SPC Journal of Environmental Sciences, 8 (1) (2026) 11-17

SPC Journal of Environmental Sciences

Website: www.sciencepubco.com/index.php/JES
https://doi.org/10.14419/dax0xw52
Research paper

Climate Classification of Nigeria Using K-Means Clustering and MODIS Data

Emmanuel M. Menegbo *

Department of Surveying and Geoinformatics, Captain Elechi Amadi Polytechnic, Port Harcourt, Nigeria
**Corresponding author E-mail: nenibarini@yahoo.com*

Received: January 22, 2026, Accepted: March 10, 2026, Published: March 14, 2026

Abstract

Understanding spatiotemporal climate patterns is essential for agricultural planning, water resource management, and biodiversity conservation, particularly in climatically diverse regions like Nigeria. Traditional climate classification systems, such as Köppen-Geiger, rely on long-term monthly averages and stationary weather station data, which often suffer from spatial discontinuities and latency. This study proposes an alternative, data-driven approach to climate zoning by applying unsupervised machine learning (K-means clustering) to high-resolution remote sensing data. Utilizing the Google Earth Engine (GEE) cloud computing platform, we processed Terra and Aqua MODIS time-series data for the year 2020, specifically the Land Surface Temperature (LST) and Normalized Difference Vegetation Index (NDVI) products. By extracting seasonal phenological signatures and thermal profiles across the Nigerian landscape, the study generated a spatially continuous climate classification map. The analysis revealed distinct clusters corresponding to Nigeria's major ecological zones: Mangrove/Swamp Forest (South), representing 25.0% of the total, Rainforest (South-West/South-East) with a percentage of 13.4%, Guinea Savannah accounting for 28.8%, Sudan Savannah at 15.9%, and Sahel Savannah at 17.0%. The results demonstrate that K-means clustering of MODIS data effectively captures the hydro-thermal gradients driving Nigeria's climate variability, offering a robust, scalable methodology for dynamic climate monitoring at a high spatial resolution (1km).

Keywords: *Climate Classification; K-Means Clustering; MODIS; Google Earth Engine; Nigeria; Remote Sensing.*

1. Introduction

Climate classification serves as the foundational framework for understanding the distribution of atmospheric conditions and their influence on terrestrial ecosystems. Historically, systems such as the Köppen-Geiger classification have provided global benchmarks for defining climatic zones based on long-term averages of temperature and precipitation (Köppen, 1936). While these systems are invaluable for broad-scale understanding, they often rely on sparse meteorological station data, leading to spatial interpolation errors, particularly in developing Nations with limited ground-based infrastructure. Furthermore, traditional classifications are static and fail to capture the rapid subannual variability induced by climate change.

Nigeria, situated in West Africa, presents a complex climatic tapestry characterized by a pronounced south-north gradient. The country transitions from the equatorial rainforests and mangroves of the south, through the central Guinea and Sudan savannahs, to the semi-arid Sahel in the extreme north. This climatic diversity is largely driven by the seasonal migration of the Intertropical Convergence Zone (ITCZ) and the movement of the West African Monsoon. Accurate delineation of these zones is critical for Nigeria's economy, which is heavily reliant on rain-fed agriculture.

Recent advancements in remote sensing and cloud computing have revolutionized climatological studies. The Moderate Resolution Imaging Spectroradiometer (MODIS) provides global data products at high temporal (daily to weekly) and moderate spatial (250m to 1km) resolutions. When processed within the Google Earth Engine (GEE) environment, petabytes of satellite imagery can be analysed efficiently. This study applies an unsupervised machine learning algorithm—K-means clustering—to MODIS Land Surface Temperature (LST) and Normalized Difference Vegetation Index (NDVI) time-series data for the year 2020.

The objective is to derive a data-driven, high-resolution climate classification map of Nigeria that captures the region's hydro-thermal dynamics and to validate the methodology against known agro-ecological zones.

