CoRE stack: Commoning for Resilience and Equality - Geospatial data layers

Draft version: February 2025

Shivani A Mehta[1], Aaditya Sinha[1], Aatif Nisar Dar[1], Aila Dutt[2], Akshay P Sarashetti[1], Akshay Pratap Singh[1], Aman Gupta[1], Aman Verma[2], Ajay Tannirkulam[4], Ananda Sreenidhi[2], Ananjan Nandi[1], Ankit[1], Ankit Kumar[2], Anamitra Singha[1], Anant Gulgulia[1], Ashima Mittal[1], Atharv Dabli[1], Athira P[3], Badrinath Padmanabhan[1], Balakumaran Ramachandran[5], Chahat Bansal[1], Chintan Sanjaybhai Sheth[1], Craig Dsouza[5], Deepak Kumar[2], Dharmisha Sharma[1], Dhruvi Goyal[1], Harshita[1], Harsh Singh Chauhan[1], Jaskaran Singh[1], Kapil Dadheech[2], Keshav Rawat[1], Ksheetiz Agrahari[2], Manvi Khatri[2], Om Krishna[1], Pooja Prasad[1], Priyadarshini Radhakrishnan[1], Rittwick Bhabak[1], Ruptirumal Sai Bodavula[1], Raman Kumar[1], Ramita Sardana[1], Ramneek Kaur[1], Saketh Vishnubhatla[1], Samitha Haldar[2], Sanjali Agrawal[1], Shruti Kumari[1], Siddharth S[1], Sparsh Sheth[1], Sukriti Kumari[2], Taru Jain[2], Vatsal Agarwal[1], Vishnu S[2], and Aaditeshwar Seth[1,2]

[1]Indian Institute of Technology, Delhi
[2]Gram Vaani
[3]Indian Institute of Technology, Palakkad
[4]Magasool
[5]Well Labs

Contact: contact@core-stack.org

Acknowledgements

None of this would have been possible without valuable insights and field support from our partners. A very special thanks to Subrata Singh, Chiranjit Guha, Himani Sharma, Pratiti Priyadarshini and the wider FES team, for their constant support, encouragement, and thought partnership. Our sincere gratitude to the teams at SUPPORT in Dumka, 4S in Gaya, PRADAN in Sirohi, Jan Jagran Kendra in Giridih, CCD in Adilabad, SPWD and JJBA in Latehar, and IIT-IIT in Kolar, for field access. Finally, our heartfelt thanks to our donors: GIZ, Common Ground, Rainmatter Foundation, Rohini Nilekani Philanthropies, Tower Research, Hellermann Tyton, Tarides, R Systems, Google, and AVPN whose generous contributions have enabled us to build the CoRE stack so far. We would also like to especially acknowledge Sharat Singh from GIZ for believing in our vision and being the first to mobilize support for the CoRE stack.

Table of Contents

1 Acknowledgements **3**

2 Introduction **13**

3 Landscape level **19**

3.1 Hydrological . . . 19
3.1.1 Boundaries . . . 19
3.1.1.1 Basins . . . 19
3.1.1.2 Subbasins . . . 20
3.1.1.3 Watersheds . . . 21
3.1.1.4 Microwatersheds . . . 22
3.1.1.5 Upstream-downstream microwatersheds . . . 23
3.1.2 Variables . . . 24
3.1.2.1 Precipitation . . . 24
3.1.2.2 Runoff . . . 25
3.1.2.3 Evapotranspiration . . . 28
3.1.2.4 Downscaled evapotranspiration . . . 29
3.1.2.5 Change in groundwater . . . 31
3.1.2.6 Change in well depth . . . 33
3.1.2.7 Aquifers . . . 35
3.1.2.8 Groundwater level in dugwells . . . 37
3.1.2.9 Groundwater level in borewells . . . 38
3.1.2.10 Stage of Groundwater Extraction . . . 39
3.1.2.11 Rivers . . . 41
3.1.2.12 Canals . . . 42
3.2 Climate variables . . . 43
3.2.1 Drought frequency and intensity . . . 43
3.2.2 Drought causality . . . 47
3.2.3 Monsoon onset . . . 49
3.3 Terrain variables . . . 51
3.3.1 Terrain classification . . . 51
3.3.2 Terrain cluster . . . 54
3.4 Land use variables . . . 56
3.4.1 Land use land cover . . . 56
3.4.2 Land use on terrain . . . 60
3.4.3 Land use changes . . . 62
3.4.4 Cropping intensity . . . 64
3.4.5 Water bodies . . . 66
3.4.6 First census of water bodies . . . 68

3.5 Forest variables . . . 70
3.5.1 Tree health monitoring . . . 70
3.5.2 Tree canopy cover density . . . 72
3.5.3 Tree canopy height . . . 73
3.5.4 Tree health changes . . . 74
3.6 Welfare variables . . . 76
3.6.1 NREGA assets categorization . . . 76
3.6.2 NREGA panchayat category . . . 79
3.6.3 NREGA MIS parameters . . . 81
3.6.4 Fairness in NREGA expenditure . . . 83
3.6.5 Caste-based inequity in NREGA allocation . . . 85
3.7 Administrative . . . 88
3.7.1 Boundaries . . . 89
3.7.1.1 State . . . 89
3.7.1.2 District . . . 90
3.7.1.3 Block . . . 91
3.7.1.4 Panchayat . . . 92
3.7.1.5 Village . . . 93

4 Site level **95**
4.1 Water structure planning layers . . . 95
4.1.1 Lineament . . . 95
4.1.2 Lithology . . . 96
4.1.3 Drainage lines . . . 97
4.1.4 Stream order raster . . . 99
4.1.5 Distance to upstream drainage line . . . 101
4.1.6 Catchment area . . . 102
4.1.7 Natural depressions . . . 103
4.2 Water structure planning algorithms . . . 104
4.2.1 CLART . . . 104
4.2.2 Runoff accumulation . . . 109
4.2.3 Upstream-downstream waterbodies . . . 110
4.3 Resource mapping . . . 112
4.3.1 Farm ponds and wells . . . 112
4.3.2 Farm boundaries and plantations . . . 115

List of Tables

2.1 Hosting specification of vector layers in CoRE Stack with spatial coverage across India . . . 14
2.2 Hosting specification of raster layers in CoRE stack with spatial coverage across India . . . 16

3.1 Categories of AMC conditions . . . 26
3.2 Categories of Stage of Groundwater Extraction . . . 39
3.3 Possible combinations of trigger and impact indicators for moderate and severe drought intensities. . . . 48
3.4 MGNREGA category and their corresponding keywords . . . 76
3.5 MIS Parameters used for welfare analysis . . . 82

4.1 Recharge potential values and their corresponding scores for recharge . . . 105

List of Figures

2.1 CoRE stack: Modular approach to build a digital public infrastructure 14

3.1 Basin (left) and subbasin (right) boundaries across India 20
3.2 Microwatershed (left) and watershed (right) boundaries across Mohanpur block, in Gaya district of Bihar 22
3.3 Upstream-downstream connectivity of micro-watersheds across Mohanpur block in Gaya district of Bihar 23
3.4 Methodology to estimate runoff 27
3.5 Visualisation of 30 meter resolution ET product (middle) predicted by the trained Random Forest Model for a region from AEZ 10 of area 2 km $\times$ 2 km against a satellite image (left) and MODIS ET image (right). 30
3.6 Spatial variation in precipitation, runoff, evapotranspiration and change in groundwater during 2022-23 across microwatersheds of Mohanpur block, Gaya district, Bihar 32
3.7 Time series of water balance variables: precipitation, runoff, evapotranspiration and change in groundwater (Deltag) on fortnight basis between 2022-2023 for a microwatershed in Mohanpur. Time series of groundwater (G) is derived by assuming the groundwater in the first fortnight as zero in the worst case. 32
3.8 Change in well depth across microwatersheds of Mohanpur during 2017-22 (left) and 2018-23 (right) 34
3.9 Principal aquifer systems of India 36
3.10 Major aquifer systems of India 36
3.11 Stage of Groundwater Extraction of assessment units (Talukas/Blocks/Mandals/Firkas) 40
3.12 Canal (left) and River (right) network across India 42
3.13 Drought intensity is determined for each monthly window (rows) in terms of mild, moderate, severe and no drought. Weekly drought intensity (columns) is determined by the maximum drought intensity that has occurred across the corresponding monthly windows in which the week is present. 45
3.14 Flow chart of determining monthly drought intensity 46
3.15 The figure shows five hydro-meteorological zones (A to E) delineated based on the duration of the hydro-meteorological seasons (monsoon, post-monsoon and pre-monsoon) 50
3.16 Scale dependency of TPI 52
3.17 Landforms classification using slope and thresholds (denoted using λ) on both small-scale and large scale TPIs 53
3.18 Landform classification of Masalia with micro-watershed boundaries in black. 53

3.19 Distribution of plains, broad slopes, steep slopes, valleys and ridges across terrain clusters: mostly plains (cluster 0), plains and slopes (cluster 1), and mix of hills, valleys, slopes and plains (cluster 2) 54
3.20 Terrain clusters assigned to microwatersheds of Pindwara block 55
3.21 Land use land cover classification of Raichur district in Karnataka for 2022 . 56
3.22 Version-2 Land use land cover classification pipeline. 57
3.23 The figure characterizes land use on plains and slopes respectively across microwatersheds of Masalia block in Dumka district of Jharkhand 61
3.24 Change detection maps of Gulguldih village in Lakshmipur block of Jamui district in Bihar. 63
3.25 For a micro-watershed in Mohanpur, we analyze the percentage of area under single, double and triple cropping (left), and corresponding cropping intensity over the years. 65
3.26 Water bodies as seen from satellite (first column) in Masalia block of Dumka district in Jharkhand and their detection using the classifier (second column) 67
3.27 Bar graph depicting the surface area of a water body (in percentage) in Mohanpur block of Gaya district in Bihar 67
3.28 The map shows geo-coordinate of waterbodies from the waterbody census data of Rajasthan, specifically zoomed into Mandalgarh block of Bhilwara district for visualization. 69
3.29 Canopy cover density and height outputs of AY 2017 and AY 2021, and tracking overall tree health change between July 2017 - June 2022 in a forest patch of Dongariya village, Balaghat district, Madhya Pradesh 70
3.30 Canopy Cover Density (CCD) & Canopy Height (CH) Distribution Based Tree Health Change Categories 74
3.31 Large and Small Tree Patches Transition Categories 74
3.32 Tree Health Change (AY 2017 – 2021) Statistics of Eastern Plateau and Hills Region (ACZ-7) 75
3.33 MGNREGA assets in Pindwara block of Sirohi district in Rajasthan 77
3.34 Metadata of a MGNREGA asset in Masalia block of Dumka district in Jharkhand 78
3.35 Categorization of SC Panchayats based on the distribution of SC population within panchayat 80
3.36 Categorization of ST Panchayats based on the distribution of ST population within panchayat 80
3.37 Panchayat-wise MGNREGS expenditure in the Masalia block in the Dumka district of the state of Jharkhand in India 84
3.38 State-level inequity indices 86
3.39 District-level inequity indices 86
3.40 Panchayat-level inequity indices 87
3.41 State, district and block boundaries (from left to right) of India 91
3.42 Panchayat (top) and village (bottom) boundaries of Mohanpur block, in Gaya district of Bihar 93

4.1 Lithology shapefile from Bhukosh for Bhilwara district of Rajasthan 96
4.2 Snapshot of drainage lines with stream orders over Jamui district in Bihar, overlaid on open street map (OSM). 98
4.3 The figure shows drainage lines and the stream order raster map of a microwatershed (highlighted in red) in the Masalia block in Dumka district of Jharkhand. 100

4.4 The figure shows distance to nearest upstream drainage line, computed at each pixel and overlaid with drainage lines of a micro-watershed (highlighted in red) in the Masalia block in Dumka district of Jharkhand. 101
4.5 The left figure shows computed catchment area of each pixel of a micro-watershed in the Masalia block in Dumka district of Jharkhand. The right figure shows the delineated catchment area (highlighted in yellow) of a drainage pixel (highlighted in red). 102
4.6 The figure shows natural depression area in black (right) computed using elevation data (left). 103
4.7 The recharge potential scores are combined with different slope percentages and land cover categories to derive treatment codes in CLART. 106
4.8 Flowchart of methodology to produce reprogrammable CLART using score assignment 107
4.9 The figure shows the component layers for generating CLART output for Mandalgarh block in Bhilawara district of Rajasthan 108
4.10 The figure in left shows average runoff accumulated in non-drought years and the figure on the right shows sensitivity of runoff accumulation during drought years expressed as percentage deviation. 109
4.11 The figure in left shows drainage lines (in shades of blue) with stream order. Using the stream order hierarchy and spatial arrangements of water bodies (in purple circle as shown on top of these drainage lines), waterbody graph is generated (using directional edges in black) to identify upstream-downstream waterbodies as shown in the figure on the right. 111
4.12 Screenshots showing geolocations of predicted assets visualized in QGIS: (a) farm ponds and (b) wells. 114
4.13 Farm-boundary pipeline integrated with LULC pipeline to create IndiaSAT v4 116
4.14 (a) Example of farm, plantation and scrubland segment, (b) Result of Hough transform when applied on a gridded plantation, (c) Illustration showing easier segments being separated into different buckets of farm, plantation and scrubland using rules (d) Using mean NDVI timeseries to bring in gray area (e) Example of a NDVI time series 117
4.15 Dividing Roi into blocks and results of farm boundaries on several areas . . . 117

Introduction

The document outlines water security related datasets in the Commoning for Resilience and Equality (CoRE) stack to build a water security plans. The CoRE stack is a collaborative effort with several academic research institutions, civil society organizations working on ecological sustainability and rural livelihood, innovation catalysts, and technology-led social enterprises. Being designed as a Digital Public Good, the CoRE stack enables an open-access co-creation network to innovate and scale digital technology solutions for ecosystem sustainability. Figure 2.1 shows the hierarchical modules of CoRE stack.

Datasets are essential to quantify and assess the state of water security. Water security requires tracking of supply-side variables such as precipitation and incidence of droughts, demand-side variables such as increasing number of cropping cycles being practiced and intervention variables such as the available surface runoff for harvesting. We leveraged remote-sensing and machine learning to temporally estimate various variables at the micro-watershed scale. A few initiatives are already attempting to provide water security related datasets [42, 38], but they tend to use out-dated or lower resolution data products and sub-optimal algorithms.

We are guided by the principles of open data and algorithms to promote transparency and foster innovation through collaborations. We publish temporally static datasets such as hydrological and administrative boundaries using drive folders, and publish block-wise temporally dynamic datasets such as hydrological and climate variables on the Landscape explorer. To generate temporally dynamic datasets for your own block, we publish our GitHub repositories for reproducibility, collaboration and contribution.

Figure 2.1: CoRE stack: Modular approach to build a digital public infrastructure

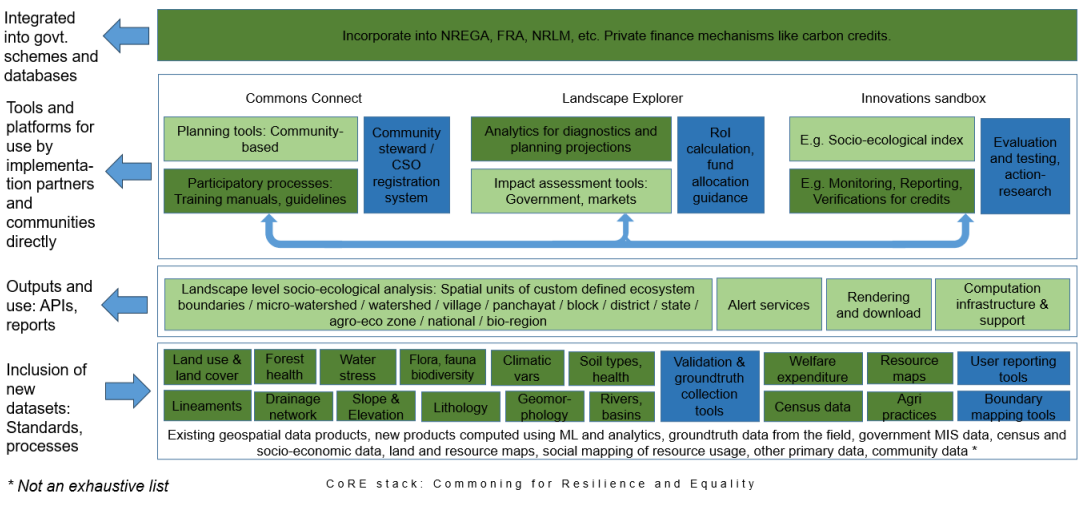


Table 2.1: Hosting specification of vector layers in CoRE Stack with spatial coverage across India

Layer	**Temporal resolution**	**Dataset**	**Codebase**	**Spatial unit of availability / computation**
Basin	static	Google drive folder	-	pan India
Sub-basin	static	Google drive folder	-	pan India
Watershed	static	Google drive folder	-	pan India
Microwatershed	static	Google drive folder	-	pan India
Upstream-downstream microwatersheds	static		Github repository	Computed at basin level
Precipitation	fortnightly	-	Github repository	Computed at microwatershed level
Runoff	fortnightly	-	Github repository	Computed at microwatershed level
Evapotranspiration	fortnightly	-	Github repository	Computed at microwatershed level
Change in groundwater	fortnightly, yearly	-	Github repository	Computed at microwatershed level

Change in well depth	five years	-	Github repository	Computed at microwatershed level
Aquifers	static	Google drive folder	-	pan India
Groundwater levels in dugwells	static	Google drive folder	-	pan India
Groundwater levels in borewells	static	Google drive folder	-	pan India
Stage of Groundwater Extraction	static	Google drive folder	-	pan India
Rivers	static	Google drive folder	-	pan India
Canals	static	Google drive folder	-	pan India
Drought intensity and frequency	yearly	-	Github repository	Computed at microwatershed level
Drought causality	yearly	-	Github repository	Computed at microwatershed level
Monsoon onset	yearly	-	Github repository	Computed at micro-watershed level
Terrain cluster	static	-	Github repository	Inferred at microwatershed level
Land use on terrain	yearly	-	Github repository	Inferred at microwatershed level
Cropping intensity	yearly	-	Github repository	Computed at micro-watershed level
Water bodies	yearly	-	Github repository	Computed at block level
First census of water bodies	static	Google drive folder	-	pan India
NREGA assets categorization	yearly	Google drive folder	Github repository	pan India, Computed at district level
State	static	Google drive folder	-	pan India
District	static	Google drive folder	-	pan India
Block	static	Google drive folder	-	pan India

Panchayat	static	Google drive folder	-	pan India
Village	static	Google drive folder	-	pan India
Lithology	static	Google drive folder	-	pan India
Drainage lines	static	Google drive folder	-	pan India
Upstream-downstream waterbodies	yearly	-	Github repository	Computed at block level
Farm ponds and wells	static	-	Github repository	Computed at block level
Farm boundaries	static	-	Github repository	Computed at block level

Table 2.2: Hosting specification of raster layers in CoRE stack with spatial coverage across India

Layer	**Spatial resolution**	**Temporal resolution**	**Dataset**	**Codebase**	**Spatial unit of availability / computation**
Downscaled evapotranspiration	30m	Landsat	-	Github repository 1	Inferred at pixel level
			-	Github repository 2	Inferred at pixel level
Terrain classification	30m	static	-	Github repository	Inferred at pixel level
Land use land cover	10m	yearly	-	Github repository	Inferred at pixel level
Land use changes	10m	yearly	-	Github repository	Computed at block level
Tree canopy cover density and height	25m	yearly	Google Earth Engine Assets	Github repository	pan India
Tree health change	25m	five years	Google Earth Engine Assets	Github repository	pan India
Lineament	194m	static	Google drive folder	-	pan India
Stream order raster	30m	static	Google drive folder	-	pan India
Distance to upstream drainage line	30m	static	Google drive folder	-	pan India
Catchment area	30m	static	Google drive folder	-	pan India

Natural depressions	30m	static	Google drive folder	-	pan India
CLART	30m	static	-	Github repository	Computed at pixel level
Runoff accumulation	30m	static	-	Github repository	Computed at pixel level

Landscape level

3.1 Hydrological

3.1.1 Boundaries

3.1.1.1 Basins

A watershed, also known as a drainage basin or catchment area, is an area of land where all the water that falls or drains into it ultimately drains into a common outlet, such as a river, lake, or ocean. This includes surface water runoff, precipitation, and even underground water flow. Watersheds play a crucial role in the hydrological cycle, influencing the distribution and movement of water across the landscape.

Based on the size, the hydrological units are termed as basin, sub-basin, catchment, sub-catchment, watershed, sub-watershed and micro-watershed. India Water Resources Information System (WRIS) divides the Indian subcontinent into 25 major river basins with major river basin of Ganga-Brahmaputra-Meghna followed by Indus, Mahanadi, Godavari and Krishna.

Basin boundaries for India were obtained from WRIS ArcGIS portal. The layers were fetched on QGIS by using the url of WRIS ArcGIS portal in the QGIS's ArcGIS rest server. Basin boundaries across India are shown in figure 3.1

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset: Google drive folder

3.1.1.2 Subbasins

The subbasin boundary layer obtained from WRIS ArcGIS portal had attribute 'sbcode' which refers to the sub-basin code to which the watershed belongs. The watershed layer was dissolved on the 'sbcode' attribute to combine all the watersheds belonging to the same sub-basin. Subbasin boundaries across India are shown in figure 3.1

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset: Google drive folder

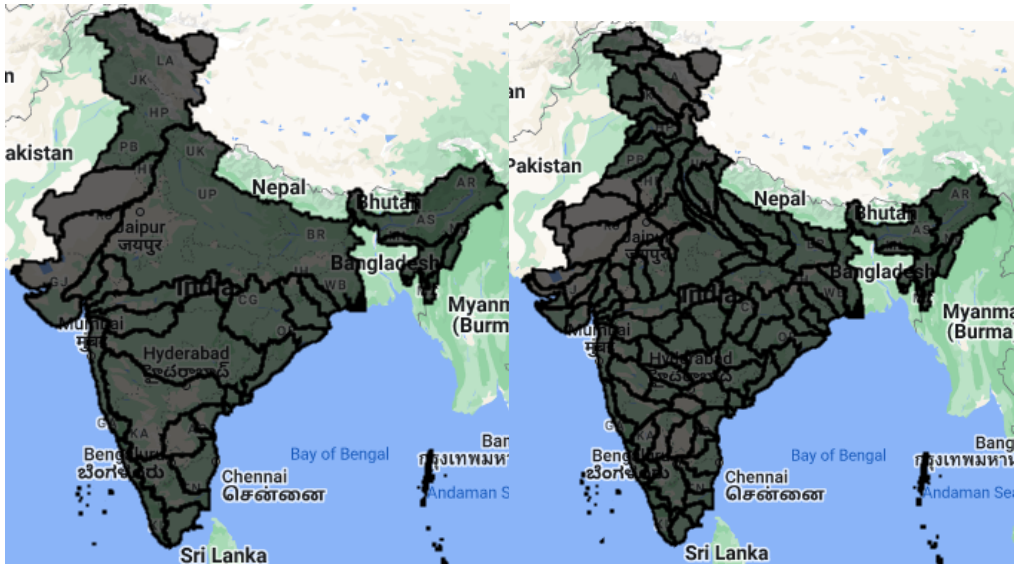


Figure 3.1: Basin (left) and subbasin (right) boundaries across India

3.1.1.3 Watersheds

Watershed boundaries for India were obtained from WRIS ArcGIS portal. The layers were fetched on QGIS by using the url of WRIS ArcGIS portal in the QGIS's ArcGIS rest server. Watershed boundaries across Mohanpur block, in Gaya district of Bihar are shown in the figure 3.2

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset: Google drive folder

3.1.1.4 Microwatersheds

Within each watershed, micro watersheds were delineated using the land elevation data. We used digital elevation model (DEM) data of Shuttle Radar Topography Mission (SRTM) for land elevation. SRTM was an international research effort that obtained digital elevation models with a spatial resolution of 1 arc-second (30 m) for global coverage. The DEM data can be obtained for a region of interest using the QGIS SRTM Downloader plugin and can be clipped for every watershed.

