Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Data-Driven Classification of Rainy Season Regimes in Ethiopia Using Unsupervised Machine Learning

Yosef Tesfaye Balcha

matyyos@gmail.com

Addis Ababa University of Ethiopia

Tadesse Terefe

Addis Ababa University of Ethiopia

Solomon Bililign

Addis Ababa University of Ethiopia

Dejene K. Mengistu

Addis Ababa University of Ethiopia

Asaminew Teshome

Addis Ababa University of Ethiopia

Teferi Demisse

Addis Ababa University of Ethiopia

Research Article

Keywords: Classification, Rainy Seasons, Unsupervised Machine Learning

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11586774/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A DATA-DRIVEN CLASSIFICATION OF RAINY SEASON REGIMES IN ETHIOPIA USING UNSUPERVISED MACHINE LEARNING

Yosef Tesfaye[1], Tadesse Terefe[2], Solomon Bililign [3], Dejene K. Mengistu [4], Asaminew Teshome[5], and Teferi Demisse[6]

Corresponding author: matyyos@gmail.com

[1]matyyos@gmail.com, Addis Ababa University, Addis Ababa, Ethiopia

[2]tadesse.terefe@aau.edu.et, Addis Ababa University, Addis Ababa, Ethiopia

[3]bililignsol@gmail.com, Addis Ababa University, Addis Ababa, Ethiopia

[4]d.mengistu@cgiar.org, Bioversity International, ILRI Campus, Addis Ababa, Ethiopia

[5] asmin2met@gmail.com, Ethiopian Meteorological Institute, Addis Ababa, Ethiopia

[6]T.Demissie@cgiar.org, Alliance of Bioversity International, Addis Ababa, Ethiopia

ABSTRACT

Rainfall seasonality varies substantially due to complex topography and diverse climatic conditions. Existing rainy-season classifications often rely on subjective criteria and predefined climatological boundaries, which may not capture rainfall-season variability. This study presents a data-driven framework for classifying rainy seasons across Ethiopia using unsupervised machine learning techniques. The study analyzed high-resolution merged climate data from the Ethiopian Meteorological Institute. The study employed two widely used clustering methods to classify homogeneous regimes based on long-term monthly rainfall, temperature, and elevation data. The optimal number of clusters was determined using the elbow method. The analysis identified ten distinct regimes, and these regimes exhibit distinct rainfall seasonality. Based on these regimes, we developed a seasonal timetable comprising dry, short rainy, and main rainy seasons. The proposed classification provides a robust statistical framework for understanding the variability of rainfall seasonality and valuable information for climate forecasting and agricultural planning.

Keywords: Classification; Rainy Seasons; Unsupervised Machine Learning.

INTRODUCTION

Rainfall seasonality is one of the most important characteristics of regional climate because it strongly influences agriculture, water resources, ecosystems, and disaster risk management (Nicholson 2018; Palmer et al. 2023). In countries like Ethiopia, where livelihoods depend largely on rain-fed agriculture, understanding the spatial and temporal distribution of rainfall seasons is essential for improving climate services and supporting sustainable development. Ethiopia has remarkable climatic diversity resulting from its complex topography, pronounced elevation gradients, and the interaction of regional and large-scale atmospheric circulation systems (Berhanu et al. 2014; Meron et al. 2021). Consequently, rainfall varies considerably across the country in terms of onset, cessation, duration, intensity, and seasonal distribution

(Segele and Lamb 2005). This substantial spatial and temporal variability has critical implications for all socioeconomic activities and food security (Adriaan et al. 2020). Several attempts have been made to classify Ethiopia into distinct rainfall zones using climatological observations, seasonality indices, and expert-based regionalization approaches (Markos et al. 2022; Alhamshry et al. 2020; Korecha and Sorteberg 2013; Tsidu 2012; Diro et al. 2011b; Haile and Yarotskaya 1987). Although these studies have substantially improved the understanding of regional climate, differences in classification criteria and methodological approaches have produced inconsistent rainfall regime maps and seasonal definitions. Most existing classifications rely on predefined climatological boundaries and expert interpretation, which may not fully capture the natural spatial classification of rainfall regimes across lowlands and highlands of Ethiopia. These inconsistencies limit the development of a unified and objective rainfall classification framework.

Recent advances in machine learning have provided powerful approaches for objectively identifying hidden structures within large climate datasets. In particular, unsupervised machine learning techniques identify natural groupings without predefined class labels, making them especially suitable for climate regionalization and rainfall classification (Xiao et al. 2025; Hussain et al. 2023; Wang and Biljecki 2022; Huber and Love 2018). Among these methods, Ward's hierarchical clustering (Ward 1963) and K-means clustering (MacQueen 1967; Hartigan and Wong 1979) have been widely applied to classify homogeneous climate regions because of their robustness, computational efficiency, and ability to identify coherent spatial patterns. Despite these methodological advances, few studies have applied objective unsupervised machine learning techniques to classify rainy-season regimes using long-term, high-resolution gridded climate data covering other countries (Urgiles et al. 2021). Consequently, the spatial organization and seasonal characteristics of rainfall regimes remain incompletely understood in Ethiopia. Addressing this knowledge gap is vital for improving seasonal forecasting, agricultural planning, climate risk assessment, and climate change adaptation. To address these limitations, this study aimed at developing an objective, data-driven framework that classifies Ethiopia’s rainy seasons into distinct types. Utilizing high-resolution gridded rainfall data from the Ethiopian Meteorology Institute for the period of 1981–2022, Ward's hierarchical clustering and K-means clustering algorithms were employed to identify homogeneous rainfall regimes, while the Adjusted Rand Index was used to evaluate the consistency of the classification. Ultimately, this framework provides an objective statistical classification that defines ten homogeneous rainy season regimes and establishes a revised seasonal timetable. These study findings could contribute directly to improved climate services, drought monitoring, water resource management, and evidence-based climate adaptation strategies. Therefore, to fulfil this general objective, the main specific objectives are to classify rainy seasons based on climate characteristics and to define homogeneous rainfall regimes to classify rainy season regimes over Ethiopia using selected clustering analysis techniques. To achieve these, the following research questions have been identified: “How can rainy seasons be statistically classified into distinct types based on their climatic characteristics using unsupervised machine learning (USML) over Ethiopia?” and “How can seasons be classified and defined based on climate characteristics in Ethiopia, and do the climate characteristics indicate the seasonal rainfall pattern?”

METHODOLOGY

Description of the study area and Climate data

Ethiopia, located in the Horn of Africa, is characterized by a highly diverse geographical landscape that includes rugged mountains, flat-topped plateaus, deep gorges, and river valleys (ESDP 2023). As illustrated in Figure 1, the country features complex topography across its northern and central highlands, bisected from the northeast to the southwest by the lowlands of the East African Rift Valley (Alhamshry et al. 2020). Consequently, the country experiences diverse climate regimes, ranging from humid conditions across the western highlands (Berhanu et al. 2014) to arid zones throughout the eastern lowlands of the Rift Valley (Viste and Sorteberg 2013; Meron et al. 2021). To effectively analyze distinct rainfall patterns monthly rainfall, monthly mean temperature, and digital elevation model (DEM) data were utilized. The dataset consists of high-resolution (4 km x 4 km) gridded data encompassing more than 66,000 points, obtained from the EMI for the temporal scale spanning 1981 to 2022.

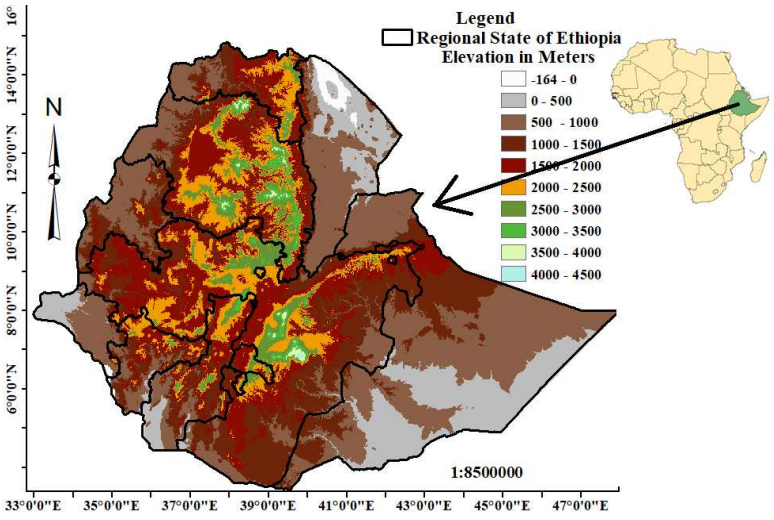


Figure 1. Topography and Location of Study Area.