Copyright © Emmanuel M. Menegbo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2. Literature Review

2.1. Traditional vs. remote sensing-based classification

Traditional climate classifications are deterministic, based on fixed thresholds of temperature and precipitation. While effective, they often suffer from the "modifiable area unit problem" and lack the granularity to differentiate micro-climates. Remote sensing offers a continuous spatial dataset that overcomes the limitations of point-based weather stations. Previous studies have demonstrated the efficacy of satellite data in climate zoning. For instance, Weiss et al. (2004) utilized MODIS data to map agro-ecological zones in West Africa, highlighting the correlation between vegetation phenology and rainfall patterns.

2.2. The role of MODIS in climate studies

MODIS instruments aboard the Terra and Aqua satellites are pivotal for climate monitoring due to their consistent global coverage. The NDVI product (MOD13Q1/MYD13Q1) serves as a proxy for vegetation health and biomass, which is directly influenced by precipitation. The LST products (MOD11A1/MYD11A1) provide surface thermal measurements, offering insights into energy fluxes and evaporative cooling. Research by Hmamou et al. (2019) showed that integrating LST and NDVI significantly improves the accuracy of aridity assessments compared to using precipitation data alone.

2.3. Unsupervised machine learning in climatology

K-means clustering is a partitioning algorithm that groups data points into k distinct clusters by minimizing the variance within each cluster (MacQueen, 1967). In the context of remote sensing, it is used to classify pixels based on their spectral-temporal signatures without prior labeling. This is particularly useful in regions where ground-truth data is scarce. Studies in the Mekong Basin and the Indian subcontinent have successfully used K-means on time-series MODIS data to identify distinct climatic zones that align with Köppen classifications (Panda et al., 2017). However, application of this methodology specifically within the Nigerian context using high-resolution 2020 data remains underexplored, particularly using the computational power of GEE to handle the temporal dimension effectively.

2.4. The Nigerian climatic context

Nigeria's climate is predominantly tropical, but the north-south rainfall gradient is steep. The southern regions (Guinea Coast) receive annual rainfall exceeding 1500 mm, while the northern Sahelian regions receive between 300 and 600 mm (Odekunle, 2004). The peak of the rainy season shifts from June in the south to August in the north. This temporal shift in rainfall onset and cessation creates distinct vegetation phenologies that satellite sensors can detect. By analyzing the LST-NDVI relationship over a complete annual cycle (2020), the study aims to capture these phenological phases

3. Methodology

3.1. Study area

The study area encompasses the Federal Republic of Nigeria, located between latitudes 4°N and 14°N and longitudes 2°E and 15°E. The country covers approximately 923,768 km². Topographically, Nigeria is relatively flat in the coastal plains and the Sokoto Basin, with highlands in the Jos Plateau and the Obudu Plateau. The climatic zones of interest include:

1) Mangrove/Swamp Forest (South): High humidity, high rainfall.
2) Rainforest (South-West/South-East): Distinct dry seasons but heavy precipitation.
3) Guinea Savannah: Transitional zone with moderate rainfall.
4) Sudan Savannah: Drier, with shorter rainy seasons.
5) Sahel Savannah: Semi-arid to arid, sparse vegetation.



Fig. 1: Map of Nigeria.

Data Sources: Google Earth Engine (GEE) was used for all data processing. GEE hosts the MODIS archive and allows for server-side computation, eliminating the need for high-performance local hardware.

- MODIS Land Surface Temperature (LST): Product MODIS/006/MOD11A1 (Terra) and MYD11A1 (Aqua). Daily LST at 1km resolution. We selected the "Daytime_LST" band. To reduce cloud contamination and noise, 8-day composites were generated by calculating the mean LST for every 8 days.
- MODIS NDVI: Product MODIS/006/MOD13Q1. 16-day composite at 250m resolution. To match the LST scale, the NDVI data were resampled to 1km using a nearest-neighbor approach and aggregated into 8-day composites.
- Temporal Scope: Full calendar year 2020 (January 1st to December 31st), resulting in 46 time steps per variable.