Watershed usually covers an area of about 20,000 to 1,50,000 ha that contains many micro- watersheds (500 to 1500 ha size). Therefore, micro-watersheds were generated within each watershed with a minimum area threshold of 5555 pixels (500 ha divided by 900 m^2, 1 pixel = 30m by 30m) as an input parameter to the r.watershed function of GRASS library in QGIS. Microwatershed boundaries across Mohanpur block, in Gaya district of Bihar are shown in the figure 3.2

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset: Google drive folder

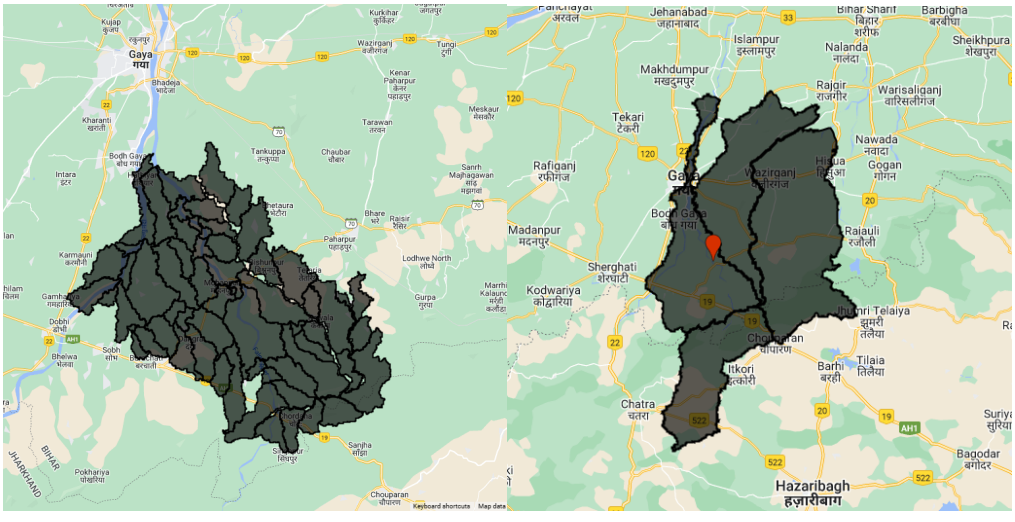


Figure 3.2: Microwatershed (left) and watershed (right) boundaries across Mohanpur block, in Gaya district of Bihar

3.1.1.5 Upstream-downstream microwatersheds

Introduction Identifying upstream and downstream microwatersheds in a basin is useful for planning water structures in an equitable manner. Upstream microwatersheds contribute water to downstream areas, affecting the surface runoff and groundwater recharge. If upstream areas are harvesting more surface runoff or over-extracting groundwater, downstream farmers may face water shortages. Conservation efforts like check dams and afforestation are usually performed in upstream areas to enhance downstream water availability.

Methodology A micro-watershed has a single pour point as its outlet connecting to a downstream micro-watershed but can receive inflow from multiple upstream micro-watersheds. We identified pour point of each micro-watershed using flow accumulation. Each pour point was mapped to its downstream micro-watershed to which it flows, resulting in a micro-watershed connectivity network in a basin. The figure 3.3 shows an instance of micro-watershed connectivity network.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Codebase: Github repository

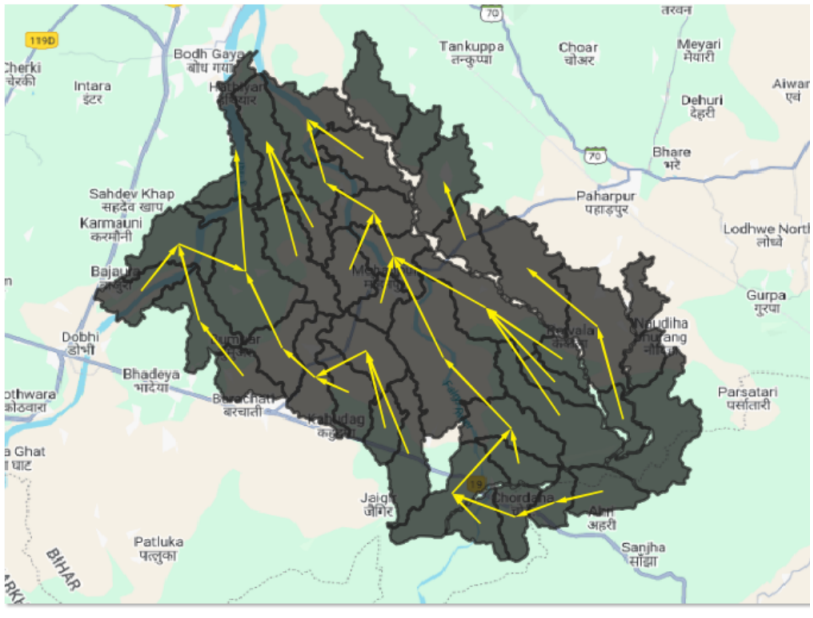


Figure 3.3: Upstream-downstream connectivity of micro-watersheds across Mohanpur block in Gaya district of Bihar

3.1.2 Variables

3.1.2.1 Precipitation

Introduction In the context of water balance, precipitation is the key input that contributes to the positive side of the water balance, representing the input of water to a particular region. Global precipitation data is available from satellite based products for last two decades which is critical for water resource management, agricultural production and addressing issues related to drought or flood assessment in a given area.

Input layers Precipitation can be calculated during an interval using the Global Satellite Mapping of Precipitation (GSMaP) dataset available on Google Earth Engine's data catalogue [17]. GSMaP provides a global precipitation in mm/hr at spatial resolution of approximately 11km. GSMaP is a product of the Global Precipitation Measurement (GPM) mission, which uses multi-band passive microwave and infrared radiometers from a constellation of satellites to estimate hourly precipitation.

Methodology Given a time duration and a region of interest, we filter the daily images of dataset between the start and end dates of the duration and add these images to get total precipitation of the duration on Google Earth Engine. The total precipitation image is clipped for the region of interest and a mean is performed across all the pixels in the region of interest to derive total precipitation of the region of interest for that duration. We derive fortnightly precipitation for each microwatershed across the five blocks. The spatial variation of precipitation is shown figure 3.6.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: fortnightly
- Codebase: Github repository

3.1.2.2 Runoff

Introduction Surface runoff refers to the flow of water that occurs when excess rainwater cannot be absorbed by the soil or vegetation and, as a result, flows over the Earth's surface. It is a major component of watewr cycle and is also known as overland flow. The spatio-temporal variation of runoff generation plays an important role in planning rainwater harvesting structures for effective water resource management, agricultural production and flood control. Remote sensing and GIS based layers such as land cover and soil type are very useful input data to runoff modelling. Runoff modelling can be grouped into three categories: empirical black-box, lumped conceptual and distributed physically based modelling [51]. Empirical models lack well defined representation of physical processes that converts a part of rainfall into runoff and distributed physically based modelling are highly data intensive. Therefore, we focus on conceptual modelling of surface runoff by using Soil Conservation Service Curve Number (SCS-CN) method. The SCS-CN method is widely used to convert rainfall to surface runoff using the curve number and involves the relationship of land cover, hydrologic soil groups, slope and 5-day antecedent rainfall.

Input layers

- **Precipitation data**: using the GSMaP dataset [17] at a spatial resolution of 11Km.
- **Soil type**: using the HYSOGs250m dataset [16] that represents a global gridded dataset of hydrologic soil groups (HSGs) with a spatial resolution of 250m, developed by USDA to support curve-number runoff modeling. Hydrologic soil groups A, B, C, and D correspond to low, moderately low, moderately high, and high runoff potential, respectively.
- **Slope**: using the NASA SRTM DEM dataset [34] at a spatial resolution of 30m.
- **Land cover**: using the Dynamic World dataset [13] which has a spatial resolution of 10m and includes nine categories of land covers such as water, trees, grass, flooded vegetation, crops, shrub and scrub, built-up, bare and snow .

Methodology The SCS-CN method uses curve number along with previous 5-day rainfall for surface runoff estimation. Curve number is dependent on the type of land cover, hydrologic soil group and slope as shown in the flowchart 3.4. Antecedent moisture condition (AMC) is calculated based on the total previous 5-day rainfall. AMC is categorized into three categories: AMC-I that refers to dry condition, AMC-II that refers to normal condition and AMC-III refers to wet condition. Curve number for AMC-II (CN2) is determined for each combination of land cover type and HSG [71]. Slope is also an important factor for determininig the movement of runoff and therefore slope adjusted CN2 values using Sharpley and Williams method [73] are derived, namely CN_2s. Curve numbers for AMC-I (CN1) and AMC-II (CN3) can be calculated using CN2 values. The higher the AMC, the higher the CN and therefore the surface runoff.

The NRCS-CN method modified by antecedent moisture (M) to estimate runoff [60] is defined as:

$$Q = \frac{(P - I_a)(P - I_a + M)}{(P + I_a + S_r + M)} \tag{3.1}$$

$$M = 0.5 * (-S_r + \sqrt{S_r^2 + 4P_5S_r}) \tag{3.2}$$

$$I_a = \lambda * S_r \tag{3.3}$$

$$S_r = (\frac{25400}{CN_s} - 254) \tag{3.4}$$

$$CN_{2s} = \frac{1}{3} * (CN_3 - CN_2)(1 - 2e^{-13.86s}) + CN_2 \tag{3.5}$$

$$CN_3 = CN_2 e^{0.00673(100-CN_2)} \tag{3.6}$$

$$CN_{1s} = \frac{4.2CN_{2s}}{10 - 0.058CN_{2s}} \tag{3.7}$$

$$CN_{3s} = \frac{23CN_{2s}}{10 + 0.13CN_{2s}} \tag{3.8}$$

where P_5 is the antecedent 5-day total rainfall, P is the total precipitation, I_a is the initial abstraction, Q is the direct run-off, S_r is the potential maximum retention, and λ is the initial abstraction coefficient ($\lambda = 0.2$). Here, CN_2 and CN_3 are the SCS curve numbers for AMC-II and AMC-III conditions, and s(mm^1) is the soil slope. CN_{1s} and CN_{3s} are slope adjusted curve numbers for AMC-I and AMC-III conditions.

The above datasets with different spatial resolutions were scaled up and scaled down to compute runoff at 30m spatial resolution on daily basis. Given a duration and a region of interest, we generate runoff rasters between the start and end dates of the duration and add these rasters to get total runoff of the duration. The total runoff raster is clipped for the region of interest and a mean is performed across all the pixels in the region of interest to derive total runoff of the region of interest for that duration. We derive fortnightly runoff for each microwatershed across the five blocks. The spatial variation of runoff is shown figure 3.6.

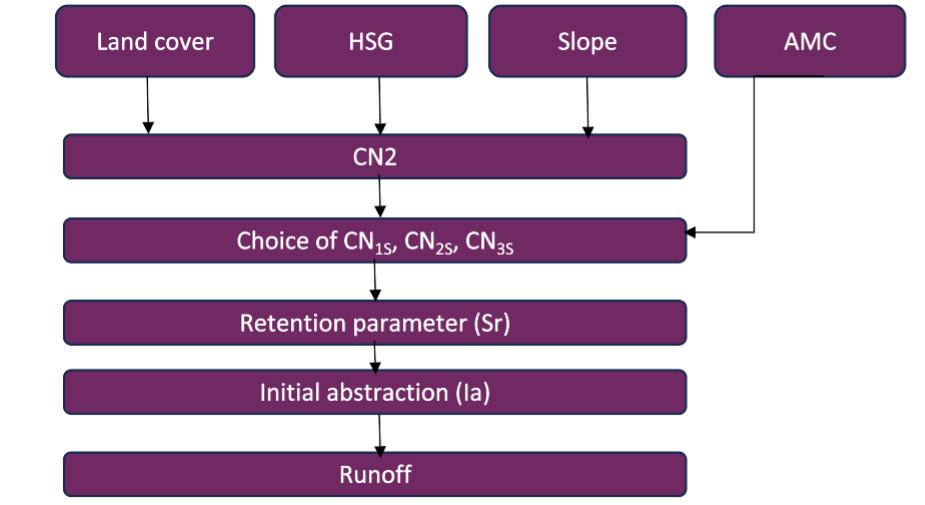
Table 3.1: Categories of AMC conditions

Antecedent rainfall	**AMC condition**
0 < rainfall previous 5days ≤ 35 mm	AMC-I
35 < rainfall previous 5 days ≤ 52.5 mm	AMC-II
rainfall previous 5days > 52.5 mm	AMC-III

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: fortnightly
- Codebase: Github repository

Figure 3.4: Methodology to estimate runoff



3.1.2.3 Evapotranspiration

Introduction Actual evapotranspiration plays a crucial role in the water balance of a region, contributing to the movement of water from the earth's surface to the atmosphere. It returns more than 60% of the total precipitation from land to atmosphere [49]. Evapotranspiration (ET) is the combination of transpiration from vegetation (Et), evaporation from bare soil (Eb), evaporation of intercepted precipitation from vegetated canopy (Ec), and evaporation from open water surfaces (Eo) [43]. Evapotranspiration is estimated using potential/reference evapotranspiration which provides the evaporative demand of atmosphere based on meteorological parameters and is independent of crop type, crop growing phase and management practices [12].

There are numerous remote sensing-based ET products that are available in the public domain and updated regularly for the user community. The products can be broadly classified into three categories [70] based on the methodology of computing ET: Penman-Monteith equation[61], Priestley-Taylor [67] equation and models that performs surface energy balance to compute evapotranspiration [75]. Remotely sensed meteorological parameters such as air temperature, net radiation and LULC are commonly used as inputs/forcings to land surface models to estimate evapotranspiration over large areas [56].

We used ET product from FLDAS Noah land Surface model [14] developed by NASA and USGS. The choice of product was determined by its spatio-temporal resolution, spatial coverage and its update frequency for data access. Noah ET is the sum of Ec, Et and Eb weighted by respective land surface coverage fractions. Noah's potential evapotranspiration (PET) is computed using Penman approach where PET is scaled by variables of canopy, transpiration and soil soil moisture availability to compute Ec, Et and Eb respectively.

Input layers We use FLDAS NOAH dataset [14] that provides the ET on a daily basis at a spatial resolution of 0.01 degrees (1.31 km).

Methodology Given a time duration and a region of interest, we filter and download the daily images of dataset between the start and end dates of the duration by using the python APIs of NSIDC DAAC's Data Access [1]. We add these images to get total evapotranspiration of the duration. The total evapotranspiration image is clipped for the region of interest and a mean is performed across all the pixels in the region of interest to derive total evapotranspiration of the region of interest for that duration on Google Earth Engine. We derive fortnightly evapotranspiration for each microwatershed across the five blocks. The spatial variation of evapotranspiration is shown figure 3.6.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: fortnightly
- Codebase: Github repository

3.1.2.4 Downscaled evapotranspiration

Introduction Over the years, the amount of cropping and by extension the water used for cropping has intensified greatly. The evapotranspiration hydrological variable is a good proxy to measure crop water usage in different regions. The coarser spatial resolutions of global ET products may not be suitable to derive field level ET estimates such as for farmlands which are irrigated using groundwater pumping. The highest resolution of operational global ET products is at 500m (MODIS). Estimates of ET at finer resolution can provide ET estimates due to different land covers such from tree canopies and from bare soil as the operational ET products provide total estimate of ET across all the land covers in a grid.

We downscaled MODIS ET at 500m to 30m using machine learning model to observe finer relative differences in ET values and facilitate local decision making. We validate the downscaled ET across multiple geographic locations and also study its correlation with in situ evapotranspiration measurements using a network of 48 lysimeter stations established by Indian Meteorological Department (IMD) across the country to measure daily ET since 2000.

Methodology We use satellite data derived from the Landsat satellite and also meteorological variables from the GLDAS system from NASA and we use these as features for our model. These are available at different spatial and temporal resolutions.

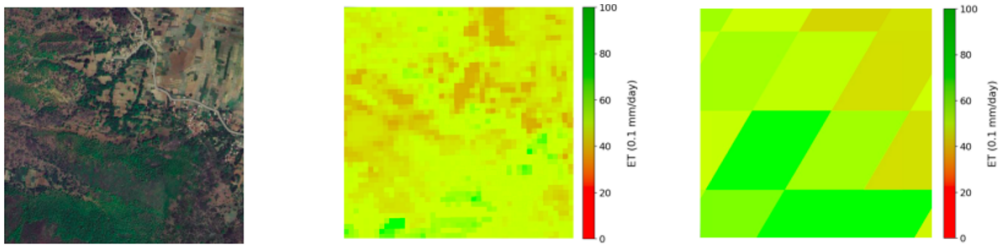
For aligning them spatially, all the variables were first brought to a resolution of 500 meters by aggregation of their rasters. The resolution of 500 meters was chosen because the MODIS dataset is available at this resolution. For aligning them temporally, we only take data points for which both MODIS and Landsat data are available and data acquisition for their rasters begin on the same day.

We then train a random forest model for each Agro-Ecological-Zones using these input features on the MODIS dataset. For inference of this trained model, we aggregate the input features to a resolution of 30 meters now. These features are provided as an input to the trained RF model which can then return an output raster image at a resolution of 30 meters.

Hosting specifications

- Layer type: raster
- Spatial resolution: 10m
- Temporal resolution: Landsat
- Codebase: Github repository 1, Github repository 2

Figure 3.5: Visualisation of 30 meter resolution ET product (middle) predicted by the trained Random Forest Model for a region from AEZ 10 of area 2 km $\times$ 2 km against a satellite image (left) and MODIS ET image (right).



3.1.2.5 Change in groundwater

Introduction Water balance in the discipline of hydrology that aims to estimate the unknown water fluxes. It is an equation which is expressed in terms of water inputs, outputs and storage in a hydrological unit such as watershed [55]. We intend to estimate the net change in groundwater on a fortnightly basis for a micro watershed by solving the water balance equation.

With the current groundwater state in hand, the objective is to improve the groundwater state through interventions such as construction of rainwater harvesting structures, change in cropping patterns and growing plantations. The interventions will affect different components of the water balance equation and will allow us to project the groundwater state in future. For example, change in cropping patterns will affect evapotranspiration in the water balance equation. Groundwater projections will facilitate scientific and participatory planning within the community.

Input layers We used the datasets of precipitation, runoff and evapotranspiration mentioned in section 4.1, 4.2 and 4.3 to compute change in groundwater as shown below for a micro-watershed in a fortnight.

Methodology Water balance equation will take precipitation (P), outgoing runoff (Q_{out}) and evapotranspiration (ET) as inputs to output change in groundwater (ΔG) as shown in the equation 9.

$$P = Q_{out} + ET + \Delta G(unknown) \tag{3.9}$$

We assume incoming runoff, change in soil moisture and change in surface water storage as zero in the water balance and intend to model them in future. Each of these water balance inputs are derived using remote sensing products in order to diagnose the groundwater state of a micro-watershed on fortnight basis as shown in the figure 3.7 and the spatial variation of change in groundwater is shown figure 3.6.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: fortnightly
- Codebase: Github repository

Figure 3.6: Spatial variation in precipitation, runoff, evapotranspiration and change in groundwater during 2022-23 across microwatersheds of Mohanpur block, Gaya district, Bihar

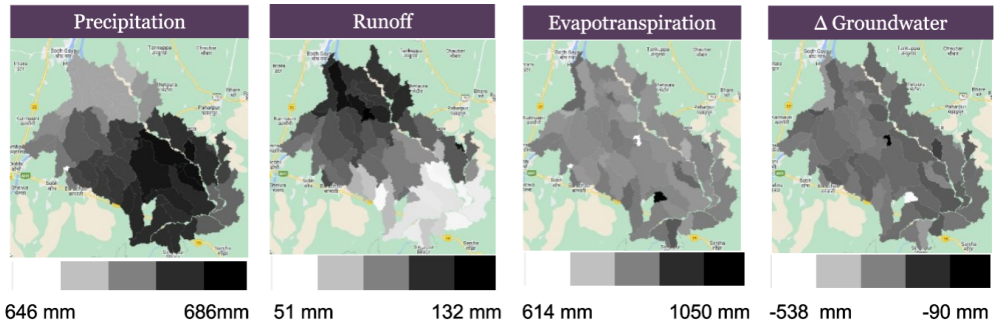
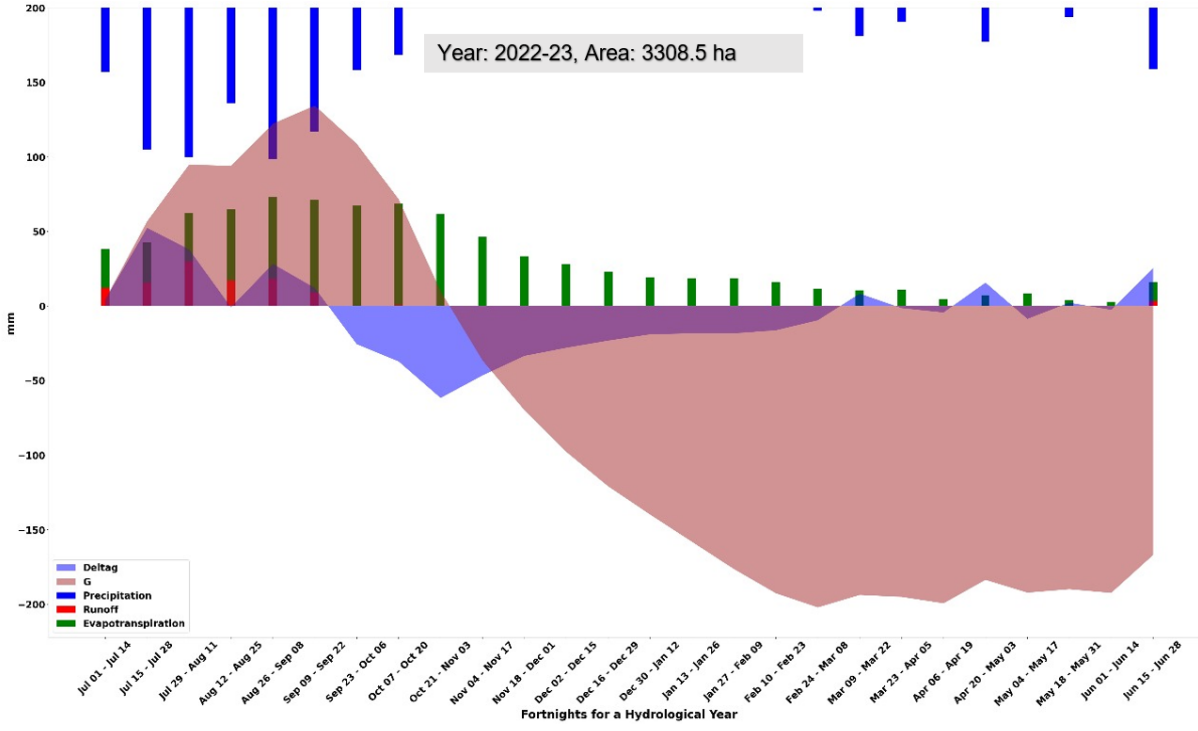


Figure 3.7: Time series of water balance variables: precipitation, runoff, evapotranspiration and change in groundwater (Deltag) on fortnight basis between 2022-2023 for a microwatershed in Mohanpur. Time series of groundwater (G) is derived by assuming the groundwater in the first fortnight as zero in the worst case.