Data-Driven Unsupervised Machine Learning Classification Methods

Unsupervised machine learning algorithms offer strong pathways for discovering underlying spatial and temporal patterns in complex climate datasets without relying on predefined categories or extensive ground-truth labels (Wang and Biljecki 2022). This data-driven approach is particularly valuable in regional climatology for identifying distinct rainfall regimes. In this study, cluster analysis was employed to classify rainy seasons by grouping spatial and temporal rainfall features from over 42-year climate dataset. Quality control and outlier analysis were performed by standardizing the mean square statistics to assess their statistical significance. To detect and correct climate change points and potential data discontinuities, the

Standard Normal Homogeneity Test (SNHT) was applied using a hierarchical procedure for multiple breaks (WMO 2017). To group homogeneous rainfall regimes, this study utilized two complementary unsupervised clustering algorithms: Ward's hierarchical clustering and K-means clustering (Ward 1963; MacQueen 1967; Hartigan and Wong 1979). Ward's agglomerative method was first used to find dense, spherical clusters by minimizing the internal variance across groups. The objective function for Ward's method is defined as:

$$w = \sum_{c=1}^{k} \sum_{j=1}^{m} \sum_{i=1}^{N_c} (y^c ij - y^c .j)\text{^}2$$

Where W is the objective function of Ward's algorithm; k is the total number of clusters; m is the number of attribute features; Nc is the number of feature vectors (data points) assigned to cluster c; ycij is the rescaled value of j-th attribute of the i-th feature vector assigned to cluster c and yc•j denotes the mean value of a j-th attribute for cluster c (i.e mean value of the j-th feature across all data points assigned to cluster c)(Ward 1963; Marzieh et al. 2020). Following the initial hierarchical grouping, K-means clustering was applied to partition the gridded data into distinct, non-overlapping rainfall zones based on monthly rainfall variables (Hartigan and Wong 1979). For high-resolution gridded datasets, this automated partitioning generates regions with stable spatial boundaries (Abreu et al. 2020b). The K-means algorithm was executed via the following five steps (Zhang et al. 2016):

- Determine the optimal number of clusters (k) using the Elbow method.
- Randomly select k points as initial cluster centroids.
- Assign each grid point's time series to the nearest centroid based on Euclidean distance.
- Recalculate the mean of all data points within each cluster to update the centroids.
- Repeat the assignment and update phases iteratively until the centroids stabilize.

To evaluate the mathematical validity of the cluster configurations, the total Within-Cluster Sum of Squares (WSS) was minimized (Huber and Love 2018). WSS computes the sum of squared distances between each grid point's time series and its corresponding cluster centroid:

$$\text{WSS} = \sum_{\text{c=1}}^{\text{k}} \sum_{\text{xi}\varepsilon\text{ck}} (\text{xi} - \mu\text{c})^2$$

Where WSS is the within-cluster sum of squared errors; k is the total number of clusters; x_i is the i-th data point; $x_i \in c$ indicates that x_i belongs to cluster c; μ_c is the centroid (mean) of cluster c; and $(x_i - \mu_c)^2$ represents the squared Euclidean distance between the data point and the centroid of its assigned cluster. All data processing, spatial analyses, and visualizations were executed using the Climate Data Tool (CDT), Python and ArcGIS 10.8 (Chung et al. 2019). Finally, the Adjusted Rand Index (ARI) was calculated to cross-validate and quantify the spatial similarity between the cluster boundaries generated by Ward's method and

the K-means algorithm (Jose et al., 2022). ARI values range from 0 to 1, where a value of 1 signifies a perfect clustering match and a value close to 0 indicates a random spatial assignment. In addition, box-and-whisker plots provided a standardized, robust method for evaluating the statistical distribution and characteristics of rainfall across each regime. To establish a standardized baseline, this study applied the criteria utilized by Yosef et al. (2024) and the FAO (1978), defining wet season onset as the point where monthly rainfall either exceeds half of the potential evapotranspiration or continuously exceeds a 60 mm threshold.

RESULT

Unsupervised Machine Learning Classification of Rainy Seasons

In case of high spatial and temporal variability of Ethiopian rainfall, it would be crucial to delineate boundaries and define the country into homogenous regimes, as the optimal zonation would be valuable for agriculture, hydrologic modeling, prediction, ecological, and climate classification (Zhang, et al. 2016). Thus, based on the two unsupervised clustering techniques applied over the study domain, ten distinct rainy season regimes (RSRs) were defined over Ethiopia. As demonstrated in Figure 2, these are grouped as clusters 1 through 10 for both Ward's method and K-means method. The results strongly indicate that optimal data stratification is achieved when 10 distinct regimes are used to capture the spatial distribution of rainfall. The regional grouping is strongly driven by Ethiopia's complex topography and seasonal monthly rainfall cycles, where sharp variations in altitude generate highly localized rainfall patterns. By combining Ward's and K-means clustering with monthly climatological rainfall, mean temperature, and DEM data, the ten identified Ethiopian RSRs were designated as: RSR_1, RSR_2, RSR_3, RSR_4, RSR_5, RSR_6, RSR_7, RSR_8, RSR_9, and RSR_10 regimes.

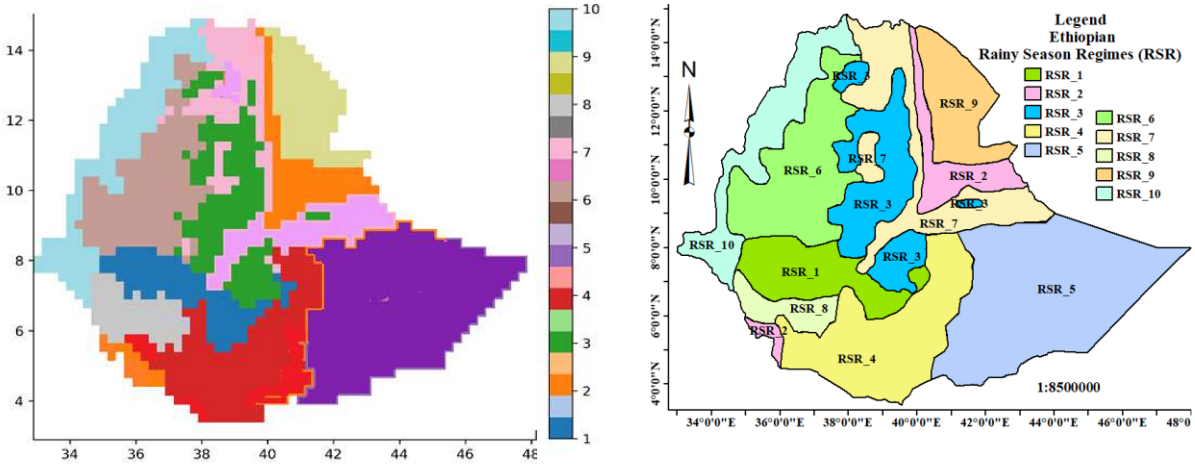


Figure 2. Homogeneous Classified Rainy Season Regimes over Ethiopia (Left: Result of Methods, Right: the shape file of classified regimes)

To evaluate and validate the spatial consistency between Ward's and K-means clustering through different months, the Adjusted Rand Index (ARI) was computed monthly (Figure 3). Adjusted Rand Index values vary throughout the year, reflecting seasonal variations in the structural agreement between the two algorithms. High ARI values (>= 0.8) in January, February, March, May, June, and December indicate a near-perfect match between the clusters generated by Ward's and K-means methods. This high consistency is likely due to well-separated, highly distinct data structures during these months. Conversely, visible drops to 0.6 and 0.7 in April, July, and August, September, October, and November, respectively indicate a slightly strong match.

