sustainability (Switzerland)*, vol. 16, no. 18, Sep. 2024, doi: 10.3390/su16187874.

[9] D. Selvida, A. F. Pulungan, and A. S. Huzaifah, “Optimization Of Garlic Cultivation Land Selection Using Pca And K-Means Approach In Spatial Intelligent System,” *Eastern-European Journal of Enterprise Technologies*, vol. 3, no. 2, pp. 54–64, 2025, doi: 10.15587/1729-4061.2025.325340.

[10] C. Zhu *et al.*, “Digital Mapping of Soil Organic Carbon Based on Machine Learning and Regression Kriging,” *Sensors*, vol. 22, no. 22, Nov. 2022, doi: 10.3390/s22228997.

[11] X. Li *et al.*, “A zoning-based machine learning framework for accurate soil organic matter prediction across Mollisol and non-Mollisol regions,” *J. Integr. Agric.*, 2026, doi: https://doi.org/10.1016/j.jia.2026.01.016.

[12] R. A. Purnamasari, “Land suitability evaluation for sugarcane cultivation based on agroecological zoning system in East Java Indonesia,” *Buitenzorg: Journal of Tropical Science*, vol. 1, no. 2, pp. 42–50, Dec. 2024, doi: 10.70158/buitenzorg.v1i2.12.

[13] A. Filintas, N. Gougoulias, N. Kourgialas, and E. Hatzichristou, “Management Soil Zones, Irrigation, and Fertigation Effects on Yield and Oil Content of Coriandrum sativum L. Using Precision Agriculture with Fuzzy k-Means Clustering,” *Sustainability (Switzerland)*, vol. 15, no. 18, Sep. 2023, doi: 10.3390/su151813524.

[14] L. Poggio *et al.*, “SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty,” *SOIL*, vol. 7, no. 1, pp. 217–240, Jun. 2021, doi: 10.5194/soil-7-217-2021.

[15] J. Muñoz-Sabater *et al.*, “ERA5-Land: a state-of-the-art global reanalysis dataset for land applications,” *Earth Syst. Sci. Data*, vol. 13, no. 9, pp. 4349–4383, 2021, doi: 10.5194/essd-13-4349-2021.

[16] C. Funk *et al.*, “The Climate Hazards Center Infrared Precipitation with Stations, Version 3,” *Sci. Data*, vol. 13, no. 1, Dec. 2026, doi: 10.1038/s41597-026-07096-4.

[17] M. Simard, M. Denbina, C. Marshak, and M. Neumann, “A Global Evaluation of Radar-Derived Digital Elevation Models: SRTM, NASADEM, and GLO-30,” *J. Geophys. Res. Biogeosci.*, vol. 129, no. 11, Nov. 2024, doi: 10.1029/2023JG007672.

[18] P. Rautenstrauch and U. Ohler, “Shortcomings of silhouette in single-cell integration benchmarking,” *Nat. Biotechnol.*, 2025, doi: 10.1038/s41587-025-02743-4.

[19] F. Ros, R. Riad, and S. Guillaume, “PDBI: A partitioning Davies-Bouldin index for clustering evaluation,” *Neurocomputing*, vol. 528, pp. 178–199, 2023, doi: https://doi.org/10.1016/j.neucom.2023.01.043.

[20] A. Rykov, R. C. De Amorim, V. Makarenkov, and B. Mirkin, “Inertia-Based Indices to Determine the Number of Clusters in K-Means: An Experimental Evaluation,” *IEEE Access*, vol. 12, pp. 11761–11773, 2024, doi: 10.1109/ACCESS.2024.3350791.

[21] L. E. Ekemeyong Awong and T. Zielinska, “Comparative Analysis of the Clustering Quality in Self-Organizing Maps for Human Posture Classification,” *Sensors*, vol. 23, no. 18, Sep. 2023, doi: 10.3390/s23187925.

[22] I. Djalovic *et al.*, “Maize and heat stress: Physiological, genetic, and molecular insights,” Mar. 01, 2024, *John Wiley and Sons Inc*. doi: 10.1002/tpg2.20378.

[23] W. Zhu, E. E. Rezaei, Z. Sun, J. Wang, and S. Siebert, “Soil-climate interactions enhance understanding of long-term crop yield stability,” *European Journal of Agronomy*, vol. 161, p. 127386, 2024, doi: https://doi.org/10.1016/j.eja.2024.127386.

[24] X. He *et al.*, “Global patterns and drivers of phosphorus fractions in natural soils,” *Biogeosciences*, vol. 20, no. 19, pp. 4147–4163, Oct. 2023, doi: 10.5194/bg-20-4147-2023.

[25] J. Delfim, A. Moreira, L. A. C. Moraes, J. F. Silva, P. A. M. Moreira, and O. F. Lima Filho, “Soil Phosphorus Availability Impacts Chickpea Production and Nutritional Status in Tropical Soils,” *J. Soil Sci. Plant Nutr.*, vol. 24, no. 2, pp. 3115–3130, 2024, doi: 10.1007/s42729-024-01738-5.