3.1.2.6 Change in well depth

Introduction The change in groundwater in an aquifer is related to the change in well depth through the specific yield of the aquifer [28]. Specific yield is a key parameter in hydrogeology that represents the fraction or percentage of groundwater that a saturated rock or soil can release under the influence of gravity. It is a measure of the storage and release capacity of an aquifer or a formation. Mathematically, specific yield is defined as the ratio of the volume of water drained from an aquifer by gravity to the total volume of the aquifer.

Input layers We used the change in groundwater layer mentioned in section 4.4 and aquifer layer to compute change in well depth for each micro-watershed. Aquifer mapping was performed in the entire country by Ministry of Jal Shakti in India under Aquifer Mapping and Management Programme (NAQUIM) [5]. NAQUIM was taken up by Central ground Water board (CGWB) to take appropriate water resource management measures and understand the underlying ground water potential and recharge. Aquifer layer for India was obtained from India Water Resources Information System (WRIS) ArcGIS portal. The CGWB has divided India across 14 principal aquifers and 42 major aquifers [5].

Methodology The equation 10 describes the relationship between change in ground water storage of an aquifer and change in groundwater level [28].

$$\Delta G = \Delta h * A * S_y \tag{3.10}$$

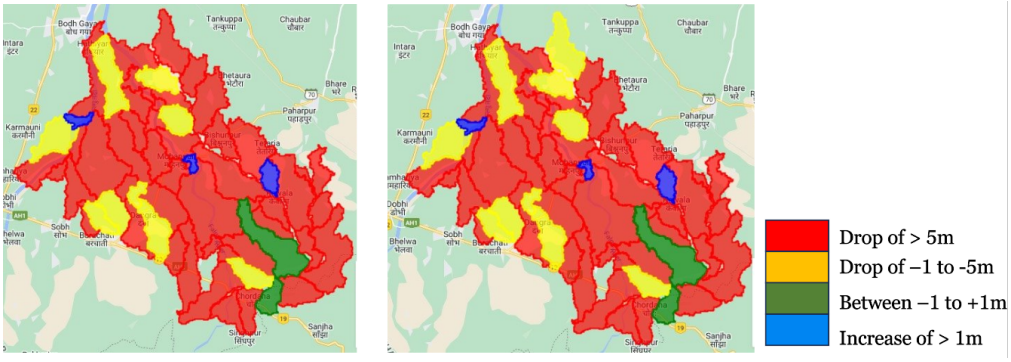
where, ΔG is change in groundwater volume, Δh is change in well depth, A is the cross-sectional area of the aquifer and S_y is the specific yield of aquifer. The ΔG in the equation assumes the change in the groundwater volume due to vertical fluxes and horizontal fluxes such as lateral flows across the boundaries of the aquifer is assumed to be zero. For example, alluvium aquifer shows fluctuation in groundwater by both lateral and vertical fluxes [28] while basalt aquifer which has medium to low permeability shows groundwater fluctuations due to vertical fluxes [28]. Therefore, to bring more accuracy for ΔG, we will try to fine-tune the water balance equation as per different aquifer types.

Currently, the aquifer layer is clipped for each microwatershed to determine its specific yield. The change in well depth is computed over five year window for each microwatershed using the change in groundwater over 5 years and its specific yield. The figure 3.8 shows the variation of change in well depth across the micro-watersheds of Mohanpur between 2017-22 and 2018-23.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: five years
- Codebase: Github repository

Figure 3.8: Change in well depth across microwatersheds of Mohanpur during 2017-22 (left) and 2018-23 (right)



3.1.2.7 Aquifers

Introduction For increasing economic growth and meeting its food security goals, India has relied heavily on groundwater by becoming largest consumer of groundwater in world. About more than 80% of rural drinking and more than 50% of agriculture's irrigation demand is met by groundwater [5]. Unplanned over exploitation of this valuable resource has led to drying of aquifers. Demarcation of aquifer systems across the country is essential to assess the ground water resource. In view of this, Central Ground Water Board (CGWB) of India carried out an exercise of mapping the lateral and vertical extent of aquifer systems across the nation and created a geospatial data that can be imported on GIS platforms. Based on the hydrogeological characteristics, the entire nation is divided into 14 principal aquifers which are in turn divided into 42 major aquifers [5] by rock types and their age as shown in the figures 3.9, 3.10 respectively.

Methodology The aquifer shapefile for India was obtained from India Water Resources Information System (WRIS) ArcGIS portal. Each geometry in the shapefile represents a major aquifer uniquely identified by its region-specific aquifer properties such as aquifer thickness, specific yield, transmissivity and whether it is a unconfined and confined aquifer. Aquifer properties and their description can be found here [4].

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset: Google drive folder

Figure 3.9: Principal aquifer systems of India

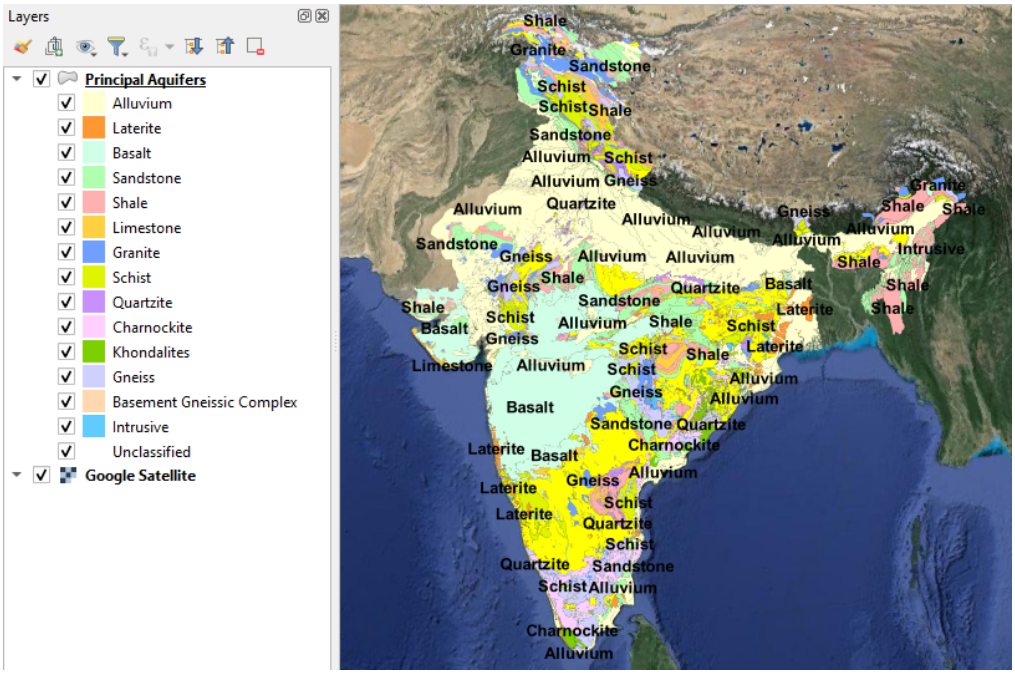
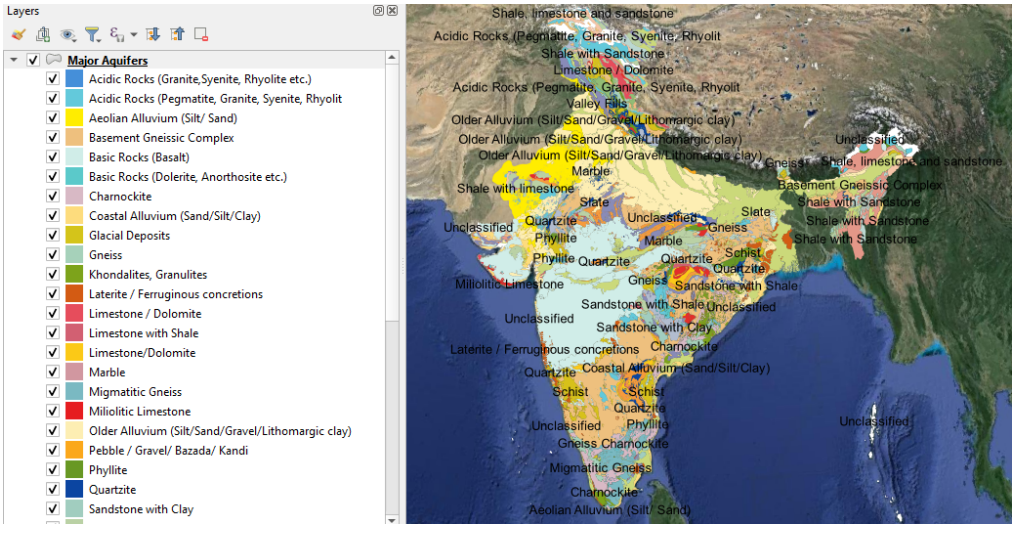


Figure 3.10: Major aquifer systems of India



3.1.2.8 Groundwater level in dugwells

Introduction Ground water levels are being measured by Central Ground Water Board (CGWB) four times a year: January (post monsoon), March/April/May (pre monsoon), August (post monsoon), November (post monsoon). The monitoring started in the year 1969 has network of 23,125 observation dug wells by 2017 located across India [10]. The data forms the basis for groundwater development and can be analyzed to study the spatial and temporal changes in ground water levels.

Methodology We downloaded the well locations and time series of groundwater level for each well location from WRIS [11].

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: ground water level data across four seasons during 2001-23.
- Dataset: Google drive folder

3.1.2.9 Groundwater level in borewells

Introduction Atal Bhujal Yojana is a central government scheme in India aimed to improve groundwater in over-exploited and ground water stressed states such as in Gujarat, Haryana, Karnataka, Madhya Pradesh, Maharashtra, Rajasthan, and Uttar Pradesh [6]. The scheme monitors groundwater levels in deeper aquifers than shallow aquifers to track depletion trends, assess the impact of interventions, and promote data-driven groundwater management. Groundwater level is also known as piezometric head and therefore piezometers and borewells are used to monitor groundwater levels in deeper aquifers [19].

Methodology Atal Bhujal Yojana has published its groundwater level data for pre and post-monsoon seasons from 2015 to 2022 on the Open Government Data Platform of India [18]. This dataset includes borewell and piezometer level time series of groundwater levels present in Gujarat, Haryana, Karnataka, Madhya Pradesh, Maharashtra, Rajasthan, and Uttar Pradesh.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: ground water level data pre and post monsoon from 2015 to 2022
- Dataset: Google drive folder

3.1.2.10 Stage of Groundwater Extraction

Introduction Groundwater resource assessment is carried out annually by State and Central Ground Water Board (SGWBs and CGWB) in India to compute annual extraction of groundwater resource and annual extractable groundwater resource at Talukas/Blocks/ Mandals/Firkas level to derive annual stage of groundwater extraction. Periodic assessment of groundwater resources is essential for securing India's agriculture and drinking water security. Groundwater levels are rapidly declining in many regions of India due to excessive withdrawal. Data on the state of groundwater resource at block level can facilitate prioritization and planning of groundwater management interventions such as planning recharge structures under MGNREGA and changes in cropping patterns (high to low water intensive crops). The Stage of Groundwater Extraction for assessment units across India for the assessment year 2023-24 is shown in the figure 3.11

Methodology The stage of groundwater extraction is defined as the ratio of annual extraction of groundwater resource and annual extractable groundwater resource [29]. The categorization is defined by Stage of Ground Water Extraction as mentioned in table 3.2. Annual extraction of groundwater resource is estimated using groundwater extraction for irrigation, domestic and industrial uses. Annual extractable groundwater resource is estimated by deducting the annual natural discharge from annual groundwater recharge. Annual groundwater recharge from rainfall accounts for approximately 61% of the annual groundwater recharge. Additional recharge comes from other sources, including canals, surface water irrigation, groundwater irrigation, tanks, ponds, and water conservation structures in both command and non-command areas.

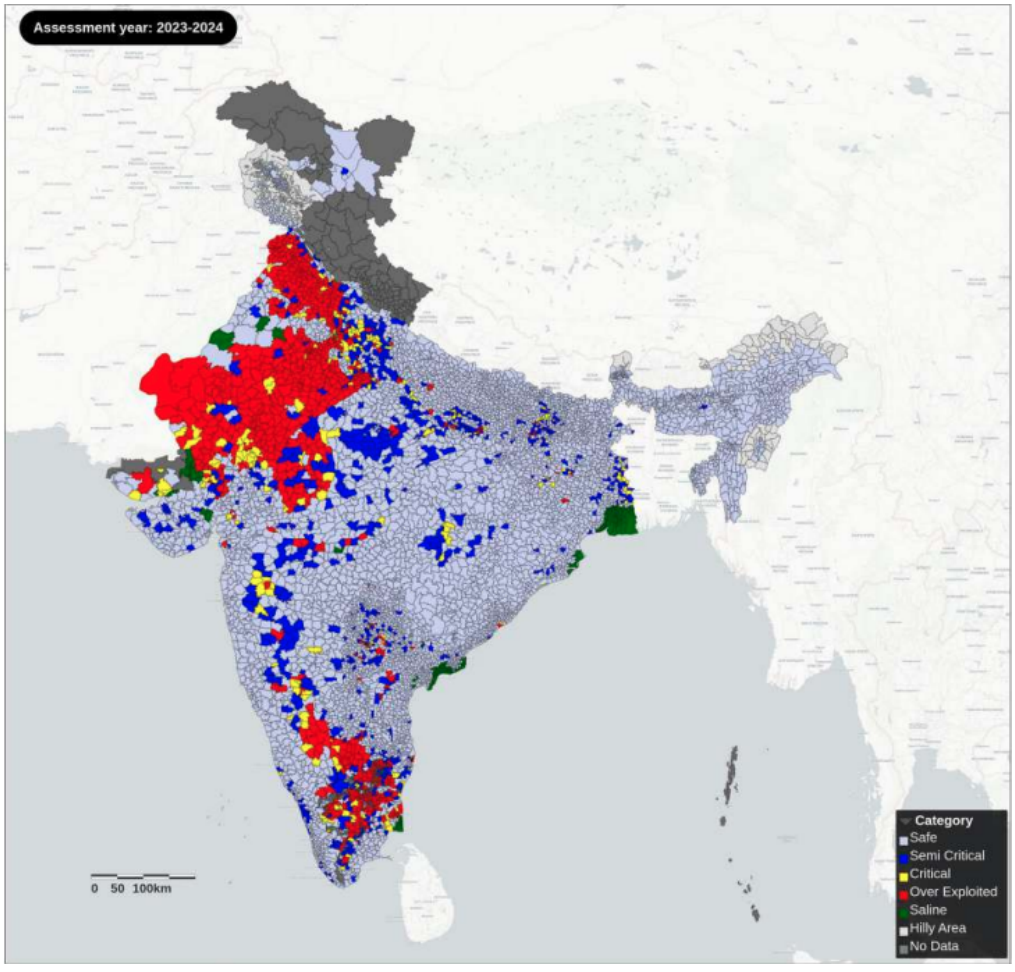
Table 3.2: Categories of Stage of Groundwater Extraction

Stage of Groundwater Extraction	**Category**
$\leq 70\%$	Safe
$> 70\%$ and $\leq 90\%$	Semi-critical
$> 90\%$ and $\leq 100\%$	Critical
$> 100\%$	Over Exploited

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: available for years 2011, 2013, 2017 and 2020
- Dataset: Google drive folder

Figure 3.11: Stage of Groundwater Extraction of assessment units (Talukas/Blocks/ Mandals/Firkas)



3.1.2.11 Rivers

Introduction The river systems of India can be classified into four groups [20]: Himalayan rivers, Deccan rivers, Coastal rivers and Rivers of the inland drainage basin. Himalayan rivers are perennial, fed by melting glaciers in summer and swollen by monsoon rains, often causing floods. The Gangetic basin, India's largest river system, drains nearly a quarter of the country. Peninsular rivers are rain-fed, with reduced flow in summer and revived during the monsoon. The Godavari basin, India's largest in the peninsula, covers nearly one-tenth of the country. Coastal rivers gush down the peaks of the western ghats into the Arabian sea during monsoon. Seasonal streams such as Sambhar in western Rajasthan and Luni in the Rann of Kutch drains into inland basins, salt lakes and salt desserts. The rivers across India is shown in figure 3.12.

Methodology The river shapefile for India was obtained from India Water Resources Information System (WRIS) ArcGIS portal.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset: Google drive folder

3.1.2.12 Canals

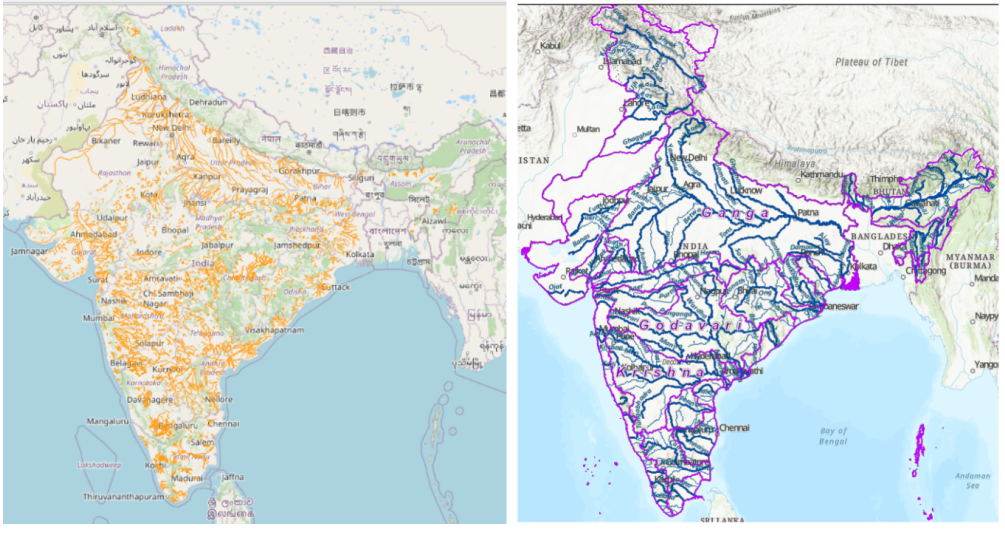
Introduction To provide irrigation to agriculture fields, either reservoirs are made on the rivers by putting dams across the rivers to store and divert river water or river water is simply diverted using canal network. In India, both reservoir-fed canals and river diversion canals are in existence. Canals are typically classified as "Main Canal," "Branch Canal," "Major Distributary," and "Minor Distributary," based on their discharge carrying capacity [27]. "Main Canal" is the primary channel diverting water from a river, followed by progressively smaller branches distributing water to fields. The canal network across India is shown in figure 3.12.

Methodology The canal shapefile for India was obtained from India Water Resources Information System (WRIS) ArcGIS portal. Each geometry in the shapefile identified by its canal type as mentioned above.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset: Google drive folder

Figure 3.12: Canal (left) and River (right) network across India



3.2 Climate variables

3.2.1 Drought frequency and intensity

Introduction The impact of drought on Indian agriculture is significant, given the country's heavy reliance on rainfall for crop production. Drought significantly affects the crop yields and crop failure in worst case resulting in significant economic losses for farmers that can exacerbate perpetuating cycles of poverty and vulnerability. Drought-related crop failures can impact food security at both local and national levels, leading to reduced availability of food supplies and potential price spikes. This can disproportionately affect vulnerable populations, exacerbating hunger and malnutrition in affected regions. In the recent past, India has witnessed drought in 1987, 2002 and 2009 with an overall rainfall deficiency of 19%, 19% and 22% respectively which affected 59–60% of the normal cropped area alone in 1987. During 2014 and 2015, the drought encompassed major agricultural states in the country causing widespread hardships among the population.

Drought is a recurrent and sporadic feature of climate, which stems from significant rainfall deviation from normal in its spatial and temporal distribution such that it adversely impact the crops in a cropping season or successive cropping seasons. The intensity of drought depends on several factors including agro-climatic features, cropping choices and patterns, socio-economic vulnerabilities of the local population etc. The Department of Agriculture, Cooperation & Farmers Welfare under the Ministry of Agriculture & Farmers Welfare, Government of India has released the "Manual for Drought Management" in 2016 for the prevention, mitigation and management of droughts in India. The manual introduces scientific indices and parameters such as Vegetation Condition Index (VCI) for more accurate determination and assessment of drought. We have tried to implement the methodology of drought manual using various remote sensing products. The choice of product was determined by its spatial coverage, its spatio-temporal resolution and its near real time availability.