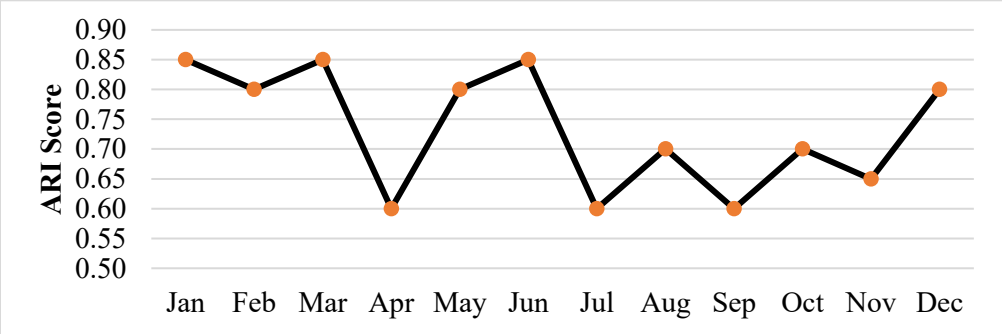


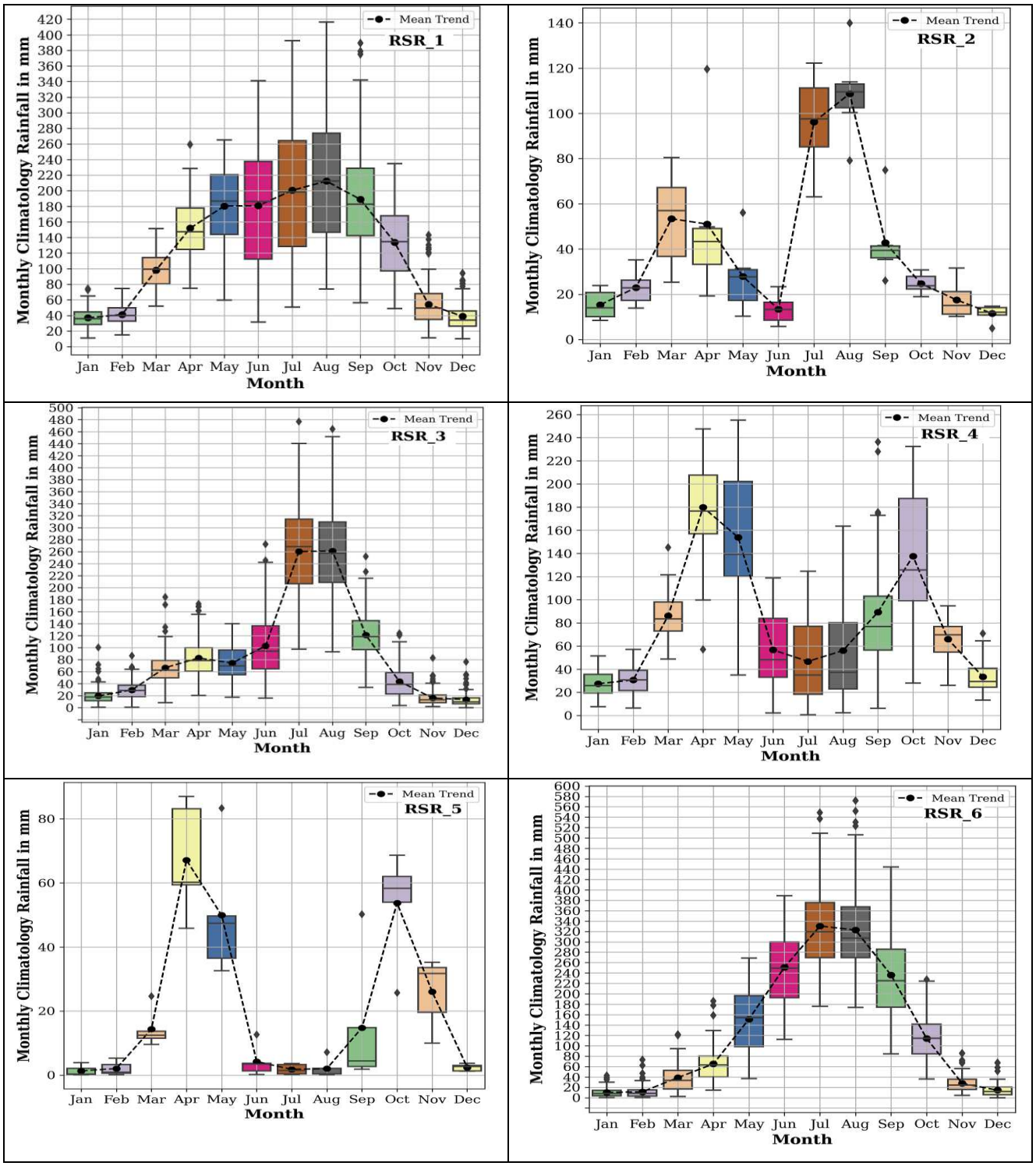
Figure 3. Adjusted Rand Index between Ward's hierarchical and K-means clustering per month

Descriptions of the Classified Rainy Season Regimes

The defined rainy season regimes (RSR_1 through RSR_10) vary geographically due to the country's intense climatic diversity. The monthly climatological rainfall cycles for each regime are represented via box-and-whisker plots in Figure 4. Comparing median horizontal lines across clusters reveals which regimes are wet versus dry, while the height of the boxes indicates rainfall stability. A tall box suggests high rainfall variability throughout the year (highly seasonal regimes), whereas a short box suggests a more consistent, uniform rainfall pattern. Anything beyond these whiskers is a statistical outlier, represented by black stars and the mean trend by a dashed black line. Any dots appearing outside the whiskers are extreme events, that is, unusually wet or dry months for that specific regime, which are critical for identifying climate anomalies in Ethiopia. Clusters in the Ethiopian highlands (RSR_1, RSR_3, and RSR_6) often show higher rainfall totals and larger boxes due to orographic effects, whereas lowland clusters display smaller, tighter boxes (Figure 4). RSR_1 comprised 4,815 (7%) of grid points. This regime is characterized by a prolonged unimodal rainy season extending continuously from March to November, with monthly rainfall amount ranging from 60 to 400 mm, with no distinct mid-season transition period. The dry season is confined to the December-January-February (DJF) winter months. The landscape is topographically diverse, with elevations ranging from 900 to 3,150 meters above mean sea level (m.a.s.l.), and it receives an annual rainfall of 750 and 2,400 mm. RSR_2 comprises 3,196 (5%) of the grid points and is a lowland regime situated at elevations

ranging from 0 to 1,350 m.a.s.l. The RSR_2 lowlands are highly arid, receiving less than 700 mm of annual rainfall. This regime is characterized by a minor, short rainy season from March to April, with monthly rainfall amounts ranging from 20 to 80 mm, followed by its primary, short-duration rainy season from July to August, with monthly rainfall amounts of 60 to 120 mm. The remaining months of the year constitute a prolonged dry season. RSR_3 covers the northern and central highlands and comprises 7,669 (12%) of the grid points and falls at elevations ranging from 1,800 to 4,500 m.a.s.l and receives an average annual rainfall of 490 to 1,900 mm. It is characterized by a well-defined bimodal rainfall pattern, with a short rainy season extending from March to May, during which monthly rainfall amounts range from 60 to 180 mm, followed by the primary rainy season from June to September (JJAS), with monthly rainfall amounts ranging from 60 to 450 mm. A cool, dry transitional period occurs from October to November, followed by the main dry season from December to February.

