Watershed Level Clustering and Know Your Landscape

March 16, 2024

Table of Contents

1 Watershed Level Clustering

1.1 Introduction

Agro-ecological zones (AEZs) are geographic areas characterized by specific combinations of climate, soil, topography, and land use patterns that influence agricultural productivity and sustainability. These zones are essential for understanding the diversity of agricultural systems and formulating tailored strategies for crop production, resource management, and environmental conservation.

1.2 Input layers

Agro-ecological zones are further divided into watersheds. The main aim of this project is to cluster the watersheds based on the detection of the underlying climatic and land use patterns that are common across the watersheds. For clustering we considered the variables for the years 2016 till 2022. We considered land use land cover variables such as the area under kharif cropping, area under single non kharif cropping, area under double cropping, area under triple cropping, barren land area and forest area. We considered the climatic variables such as Evapotranspiration, Total Rainfall, Normalized Difference Vegetation Index (NDVI) across the cropping seasons Kharif, Rabi and Zaid. We considered the variables related to drought such as onset of monsoon, frequency and Intensity of severity of the drought i.e No, Mild, Moderate and Severe Drought.

At the first level we had done the clustering by observing the cropping patterns across the watersheds. Averaging the land use and land cover (LULC) variables across multiple years 2016 to 2022, provides a more comprehensive understanding of agro-ecological zones. By considering data over a longer time period, we can capture the variability and stability of land use patterns and environmental conditions, leading to more robust delineation of agro-ecozones. This smooths out short-term fluctuations and better identifies persistent patterns and trends that characterize different watersheds in the agro-ecological zone.

1.3 Methodology

1.3.1 Data Preprocessing

Step 1: The input file contains column names such as prop1, prop2..., since the file originates from Google Earth Engine. In Earth Engine, each column's first 10 characters must be unique, but there are several columns in the data that share the same first 10 characters. To address this issue, the column names are mapped to the {column_name : prop_number } format, and this mapping is utilized to revert the column names back in the Python file. This mapping will be unique for each output file.

Step 2: For each watershed create new LULC variables which are the average of the LULC variables across the years 2016 till 2022.

Step 3: Now divide each of the newly created variables with the area of the watershed.

Step 4: Min Max normalization is done on the dataset to ensure all the features are on the same scale.

1.3.2 Feature Selection

Consider the newly created LULC variables as features and the data points are the watersheds in an agro ecological zone.

1.3.3 Clustering

We use hierarchical clustering technique for achieving unsupervised clustering. Hierarchical clustering is a method used to cluster data points based on their similarity or dissimilarity. It creates a hierarchy of clusters by iteratively merging or splitting clusters until a stopping criterion is met. Used Ward's method as the linkage criteria which is commonly used in hierarchical clustering algorithms.

Step 1: The distance metric is created by calculating the distance or dissimilarity between each pair of data points.

Step 2: Initially, each data point (watershed here) is considered as a separate cluster. The distance between clusters is computed using the distance metric.

Step 3: Ward's method is a type of linkage criterion used to determine the distance between clusters. It minimizes the variance when merging two clusters. Ward's method calculates the sum of squared differences within all clusters before and after merging and selects the merge that minimizes this sum.

Step 4: Hierarchical clustering proceeds iteratively by merging the two closest clusters based on the Ward's linkage criterion. This process continues until a predefined number of clusters is reached.

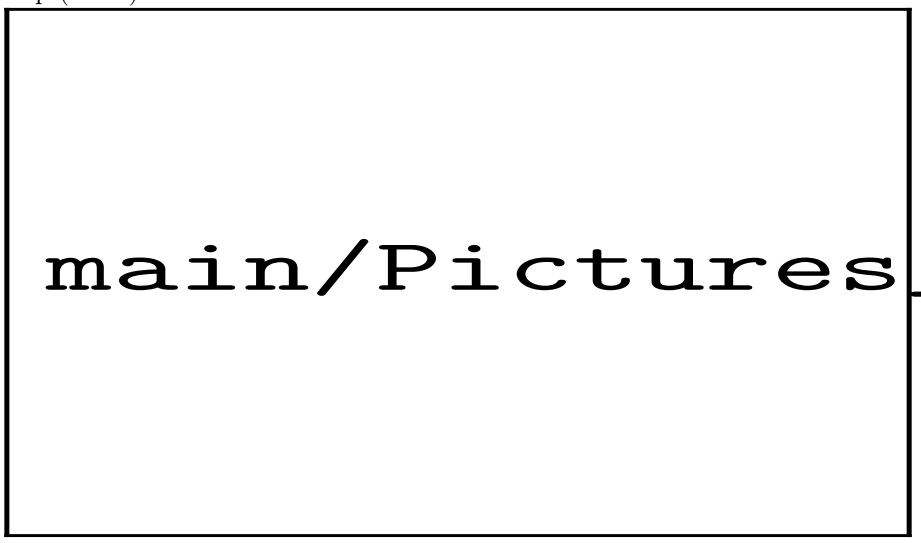
1.4 Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset:
- Codebase:

1.5 Results

We analyzed the results by considering the main cases where the number of clusters as 4 and 5. The patterns are observed for each case by plotting mean values of each variable for each cluster. The observation depicted that 4 clusters are sufficient enough for capturing the data across the watersheds for classification. Below are the attached figures for the behavior of each variable in each cluster.

Figure 1: Snapshot of drainage lines with stream orders overlaid on open street map (OSM) for validation with its rivers.



2 Know Your Landscape

2.1 Introduction

The "Know Your Landscape" project is a crucial initiative aimed at comprehending the behavior of extraneous and constructed time series variables within a micro-watershed. This project deals with the time series variables which spans from the years 2003 to 2022, and are essential for understanding environmental dynamics.

- We considered extraneous variables such as total rainfall in the Kharif season, rainfall deviations, monsoon onset dates, and their respective deviations, from these users will gain deep insights into meteorological patterns and trends specific to the micro-watershed.

- Comprehensive assessment of drought frequencies and intensities, ranging from no drought to severe drought, along with delineated paths based on various environmental parameters. This information is invaluable for drought monitoring and mitigation strategies.
- Utilizing historical data, the project identifies monsoon due dates and its deviations for different micro watersheds in India. This knowledge aids in planning and preparedness for seasonal variations and climate-related challenges.
- Users can delve into derived variables such as cropping intensity and the number of NREGA works undertaken across different categories within the region. These derived variables offer additional perspectives on land use patterns and socio-economic activities.

The methodology employed in this project ensures robust data collection, analysis, and interpretation, guided by established scientific principles. By leveraging advanced computational techniques the project delivers reliable and actionable insights crucial for informed decision-making in environmental management and agricultural planning.

2.2 Cropping Intensity

2.2.1 Introduction

Cropping intensity measures the level of crop cultivation within a specific region. It encompasses single, double, and triple cropping cycles occurring in the region, and the determination of cropping intensity is based on the proportion of land utilized for each type of cropping.

2.2.2 Input layers

2.2.3 Methodology

$$\text{Cropping Intensity} = \frac{\text{Single Cropped Area} + (2 \times \text{Double Cropped Area}) + (3 \times \text{Triple Cropped Area})}{\text{Total Cropable Area}} \times 100$$

2.2.4 Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset:
- Codebase:

2.2.5 Results

We analyzed the results by considering the main cases where the number of clusters as 4 and 5. The patterns are observed for each case by plotting mean values of each variable for each cluster. The observation depicted that 4 clusters are sufficient enough for capturing the data across the watersheds for classification. Below are the attached figures for the behavior of each variable in each cluster.

2.3 Total Rainfall and its Deviation

2.3.1 Introduction

total rainfall in each year kharif season, Rainfall deviation in each year kharif season from past 30 years,

2.3.2 Input layers

2.3.3 Methodology

- Total Rainfall: Total Rainfall is obtained by summing up the total amount of rainfall occured on each day in the Kharif season.
- Rainfall Deviation: We get the mean value of the past 30 years total rainfall till the current year which we refer as Rfn and the current year total rainfall as Rfi. Now we obtain the Rainfall deviation of current year by subtracting Rfn from Rfi and dividing the entire expression with Rfn. Mathematical Equation (Rfi-Rfn)/Rfn.

2.3.4 Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset:
- Codebase:

2.3.5 Results

2.4 Onset of Monsoon and its Deviaition

2.4.1 Introduction

Understanding the monsoon onset and its deviations is crucial for effective drought monitoring, agricultural planning, water resource management, and climate resilience. These insights enable informed decision-making, aiding in

mitigating risks associated with extreme weather events and optimizing resource allocation for sustainable development.