Input layers Based on the wide consultation with domain specialists, the drought manual [21] recommends five categories of indices [21] for determining drought incidence and severity: Rainfall, Vegetation, Water, Crop and others. Rainfall related indices include rainfall deviation, dry spell and standard precipitation index (SPI). Vegetation based indices such as Vegetation Condition Index (VCI) or NDVI/NDWI deviations from normal are derived from remote sensing data. Crop situation related indices include area under sowing and soil moisture based indices such as Percent Available Soil Moisture (PASM) and Moisture Adequacy Index (MAI). Hydrological indices include Reservoir Storage Index (RSI), Groundwater Drought Index (GWDI) and Stream-Flow Drought Index (SFDI). Other factors include socio-economic indicators such as prices of essential commodities as compared to normal prices, scarcity of drinking water supply, agricultural wages etc. The following section outlines the methodology of using the above variables in the determination of drought at microwatershed level subject to the availability of data. We use the following publicly available datasets for building drought indicators:

- **Rainfall data** - Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) rainfall dataset with a spatial resolution of 5.566km and a temporal resolution of 1 day.
- **Vegetation data**- NDVI and NDWI are derived from Landsat dataset with a spatial resolution of 30m and a temporal resolution of 14 to 26 days.
- **Crop data**- Evapotranspiration and potential evapotranspiration required for computing MAI indicator were derived from MODIS dataset with a spatial resolution of 500m

and a temporal resolution of 8 days.

- **Land use land cover**- as described in section 5.0.2

Methodology We compute weekly and monthly rainfall deviation using a long period average of that particular week or month across 30 years as shown in equation 11 and 12. The departure of rainfall from its long period average is considered as a credible indicator of drought as shown in table in figure 3.14. Weekly rainfall deviations are used to determine dry spell. The occurrence of dry spell is defined across four consecutive weeks with each week of these four weeks incurring a rainfall deviation of less than 50%. Prolonged dry spells can lead to significant reduction in crop sown area. SPI is computed on monthly scale and expresses the current monthly rainfall as a standardized departure with respect to rainfall probability distribution function. Positive SPI indicate deviations greater than median precipitation and negative SPI indicate deviations less than median precipitation as shown in table in figure 3.14. Rainfall indicators such as rainfall deviation, dry spell and SPI mentioned above are considered to be mandatory [21] for the first drought trigger to set off as shown in table in figure 3.14.

In the event where the first drought trigger is set off, the intensity of drought is assessed through its impact on agricultural area, crop condition, soil moisture and hydrological resources. Agricultural area is linked with the extent of sowing in the monsoon season. If the sowing fails due to delayed onset of monsoon or its deviation from normal, then it provides reliable information on the availability of water for agricultural operations. Impact on agricultural area is computed as the percentage of cropping area sown in Kharif season of that year. We compute the cropping area in kharif and total cropping area using the land use land cover classes: single kharif, single non-kharif, double and triple. Impact of drought on crop condition is often determined by spectral vegetation indices such as Normalized Difference Vegetation Index (NDVI) and Normalized Difference Wetness Index (NDWI), derived from remote sensing data. NIR and Red are the reflectance in visible and near infrared channels. NDVI values from 0.2 to 0.6 are associated with greater green leaf area and biomass while high NDWI is associated with more surface wetness. Observed NDVI and NDWI range is compared with historical range (using maximum and minimum) to derive monthly VCI for NDVI and NDWI respectively as shown in the equations 13, 14 and 15. VCI provides an idea about where the observed value is placed in the historical range as shown in table in figure 3.14. While combining VCI of NDVI and NDWI, the minimum of the two values is taken. Availability of moisture in soil is a very relevant indicator for drought as shown in table in figure 3.14 and is expressed as the ratio of Actual Evapo-transpiration (AET) to the Potential/Reference Evapo-transpiration (PET or RET) on monthly scale for MAI. The intensity of the drought will be contingent upon the values of at least three out of four impact indicators viz, agriculture, crop condition, soil Moisture and hydrology in the following manner:

1. Severe: if all the above 3 impact indicators are in 'Severe' category.
2. Moderate: if two of the above 3 impact indicators are in 'Moderate' or 'Severe' class.
3. Mild: for all other cases.

We divide the south-west monsoon season into weeks and then into months (each of 4 weeks), starting from the first week with a stride of one week. For example, if there are 'n' weeks in the monsoon season, then there are 'n-3' months in the same monsoon season. Drought trigger is examined for each month using the dry Spell, monthly rainfall deviation and monthly SPI values as shown in table 3.14. In case of drought trigger, the intensity of drought is determined by the impact indicators of that month as shown in the flow chart

3.14. Weekly drought intensity is determined by the maximum drought intensity across all the months in which the week is considered as shown in the figure 3.13.

Monthly rainfall deviation in i^{th} month

$$RFdev_{monthly} = \frac{RF_i - RF_n}{RF_n} \tag{3.11}$$

where RF_n is the long term mean rainfall in i^{th} month and RF_i is the current rainfall in month i.

Weekly rainfall deviation in i^{th} week

$$RFdev_{weekly} = \frac{RF_i - RF_n}{RF_n} \tag{3.12}$$

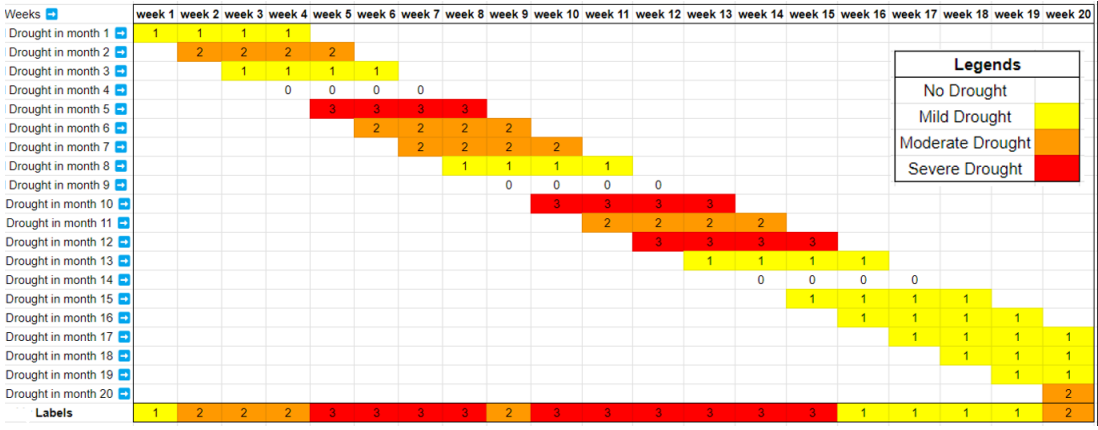
where RF_n is the long term mean rainfall in i^{th} week and RF_i is the current rainfall in week i.

$$VCI_{NDVI} = \frac{NDVI_{current} - NDVI_{minimum}}{NDVI_{maximum} - NDVI_{minimum}} \tag{3.13}$$

$$VCI_{NDWI} = \frac{NDWI_{current} - NDWI_{minimum}}{NDWI_{maximum} - NDWI_{minimum}} \tag{3.14}$$

$$VCI = minimum(VCI_{NDVI}, VCI_{NDWI}) \tag{3.15}$$

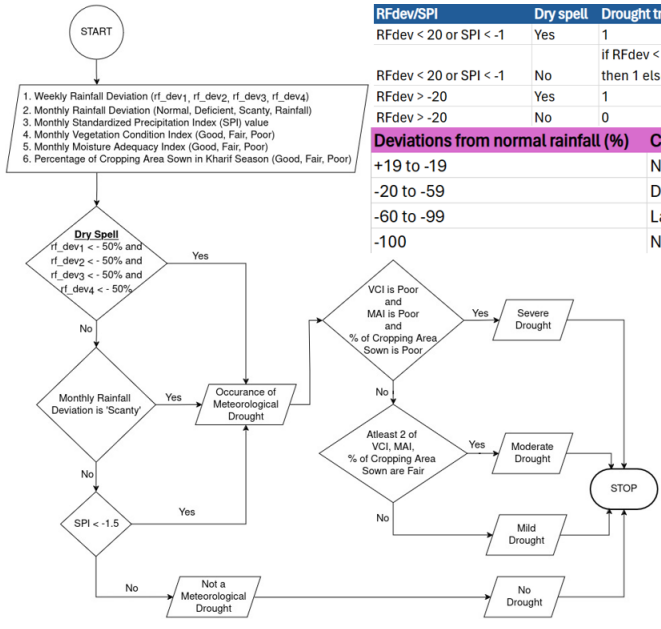
Figure 3.13: Drought intensity is determined for each monthly window (rows) in terms of mild, moderate, severe and no drought. Weekly drought intensity (columns) is determined by the maximum drought intensity that has occurred across the corresponding monthly windows in which the week is present.



Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: yearly
- Codebase: Github repository

Figure 3.14: Flow chart of determining monthly drought intensity



RFdev/SPI	Dry spell	Drought trigger
RFdev < 20 or SPI < -1	Yes	1
RFdev < 20 or SPI < -1	No	if RFdev < -60 or SPI < -1.5 then 1 else 0
RFdev > -20	Yes	1
RFdev > -20	No	0

Deviations from normal rainfall (%)	Category
+19 to -19	Normal
-20 to -59	Deficient
-60 to -99	Large deficient
-100	No rain

SPI value	Category
< -2	Extremely dry
-1.99 to -1.5	Severely dry
-1.49 to -1	Moderately dry
-0.99 to 0	Mildly dry
0 to 0.99	Mildly wet
1 to 1.49	Moderately wet
1.5 to 1.99	Severely wet
> 2	Extremely wet

MAI (%)	Category
76-100	No drought
51-75	Mild
26-50	Moderate
0-25	Severe

VCI (%)	Vegetation condition
60-100	Good
40-60	Fair
0-40	Poor

Crop area sown (%)	Category
66-100	Mild
33-66	Moderate
0-33	Severe

3.2.2 Drought causality

Introduction Identifying the triggers and impact of drought in case of drought declaration is crucial for the rural communities to design climate adaptation strategies. Drought declaration may happen for a year when the total number of moderate drought weeks and severe drought weeks is greater than 5 as per the "Manual for Drought Management" released by the Government of India in 2016. The final decision of drought declaration is usually made after field verification. Rainfall indicators such as rainfall deviation, dry spell and SPI act as triggers to set off drought while impact of drought assessed on agricultural area, crop condition, soil moisture and hydrological resources as mentioned in section 3.2.1. If some rainfall indicators are the source of trigger, rainwater harvesting may help. If soil moisture decline is the drought impact, then altering cropping patterns or promoting drought-resistant crops may be more effective. Similarly, poor crop health and reduction in cropping area were drought impacts then rural communities can prepare themselves with long-term adaptation strategies, such as diversifying income sources.

Input layers We compute drought triggers and drought impact indicators such as weekly and monthly rainfall deviation, monthly SPI, monthly VCI, monthly MAI and percentage of cropping area in kharif season as mentioned in section 3.2.1 and save them for each monsoon week. Based on the triggers and impact, each week may experience of the drought intensity such as no drought, mild drought, moderate drought or severe drought.

Methodology According to the drought methodology shown in figure 3.14, we derive the possible combinations of drought triggers and drought impact indicators that determine whether a week is classified as having moderate and severe drought intensity as shown in table 3.3. In the event that a year is declared a drought year, then we count the number of unique combinations of trigger and impact indicators for moderate and severe weeks and identify the top three most frequent combinations.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: yearly
- Codebase: Github repository

Table 3.3: Possible combinations of trigger and impact indicators for moderate and severe drought intensities.

Drought intensity	**Triggers**	**Impacts**
Severe	Dry spell	VCI under Poor category, MAI under Severe category and Percentage of cropping area under Severe category
	Rainfall deviation under Large deficient category	VCI under Poor category, MAI under Severe category and Percentage of cropping area under Severe category
	SPI under Severely dry or Extremely dry category	VCI under Poor category, MAI under Severe category and Percentage of cropping area under Severe category
Moderate	Dry spell	VCI under Fair or Good category, MAI under Severe category and Percentage of cropping area under Severe category
		VCI under Poor category, MAI under Moderate or Mild or No drought category and Percentage of cropping area under Severe category
		VCI under Poor category, MAI under Poor category and Percentage of cropping area under Moderate or Mild category
Moderate	Rainfall deviation under Large deficient category	VCI under Fair or Good category, MAI under Severe category and Percentage of cropping area under Severe category
		VCI under Poor category, MAI under Moderate or Mild or No drought category and Percentage of cropping area under Severe category
		VCI under Poor category, MAI under Poor category and Percentage of cropping area under Moderate or Mild category
Moderate	SPI under Severely dry or Extremely dry category	VCI under Fair or Good category, MAI under Severe category and Percentage of cropping area under Severe category
		VCI under Poor category, MAI under Moderate or Mild or No drought category and Percentage of cropping area of Severe category
		VCI under Poor category, MAI under Poor category and Percentage of cropping area under Moderate or Mild category

3.2.3 Monsoon onset

Introduction For rural communities in India, the onset of the monsoon is a crucial event because majority of rural farmers depend on rainfed agriculture. Early or delayed onset affects crop selection, sowing time, and yield potential. If the monsoon is delayed, farmers may opt for short-duration or drought-resistant crops to reduce risks. Delayed onset may necessitate preparation for water storage and irrigation planning.

Input layers We use Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) rainfall dataset with a spatial resolution of 5.566km and a temporal resolution of 1 day.

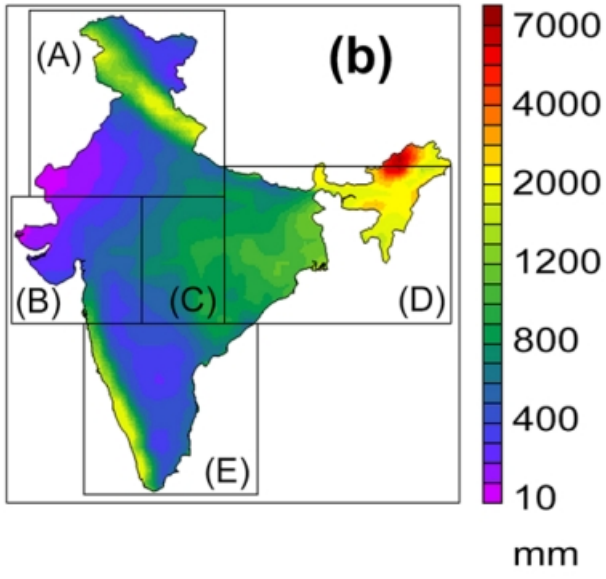
Methodology India is divided into five Hydro-Meteorological Zones (HMZs), based on historical trends of the hydro-meteorological seasons (monsoon, post-monsoon and pre-monsoon); [3, 41]. To determine the monsoon onset for a specific region, we first identify the HMZ it belongs to and retrieve the corresponding threshold value. The total weekly rainfall is then calculated for each week of the year. Beginning from the second week of May, we check for the first week where rainfall exceeds the threshold, marking that as the monsoon week. The monsoon onset date is then determined as the first rainy day within that week.

To quantify monsoon onset deviation, we analyze historical monsoon onset dates and compute the long-term mean as the average onset date across all past years. The deviation for any given year is obtained by calculating the difference between the current year's onset date and the long-term mean. A positive deviation indicates an earlier-than-average monsoon, while a negative deviation signifies a delayed onset. For further classification, weekly deviation bins are created. If the absolute deviation is less than seven days, it is categorized as Week 0 (zero weeks deviation). If the deviation falls between 7 to 14 days, it is classified as Week 1 (one-week deviation), and so on.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: yearly
- Codebase: Github repository

Figure 3.15: The figure shows five hydro-meteorological zones (A to E) delineated based on the duration of the hydro-meteorological seasons (monsoon, post-monsoon and pre-monsoon)



3.3 Terrain variables

3.3.1 Terrain classification

Introduction The Ministry of Rural Development, Government of India under the project of "Environmental benefits of MGNREGA", prescribes watershed development to create durable and sustainable assets [38]. The watershed development follows ridge to valley approach where the watershed is segregated into areas of different terrain for the construction of terrain specific assets to enhance the capacity of the assets for recharge and surface storage.

The CSO volunteers, following the ridge-to-valley approach identifies uplands, midlands and low lands in their landscape to intervene with specific type of assets in each of them. For example, Water Absorption Trenches (WATs) are constructed in uplands to harvest surface runoff, while farm ponds are constructed on lowlands to arrest surface and subsurface runoff for protective irrigation. The current process of identification is through visual inspection by observing the direction of flow of water during monsoon and assigning each farm plot on the cadastrap map with upland, midland or lowland. Cadastral Maps are digital land records maintained by the state government. An automated classification of landscapes into plains, broad slopes, steep slopes, ridges and valleys along with contour lines using digital elevation model may help in precision and efficiency of the process.

Terrain classification involves the categorization of Earth's physical features into distinct landforms such as hilltops, valley bottoms, ridges, plains, and slopes. Many natural processes such as hydrological balance and soil erosion are significantly correlated with the land forms.

Input layers We used NASA's SRTM Digital Elevation Model [34] at 30m resolution to generate landform classification.

Methodology We use landform classification method [77] by The Nature Conservancy to identify pixel level landform using Topographic Position Index (TPI) and slope at each pixel. The Topographic Position Index (TPI) compares the elevation of each pixel in a DEM to the mean elevation of a specified neighborhood around that pixel. Positive TPI values indicate positions that are elevated compared to the average of their neighboring terrain, such as ridges. Conversely, negative TPI values signify positions that are lower in elevation relative to their surroundings, like valleys. TPI values close to zero typically denote either flat areas or regions with consistent slopes.

The distance/scale at which the neighbourhood of TPI is defined, plays an important role in identifying the landform. For example, as shown in the figure 3.16, the TPI is near zero in A, while it is positive in B and negative in C. Therefore, both small-scale TPI and large-scale TPI are necessary to identify landform at a pixel. We threshold the range of small-scale TPI and large-scale TPI in a microwatershed by λ standard deviation and combine it with slope to determine landform classes as shown in the figures 3.17. We define λ using the equation 3.16 after fine-tuning the TPI algorithm for Indian landscapes such as Nalanda, Mohanpur, Masalia, Pindwara, Mandalgarh and Angul. We perform landform classification using 11 landform classes as shown in the figure 3.18.

$$\lambda = max(3 - log(\sigma_{dc} + 1), 0.3) \tag{3.16}$$

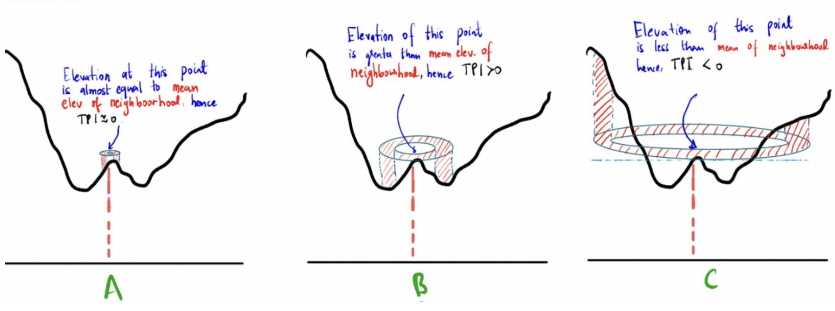
where σ_{dc} is the standard deviation of elevation in a microwatershed.

Hosting specifications

- Layer type: Raster

- Spatial resolution: 30m
- Temporal resolution: Static
- Codebase: Github repository

Figure 3.16: Scale dependency of TPI



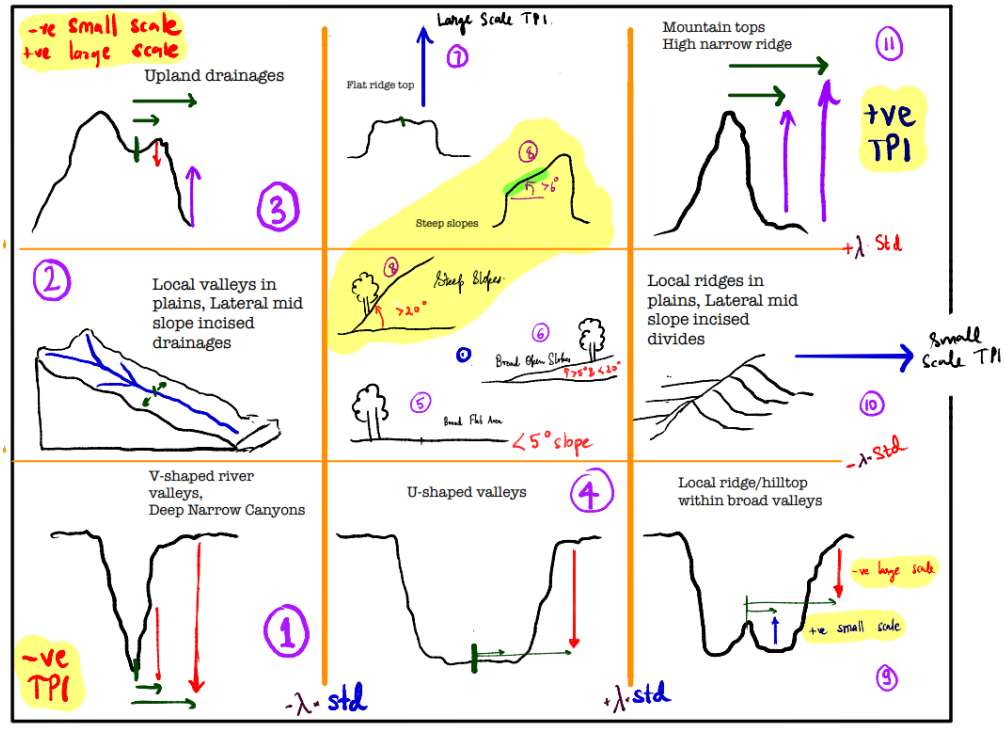


Figure 3.17: Landforms classification using slope and thresholds (denoted using λ) on both small-scale and large scale TPIs

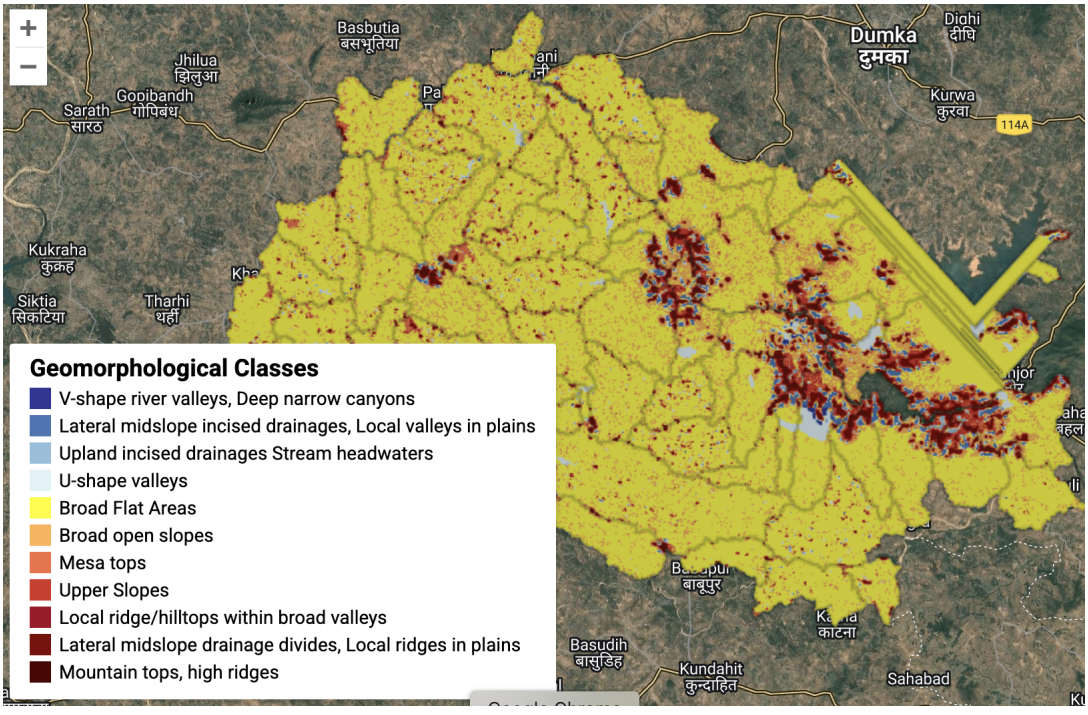


Figure 3.18: Landform classification of Masalia with micro-watershed boundaries in black.