RSR_4 encompasses the southern portion of the country and comprising 8,862 (13%) of the grid points; this regime occurs at elevations ranging from 450 to 1,800 m.a.s.l and receives 290 to 1,650 mm of annual rainfall. It exhibits a distinct bimodal rainfall pattern with two rainy seasons: the primary rainy season during March to May (MAM), with monthly rainfall amounts ranging from 60 to 250 mm and a secondary rainy season during September to November (SON), with monthly received rainfall ranging from 60 to 230 mm. It also experiences two dry periods: the primary dry season during December to February (DJF) and a secondary, relatively dry period during June to August (JJA), when total seasonal rainfall is less than 150 mm. RSR_5 is the feature of the southeastern most corner of the country and comprises 14,316 (22%) of the grid points extending from low-to-mid-elevation regime in the range of 0 to 1,350 m.a.s.l. It exhibits a bimodal rainfall pattern, with the primary rainy season occurring during April and May, with monthly rainfall amounts ranging from 40 to 90 mm, followed by a weaker secondary rainy season during October and November, with monthly rainfall amounts ranging from 20 to 70 mm. Prolonged dry seasons occur during both the December–March (DJFM) and June–September (JJAS) periods. Overall, this regime is highly arid, receiving less than 400 mm of annual rainfall.


RSR_1
RSR_2
RSR_3
RSR_4
RSR_5
RSR_6
Mean Trend
Monthly Climatology Rainfall in mm
Month
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

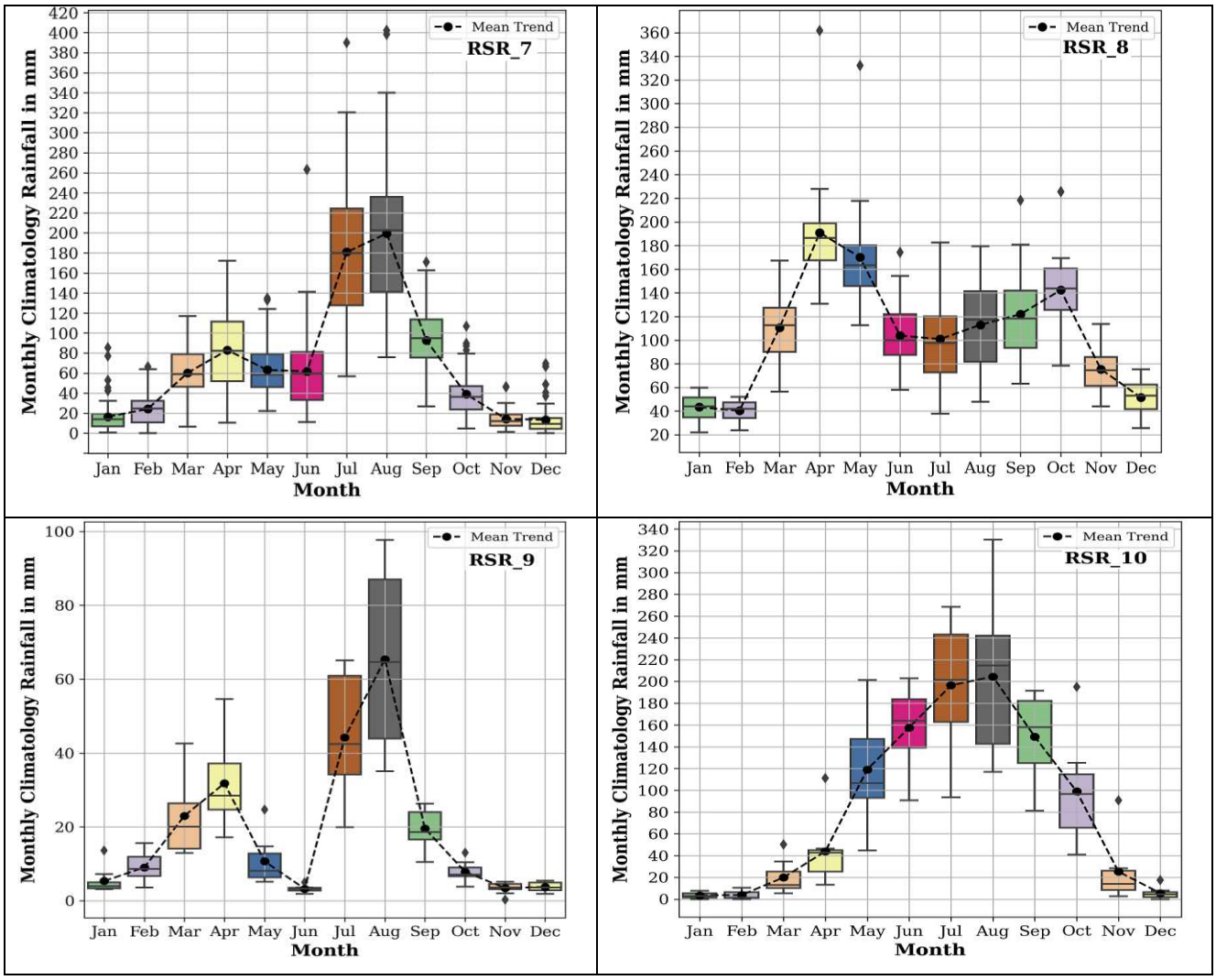


Figure 4. Boxplots of Monthly Climatology Rainfall for Defined Rainy Season Regimes; where the mean is denoted by a black dot, the line inside the box is the median rainfall.

As shown by Figure 2, Table 1, and Figure 4, RSR_6, comprising 7,925 (12%) of the grid points, is characterized by highly varied terrain, with elevations ranging from 900 to 3,150 m.a.s.l. It features a prolonged unimodal rainy season extending from April to October, with monthly received rainfall ranging from 60 to 550 mm. The distinct dry season extends from November to March. This regime receives abundant rainfall, with annual totals ranging from 800 to 2,700 mm. On the other hand, RSR_7 covers the northern region, west of the Afar region and the eastern highlands of the country. It comprises 7,797 (12%) of the grid points in the elevation ranges of 900 to 1,800 m.a.s.l and receives an average annual rainfall of 400 to 1,700 mm. Furthermore, it exhibits a bimodal rainfall pattern, with a short rainy season extending from March to May, during which monthly rainfall amounts range from 60 to 180 mm, followed by the primary rainy season from June to September (JJAS), with monthly rainfall amounts ranging from 60 to 350 mm. A cool, dry transitional period occurs from October to November, followed by the main dry season from December to

February. RSR_8, featuring southwestern Ethiopia, covers 1,577 grid points (3% of the total) and occurs at elevations ranging from 450 to 2,700 m.a.s.l. The regime receives between 800 and 2,050 mm of annual rainfall and exhibits a bimodal rainfall pattern. The primary rainy season of the regime occurs from March to May (MAM), with monthly rainfall ranging from 60 to 230 mm, while a prolonged secondary rainy season extends from June to November (JJASON), receiving 60 to 180 mm of rainfall per month. The main dry season occurs during December to February (DJF). RSR_9 encompasses 4,602 grid points (7% of the total) and is predominantly associated with lowland areas ranging from164 meters below sea level to 900 m.a.s.l. This highly arid regime receives less than 400 mm of annual rainfall. It is characterized by a minor rainy season during March and April, with monthly rainfall amounts of 20 to 50 mm, followed by a primary rainy season from July to August, during which monthly rainfall ranges from 30 to 90 mm. The remainder of the year is dominated by prolonged dry conditions. RSR_10 covers 5,241 grid points (8% of the total) and extends across parts of western and northwestern Ethiopia along the country's border regions. Occurring at elevations between 450 and 1,800 m.a.s.l., this regime receives substantial annual rainfall ranging from 650 to 1,550 mm. It is characterized by a prolonged unimodal rainy season from May to October, with monthly rainfall totals of 60 to 350 mm, followed by a distinct dry season extending from November to April.

Seasonal Timetable for Rainy Season Regimes

When aggregated across the country, the results of the unsupervised machine learning algorithm validate the localized seasonal timelines (calendar). Accordingly, several variations in monthly classifications were calculated and summarized in Figure 4 and Table 1, collectively referred to as the “seasonal timetable.”