2.4.2 Input layers

2.4.3 Methodology

- Monsoon Onset: As entire india is divided into five Hydro Meteorological Zones (HMZ) and each HMZ has the same onset date over the years. So fixed thresholds have been defined for each HMZ.
  - Initially check to which HMZ our region of interest belongs to and get the threshold value.
  - calculate the total rainfall for each week in a year.
  - Now starting from second week of May we will check which week has the value above threshold and we mark that week as a monsoon week.
  - In order to obtain the monsoon date we check the first rainy day of that week
- Monsoon Onset Deviation: The Deviation of monsoon onset for each year can be captured by following the below steps
  - Capture the monsoon onset day of all the past years.
  - Capture the long term mean by average of all the past years onset days.
  - Difference between the mean value and current year onset day of the year captures the deviation

  Positive deviation indicates the early monsoon from the longterm mean. Negative deviation indicates the late monsoon from the longterm mean.

  Further the weekly deviation bins are generated where if the absolute value of the deviation is less than 7 days then the deviation is captured as week0 deviation (zeros weeks deviation) and similarly if the deviation is between 7 to 14 days then it is captured as week1 devaition(one week deviation) and so on.

2.4.4 Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset:
- Codebase:

2.4.5 Results

2.5 Drought

2.5.1 Introduction

Understanding the frequency and intensity of drought is vital for this project as it offers valuable insights into how often and how severely droughts occur. This information is crucial for managing environmental resources, especially water, and for implementing effective strategies to mitigate the impacts of drought on ecosystems and communities. By analyzing drought patterns i.e. the causality, we can develop sustainable solutions that address the challenges posed by water scarcity and ensure resilience in the face of changing climatic conditions.

- Variables related to Frequency and Intensity of Drought
  - Frequency of no drought
  - Frequency of mild drought
  - Frequency of moderate drought
  - Frequency of severe drought
  - Intensity of no drought
  - Intensity of mild drought
  - Intensity of moderate drought
  - Intensity of severe drought
- Variables related to Severe Drought
  - Severe Drought Path1 (Number of weeks which are in dryspell and with poor vci, poor mai and poor cropping area sown)
  - Severe Drought Path2 (Number of weeks which are in scanty rainfall and with poor vci, poor mai and poor cropping area sown)
  - Severe Drought Path3 (Number of weeks which are in extremely dryness and with poor vci, poor mai and poor cropping area sown)
- Variables related to Moderate Drought
  - Moderate Drought Path1 (Number of weeks which are in dryspell and with fair vci, poor mai and poor cropping area sown)
  - Moderate Drought Path2 (Number of weeks which are in dryspell and with good vci, poor mai and poor cropping area sown)
  - Moderate Drought Path3 (Number of weeks which are in dryspell and with poor vci, fair mai and poor cropping area sown)
  - Moderate Drought Path4 (Number of weeks which are in dryspell and with poor vci, good mai and poor cropping area sown)

  - Moderate Drought Path5 (Number of weeks which are in dryspell and with poor vci, poor mai and fair cropping area sown)
  - Moderate Drought Path6 (Number of weeks which are in dryspell and with poor vci, poor mai and good cropping area sown)
  - Moderate Drought Path7 (Number of weeks which are in scanty rainfall and with fair vci, poor mai and poor cropping area sown)
  - Moderate Drought Path8 (Number of weeks which are in scanty rainfall and with good vci, poor mai and poor cropping area sown)
  - Moderate Drought Path9 (Number of weeks which are in scanty rainfall and with poor vci, fair mai and poor cropping area sown)
  - Moderate Drought Path10 (Number of weeks which are in scanty rainfall and with poor vci, good mai and poor cropping area sown)
  - Moderate Drought Path11 (Number of weeks which are in scanty rainfall and with poor vci, poor mai and fair cropping area sown)
  - Moderate Drought Path12 (Number of weeks which are in scanty rainfall and with poor vci, poor mai and good cropping area sown)
  - Moderate Drought Path13 (Number of weeks which are in extremely dryness and with fair vci, poor mai and poor cropping area sown)
  - Moderate Drought Path14 (Number of weeks which are in extremely dryness and with good vci, poor mai and poor cropping area sown)
  - Moderate Drought Path15 (Number of weeks which are in extremely dryness and with poor vci, fair mai and poor cropping area sown)
  - Moderate Drought Path16 (Number of weeks which are in extremely dryness and with poor vci, good mai and poor cropping area sown)
  - Moderate Drought Path17 (Number of weeks which are in extremely dryness and with poor vci, poor mai and fair cropping area sown)
  - Moderate Drought Path18 (Number of weeks which are in extremely dryness and with poor vci, poor mai and good cropping area sown)