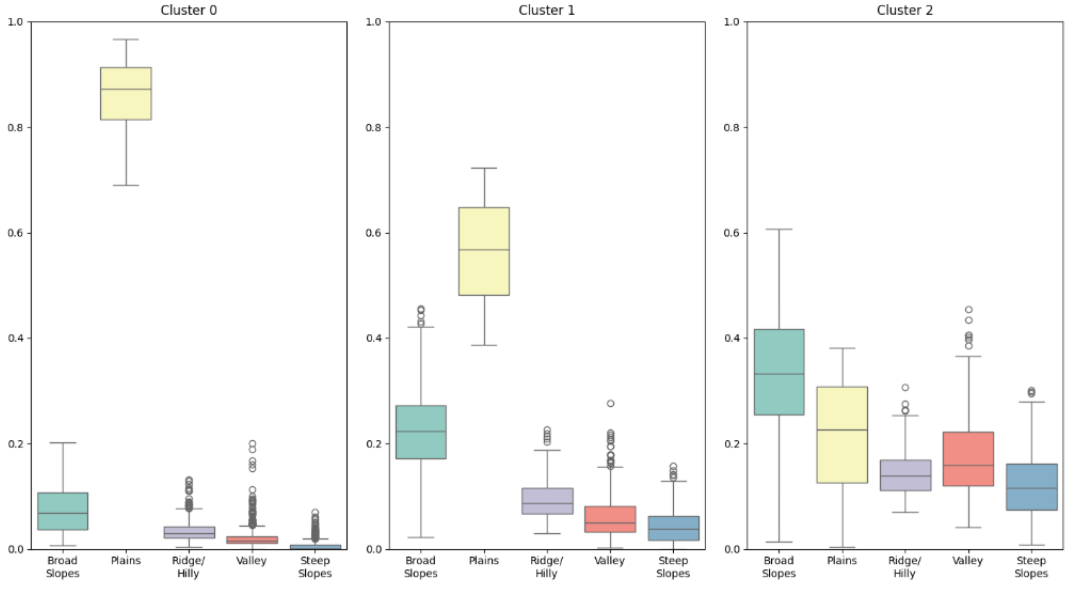
3.3.2 Terrain cluster

Introduction We observed that CSO field staff and volunteers familiar with the ridge-to-valley watershed development approach wanted to identify plain areas in uplands, midlands, and lowlands, to relate it with the type of assets suitable for these areas [38]. For example, Water Absorption Trenches (WATs) are constructed in uplands to harvest surface runoff, while farm ponds are constructed on lowlands to arrest surface and subsurface runoff for protective irrigation. To further guide in which terrains would such a ridge-to-valley approach be suitable, we classify the terrain of micro-watersheds using digital elevation.

Input layers We use landform classification generated in section 3.3.1 as input to derive terrain clusters.

Methodology Landform classification generated in section 3.3.1 produces a raster image with each pixel classified into one of the 11 landforms. We group these 11 landforms to five broad categories: plains, broad slopes, steep slopes, valleys and ridges. We compute the percentage of area under each category to add them as features for a microwatershed. We took microwatersheds across diverse terrains such as Masalia, Angul, Mandalgarh, Pindwara, and Mohanpur blocks to perform clustering. We perform K-Means clustering to obtain the following optimal clusters: mostly plains, plains and slopes, and mix of hills, valleys, slopes and plains. The precomputed cluster centroids are then used to assign terrain clusters to new microwatersheds.

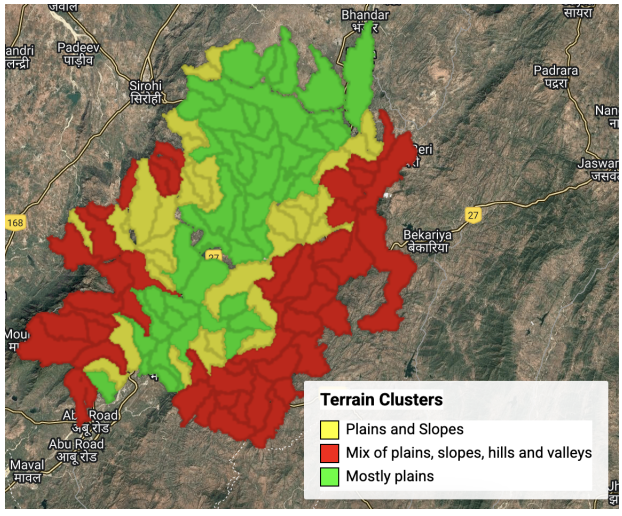
Figure 3.19: Distribution of plains, broad slopes, steep slopes, valleys and ridges across terrain clusters: mostly plains (cluster 0), plains and slopes (cluster 1), and mix of hills, valleys, slopes and plains (cluster 2)



Hosting specifications

- Layer type: Vector

Figure 3.20: Terrain clusters assigned to microwatersheds of Pindwara block



- Spatial resolution: NA
- Temporal resolution: Static
- Codebase: Github repository

3.4 Land use variables

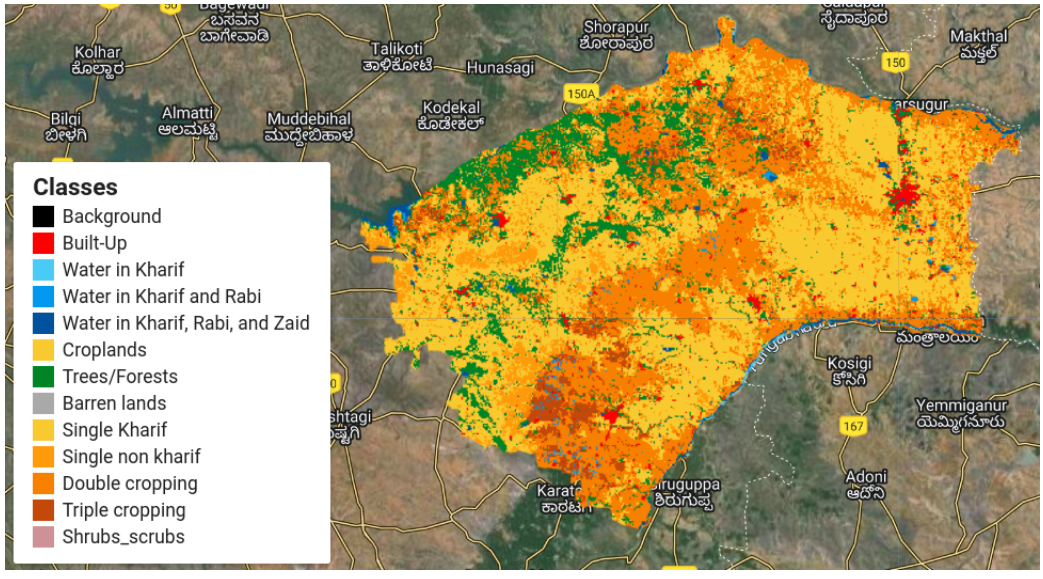
3.4.1 Land use land cover

Introduction Land use land cover (LULC) classification that indicates different features on the Earth's surface, such as forests, rivers, croplands, or buildings, is used actively for monitoring and planning of land-use. For example, tracking of anthropogenic activities such as environmental degradation through deforestation and loss of farmlands are done through satellite maps annotated with LULC categories [69], and used by development practitioners to design sustainable land management policies [45].

Most global LULC products produce annual outputs and do not capture within-year dynamics. The recent Dynamic World product from Google produces a near real-time output each fortnight, but the methodology has a limitation in being cross-sectional without considering past temporal values in each classification. For this reason, within-year variations are not captured well in this product.

We have developed an intra-annual LULC classifier that uses time-series data across all seasons to classify an area (10m x 10m pixel) in terms of its cropping intensity (accuracy 83%). Further, the model is trained on regional data sampled from across all agro-climatic zones of the country and produces a higher accuracy than other data products even on standard LULC labels such as surface water, trees, crops, etc. - accuracy 94%). We further have an ongoing groundtruth data collection effort to validate the outputs in out-of-sample geographies. We use Google Earth Engine to produce annual outputs for any area of interest, since 2003 to the current date as shown in the figure 3.21.

Figure 3.21: Land use land cover classification of Raichur district in Karnataka for 2022



Input layers To generate land use and land cover maps, we use multiple data sources as input-

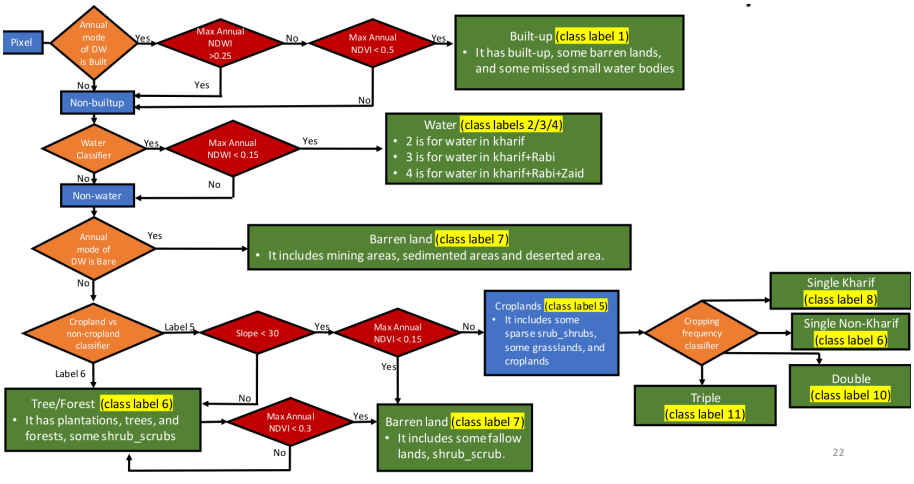
- Satellite data- We use multi-spectral data from Landsat-7 [35], Landsat-8 [36], Sentinel-2 [33], MODIS [26], and Sentinel-1 [32] satellite constellations. Different combina-

tions of this data are used to create feature vectors that serve as input to different machine learning algorithms. We use Google Earth Engine to access these datasets.

- Dynamic World- Google's land use and land cover product is used to identify certain land cover classes like built-up, water, etc. We access the Dynamic World data repository from Google Earth Engine [13].
- Open Street Maps- This data is used in the construction of the IndiaSAT groundtruth [30].
- Input Shapefiles- To run LULC for any region of interest, you require to give its shapefile (uploaded as a GEE asset) as input. For example- you can access different administrative boundaries in India at [2].

Methodology We perform a series of binary classifications followed by error-correction layers tailored for each of the final LULC classes under consideration, as shown in Figure 3.22. Each level of classification uses a different set of input features, datasets and methods.

Figure 3.22: Version-2 Land use land cover classification pipeline.



- **Dynamic World based Classification**: We use Dynamic World to identify built-up regions, barren lands, and shrubs and scrubs. Segmentation-based classification methods perform well in identifying these classes and thus we use dynamic world as a proxy in our classification pipeline.
- **Surface Water Body Classification**: Our goal is to predict both seasonal and perennial water bodies and also give information on which agricultural season do these water bodies have water in. Existing LULC products like Global Surface Water and Dynamic World use optical satellite data to identify water bodies which under-performs in the Kharif season due to lower quality of optical data in monsoon season. To capture the seasonality of water bodies, we use SAR data from Sentinel-1 to identify pixels having water in kharif season and Dynamic World for Rabi and Zaid. A threshold based method on number of months predicting water in a particular agricultural season is used to identify water in all three seasons.
- **Cropland vs Tree Classification**: We classify the non-classified pixels further into 2 classes- Croplands and Forests/Trees. We use Sentinel-1 SAR data time series to

perform this classification. The training data at this level is partly taken from the IndiaSAT groundtruth used in level-1 and is partly marked manually through visualization on Google Earth Pro. It is ensured that the groundtruth for both categories is geographically well distributed across different agro-climatic zones in India. We use a random forest classifier with 100 trees as the classification model that takes as input an annual 16-day time series of VV and VH bands from Sentinel-1. We further use Slope information from SRTM DEM [34] and use a threshold of 30 degrees to correct misclassifications in croplands, if any.

- **Cropping Frequency Classification**: To focus on agricultural sustainability applications, we classify cropland pixels into 4 categories based on their cropping frequency- Single Kharif, Single Non-Kharif, Double, and Triple cropping. To perform this classification, we did not have access to any training data. So, we used unsupervised classification using K-nearest neighbour algorithm. We randomly sampled cropland pixels from all agro-climatic regions in India (marked from the cropland groundtruth at level-2). The feature vector for this classification is a 16-day NDVI time series that is derived from a combination of Landsat-7, Landsat-8, Sentinel-2, and MODIS data. This time series is generated on the lines of GCI-30 paper [78]. We do not perform Whittaker smoothing to avoid missing out important peaks in the time series. Initially we created 16 clusters using KNN method and then hierarchically split them into 2 whenever the distortion exceeded the threshold of 0.23. We manually labeled each cluster by interpreting the spread of time series that belonged to that cluster. These clusters are then used to assign the classification label based on euclidean distances to their centroids.

Version-3 Enhancements in Land Use and Land Cover Maps In Version-3, a new **Shrubs and Scrubs** class has been introduced. To refine the classification of **single kharif croplands**, we assess their potential misclassification by leveraging **Dynamic World** outputs. If a single kharif pixel is consistently classified as **Shrubs and Scrubs** in the **annual mode** of Dynamic World, it is reassigned to that class.

Additionally, a **temporal correction framework** has been implemented to enhance the stability of multi-temporal outputs. Temporal inconsistencies may arise due to variations in satellite data quality, missing data at specific timestamps, salt-and-pepper classification errors from pixel-based methods, or limited generalizability of classification models over time.

To ensure a consistent and reliable land cover representation, the temporal correction framework systematically detects and corrects inconsistencies. The process involves:

1. *Identifying Recurring Inconsistencies*
    - A three-year sliding temporal window is used to detect patterns of misclassification.
    - Pixels exhibiting inconsistent classifications are flagged for correction.
2. *Rule-Based Reclassification*
    - A set of handcrafted rules is applied to reclassify misclassified pixels accurately.
3. *Structured Algorithm Implementation*
    - The correction algorithm operates within the sliding-window framework.
    - Logical land cover transitions are enforced while minimizing unnecessary changes.

This approach ensures temporal stability and improves classification accuracy across multi-temporal datasets.

Hosting specifications

- Layer type: raster
- Spatial resolution: 10m
- Temporal resolution: yearly
- Codebase: Github repository

3.4.2 Land use on terrain

Introduction Characterizing land use (cropping, trees, barren, and shrubs) separately for slopes and plains is useful to prioritize interventions under MGNREGA planning. Slope areas are prone to soil erosion, rapid runoff, and limited water retention. If slopes are covered with trees, they play a crucial role in stabilizing soil and reducing erosion. In contrast, barren slopes are more vulnerable to degradation and can benefit from contour bunds, stone check dams, or gabions to slow down runoff and promote infiltration. In plains, land use characterization helps determine the most effective locations for water harvesting and groundwater recharge structures. Cropping areas benefit from well-placed farm ponds and percolation tanks that provide supplemental irrigation and improve water security. Barren or shrub-covered plains indicate opportunities for land development through rainwater harvesting and soil improvement measures.

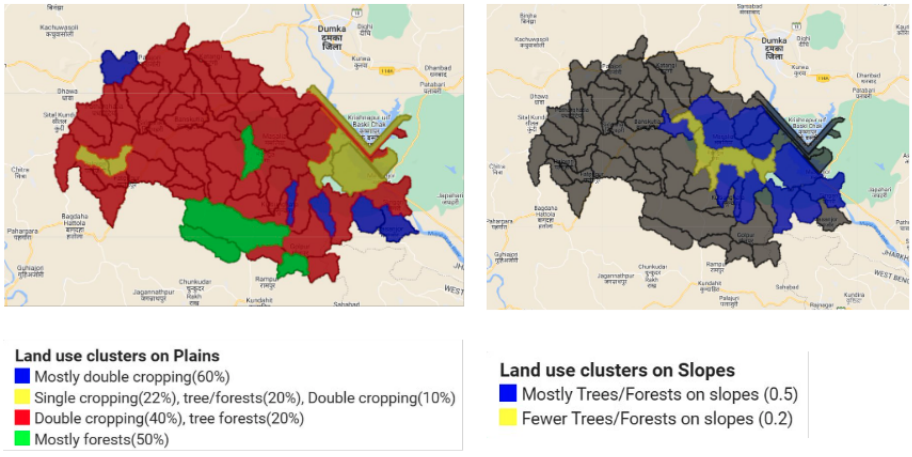
Input layers We use land use land cover maps mentioned in section 3.4.1 from 2017-23, terrain classification maps from section 3.3.1 and agro-ecological zones (AEZs) of India as input to derive AEZ specific characterization of land use on slopes and plains.

Methodology For each agro-ecological zone (AEZ) in India, we performed stratified sampling of microwatersheds each assigned with a terrain cluster. We took temporal composite (mode) of land use land cover maps from 2017-23 to identify dominant land use patterns in the microwatersheds. The terrain classification of these microwatersheds with 11 landforms classes were converted to five broad terrain classes: plains, broad slopes, steep slopes, valleys, and ridges. We compute percentage of area under cropping (single kharif, single non-kharif, double and triple), barren land, shrubs and scrubs, and trees in plains and them add them as features for clustering similar microwatersheds to identify dominant land use patterns in plains. Similarly, we perform the above exercise in slopes as well to identify dominant land use patterns in slopes.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: annual
- Codebase: Github repository

Figure 3.23: The figure characterizes land use on plains and slopes respectively across microwatersheds of Masalia block in Dumka district of Jharkhand



3.4.3 Land use changes

Introduction Tracking land use changes is crucial for planning water structures under MGNREGA as it directly influences demand and supply of water availability. Agricultural expansion, increasing cropping intensity may require more water structures such as check dams, ponds, and recharge pits. Land use changes, such as deforestation or urbanization, can alter surface runoff.

Input Layers We use LULC images of six years or more as an input. This LULC images are produced using our India-SAT v3 pipeline which include temporal correction.

Methodology To compute changes over a period of six years (Let's assume 2017-2022) we will first generate the LULC images for all these six years. Then we create two composite images using the mode of the first three years (2017-2019) and the last three years (2020-2022). For example,

$$\begin{aligned} earlier &= mode(Img_{2017}, Img_{2018}, Img_{2019}) \\ now &= mode(Img_{2020}, Img_{2021}, Img_{2022}) \end{aligned} \tag{3.17}$$

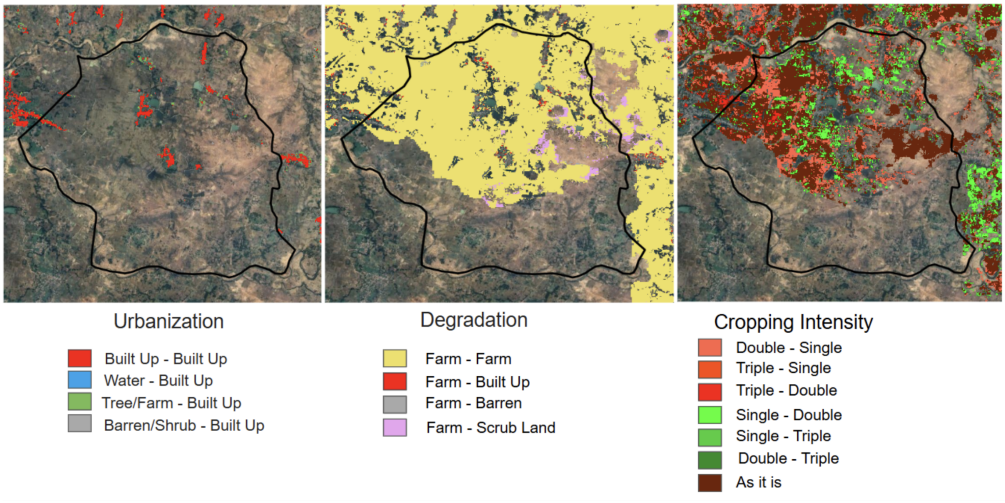
A KxK transition matrix is generated, where K represents the number of LULC classes. Each (i,j)th entry in the matrix represents the number of pixels that changed from class i to class j. Change maps and Sankey plots are generated to visualize the transitions. Following transitions are computed using these earlier and now image as shown in the figure 3.24.

- **Urbanization** refers to the transition from rural to urban areas, affecting society, infrastructure, and the environment. Urbanization include the transition of built-up, green, water, and barren areas to built-up.
- **Degradation** Land degradation results from erosion, pollution, and unsustainable practices, reducing soil fertility and biodiversity. Degradation include the transition of farm area into built-up, barren and shrubs.
- **Deforestation** is the permanent removal of trees, leading to biodiversity loss and climate change. Deforestation include the transition of trees to built-up, barren, shrubs and farms.
- **Afforestation** involves planting trees in non-forested areas, aiding carbon sequestration and ecosystem restoration. Afforestation include the transition of built-up, barren, shrubs and farms to trees.
- **Cropping Intensity Change** Changes in cropping intensity impact agricultural productivity and sustainability. The transitions include single to double cropping, single to triple cropping, double to single cropping, double to triple cropping, triple to double cropping and triple to single cropping.

Hosting specifications

- Layer type: raster
- Spatial resolution: 10m
- Temporal resolution: annual
- Codebase: Github repository

Figure 3.24: Change detection maps of Gulguldih village in Lakshmipur block of Jamui district in Bihar.



3.4.4 Cropping intensity

Introduction Being able to classify regions in terms of cropping intensity is relevant for water security to understand changes in crop water usage over time. Most of India has three agricultural seasons in a year - during the monsoons, post-monsoon, and summer. The monsoon season is rain-fed in a large part of the country, whereas post-monsoon and summer reasons require irrigation from groundwater and other sources such as canals.

Input layers We use annual land use land cover (LULC), to identify areas under single cropping, double cropping and triple cropping using pixels which are classified as single kharif, single non-kharif, double and triple classes of LULC classifier to determine cropping intensity.