Dry Season

The RSR_1 and RSR_8 regimes experience a dry period crossing December to February(DJF), which is recorded as less than 60 mm of rainfall, while the RSR_6 regime features an extended dry period spanning November to March, with less than 40 mm of rainfall. For both RSR_3 and RSR_7 regimes, a shorter dry window is manifested during the period of October to February. For RSR_5, distinct dry periods occur during December to March and June to September (locally designated as Xagaa, aligning with Zerihun and Abel 2020). The longest extended dry season was identified in the RSR_2, RSR_9 (locally referred to as Gilal) and RSR_10 lowlands. Thus, over RSR_2 during DJF, May to June, September to November and over RSR_10 during DJF, March, April and November less than 40 mm. RSR_5 during DJFM, JJAS and over RSR_9 during DJF, May, June and SON 20 mm long term rainfall observed. While RSR_4 experiences its primary dry season during DJF, it encounters a secondary dry spell from June to August (JJA) where cumulative seasonal rainfall remains strictly below 150 mm. (Figure 4 and Table 1).

Short Rainy Season

A distinct short with small rainfall amount rainfall season occurs during SON for RSR_4 and October and November for RSR_5 (locally known as Deyr) rainfall amount 60-250 and 40-90 respectively. For the RSR_3 highland regime, the short rainy season is confined to spring (MAM), representing traditional Belg rain. In the RSR_2 and RSR_9 lowlands, it occurs briefly in March and April (locally termed Segum). Conversely, western and southwestern mono-modal rainfall regimes (RSR_1, RSR_6, and RSR_10) do not experience a distinct short wet season. A distinct short rainy season occurs during MAM for RSR_7 and RSR_8 rainfall amount 60-180 and 60-240mm, respectively.

Table 1. Rainfall Season Timetable as per classified Rainy Season Regimes

Key: Dry Season | Wet Season | Short Wet Season

Regime	Winter Season			Spring Season			Summer Season			Autumn/Tseday Season			Annual Rainfall in mm
	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	
RSR_1				60-400 mm									750-2400
RSR_2				40-80 mm				60-120 mm					290-700
RSR_3				60-180 mm			60-450 mm						490-1900
RSR_4				60-250 mm						60-230 mm			290-1650
RSR_5					40-90 mm						20-70 mm		170-350
RSR_6					60-550 mm								800-2700
RSR_7				60-180 mm			60-350 mm						400-1700
RSR_8				60-240 mm			60-180 mm						800-2050
RSR_9				20-50 mm				30-90 mm					130-400
RSR_10						60-350 mm							650-1550

Wet Season

RSR_1 has an even longer wet season, beginning in March and extending through November. In addition, RSR_6 and RSR_10 have long wet seasons that start in April and May and end in October, respectively. This extended wet-season pattern in southwestern Ethiopia is consistent with the findings of Abiy et al. (2014). The main wet season for RSR_3 occurs from June to September, with rainfall amounts ranging from 60 to 450 mm. In the lowlands of RSR_2 and RSR_9, the primary wet season is limited to July and August, a period locally known as *Karma*. The dominant wet season for RSR_4 occurs during March to May, with rainfall amounts ranging from 60 to 250 mm, whereas RSR_5 experiences rainfall primarily during April and May, with amounts ranging from 40 to 90 mm; this period is locally referred to as *Gu*. The main wet seasons for RSR_7 and RSR_8 are concentrated from June to September and June to November, respectively, with rainfall amounts ranging from 60 to 350 mm and 60 to 180 mm, respectively (Figure 4).

Tseday Season (autumn)

The classified regimes, the SON period consistently serves as main Tseday transitional window. The long continuous Tseday transition is observed in RSR_1, RSR_4, RSR_5, RSR_6, RSR_8, and RSR_10, lasting from September through the end of November.

Regime Structural Validity and Wet Day Analysis

Overall, the structural validity of these rainfall regimes is well supported by the boxplot presentation of the statistical results. The RSR_1, RSR_6, and RSR_10 regimes show clear outliers across multiple seasons and exhibit highly skewed mean trends. The median values for RSR_1 and RSR_6 are relatively high, and their rainfall distributions are among the most uniform. In contrast, the lowest median rainfall values are observed in the arid RSR_2, RSR_5, RSR_7, and RSR_9 clusters. Specifically, the lowest cluster median values occur in April and September for the RSR_5 regimes and in March and July for the RSR_9 regimes.

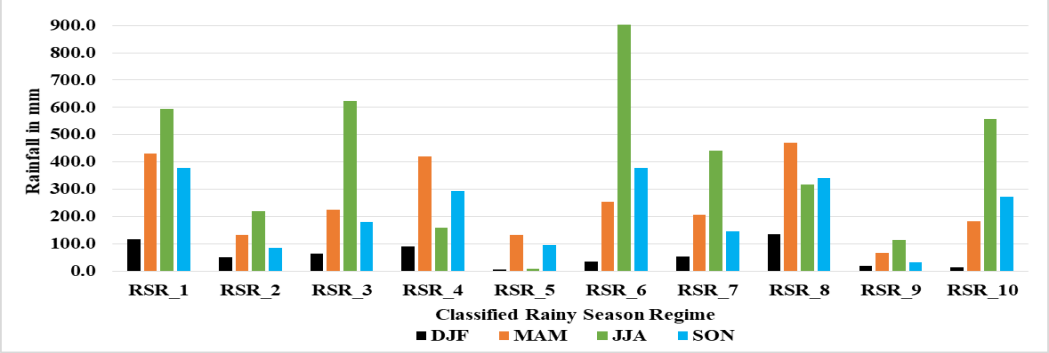


Figure 5. Long-Term Seasonal Rainfall of the Ten Classified RSRs.

A key observation across most regimes is that the upper whiskers are much longer than the lower whiskers, with a large number of outliers extending toward the upper limits. This indicates that the monthly rainfall distribution across Ethiopia is predominantly right-skewed (Figure 4). Throughout MAM, JJA, and SON, RSR_1, RSR_3, RSR_6, RSR_8, and RSR_10 receive more rainfall than the other regimes. During JJA, RSR_6 has the highest overall seasonal average, whereas during MAM, RSR_8 has the highest average (Figure 5).