Data Preprocessing

1) Masking and Cleaning: Cloud contamination affects LST and NDVI retrieval. We applied the corresponding pixel reliability layers to mask out low-quality pixels (clouds, shadows). Additionally, values outside the physical range (e.g., LST > 340 K or LST < 270 K; NDVI <- 2000 or NDVI > 10000) were masked.
2) Temporal Aggregation: To reduce data volume and noise while preserving the phenological cycle, daily LST and 16-day NDVI data were aggregated into 8-day intervals. The median value was used to minimize the impact of outliers.
3) Normalization: K-means is sensitive to the magnitude of input variables. LST values (ranging ~300K) have a much larger magnitude than NDVI (ranging -1 to 1). Before clustering, the LST and NDVI time series were normalized using Z-score standardization (subtracting the mean and dividing by the standard deviation) to ensure equal weighting.
4) Stacking: For every pixel, the normalized 8-day LST and 8-day NDVI values were stacked to create a 2-dimensional vector (92 bands: 46 bands LST + 46 bands NDVI).

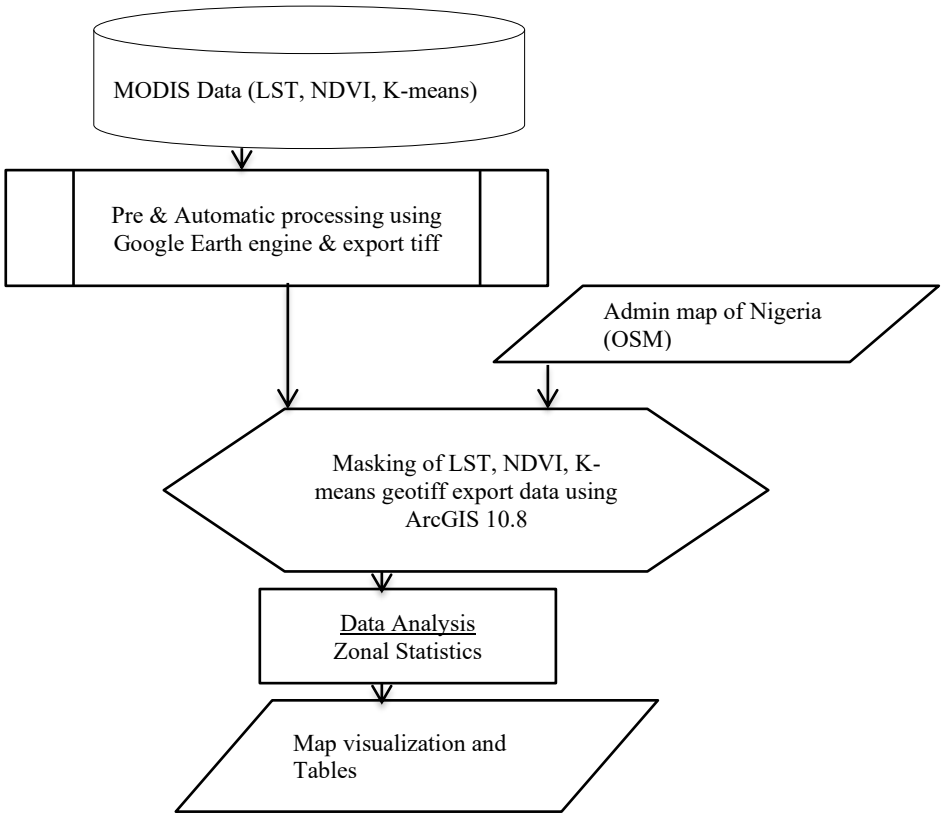


Fig. 2: Methodology Flowchart of Climate Classification Systems MODIS Data.

K-Means Clustering The K-means algorithm was implemented using the ee. Clusterer library in GEE.

1) Training: The algorithm was trained on a stratified random sample of 50,000 points distributed across Nigeria. This sampling strategy ensured representation from all regions, from the densely forested south to the arid north.
2) Selection of k: To determine the optimal number of clusters, the "elbow method" was employed. The Within-Cluster Sum of Squares (WCSS) was calculated for k values ranging from 2 to 10. The analysis revealed an inflection point at k=5, indicating that five clusters provided the best trade-off between variance reduction and computational complexity.
3) Classification: The trained K-means model was applied to the entire Nigerian territory to assign each pixel to a cluster (0 through 4).