Methodology Using the LULC of a micro-watershed, we compute cropping intensity for a microwatershed in a year y using the following equation:

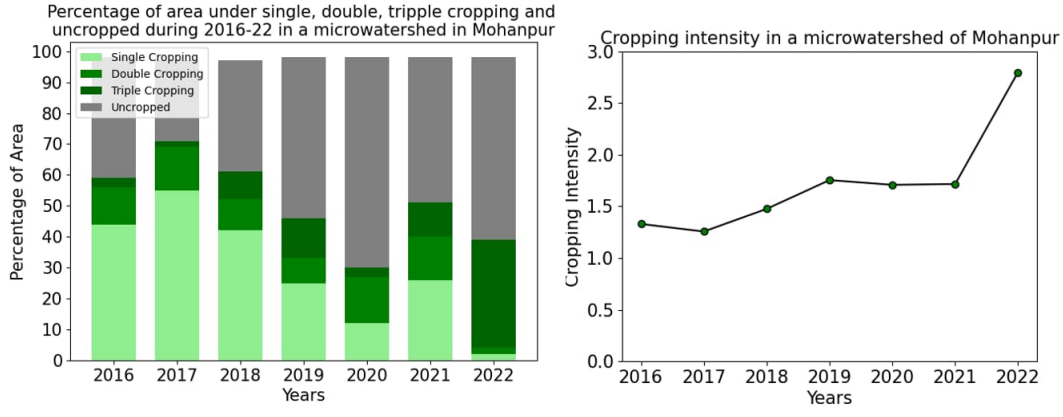
$$\begin{aligned} \text{Cropping intensity} = {} & 1 \times \% \text{ of area under single cropping} \\ & + 2 \times \% \text{ of area under double cropping} \\ & + 3 \times \% \text{ of area under triple cropping} \end{aligned} \tag{3.18}$$

where % of area under single cropping is computed as the summation of area under single kharif cropping and area under single non-kharif cropping divided by the total cropped area. Total cropped area is the union of cropping areas across the years. Area under a particular class is computed by the no. of pixels of that class multiplied by the pixel area. Similarly, for % of area under double cropping and % of area under triple cropping. 3.25

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: annual
- Codebase: Github repository

Figure 3.25: For a micro-watershed in Mohanpur, we analyze the percentage of area under single, double and triple cropping (left), and corresponding cropping intensity over the years.



3.4.5 Water bodies

Introduction Knowledge about existing water storage structures, their capacity, water availability in different seasons, siltation and repair requirements, can give important indications about the resilience of a landscape to droughts and floods, especially with climate change.

Global Surface Water (GSW) [66] maps provide information about the surface area under water in different months, but we have found that GSW maps are effective only for large water bodies and they miss out on sub-5000m2 water bodies. Some years back a waterbody census was conducted in India and a dataset has been recently released, but this is a static dataset and does not provide any real-time indications of the water availability in different seasons.

The LULC method we have developed is able to accurately capture even 1000-5000m2 water bodies, and many sub-1000m2 water bodies as well, with seasonal water availability. We are intersecting this with the waterbody census data and will then have an accurate waterbody capacity map as well. Using this, we will develop indicators for water resilience of a landscape.

Input layers To generate the water bodies vector layer for a region of interest, we use the LULC raster layer generated for that region to detect pixels with water class.

Methodology The hydrological year is divided into 3 agricultural seasons - Kharif (July-October), Rabi(November-February), Zaid (March-June). We use Sentinel-1 (SAR data) VV band for water pixel detection in Kharif season and Dynamic World to detect water pixels in Rabi and Zaid seasons. A threshold based method on number of months predicting water in a particular agricultural season is used to identify water in all three seasons leading to following classes per pixel as shown in the figure 3.26. Error correction was performed on top of it using Sentinel-2 bands to remove false positives of water pixels in croplands and barren lands. We perform the raster to vector conversion to generate a vector layer of water bodies with the following metadata for each water body: unique id, surface water area under Kharif, surface area under Rabi and surface area under Zaid, total area of the water body as shown in the figure 3.27.

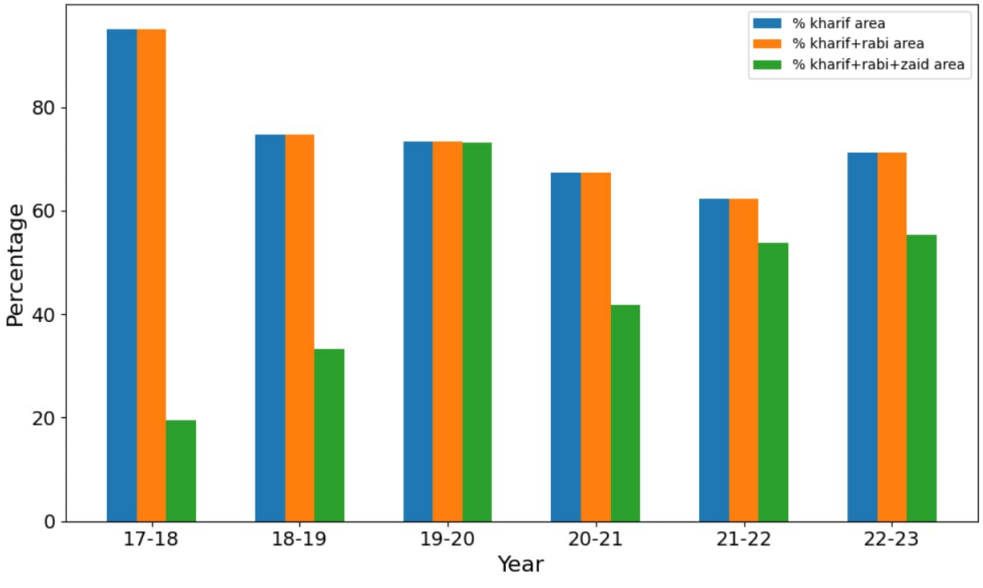
Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: yearly
- Codebase: Github repository

Figure 3.26: Water bodies as seen from satellite (first column) in Masalia block of Dumka district in Jharkhand and their detection using the classifier (second column)



Figure 3.27: Bar graph depicting the surface area of a water body (in percentage) in Mohanpur block of Gaya district in Bihar



3.4.6 First census of water bodies

Introduction The Ministry of Jal Shakti under Government of India performed the first census of water bodies [37] in 2017-18 with an objective develop a national database of all water bodies and their important properties such as storage capacity, status on whether they are functional, water spread area and type of use. However, the dataset is static and does not provide any real-time indications of the water availability in Kharif, Rabi and Zaid.

The LULC classification method mentioned in section 5.02 is able to accurately capture even 1000-5000m^2 water bodies, and many sub-1000m^2 water bodies as well, with seasonal water availability in Kharif, Rabi and Zaid. We are intersecting this with the waterbody census data to accurately generate waterbody capacity map as well. Using this, we will develop landscape-level indicators of water resilience. The capacity column of the water body can also be used to build site-level indicators of harvested runoff and infiltration.

Input layers The waterbody census data is published state-wise on Open Government Data (OGD) platform of India [15] in the form of CSV file.

Methodology We converted the state-wise CSV files of waterbody census into shapefiles using the geo-coordinate (latitude, longitude) of each waterbody as shown the figure 3.28.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static as of 2017-18
- Dataset: Google drive folder

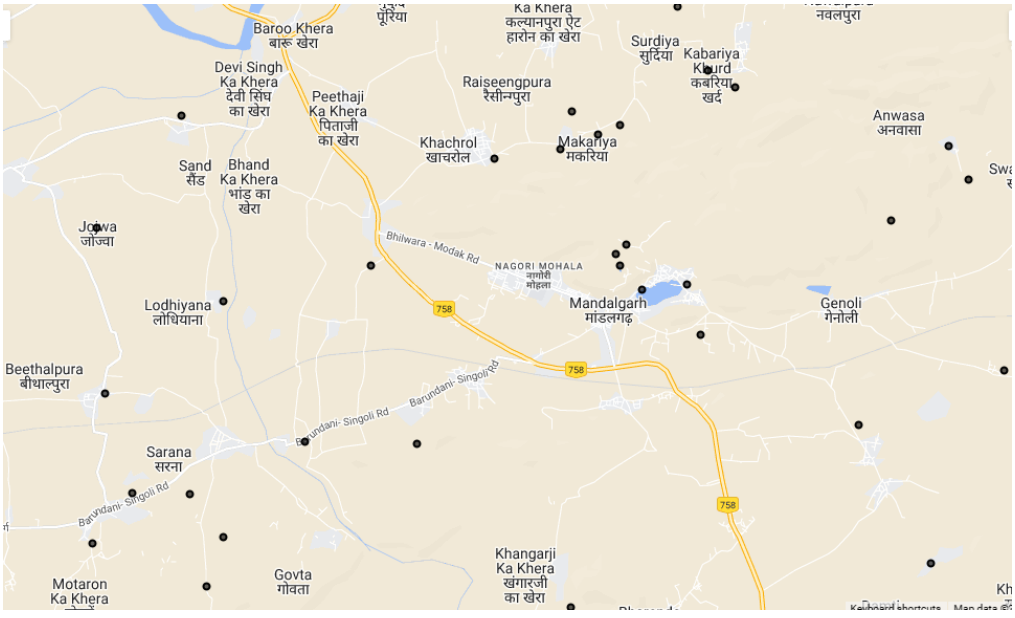


Figure 3.28: The map shows geo-coordinate of waterbodies from the waterbody census data of Rajasthan, specifically zoomed into Mandalgarh block of Bhilwara district for visualization.

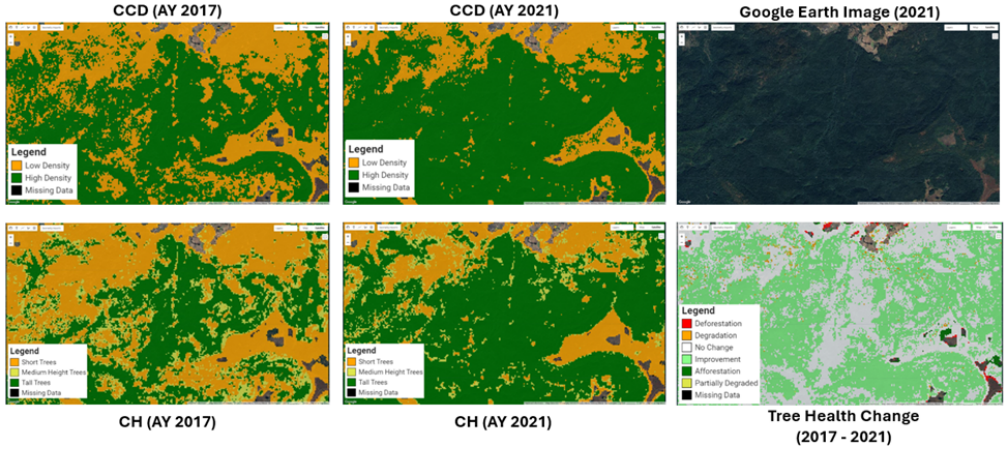
3.5 Forest variables

3.5.1 Tree health monitoring

Introduction Forests and plantations are the backbone of our planet's ecological balance, offering habitats, absorbing carbon dioxide, regulating climate, and sustaining biodiversity. Beyond traditional forests, small-scale plantations also play a vital role in rural livelihoods by enhancing agroforestry productivity and generating carbon credits. While restoration and conservation initiatives are underway, they require granular and timely insights into tree canopy health. However, tracking tree health indicators like canopy density and height over time has remained a challenging task, particularly in a country as ecologically diverse and vast as India. Traditional tree monitoring methods have often been limited by their manual, time-intensive nature. In an era of climate urgency, this delay can limit the effectiveness of conservation and restoration initiatives. Modern remote sensing technologies allow us to move from static snapshots to near real-time, actionable insights, enabling stakeholders to implement timely interventions. We introduce a novel, scalable, and open-source framework for monitoring tree health across India using advanced remote sensing and machine learning techniques at a fine spatial resolution of 25m.

We have developed Agro-Climatic Zone (ACZ) specific spatiotemporally robust models, which outperform global models by tailoring insights to local conditions, and use openly available data from the Sentinel-1, Sentinel-2 and Global Ecosystem Dynamics Investigation (GEDI) satellite systems. A sample tree canopy cover density and height output, along with overall tree health change between AY 2017 - AY 2021 for Dongariya forest patch in Madhya Pradesh is shown in Figure 3.29. All our canopy density and height outputs along with the tree health change outputs can be found on the GEE app. The outputs of our model can especially help forest departments, policymakers, and local communities to identify areas in need of restoration and to assess the effectiveness of ongoing conservation activities.

Figure 3.29: Canopy cover density and height outputs of AY 2017 and AY 2021, and tracking overall tree health change between July 2017 - June 2022 in a forest patch of Dongariya village, Balaghat district, Madhya Pradesh



Methodology At the core of our research lies the integration of diverse satellite datasets and advanced machine learning techniques to predict tree health indicators: Canopy Cover

Density (CCD) and Canopy Height (CH) Distribution. The methodology includes:

- **Data Integration:** We combined GEDI LiDAR datasets with Sentinel-1 SAR and Sentinel-2 optical imagery. The GEDI dataset serves as our ground truth for canopy metrics, while Sentinel-1 (SAR) and Sentinel-2 (optical) data provided high-resolution spatial coverage across years and is used to predict the tree canopy indicators. We utilized Agricultural Year (AY) 2019 and 2020 data for modelling, and AY 2021 data was out-of-sample dataset which was used to test the temporal robustness of models.
- **Machine Learning Models:** We trained XGBoost models for predicting CCD and CH distributions (median, 75th percentile, and maximum heights). These models were trained using 80% of the data (AY 2019-2020), validated on the remaining 20%, and tested on out-of-sample temporal datasets (AY 2021).
- **Temporal Stability Analysis:** We developed a correction mechanism to mitigate temporal output fluctuations caused by differences in data availability or quality, ensuring reliable long-term monitoring.
- **Change Analysis Framework:** Beyond yearly predictions, we implemented a systematic framework to detect and classify changes in tree health over time, distinguishing between deforestation, afforestation, improvement, and degradation.

Hosting specifications

- Layer type: raster
- Spatial resolution: 25m
- Temporal resolution: Agriculture Year
- Spatial coverage: Pan India (except ACZ 12-15)
- Codebase: Github repository

3.5.2 Tree canopy cover density

Introduction Tree canopy cover density refers to the proportion of a ground area that is covered by the vertical projection of tree crowns, essentially measuring how much of the ground is shaded by the leaves and branches of trees. It indicates the density of tree cover in a given area, with higher canopy cover percentages signifying a denser canopy and more shade on the ground, and lower percentages signifying low density.

Hosting specifications

- Layer type: raster
- Spatial resolution: 25m
- Temporal resolution: Agriculture Year
- Spatial coverage: Pan India (except ACZ 12-15)
- Dataset: Google Earth Engine Assets

3.5.3 Tree canopy height

Introduction We predict the tree canopy height distribution in the 25m * 25m plot. For each ACZ, three separate binary classifiers were trained to predict (1) the median, (2) 75th percentile, and (3) maximum tree heights as the 98th percentile of GEDI data, for 25m resolution tree plots in a given AY. These individual tree height distribution classes were then combined into comprehensible classes (short, medium and tall trees) that can be explained easily to potential users of the output.

Hosting specifications

- Layer type: raster
- Spatial resolution: 25m
- Temporal resolution: Agriculture Year
- Spatial coverage: Pan India (except ACZ 12-15)
- Dataset: Google Earth Engine Assets

3.5.4 Tree health changes

Introduction Based on the change in tree cover and the tree canopy density and height outputs across different years, we categorize the tree health changes into afforestation, deforestation, improvement, degradation and no change. The table in Figure 3.30 shows the different tree health change classes along with their descriptions.

We further categorized tree cover into Large Tree Patches (LTP) of 1 hectare of tree presence and Small Tree Patches (STP) of less than 1 hectare of tree presence, and disaggregated tree health changes to better understand whether improvements or degradations are happening in forests or trees outside forests. The LTP and STP based change classification is shown in Figure 3.31. An example of the detailed tree health change statistics can be observed for ACZ-7 in Figure 3.32. The detailed results for all the ACZ can be found in the slides.

Figure 3.30: Canopy Cover Density (CCD) & Canopy Height (CH) Distribution Based Tree Health Change Categories

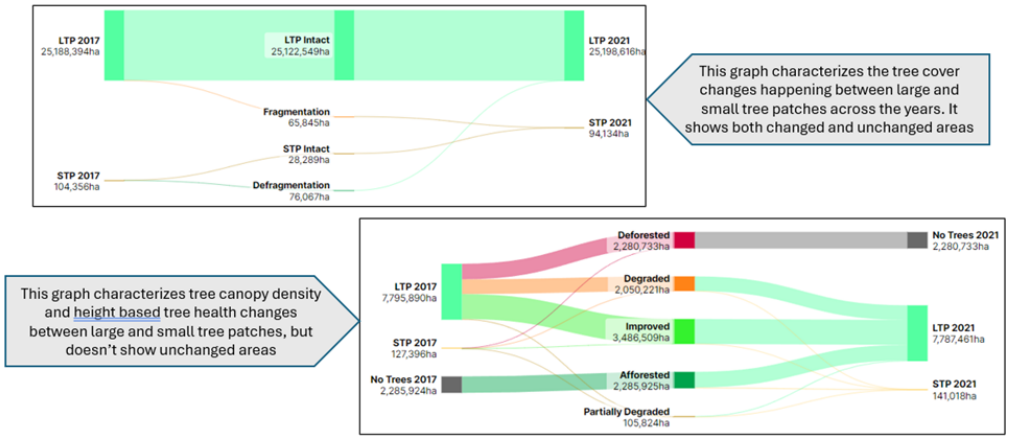
Overall Tree Health Change Category	Description
Deforested	When trees are cut, and a site goes from having tree cover to no more tree cover
Afforested	When trees are planted, and a site goes from having no tree cover to trees
Improved	When there are trees between 2 years, and their CCD/CH improves over time
Degraded	When there are trees between 2 years, and their CCD/CH deteriorates over time
Partially Degraded	When there are trees between 2 years, and CCD improves but CH deteriorates over time, or vice versa

Figure 3.31: Large and Small Tree Patches Transition Categories

Tree Category (Year-1)	Tree Category (Year-2)	Tree Patch Change from Year-1 to Year-2
LTP	LTP	LTP Intact
LTP	STP	Fragmentation
STP	LTP	Defragmentation
STP	STP	STP Intact

Input layers The Dynamic World tree cover LULC, along with the tree canopy cover density and canopy height outputs as described in Sections 3.5.2 and 3.5.3, respectively, are the inputs for the tree health change computation.

Figure 3.32: Tree Health Change (AY 2017 – 2021) Statistics of Eastern Plateau and Hills Region (ACZ-7)



Methodology First, a site is assessed for any afforestation or deforestation based on the tree cover LULC mask, and if the tree cover remains intact, then the site is assessed for degradation or improvement based on changes in the CCD and CH distribution. All the change analysis is done after temporally correcting the outputs for any fluctuations across years. The modal values of the current, previous, and next year outputs for a year are considered as the final health output for that year and are considered for detecting tree health changes.

Hosting specifications

- Layer type: raster
- Spatial resolution: 25m
- Temporal resolution: Agriculture Year
- Spatial coverage: Pan India (except ACZ 12-15)
- Dataset: Google Earth Engine Assets

3.6 Welfare variables

3.6.1 NREGA assets categorization

Introduction The Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) in India provides guaranteed wage employment to rural households for various types of public works. MGNREGS Master Circular for the financial year 2018–19 by Ministry of Rural Development, specifies 260 permissible works [22], including the construction and maintenance of water conservation assets. We categorized (as shown in table 3.4) 260 works from the lens of Natural Resource Management (NRM) where the works can provide site level and landscape level impact on agriculture, and their ability to generate livelihood. The categorization of existing works will be helpful in examining the current distribution of assets and can facilitate the planning of new assets to ensure water security and sustainable livelihoods in rural India.

Table 3.4: MGNREGA category and their corresponding keywords

Category	**Keywords**
Soil and water conservation	Bund, canal, channel, check dam, drainage structures, dykes, embankments, gully plugs, recharge pits, soak pit, soakage channels, spurs, terrace, trench, water courses, water drain
Irrigation on farms	Bunds, farm ponds, harvesting ponds, open well, percolation tank, stabilization pond
Other on-farm works	Sand filter
Plantations	Farm forestry, forest, horticulture, nursery, sericulture, shelter belt
Land restoration	Azola cultivation, biomanure, chaur land, compost, fallow land, levelling, storage building, waterlogged land
Off-farm livelihood assets	Cattle shelters, drying yards, fisheries pond, goat shelter, piggery shelter, poultry shelter, workshed
Community assets	Building, compound wall, crematorium, haat, play fields, roads, shelter, toilets

Input layers

- MGNREGS permissible work list [25, 24].
- MGNREGA assets: We obtained geotagged MGNREGA assets from Bhuvan [8]. The geotagging of existing NREGA works began in September 2016. The data contains the assets built from January 2005 to November 2022. The metadata of assets such as work name, work type, expenditure and total person days employment provided for work was obtained from NREGA MIS[23]. Figure 3.34 shows metadata for available for an asset. Details of downloading and processing the asset data can be found here [31].

Methodology We employ keyword matching algorithm to categorize geotagged MGNREGA assets. The MGNREGS permissible work list uniquely identifies each work with its one line description [25]. For example, “Construction of Farm Ponds for Individuals" is a uniquely identified work in the permissible work list. For each permissible work, we extract

keywords from its description and assign them to a category. The assignment results in a bag of words for each category as shown in table 3.4. In the above example, we mapped the “pond" keyword to “Irrigation on farms" category. The assigned keywords of each category are used to categorize a geotagged asset. The keywords present in the work type, work name and asset name columns (in order) of a geotagged asset were used to match the assigned keywords of each category. We further refined the bag of words for each category by adding stem words of the keywords and some region specific keywords in Hindi such as “pokhar" in Bihar which is used for pond while geotagging the asset. We performed random sampling of categorized geotagged assets from each category to manually examine the performance and the performance was found to be satisfactory. Figure **??** shows geotagged assets in Mohanpur block of Gaya district in Bihar, color coded based on their category.