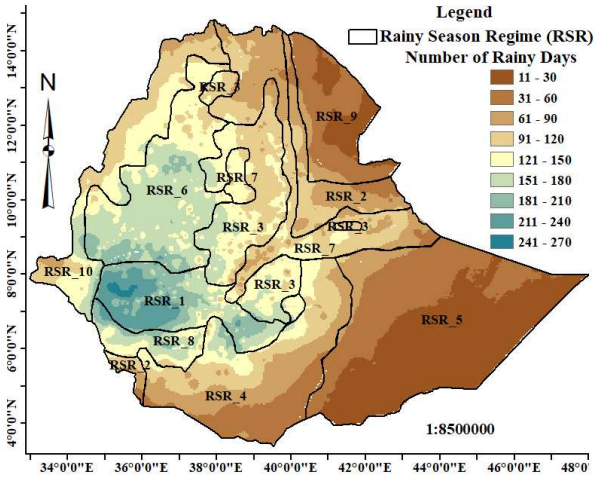


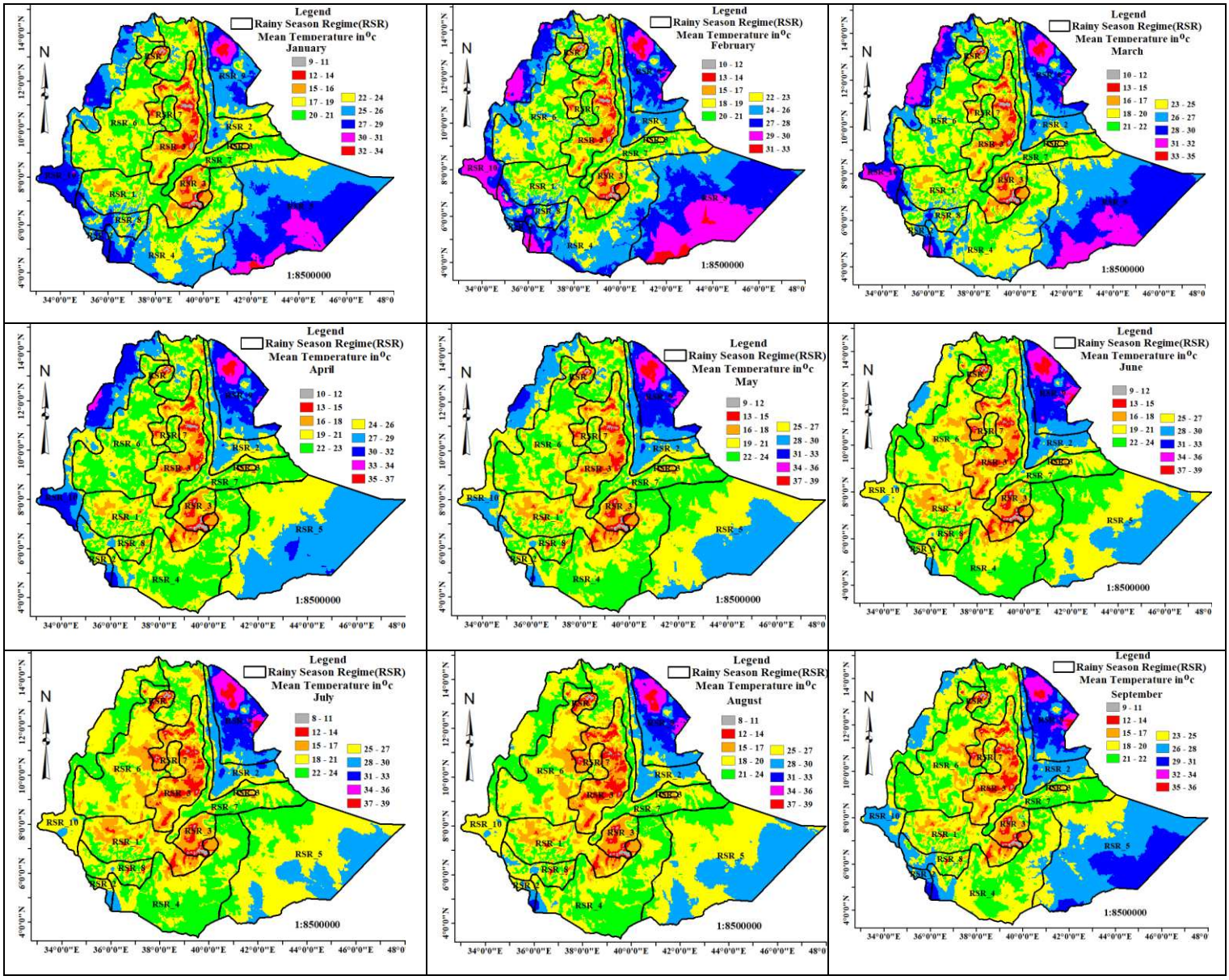
Figure 6. Long-Term Rainy Days for Classified Rainfall Regimes

Meteorologically, a rainy day is defined as any day with rainfall exceeding 1.0 mm. The number of wet days varies across the regimes. RSR_1 and RSR_8 range from 121 to 270 days and from 121 to 180 days within the rainy seasons, respectively. RSR_10 ranges from 61 to 150 days. As shown in Figure 6, the mean number of wet days in the high-yielding area ranges from 121 to 210 days for RSR_6. For the RSR_3 and RSR_7 regimes, the mean number of wet days' ranges from 91 to 180 days and from 61 to 150 days, respectively. In the RSR_4 regimes, the mean number of wet days' ranges from 61 to 150 days. The RSR_2, RSR_5, and RSR_9 regimes have mean numbers of wet days below 90 days.

Monthly Long-Term Mean Temperature for Classified Rainy Season Regimes

Climate change is causing significant increases in temperature variability across the different regimes. Based on the data, long-term monthly mean temperatures show considerable variation across the ten regimes (Figure 7). In the RSR_1 regime, mean temperatures range from 14°C to 28°C during December–February (DJF), 13°C to 29°C during March–May (MAM), 12°C to 27°C during June–August (JJA), and 12°C to 28°C during September–November (SON), making SON the coolest and MAM the warmest season. In RSR_2, mean temperatures range from 21°C to 31°C during DJF and SON, 22°C to 32°C during MAM, and 25°C to 33°C during JJA, making DJF and SON the coolest seasons. The highest seasonal temperatures are observed in the lowlands of RSR_2, south of RSR_4, and in RSR_5, RSR_8, RSR_9, and RSR_10, with temperatures reaching 25°C to 33°C, 20°C to 30°C, 24°C to 33°C, 19°C to 30°C, 25°C to 39°C, and 26°C to 34°C, respectively. The ranges of minimum mean temperatures are most pronounced in the RSR_1, RSR_3, and RSR_6 regimes, varying from 12°C to 15°C, 8°C to 10°C, and 14°C to 19°C, respectively. In the RSR_9 regimes, the lowest mean temperatures occur during DJF, whereas the highest temperatures are recorded during May to August. Conversely, the RSR_5 regime experiences its lowest temperatures during JJSON and its highest during December to March. Finally, temperatures in the RSR_4 and RSR_7 regimes range

from 16°C to 30°C and 16°C to 28°C, respectively. The lowest mean temperatures in the RSR_4 and RSR_3 regimes are observed during May to July and October to January, respectively, whereas the highest temperatures occur during May and the January to March season, respectively (Figure 7).



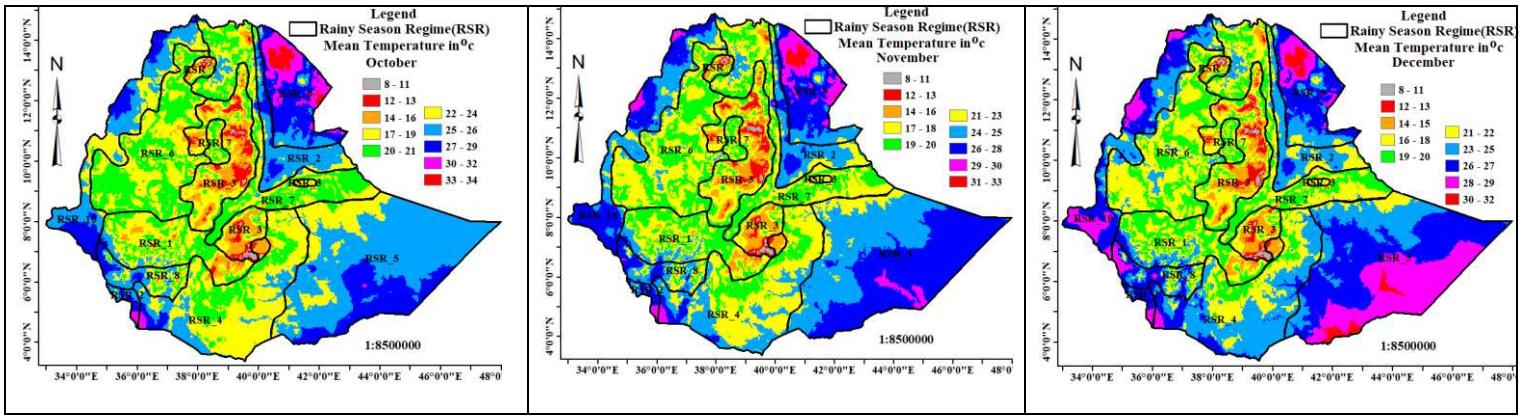


Figure 7. Long-term mean temperature in °c for Classified Rainfall Regimes

DISCUSSION

Rainy Season Regimes over Ethiopia

Clustering techniques in unsupervised machine learning, specifically Ward's hierarchical and K-means algorithms, proved to be highly efficient grouping mechanisms for processing large volumes of spatial climate data within a rapid computational timeframe. Understanding these newly classified and defined rainy season regimes provides vital insights into temporal climate variability across the regimes. The elbow method was evaluated to determine the appropriate cluster numbers, alongside traditional diagnostic criteria. Because this study isolated monthly climatology rainfall, mean temperature, and a digital elevation model as the primary clustering variables, both Ward's and K-means algorithms converged on highly consistent spatial structures. The unsupervised cluster analysis successfully identified ten homogeneous rainy season regimes across Ethiopia (Figure 2). These ten distinct rainfall regimes were chosen because their localized rainfall cycles reflect geographic realities on the ground, stratifying the study area into clusters 1 through 10. While a similar clustering approach was applied by Markos et al. (2022), the spatial boundaries and seasonal timelines recognized in this study diverge regarding the specific onset, cessation month, internal rainfall distributions, and regionalization. The structural definitions of the regimes illustrate high spatial agreement across both algorithms, diverging only slightly in minor localized pockets, which were successfully resolved by incorporating long-term monthly rainfall climatological cycles. In contrast, early foundational studies by Haile and Yarotskaya (1987) and Tsidu (2012) classified Ethiopian climate zones into different macro-scale rainfall regimes based on broad inter-annual patterns. Haile and Yarotskaya (1987) provided a rigid seasonal timeline dividing the year into long rains (*Kiremt*; JJAS), Bega season (ONDJ), and short rains (*Belg*; FMAM). Similarly, Kassahun (1999) characterized the JJAS period as the dominant rainy season for the entire country, excluding only the southern and southeastern lowlands. However, the high-resolution empirical timeline produced by this study demonstrates that the newly classified RSR_1 through RSR_10 regimes exhibit complex localized rainfall characteristics that depart significantly from these historical