Validation Ground truth validation data for "climate" is subjective. However, the results were validated against:

1) Agro-ecological Zone Maps: Comparing cluster distributions with the known boundaries of the Sudan, Guinea, and Derived Savannah zones as defined by the Federal Ministry of Agriculture.
2) Statistical Analysis: Calculating the mean annual temperature for each cluster to assess internal consistency.

4. Results

4.1. Spatial distribution of clusters

The K-means classification successfully delineated five distinct climatic zones across Nigeria in 2020 (Figure 3 & 4 & Table 1). The spatial distribution exhibits a clear latitudinal gradient.

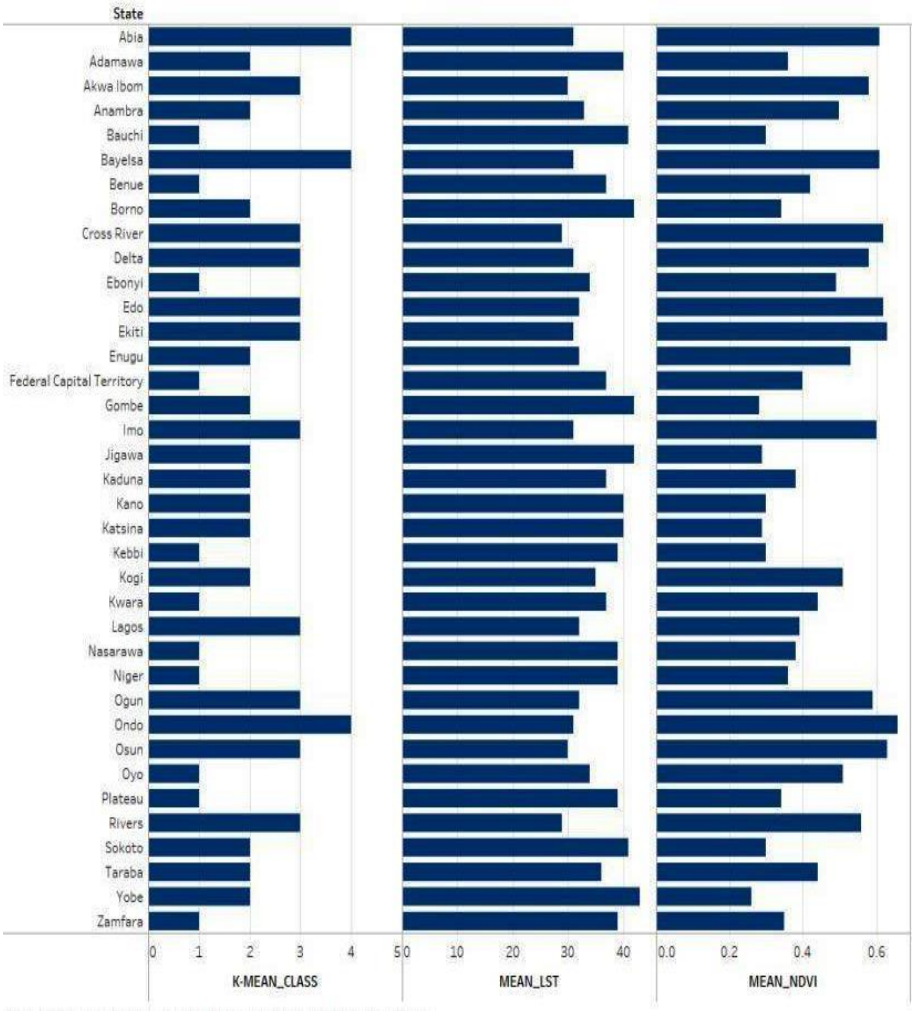


Fig. 3: A Map Showing States' Distributions of K-Mean Class, LST & NDVI.