- Variables related to Mild Drought
  - Mild Drought dryspell score
  - Mild Drought rainfall deviation score
  - Mild Drought spi score
  - Mild Drought vci score
  - Mild Drought mai score
  - Mild Drought cropping area sown score

2.5.2 Input layers

2.5.3 Methodology

- Frequency and Intensity of Drought: We captured the following variables from drought asset for better understanding go through the core app v1 layers manual.
  - Frequency and Intensity of Mild Drought
  - Frequency and Intensity of Moderate Drought
  - Frequency and Intensity of Severe Drought
  - Frequency and Intensity of No Drought
- Causuality of Drought: Now we capture all the below intermediate variables in the code during the computation of drought which eventually helps in capturing the causality of the severe, moderate, mild droughts.
  - Intermediate variables in the computation
  - Weekly values of dry spells for the year
  - Weekly values of rainfall deviation for the year
  - Weekly values of meteorological drought for the year
  - Weekly values of SPI for the year
  - Weekly values of VCI for the year
  - Weekly values of MAI for the year

  Now we follow the below steps for capturing causality of drought.
  - we check if there is a dryspell or scanty rainfall (i.e from rainfall deviation) or an extremely dryness (i.e from spi) then there is meteorological drought, else no drought.
  - Causuality of Severe Drought: Now if there is hit in meteorological drought, if all the VCI, MAI and Cropping Area Sown condition is poor then the type of drought hit is severe drought so we increase the count of paths that corresponds to severe drought.
  - Causuality of Moderate Drought: Now if there is hit in meteorological drought, if any of the two of VCI, MAI and Cropping Area Sown condition is poor then the type of drought hit is moderate drought so we increase the count of paths that corresponds to moderate drought.
  - Causuality of Mild Drought: Now, if there is a hit in meteorological drought and if it is not a severe drought and not a moderate drought then it is a mild drought. Each of the variable such as SPI, Rainfall Deviation, Dryspell, VCI, MAI and Cropping Area Sown score is updated using the below the score.

  Variable score is calculated using the below mentioned formulae

- DrySpell Score = Number of weeks there is a drySpell for the weeks under mild drought.
- Rainfall Deviation Score = Number of weeks dry Spell is in scanty for the weeks under mild drought.
- Spi Score = Number of weeks of SPI in extremely dryness for the weeks under mild drought.
- VCI, MAI, Cropping Area Sown Score = (1*x + 2*y + 3*z)/6
  variable is vci, mai, cropping area sown
  Where x is the number of weeks variable is in good condition for the weeks under mild drought.
  Where y is the number of weeks variable is in fair condition for the weeks under mild drought.
  Where z is the number of weeks variable is in poor condition for the weeks under mild drought.

2.5.4 Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset:
- Codebase:

2.5.5 Results

2.6 NREGA Work Count

2.6.1 Introduction

This section comprises the count of NREGA works in different categories, in each micro watershed for each year, starting from 2005.

2.6.2 Input layers

2.6.3 Methodology

Step 1: Given a region, and its district, filter out the nrega works of that region from the source file of the district's nrega file. (part 1)

Step 2: Using the 'Work start date' column find out the year in which the work has been done. The date can be present in different formats, if the year is not found then assign '1888' as a dummy year for that work

Step 3: Save those works in a csv file

Step 4: Use that csv file. Using work category and 'year' column find the works done in each year in each category.

2.6.4 Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset:
- Codebase:

2.6.5 Results

We analyzed the results by considering the main cases where the number of clusters as 4 and 5. The patterns are observed for each case by plotting mean values of each variable for each cluster. The observation depicted that 4 clusters are sufficient enough for capturing the data across the watersheds for classification. Below are the attached figures for the behavior of each variable in each cluster.

References