macro-scale generalizations. In addition, Gissila et al. (2004) grouped Ethiopia into four clusters based on seasonal rainfall cycles and homogeneity of interannual rainfall at 24 rain gauges. In contrast, Diro et al. (2009) delineated five homogeneous summer rainfall regions (June–September) and six spring zones (February–May) based on data from 33 rainfall stations and the interannual variation of seasonal rainfall. Furthermore, Korecha and Sorteberg (2013) indicated that the region divides into eight homogenous zones depending on the main systems of atmospheric circulation and rain production. As noted by Alhamshry et al., (2020), homogenous rainfall zones over Ethiopia obtained from correlation analyses using monthly rainfall variability, with summer rainfall peaking from July to August and spring rainfall peaking from April to May. Despite following a methodology similar to the latter study, both this study and Markos et al. (2022), produced different results. Therefore, the classification of Ethiopia into homogenous zones can vary from one study to another, due to the high variability of large-scale forces producing rainfall across the country.

Seasonal variations and structural drops are typically caused by transitions in data structure, increased climate noise, or the emergence of non-spherical clusters during peak and transitional rainfall periods. This study presents the first statistical classification of rainy seasons across Ethiopia utilizing high-resolution monthly rainfall, mean temperature climatology, and DEM. While these ten regimes correspond broadly to conventional meteorological regimes, they reveal considerable localized boundary differences. The statistics of these cluster regimes were additionally verified using boxplot analysis. Based on gridded rainfall data covering 66,000 grid points across Ethiopia, the country was categorized into specific peak rainfall regimes. Specifically, six regimes were identified as summer peak (July and August) profiles (RSR_1, RSR_2, RSR_3, RSR_6, RSR_7, and RSR_10), one as a late-summer peak (August) regime (RSR_9), and one as a spring peak (April and May) regime (RSR_8). In addition, the classification identified two spring/autumn-influenced peak regimes: RSR_4 (peak months: April, May, and October) and RSR_5 (peak month: April). Furthermore, Figure 4 shows the monthly climatological rainfall characteristics calculated for each regime. Notably, the RSR_1, RSR_6, and RSR_10 regimes display a unimodal rainfall cycle, which is consistent with findings reported by Diro et al. (2011a), while the others exhibit a bimodal rainfall cycle. The ten regimes classification aligns with the foundational principles of cluster analysis outlined by Ward (1963) and Hartigan and Wong (1979), as well as the monthly rainfall cycle frameworks described by the FAO (1978).

Synthesis of Regional Climate and Environmental Dynamics

Historical climate studies (Haile and Yarotskaya 1987) frequently simplify Ethiopian weather into three fixed seasons. However, there is significant variation in the timing, onset, and cessation of seasons across sub-regions of Ethiopia. Using generalized national season terms can lead to confusion about the length of each season. While long-term rainfall is the main factor used to distinguish seasons, long-term average seasonal temperatures and elevation also play key roles in shaping the underlying climate patterns. Our results closely match threshold-based modeling, which mapped the onset and cessation of wet seasons for RSR_1, RSR_3, RSR_6, RSR_7, and RSR_8. Geographically, the clear separation of these regimes follows distinct north–

south and west–east patterns, showing that topography and latitude are the main physical factors controlling monthly rainfall distribution, except for specialized lowland systems such as RSR_2, southern RSR_4, RSR_5, RSR_9, and RSR_10. On a larger scale, these regimes are influenced by global atmospheric circulation patterns, which are driven by differences in solar heating, including seasonal changes in day length and the angle of the sun, which influence the movement of the Intertropical Convergence Zone. Incorporating a refined, data-based timetable into national weather forecasting operations can greatly improve local seasonal descriptions and strengthen early warning systems.

CONCLUSION

This study developed a robust, data-driven classification of Ethiopia's rainfall seasonality using unsupervised machine learning applied to 42 years of high-resolution rainfall, temperature, and digital elevation model data. Ten distinct Rainy Season Regimes (RSRs) were identified, each exhibiting unique patterns of rainfall onset, cessation, duration, and intensity. The strong agreement between Ward's hierarchical clustering and K-means clustering confirms the reliability of the classification. The results demonstrate that Ethiopia's rainfall seasonality is far more spatially heterogeneous than represented by the conventional Bega, Belg, and Kiremt seasonal classifications, highlighting the need for localized climate characterization. The identified rainfall regimes provide an objective framework for understanding climate variability and supporting evidence-based decision-making. Thus, this study addresses the existing research gap by developing a statistical classification of rainy season using unsupervised machine learning analysis. This approach contributes to the broader knowledge and understanding of rainfall seasonality across Ethiopia. Finally, the study delivers an understanding into the complexity of seasonality across Ethiopia, as well as direction for climate-relevant research with rainfall, temperature and elevation data as the primary input. The most significant contribution of this study is the newly proposed seasonal timetable using unsupervised machine learning. The knowledge presented here is crucial for agricultural practices, natural resource management, and other climate-dependent activities, especially to develop adaptive strategies in monitoring seasonal changes in rain and temperatures under climate change.

The study recommends integrating the rainfall regime framework into operational seasonal forecasting, agricultural advisory services, drought monitoring systems, and water resource planning. The regime-specific seasonal timetable can improve crop and livestock management, enhance early warning systems, and support more targeted climate adaptation interventions. Future research should examine the relationships between the identified regimes and large-scale climate drivers, such as ENSO, IOD, and the ITCZ; assess the impacts of climate change on regime stability; and incorporate rainfall seasonality indices to further strengthen the scientific basis and operational utility of the classification framework.

Acknowledgements: I would like to thank the Ethiopian Meteorological Institute for providing access to high-resolution rainfall and temperature data from 1981 to 2022 for this study. This study was not sponsored.

I gratefully acknowledge my wife Marta and my friend Zenna for their contributions, without which this paper would not have been possible. Lastly but not least I would like to thank Dr. Sintayehu Workneh.

Competing interests: We declare that there are no competing interests.

Authors' contributions: The corresponding author, Yosef Tesfaye, was primarily responsible for the analysis and generation of the dataset and contributed to the conceptualization, methodology, formal analysis, and writing of the original draft. All authors have read and reviewed the manuscript and approved this Data Descriptor.

Data availability: Data are owned by the Ethiopia Meteorological Institute and can be obtained from them on request.

REFERENCES

Abiy Gebremichael, Shoa Quraishi, and Girma Mamo (2014) Analysis of Seasonal Rainfall Variability for Agricultural Water Resource Management in Southern Region, Ethiopia. Journal of Natural Sciences Research Vol.4, No.11.

Abreu MC, Souza A, Lyra GB, Pobocikova I, Cecilio RA (2020b) Analysis of monthly and annual rainfall variability using linear models in the state of Mato Grosso do Sul, Midwest of Brazil. Int. J. Climatology Sep.; 41(S1): E2445-E2461.