Table 1: K-Means Classification of Distinct Climatic Zones

Climatic zones class	Area (km²)	Zones class %
0	345699	25.0
1	185099	13.4
2	398185	28.8
3	220412	15.9
4	235021	17.0

- Cluster 0 (Southern Rainforest & Mangrove): This cluster covers the southernmost states (Akwa Ibom, Bayelsa, Rivers, Lagos) and parts of the Cross River region. It is characterized by high NDVI values year-round (mean > 0.6) and relatively stable LST, with minimal seasonal fluctuation. The lack of a distinct dry season is evident in the vegetation persistence.
- Cluster 1 (South-West Agro-Forest): Located primarily in Ogun, Oyo, and Ondo states. While this zone has high annual rainfall, the 2020 data show a slight dip in NDVI during the minor dry season (August), distinguishing it from the evergreen cluster in the deep southeast.
- Cluster 2 (Central Guinea Savannah): This cluster occupies the middle belt of Nigeria (Benue, Kogi, and Niger states). It exhibits a strong seasonality: high NDVI during the rainy season (peak ~0.5) and a pronounced decline during the dry season (drop to ~0.2). LST varies significantly, with higher values recorded in the dry season (Feb-March) and lower values during cloudier rainy months.
- Cluster 3 (Northern Sudan Savannah): Covering areas like Kaduna, Bauchi, and southern Yobe. This zone is characterized by lower baseline NDVI and higher LST. The 2020 data reveal a very short growing season. Vegetation green-up is rapid following the onset of rains, but browning occurs quickly after September.
- Cluster 4 (Sahel Savannah): Located in the northernmost states (Borno, Yobe, Jigawa, Sokoto). This cluster shows the lowest NDVI values throughout the year (mostly < 0.3) and the highest mean LST. The temporal profile indicates a hyper-arid regime with a very narrow window of vegetation activity (July–September).

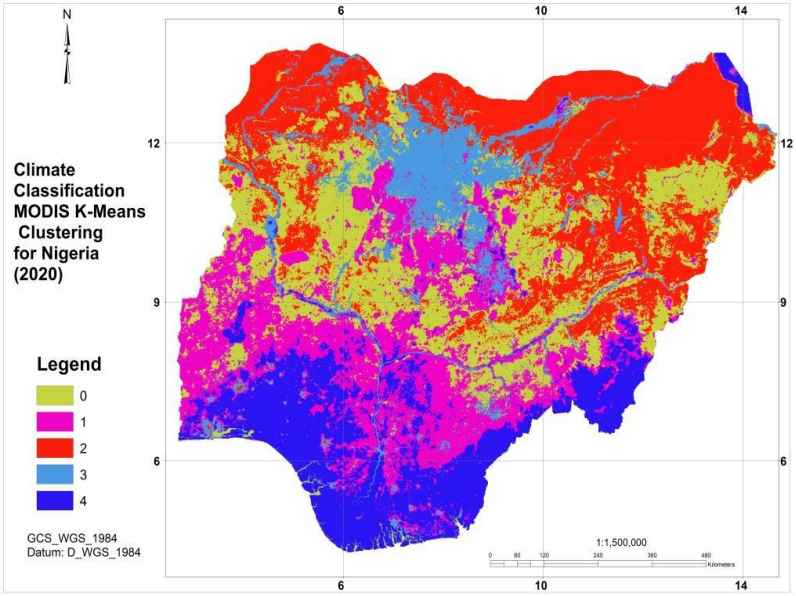


Fig. 4: A Map Showing Spatial Distributions, K-Means Classification Climatic Zones.

Hydro-Thermal Profiles The mean time-series profiles for NDVI and LST for each cluster were extracted (Figures 5 & 6).

- NDVI Dynamics: Clusters 0 and 1 maintained greenness throughout 2020, with slight depressions during the "August break" in the south. Cluster 2 showed a classic unimodal peak centered on August/September. Clusters 3 and 4 displayed suppressed amplitudes, with the Sahel (Cluster 4) barely crossing the 0.3 NDVI threshold, indicating sparse grassland/savannah cover.