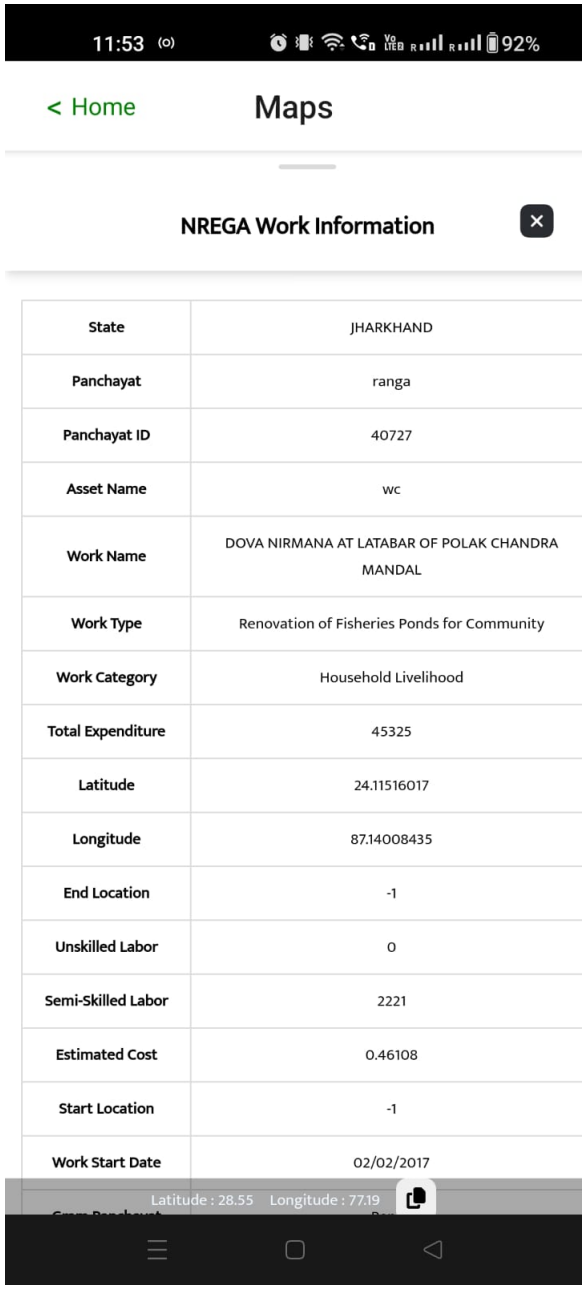
Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: yearly
- Dataset: Google drive folder
- Codebase: Github repository

Figure 3.33: MGNREGA assets in Pindwara block of Sirohi district in Rajasthan



Figure 3.34: Metadata of a MGNREGA asset in Masalia block of Dumka district in Jharkhand



3.6.2 NREGA panchayat category

Introduction India is a country marked by significant diversity, with wide disparities in the economic conditions across different sections of society. Historically, caste has played a major role in contributing to these disparities. Social welfare programs are designed to improve the livelihoods of the poor and marginalized. The Panchayat, being the smallest administrative unit, plays a crucial role in planning and implementing social and economic justice in rural India. However, it has been observed that the effectiveness of welfare programs varies across different Panchayats. To understand this variation, we have categorized the Panchayats based on the distribution of marginalized populations and analyzed the differing impacts of social welfare benefits across these units.

Methodology We follow a two-step approach to categorize panchayats. First, we categorize them based on the population percentage of marginalized castes as *significant SC panchayat population* and *significant ST panchayat population* depending on whether the percentages of SC and ST populations is more than the corresponding average[1], the average ST population percentage in panchayats across India. The remaining panchayats are classified as *others*. Next, we calculate the average SC population percentage within *significant SC population panchayats*[2]. We also calculate an uneven distribution indicator to measure unevenness in the distribution of the SC population in different villages within a panchayat[3]. We categorize *significant SC population panchayats* as having *consolidated significant SC population villages* if the SC population percentage in the panchayat is above the category-threshold and the unevenness distribution indicator is lower than the unevenness threshold value; *mixed SC population villages* if the unevenness distribution indicator is lower than its threshold value and the percentage of SC population is below the category-threshold; and *segregated significant SC population villages* if the unevenness distribution indicator is above its threshold value for the given panchayat. Similarly, we subcategorize *significant ST population panchayats* as consisting of *consolidated significant ST population villages*, *mixed ST population villages*, and *segregated significant ST population villages*.

- Dataset: Google drive folder

Different sub-categories of *Significant SC population panchayat* are presented in Figure 3.35 and of *Significant ST population panchayats* in Figure 3.36, for a sample block.

[1]The threshold used for identifying *significant SC population panchayats* is 18.1%, the average SC population percentage in panchayats across India, and is referred to as SC_PR. Similarly, the threshold used for identifying *significant ST population panchayats* is 16.3% and is called ST_PR

[2]The average SC population percentage within *significant SC population panchayats* is 33%. We call this SC_avg and it is used to identify further subcategories. The corresponding threshold for *significant ST population panchayats* ST_avg is 63%.

[3]We use Gini coefficient to measure unevenness in population distribution. The average value of the indicator for uneven population distribution amongst *significant SC population panchayats* avg(UE) is 16% which is used as a threshold. The corresponding threshold for the uneven population distribution indicator amongst *significant ST population panchayats* is 15%.

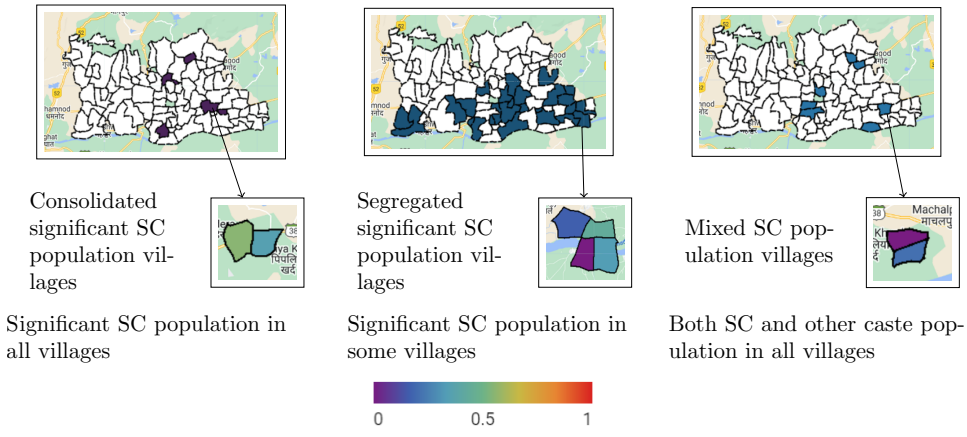


Figure 3.35: Categorization of SC Panchayats based on the distribution of SC population within panchayat

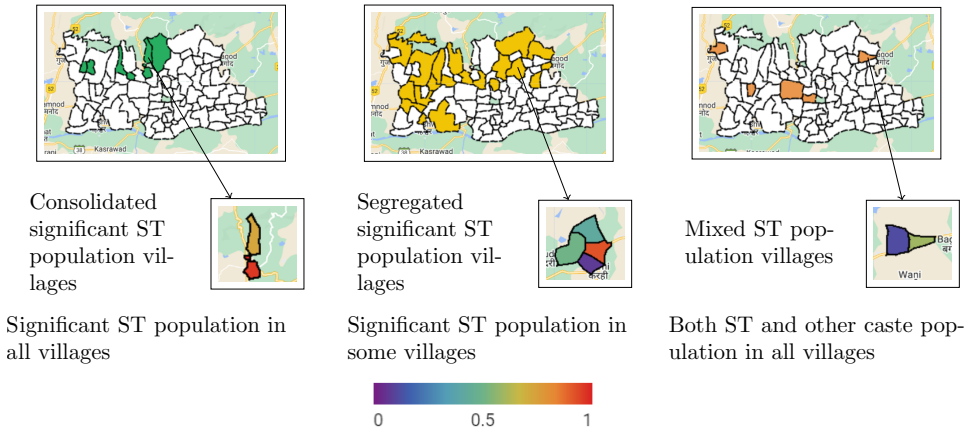


Figure 3.36: Categorization of ST Panchayats based on the distribution of ST population within panchayat

3.6.3 NREGA MIS parameters

The government of India maintains an extensive publicly accessible MIS website for MGNREGS to maintain data at the state, district, and panchayat levels. We wrote web crawlers to automatically download and parse this data. We obtained the data at the panchayat level for 293,451 panchayats, since the genesis of MGNREGS. We noticed that data specifically for the year 2013 had many missing values, and therefore restricted our analysis to 2014 onward, from the 2014/15 financial year to the 2020/21 financial year. The list of parameters used in this study are given in Table 3.5. We computed the annual average value of each parameter for our analysis.

- Dataset: Google drive folder

Variable	Description	Min	Max	Mean	Std Dev.
employment _demanded _household	Number of households demanding work under MGNREGS in the panchayat.	0	8067	189	295
employment _provided _household	Number of households provided work in the panchayat.	0	7408	170	275
employment _provided _total _persondays	Number of person days of work provided in the panchayat.	0	518256	8011	14371
employment _provided _hh_sc	Number of SC households provided work in the panchayat.	0	3989	41	103
employment _provided _hh_st	Number of ST households provided work in the panchayat.	0	8803	32	97
employment _provided _hh_oth	Number of general caste households provided work in the panchayat.	0	6386	117	214
employment _provided _persondays_sc	Number of person days of work provided to SC persons in the panchayat.	0	214462	1886	5214
employment _provided _persondays_st	Number of person days of work provided to ST persons in the panchayat.	0	2	0.7751	0.3036
employment _provided _persondays_oth	Number of person days of work provided to general caste persons in the panchayat.	0	2	0.7751	0.3036
empDemand_per _household	Number of households demanding work under MGNREGS relative to total number of household in the panchayat.	0	2	0.345	0.345
empProvided_per _empDemand	Number of households provided work relative to the number of household demanding work under MGNREGS in the panchayat.	0	2	0.7751	0.3036
average_wage	Average wage per person day. The Central and state governments revise the wage rate annually.	0	344	178	40
DCapproved_per _payable	The delay compensation to be paid by the government needs to be approved. Authorities may not deem all delays as worthy of compensation, and this variable indicates the fraction of compensation payable which was approved by the government.	0	2	0.262	0.416

Table 3.5: MIS Parameters used for welfare analysis

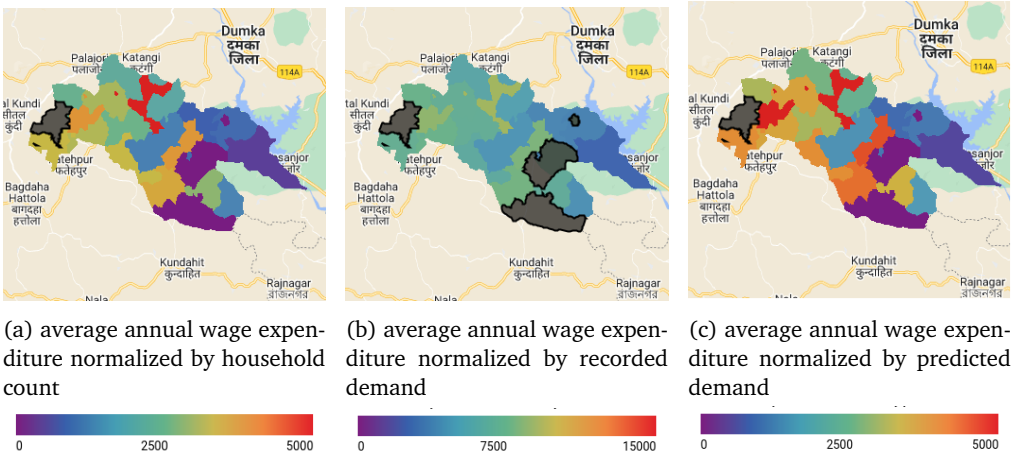
3.6.4 Fairness in NREGA expenditure

Introduction MGNREGA 2005 makes it mandatory to fulfill the demand for NREGA work or pay unemployment compensation. The official MIS data shows that 98 – 99% demand for work is met resulting in negligible cases of unemployment compensation. However, field research shows that discouragement or apprehensions in raising and/or recording employment demand are responsible for the high success rate visible in the MIS [65, 52, 72, 64, 63]. We have developed a demand model using multi-layer perceptron to predict the actual demand for work that might go unrecorded or may not even be raised. To build a model for this actual demand, we leverage the fact that both federal and state governments jointly own MGNREGS implementation and there is a wide variance in the efficiency and utilization of the scheme in different areas; being able to learn a model "good" areas can be used to predict the actual demand in other areas. The relevant features for this model are selected through a literature survey to identify factors that should influence MGNREGS demand. MGNREGS demand is known to be influenced by factors such as age, gender, land holdings [46], literacy rate, and caste [46, 44], along with the quality of scheme implementation which serves as an encouragement to demand work in the absence of alternative sources of employment [54] followed by selection of features with a correlation less than 0.7. The final set of features used to train the demand model is shown in Table 3.5.

Methodology We use a multilayer perceptron with one hidden layer to build a supervised learning based demand prediction model. We identify "good" panchayats that have a high percentage of households demanding work and a corresponding high percentage of this demand being met. Data from this set of panchayats is used to train the demand model. To identify these panchayats for model training, we find the average and standard deviation of the percentage of households demanding work in panchayats across the country, and use the average plus one standard deviation as a threshold. A second threshold is then also applied, to identify panchayats that are able to meet more than the average demand for employment across panchayats in India.

The predicted demand is then used as a normalization factor to estimate normalized expenditure across various panchayats. We have presented NREGA expenditure normalized by household count, normalized by demand recorded in MIS, and predicted demand in Figure 3.37. The variables used to predict demand, predicted demand, demand recorded in the MIS are available in "panchayat_level_predicted_demand_expenditure.csv" with the description in "Panchayat-level predicted demand-and-expenditure.pdf" in the dataset shared below.

- Dataset: Google drive folder



(a) average annual wage expenditure normalized by household count

(b) average annual wage expenditure normalized by recorded demand

(c) average annual wage expenditure normalized by predicted demand

Figure 3.37: Panchayat-wise MGNREGS expenditure in the Masalia block in the Dumka district of the state of Jharkhand in India

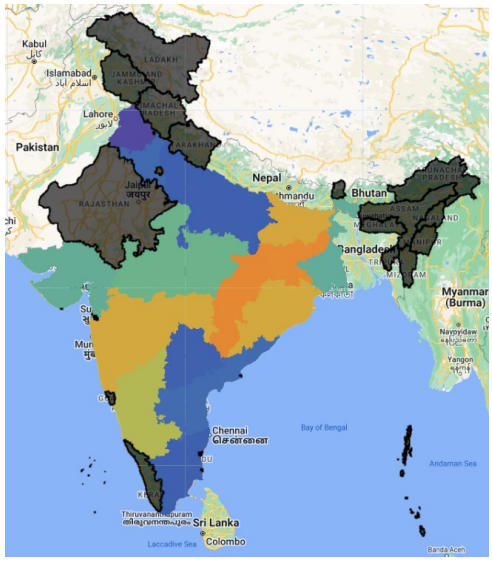
3.6.5 Caste-based inequity in NREGA allocation

Introduction India has a long history of caste-based discrimination, and social welfare schemes aim to mitigate its impact. The Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) of 2005 includes provisions to prioritize benefits for marginalized castes. However, implementation on the ground may differ from policy intentions. To evaluate disparities faced by marginalized castes across panchayats, it is crucial to compare key outcome variables related to the management and utilization of MGNREGS funds in panchayats with significant SC/ST populations against those in other panchayats. Simultaneously, accounting for confounding variables is essential to isolate the specific effect of caste (the treatment variable) on MGNREGS allocations.

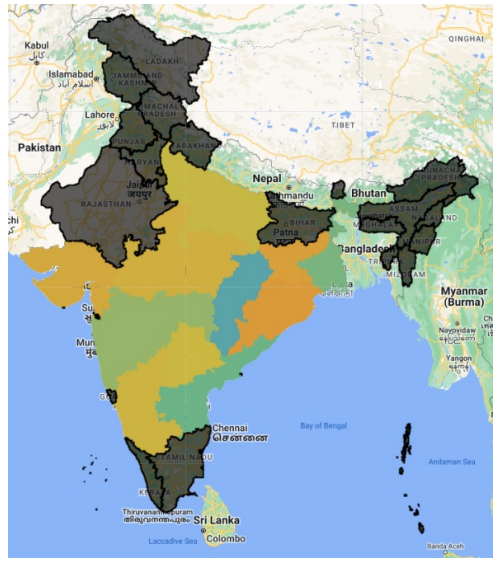
Methodology We have used a balancing score called the Propensity score to pair treated units (panchayats with a significant percentage of marginalized caste population) with counterfactual units (panchayats not with a significant percentage of marginalized castes), such that the treated and counterfactual pair are similar on other features except for the treatment variable itself [39, 68, 53]. The average of the treatment effects across all such panchayats is then called the average treatment effect and can indicate if these panchayats get more positive treatment than their counterfactuals. To check whether this average treatment effect is statistically significant, a paired t-test is further done on the treated and counterfactual pairs [39, 53].
We have computed equity scores at state, district, and panchayat level in terms of i) equity in providing employment to percentage of marginalized households as compared to the percentage of general caste households, ii) equity in percentage of marginalized households demanding employment as compared to general caste household iii) equity in persondays of employment to marginalized households as compared to general caste households in a) between panchayats with significant marginalized population as compared to panchayats without significant marginalized population b) and equity in terms of percentage of household getting employed and persondays of employment towards marginalized castes within panchayat.
We have computed equity scores for assessing equity towards scheduled castes (SC) and scheduled tribes (ST). We have also developed an aggregate metric to represent overall inequity towards SC and ST at every level. The state-level SC and ST inequity indices can be viewed in Figure 3.38, the district-level inequity indices are available in Figure 3.39. The inequity indices for various panchayats in Mahedhwar sub-district of West Nimar district in Madhya Pradesh are available in Figure 3.40.

The metrics are available in different CSVs. The details of equity and inequity metrics at panchayat, district, and state level are available in "Equity metrics and inequity indices.pdf" The metrics are available in the dataset shared below.

- Dataset: Google drive folder

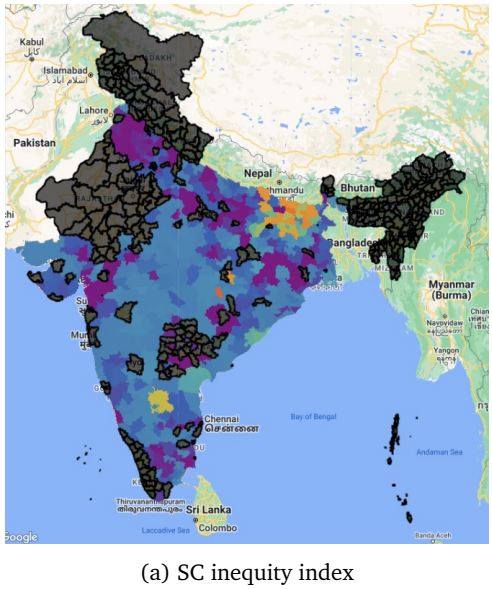


(a) SC inequity index

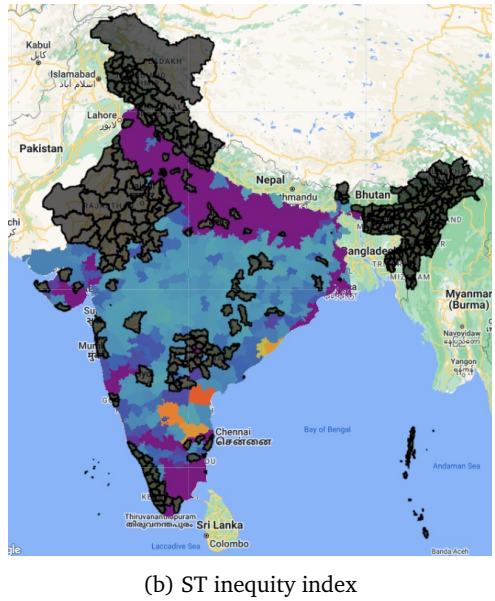


(b) ST inequity index

Figure 3.38: State-level inequity indices

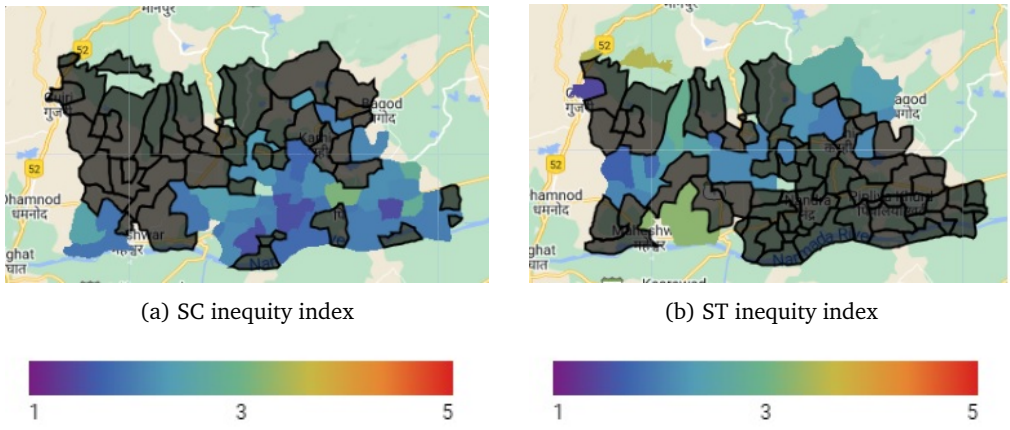


(a) SC inequity index



(b) ST inequity index

Figure 3.39: District-level inequity indices



(a) SC inequity index

(b) ST inequity index

Figure 3.40: Panchayat-level inequity indices

3.7 Administrative

Administrative boundaries in India refer to the demarcated lines that separate different administrative regions within the country, established to facilitate the governance and administration of various administrative units such as states and districts. The administrative units of India is hierarchical where India is divided into 28 states and 8 union territories. States are subdivided into districts, districts are sub-divided into blocks, blocks are sub-divided into panchayats and panchayats being further sub-divded into villages.