Adriaan J. van der Walt; Jennifer M. Fitchett (2020) Statistical classification of South African seasonal divisions on the basis of daily temperature data. S Africa J. Science; 116(9/10), Art. #7614, 15 pages. https://doi.org/10.17159/ sajs.2020/7614.

Alhamshry, Asmaa, Ayele Almaw Fenta, Hiroshi Yasuda, Reiji Kimura, and Katsuyuki Shimizu (2020) Seasonal Rainfall Variability in Ethiopia and Its Long-Term Link to Global Sea Surface Temperatures. Water 12, 1:55 https://doi.org/10.3390/w12010055.

Berhanu, Belete; Seleshi, Yilma; Melesse, Assefa M. Melesse, Assefa M.; Abtew, Wossenu; Setegn, Shimelis G. (2014) "Surface Water and Groundwater Resources of Ethiopia: Potentials and Challenges of Water Resources Development", Nile River Basin, Cham: Springer.97–117, doi:10.1007/978-3-319-02720-3_6.

Chung, Sang Yong; Venkatramanan, S.; Elzain, Hussam Eldin; Selvam, S.; Prasanna, M. V. (2019) Supplement of Missing Data in Groundwater-Level Variations of Peak Type Using Geostatistical Methods. GIS & Geostatistical Techniques for Groundwater Science.Elsevier.33–41. doi:10.1016/b978-0-12-815413-7.00004-3.

Diro, G.T., D.I.F. Grimes, and E. Black (2011b) Large Scale Features Affecting Ethiopian Rainfall. Earth System Physics Section, the Abdus Salam ICTP, Trieste, Italy. DOI 10.1007/978-90-481-3842-5_2.

Diro, G.T., Grimes, D.I.F. & Black, E. (2011a) Teleconnections between Ethiopian summer rainfall and sea surface temperature: part II. Seasonal forecasting. Clim Dyn 37, 121–13. https://doi.org/10.1007/s00382-010-0896-x

Diro, G.T.; Grimes, D.I.F.; Black, E.; O'Neill, A.; Pardo-Iguzquiza, E (2009) Evaluation of reanalysis rainfall estimates over Ethiopia. Int. J. Climatology, 29, 67–78.

Ethiopian Statistical Development Program (ESDP) EFY 2016-2018 (2023) Ministry of Planning and Development, Ethiopian Statistical Service. Addis Ababa, Ethiopia.

Feipeng Xiao, Zhitao Zhang, Zichao Wu, Wentao He, Jin Li (2025) Machine learning-based climate zoning and asphalt selection for pavement infrastructure under changing climate: A focused study of Ningxia, China. Tongji University Press. Publishing Services by Elsevier. https://doi.org/10.1016/j.ijtst.2024.10.001.

Food and Agriculture Organization (FAO) (1978) Report on the Agro ecological zones project volume1: Methodology and Results for Africa World Soil Resources Report No. 48. FAO, Rome.

Gissila, T.; Black, E.; Grimes, D.; Slingo, J (2004) Seasonal forecasting of the Ethiopian summer rains. Int. J. Climatol. 24, 1345–1358.

Gizaw Mengistu Tsidu (2012) High-Resolution Monthly Rainfall Database for Ethiopia: Homogenization, Reconstruction, and Gridding. American Meteorological Society, journal of climate volume 25. DOI: 10.1175/JCLI-D-12-00027.1.

Haile, T. and Yarotskaya, L (1987) Onset and cessation of rains in Ethiopia. NMSA, Addis Ababa, Ethiopia.

Hartigan, J. a., Wong, M. a. (1979) Algorithm AS 136: A K-Means Clustering Algorithm. J. R. Stat. Soc. C 28, 100–108. DOI: 10.2307/2346830.

Huber, W., & Love, M. I (2018) Modern Statistics for Modern Biology. Cambridge University Press. https://www.huber.embl.de/msmb/05-chap.html.

Hussain, I., Sinaga, K.P., Yang, M.S. (2023) Unsupervised multi view fuzzy C-means clustering algorithm. Electronics 12, 4467. https://doi.org/ 10.3390/electronics12214467.

J. B. MacQueen (1967) Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297.

Jing Wang, Filip Biljecki (2022). Unsupervised machine learning in urban studies: A systematic review of applications. Elsevier, Department of Architecture, National University of Singapore, Singapore. https://doi.org/10.1016/j.cities.2022.103925.

Jose E. Chacon · Ana I. Rastrojo (2022) Minimum adjusted Rand index for two clustering of a given size. Advances in Data Analysis and Classification. https://doi.org/10.1007/s11634-022-00491-w.

Kassahun, Bokretsion (1999) Climate Change and Forecast in Ethiopia, in Amharic. A discussion organized by the DPPC on Climate Change, Drought and Disaster Prevention in Ethiopia, Addis Ababa, Ethiopia.

Korecha, D and A. Sorteberg (2013) Validation of operational seasonal rainfall forecast in Ethiopia. Water Resour. Res., 49, 7681–7697, DOI: 10.1002/2013WR013760.

Markos Budusa Ware, Paolo Mori1, Kirsten Warrach-Sagi, Mark Jury, Thomas Schwitalla1, Kinfe Hailemariam Beyene and Volker Wulfmeyer (2022) Climate regionalization using objective multivariate clustering methods and characterization of climatic regions in Ethiopia. University of Hohenheim, Institute of Physics and Meteorology, Germany. DOI 10.1127/metz/2022/1093.

Marzieh Hasanzadeh Saray, Seyed Saeid Eslamian1 Björn Klöve and Alireza Gohari (2020) Regionalization of potential evapotranspiration using a modified region of influence. Theoretical and Applied Climatology 140:115–127. https://doi.org/10.1007/s00704-019-03078-2.

Meron Teferi Taye Ellen Dyer, Katrina J. Charles, Linda C. Hirons (2021) Potential predictability of the Ethiopian summer rains: Understanding local variations and their implications for water management decisions. Science of the Total Environment. https://doi.org/10.1016/j.scitotenv.142604.

Nicholson, Sharon E (2018) The ITCZ and the Seasonal Cycle over Equatorial Africa. Bulletin of the American Meteorological Society. 99 (2): 337–348.

Palmer, Paul. I., Wainwright, C.M., Dong, B. (2023) Drivers and impacts of Eastern African rainfall variability. Nat Rev Earth Environ 4, 254–270.

Segele ZT, Lamb PJ (2005) Characterization and variability of Kiremt rainy season over Ethiopia. Meteorology Atmos. Physic 89:153–180.doi:10.1007/s00703-005-0127.

Urgiles, G.; Célleri, R.; Trachte, K.; Bendix, J.; Orellana-Alvear, J (2021) Clustering of Rainfall Types Using Micro Rain Radar and Laser Disdrometer Observations in the Tropical Andes. Remote Sens. 13, 991. doi.org/10.3390/rs13050991.

Viste, E., and Sorteberg, A (2013) The effect of moisture transport variability on Ethiopian summer precipitation. International Journal of Climate, 33, 3106–3123.

Ward Jr. Joe H (1963) Hierarchical Grouping to Optimize an Objective Function. American Statistical Association Journal, 58, 236-244. https://doi.org/10.1080/01621459.1963.10500845.

World Meteorological Organization (WMO) (2017) Climate Data and Monitoring WCDMP-No. 85. Geneva, Switzerland.

Ying Zhang, Semu Moges and Paul Block (2016) Optimal Cluster Analysis for Objective Regionalization of Seasonal Precipitation in Regions of High Spatial–Temporal Variability: Application to Western Ethiopia. American Meteorological Society. DOI: 10.1175/JCLI-D-15-0582.1.

Yosef Tesfaye, Aklile Assefa and Tofikk Redi (2024) Kiremt Rainfall Characteristics over Ethiopia. Joint Scientific Journal of EMI and EtMS. 1K Printing Press, ISBN 978-99990-53-95-2. Volume 3. Addis Ababa, Ethiopia.

Zerihun Hailemariam and Abel Sisay (2020) Season Classification and Integrated Impact of ENSO and IOD on Seasonal Rainfall over Ethiopian Lowlands. Joint Scientific Journal of EMI and EMI-EtMS, Israel printing press. 1, 97-127, Addis Ababa, Ethiopia.