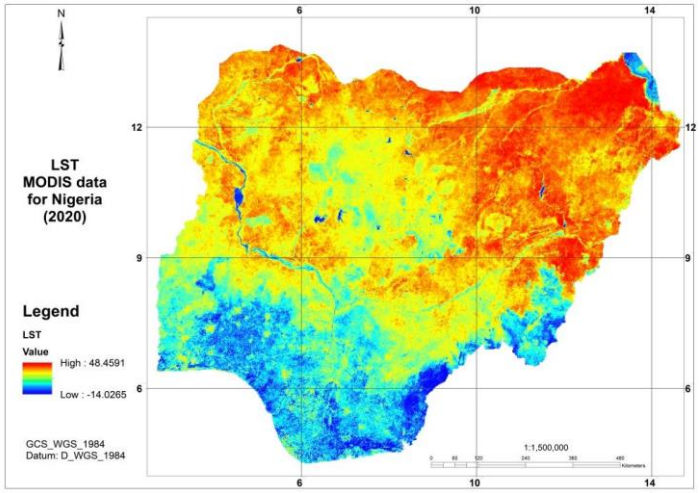


Fig. 5: A Map Showing Spatial Distributions of LST Dynamics Over Nigeria.

LST Dynamics: The LST profiles revealed a lag effect. In the north (Clusters 3 and 4), peak LST occurred in April/May (pre-monsoon) before the onset of cloud cover and rainfall. In the south (Cluster 0), LST remained relatively constant, moderated by persistent cloud cover and evapotranspiration.

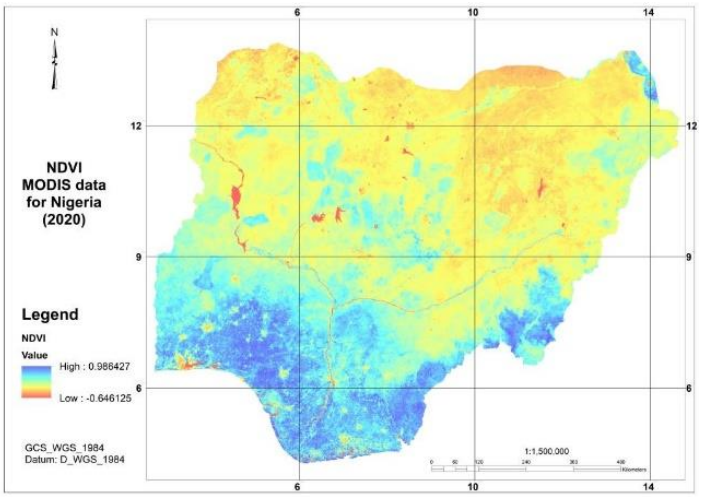


Fig. 6: A Map Showing Spatial Distributions of NDVI Dynamics Over Nigeria.

These ranges align with established rainfall isohyets in Nigeria (Odekunle, 2004), confirming that the unsupervised classification based solely on LST and NDVI effectively captured precipitation gradients.

5. Discussion

5.1. Alignment with established climatic zones

The K-means classification results correspond well with the traditional Köppen classification of Nigeria, yet offer a data-driven refinement.

- Tropical Rainforest (Af): Corresponds to Cluster 0. The consistent LST and high NDVI confirm the absence of a dry month.
- Tropical Savanna (Aw): Corresponds to Clusters 2, 3, and 4. The algorithm successfully differentiated the gradation within the Savanna biome—distinguishing the humid Guinea Savannah (Cluster 2) from the semi-arid Sudan (Cluster 3) and arid Sahel (Cluster 4). The K-means algorithm detected subtle differences in the length of the dry season and the amplitude of vegetation seasonality that might be missed by coarse global climate maps.

Impact of the 2020 Climate Context: The year 2020 was notable for meteorological anomalies globally. In West Africa, rainfall patterns were influenced by specific teleconnections (Atlantic Niño). The data suggest that the northern boundary of the Sudan Savannah (Cluster 3) may have shifted slightly southward compared to long-term averages, as indicated by the reduced NDVI amplitude in the middle-belt regions. This highlights the utility of the proposed methodology for annual dynamic classification, allowing for the monitoring of climate variability rather than relying on static historical averages.

5.2. Advantages of the GEE-MODIS-kmeans approach

1) Spatial Continuity: Unlike station-based data, the resulting map provides continuous coverage, capturing microclimates influenced by topography (e.g., the Jos Plateau).
2) Computational Efficiency: Processing the MODIS time-series for a country the size of Nigeria would be prohibitive on local machines. GEE enabled the extraction of statistical metrics (mean, max, min) for 92 bands across millions of pixels within seconds.
3) Objectivity: The unsupervised nature of K-means eliminates human bias in defining zone boundaries.

Limitations

- Surface vs. Air Temperature: MODIS LST measures the kinetic temperature of the land surface, which can be significantly higher than the near-surface air temperature, especially in arid regions (Cluster 4). This thermal discrepancy might influence the clustering, though the strong correlation with precipitation suggests the dominant signal is still climatic.
- Spatial Resolution: At 1km resolution, mixed pixels are inevitable in heterogeneous landscapes (peri-urban areas). This might smooth sharp transitions between zones.
- No Precipitation Input: While NDVI serves as a proxy for rainfall, direct integration of satellite precipitation data (CHIRPS) into the feature space could potentially enhance cluster accuracy, particularly in distinguishing between years with similar total rainfall but different distribution patterns.

5.3. Implications for agriculture and policy

Accurate climate zoning is vital for Nigeria's agricultural sector. The differentiation between the Sudan and Guinea savannahs (Clusters 2 and 3) is crucial for crop selection.

- Cluster 2 (Guinea Savannah): Suitable for maize, cassava, and yams, which require longer growing seasons.
- Cluster 3 (Sudan Savannah): Better suited for drought-tolerant crops like millet, sorghum, and cowpea.
- Cluster 4 (Sahel Savannah): Primarily pastoral, with limited crop potential.

Policy makers can utilize these high-resolution maps to target climate adaptation strategies. For instance, irrigation projects in the northern zones (Clusters 3 and 4) must be prioritized to mitigate the short growing season identified in the 2020 data

6. Conclusion

This study successfully demonstrated the application of unsupervised machine learning (K-means clustering) for climate classification in Nigeria using MODIS time-series data and Google Earth Engine. By processing a comprehensive dataset of Land Surface Temperature and NDVI for the year 2020, we derived a high-resolution classification map that effectively delineates Nigeria's climatic zones: from the humid rainforests of the south to the arid Sahel of the north.

The results confirm that K-means clustering captures the hydro-thermal gradients that define Nigeria's ecology, aligning closely with established agro-ecological zones. The methodology offers a robust, cost-effective, and dynamic alternative to traditional static climate maps. While the study is limited to a single year (2020), it establishes a reproducible framework for continuous climate monitoring. Future work should extend this temporal analysis to multi-decadal scales to assess long-term climate trends and shifts in agro-ecological boundaries due to climate change.

References

[1] Adebayo, A. A., & Tijani, M. N. (2019). Climate Change and Nigerian Agriculture: Impacts and Adaptation Strategies. University Press.
[2] Hmamou, A., Hilali, A., & Ezzahid, E. (2019). Mapping of arid areas in Morocco using MODIS data and K-means clustering. Journal of Arid Environments, 165, 45-54.
[3] Köppen, W. (1936). Das geographische System der Klimate. In: Handbuch der Klimatologie (Vol. 1). Borntraeger.
[4] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1(14), 281-297.
[5] Odekunle, T. O. (2004). Rainfall and the length of the growing season in Nigeria. International Journal of Climatology, 24(4), 467-479. https://doi.org/10.1002/joc.1012.
[6] Panda, S. S., Panigrahi, N., & Mohanty, S. (2017). Climate classification of the Mekong Basin using K-means clustering of MODIS data. GIScience & Remote Sensing, 54(5), 687-705.

[7] Weiss, E., Marsh, S. E., & Pfirman, E. S. (2004). Application of MODIS NDVI and LST for the identification of agro-ecological zones in the Sahel. International Journal of Remote Sensing, 25(22), 4515-4531.
[8] Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Steinberg, D. (2008). Top 10 algorithms in data mining. Knowledge and Information Systems, 14(1), 1-37. https://doi.org/10.1007/s10115-007-0114-2.
[9] Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C., & Gao, F. (2003). Monitoring vegetation phenology using MODIS. Remote Sensing of Environment, 84(3), 471-475. https://doi.org/10.1016/S0034-4257(02)00135-9.