3.7.1 Boundaries

3.7.1.1 State

The shapefile represents the administrative boundary of Indian states. The Survey of India provides data upto district level. The Survey of India is the country's National Survey and Mapping agency under the Department of Science and Technology, Govt. of India. The dataset can be downloaded from here. State boundaries of India are shown in the figure 3.41

3.7.1.2 District

The shapefile represents the administrative boundary of Indian districts. The Survey of India provides data upto district level. The Survey of India is the country's National Survey and Mapping agency under the Department of Science and Technology, Govt. of India. The dataset can be downloaded from here. District boundaries of India are shown in the figure 3.41

3.7.1.3 Block

The block shape files for 15 Indian states were generated by performing a union over the constituent village shape files for each block and can be found here. The village shapefiles for the year 2001 were obtained from the NASA SEDAC website. The village to block mapping was obtained from the Local Government Directory data. The data was processed to store the mapping of villages to a block as shown here. Block boundaries of India are shown in the figure 3.41

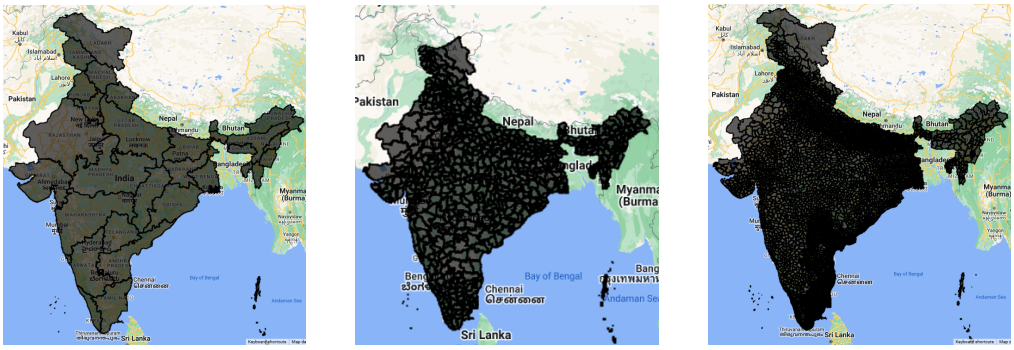


Figure 3.41: State, district and block boundaries (from left to right) of India

3.7.1.4 Panchayat

The panchayat shape files for 15 Indian states were generated by performing a union over the constituent village shape files for each panchayat and can be found here. The village shapefiles for the year 2001 were obtained from the NASA SEDAC website. The village to panchayat mapping was obtained from the Local Government Directory data. The data was processed to store the mapping of villages to a panchayat as shown here. Panchayat boundaries of Mohanpur block, in Gaya district of Bihar are shown in the figure 3.42

3.7.1.5 Village

Shapefiles at the village level for the year 2001 were obtained from the NASA SEDAC website. Village shapefiles hosted on NASA SEDAC website were generated by digitizing the village boundaries from the official analog maps published by the Survey of India [59] for 2001 to compile the socio-economic variables [58] collected during the Indian Census in 1991 and 2001 at village level.

The folder contains state wise village shapefiles compiled with village level household variables of census 2011 (as shown here) and Aggregate Development Index (ADI) [50] of 2011 and 2019. A socio-economic development index called the Aggregate Development Index (ADI) [40] was developed, which is composed of indicators related to asset ownership, primary source of drinking water, primary source of lights, access to bathroom facilities, and literacy rate [50] at a village level. ADI ranges from 5 to 15, with a higher ADI indicating better socio-economic development. Village boundaries of Mohanpur block, in Gaya district of Bihar are shown in the figure 3.42

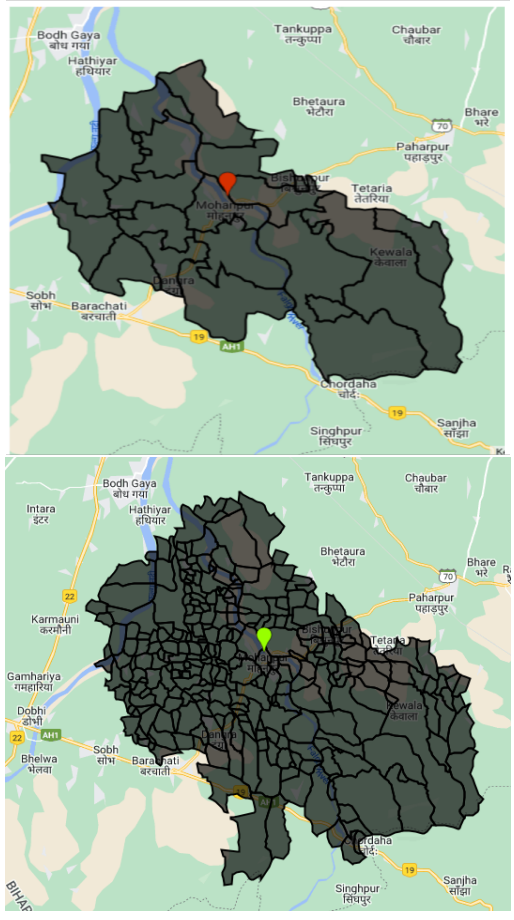


Figure 3.42: Panchayat (top) and village (bottom) boundaries of Mohanpur block, in Gaya district of Bihar

Site level

4.1 Water structure planning layers

4.1.1 Lineament

Introduction In geology, lineaments refer to linear features on the Earth's surface such as faults, fractures, and joints. Lineament is also an important hydrological variable because of its potential to provide groundwater recharge [47]. The groundwater recharge occurs due to the infiltration of rainfall and runoff through the linear features [48].

Input Layers The state-wise lineaments data has been taken from Bhuvan [9] by sending out a WMS request and generating raster files from the packets.

Methodology The raster files of lineament are processed to generate a proximity mask. The Proximity (Raster Distance) algorithm on QGIS computes the distance from the center of each pixel to the center of the nearest pixel on a target pixel (a lineament in this case). The generated raster proximity raster is now scanned for pixels with a distance of less than a buffer value (currently 2m) and selected pixels are exported as raster to form a lineaments buffer mask. This mask is now georeferenced to the respective state for ease of visualization. Lineaments of Mandalgarh block is shown in the figure 4.9a.

Hosting specifications

- Layer type: raster
- Spatial resolution: 194m.
- Temporal resolution: static
- Dataset: Google drive folder

4.1.2 Lithology

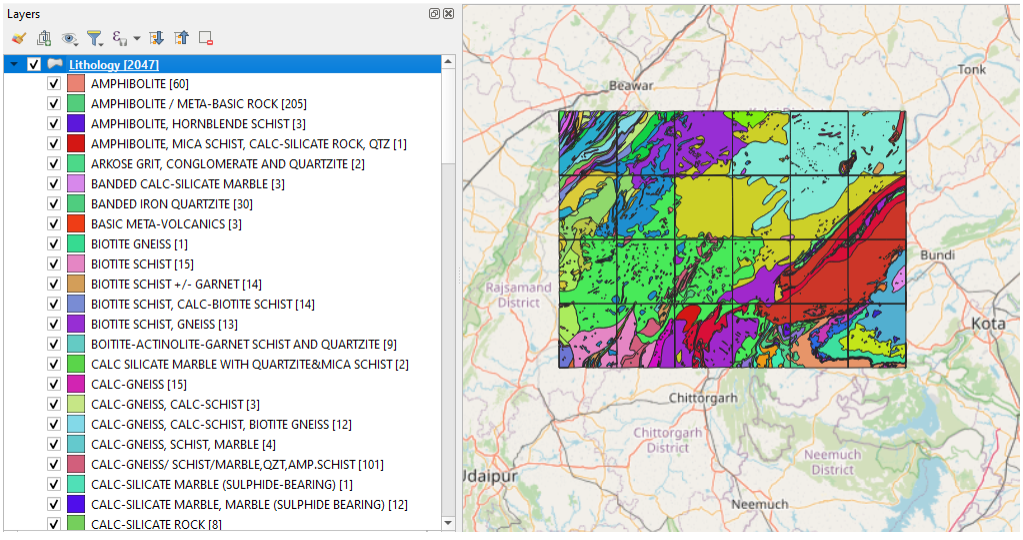
Introduction Lithology is a branch of geology that encompasses the study of various types of rocks, based on their composition, texture, structure, and other properties. The porosity and permeability of rocks influences the groundwater recharge [74]. For example, sandstone has much higher porosity and hence have higher groundwater prospect as compared to granulite rock which is hard rock and has compact structure [62].

Input Layers The state-wise lithology data has been downloaded manually from Bhukosh [7] hosted by Geological Survey of India in the form of shapefile. The lithology shapefile from Bhukosh for Bhilwara district of Rajasthan is shown in the figure **??**.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA.
- Temporal resolution: static
- Dataset: Google drive folder

Figure 4.1: Lithology shapefile from Bhukosh for Bhilwara district of Rajasthan



4.1.3 Drainage lines

Introduction Drainage lines in hydrology refer to the visible channels through which water flows from higher elevations to lower elevations on the surface of earth such as streams and rivers. These lines play a crucial role in the movement of surface water. They collect runoff from precipitation events and transport it downstream, eventually flowing out from the outlet of the watershed.

Input layers Drainage lines uses digital elevation model (DEM) raster dataset as input. The DEM provides pixel level information on the elevation of the terrain, which is used to determine the flow of water across the landscape. We use Shuttle Radar Topography Mission (SRTM) digital elevation model [34] generated by NASA and the National Geospatial-Intelligence Agency (NGA) in February 2000.

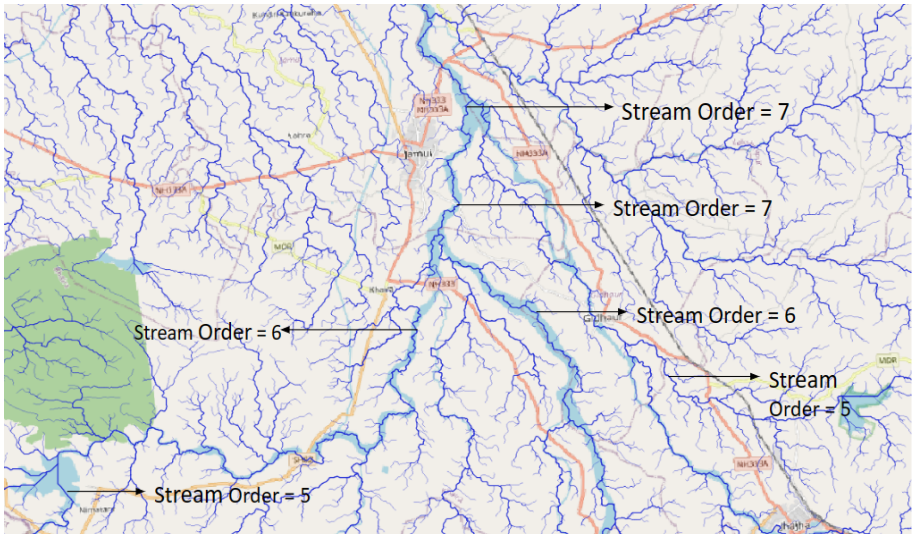
Methodology The drainage lines are generated using the *r.stream.extract* function of GRASS tools in QGIS. The function extract stream channels from the DEM based on flow accumulation threshold. The flow accumulation threshold determines the minimum accumulated flow required to consider a pixel as part of a stream. The function outputs a binary raster map representing the extracted stream channels where pixels are classified as streams with a value of 1, while non-stream pixels are assigned a value of 0. We computed the drainage lines with a flow accumulation threshold of 100.

Stream order provides insight into the hierarchical structure of stream network. Each extracted stream channel is assigned stream order based on its relative hierarchy in the stream network. Streams without any tributaries are classified as first-order streams. When two first-order streams converge, they form a second-order stream. When two second-order streams meet, they create a third-order stream and so on. We assign stream orders based on the graph generated from stream network where the edges are streams and the nodes are intersection of streams. Each node has either degree of 1 (0 incoming – 1 outgoing edge) or 3 (2 incoming – 1 outgoing edge). Corresponding stream/edge of a node with a degree of 1 is assigned with a stream order of 1. While for the remaining nodes, the stream order of outgoing stream/edge is determined based on the stream orders of incoming streams/edges. If both the incoming edges have the same stream order, then the stream order of outgoing stream/edge is assigned as the stream order of the incoming edge + 1. And if the incoming edges have distinct stream orders, then the the stream order of outgoing stream/edge is assigned as the maximum of the two stream orders.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: static
- Dataset: Google drive folder

Figure 4.2: Snapshot of drainage lines with stream orders over Jamui district in Bihar, overlaid on open street map (OSM).



4.1.4 Stream order raster

Introduction While drainage lines are essential to plan assets such as check dams located on the drainage lines, it is also important to identify areas outside of the drainage lines that are under the influence of drainage lines [57]. For example, areas under the influence of stream orders 1-3 can be used to build trenches to slow down low volume but fast runoff, while areas under the influence of stream orders 6-7 can be used to build ponds that see high volumes but slow runoff.

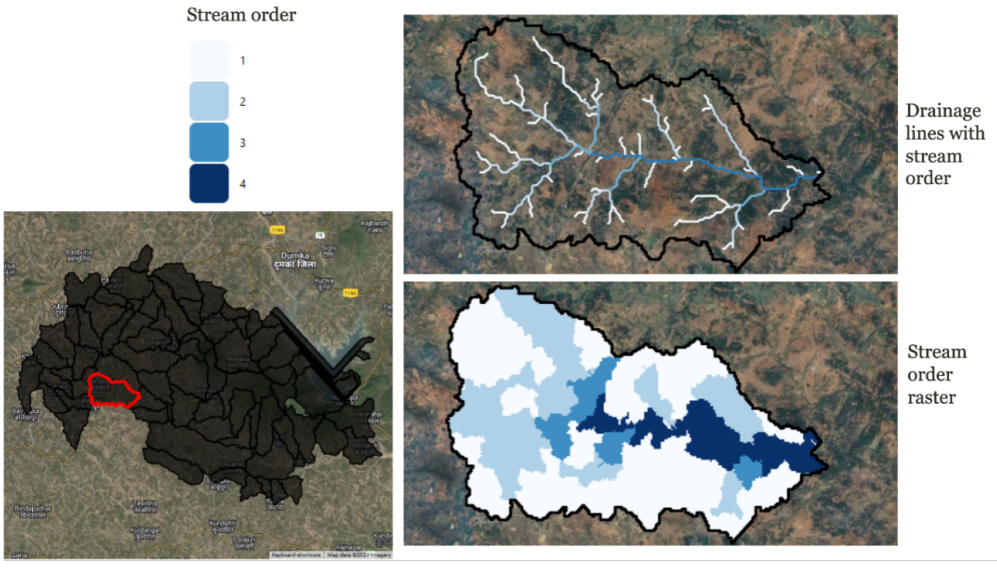
Input layers We use digital elevation model (DEM) data that provides pixel level information on the elevation and drainage lines (with stream orders) as inputs to compute stream order rasters.

Methodology We preprocessed the elevation data to generate a depression-less elevation data. Depression-less elevation data is used to compute flow direction and flow accumulation. A flow accumulation threshold of 100 pixels was used to delineate catchments at the intersection of drainage lines using `r.watershed()` tool of GRASS GIS. The drainage lines were rasterized to embed stream order information on drainage pixels. The catchments were overlaid with drainage raster to assign each catchment with the maximum stream order value present inside the catchment. The catchments embedded with their respective stream order were then rasterized to generate stream order raster with each pixel reflecting the stream order under its influence. The stream order rasters were computed at sub-basin level for all of India.

Hosting specifications

- Layer type: raster
- Spatial resolution: 30 m
- Temporal resolution: static
- Dataset: GEE asset, Google Drive

Figure 4.3: The figure shows drainage lines and the stream order raster map of a micro-watershed (highlighted in red) in the Masalia block in Dumka district of Jharkhand.



4.1.5 Distance to upstream drainage line

Introduction Experienced CSO field staff and volunteers on field indicated that in addition to seeing the recharge potential at a particular site, as indicated by the CLART layer, site-level indicators can be useful to plan MGNREGA water structures such as the distance to nearest upstream drainage line [57].

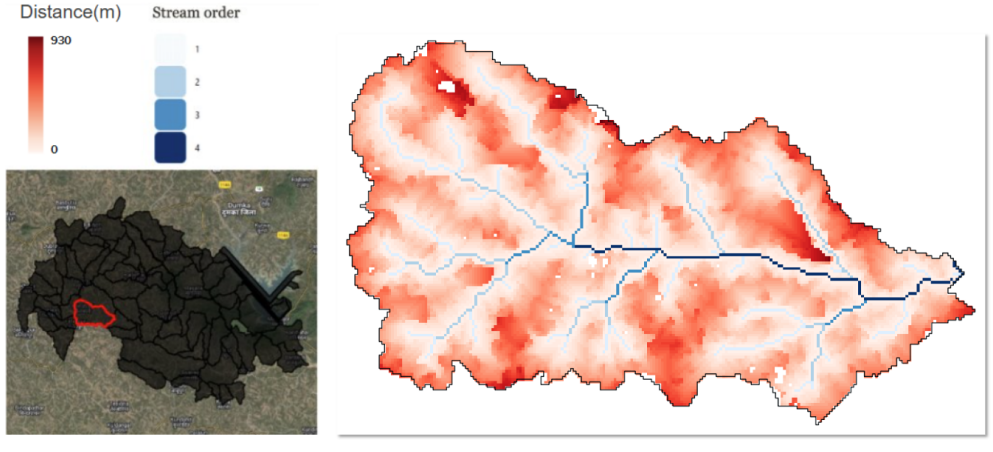
Input layers We use a digital elevation model (DEM) that provides pixel-level elevation information. Additionally, we rasterized drainage lines to create a drainage raster, embedding stream order information in the drainage pixels.

Methodology We leverage spatial indexing technique such as k-d tree to compute the distance of each non-drainage pixel to its nearest upstream drainage line. A k-d tree was constructed using the coordinates of all drainage pixels. The spatial index of k-d tree allows for rapid nearest-neighbor searches that significantly reduces the computational complexity. A drainage pixel is classified as a valid upstream candidate for the non-drainage pixel in concern if its elevation is greater than or equal to the elevation of the non-drainage pixel. If such valid upstream drainage pixels exist, a new k-d tree is constructed using only these pixels, and a nearest-neighbor search is performed to determine the closest upstream drainage pixel. The computed Euclidean distance between the non-drainage pixel and the nearest upstream drainage pixel is then stored in a matrix representing the minimum distances for all non-drainage pixels.

Hosting specifications

- Layer type: raster
- Spatial resolution: 30 m
- Temporal resolution: static
- Dataset: Google drive folder

Figure 4.4: The figure shows distance to nearest upstream drainage line, computed at each pixel and overlaid with drainage lines of a micro-watershed (highlighted in red) in the Masalia block in Dumka district of Jharkhand.



4.1.6 Catchment area

Introduction We found several instances where a more scientific approach to site selection to create new water structures could have been helpful. We observed in Pindwara (Rajasthan) that a check dam had been constructed quite upstream on a drainage line and had a very small catchment area; enough water did not accumulate to make the check dam useful and this resulted in both reducing the amount of water that would otherwise flow downstream, and did not bring any benefit to the neighboring fields either [57]. Therefore, site-level indicators such as an estimate of the catchment area draining into the point could be useful.

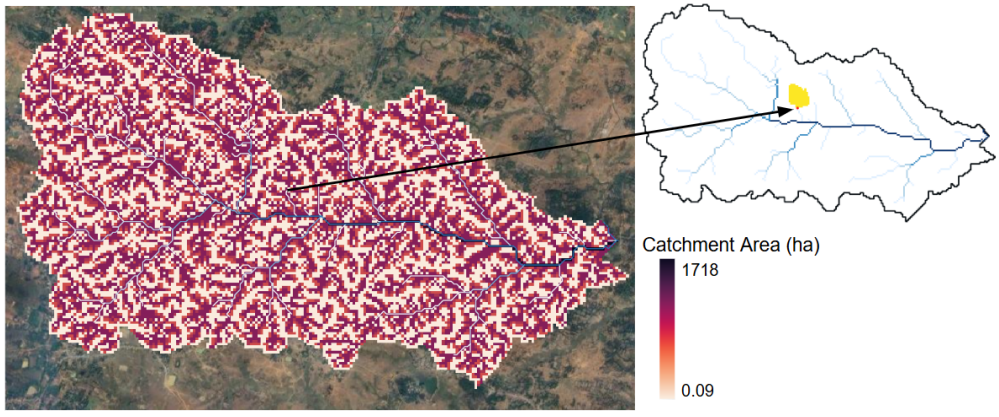
Input layers We use a digital elevation model (DEM) that provides pixel-level elevation information.

Methodology We preprocessed the elevation data to generate a depression-less elevation data. Depression-less elevation data was used and input into `r.watershed` tool of GRASS GIS to compute flow accumulation. Flow accumulation represents the number of upstream pixels flowing to a pixel. The upstream pixels flowing to a pixel forms the catchment area of that pixel. The catchment area for each pixel was calculated in hectares by multiplying its flow accumulation value with the area of a single pixel, based on the 30m spatial resolution of the DEM.

Hosting specifications

- Layer type: raster
- Spatial resolution: 30 m
- Temporal resolution: static
- Dataset: Google drive folder

Figure 4.5: The left figure shows computed catchment area of each pixel of a micro-watershed in the Masalia block in Dumka district of Jharkhand. The right figure shows the delineated catchment area (highlighted in yellow) of a drainage pixel (highlighted in red).



4.1.7 Natural depressions

Introduction Natural depressions are valuable for planning water structures under MGNREGA because they naturally collect and store water, making them ideal sites for various water conservation interventions [57]. Identifying depressions using elevation data helps in reducing the need for extensive excavation for structures such as farm ponds, percolation tanks, and check dams that improve groundwater recharge and surface water availability.

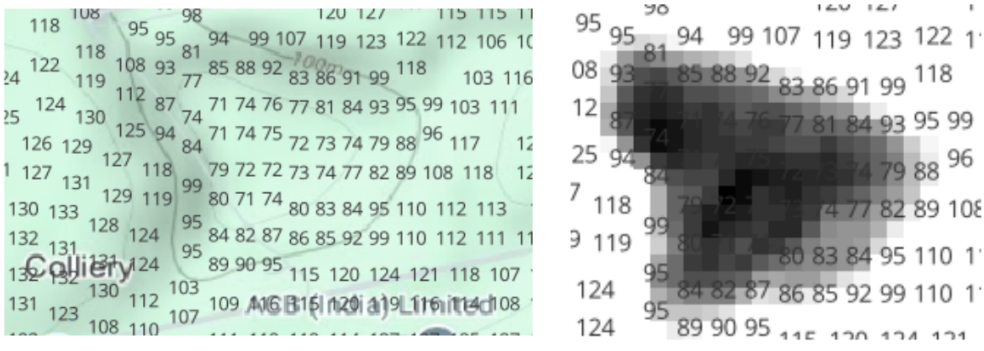
Input layers We use a digital elevation model (DEM) that provides pixel-level elevation information.

Methodology We processed the elevation data to generate a depression-less elevation data. Depression-less elevation data was generated using *r.fill.dir* algorithm of GRASS GIS. The algorithm identifies and fills sinks to ensure that water keeps flowing across the landscape. Sinks in an elevation matrix are pixels where water would get trapped because all surrounding pixels have a higher elevation. Group of spatially contiguous sinks can form a natural depression that help retain water. We iteratively run the algorithm to identify and fill sinks until no further sinks are detected. The original elevation data is then subtracted from the final processed elevation data to extract areas under natural depressions.

Hosting specifications

- Layer type: raster
- Spatial resolution: 30 m
- Temporal resolution: static
- Dataset: Google drive folder

Figure 4.6: The figure shows natural depression area in black (right) computed using elevation data (left).



4.2 Water structure planning algorithms

4.2.1 CLART

Introduction The Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), promises unskilled manual work to every adult for at least 100 days in rural parts of the country. To ensure water security in the rural landscape assets such as check dams, percolation tanks, and irrigation channels are constructed and maintained. The overall focus of the scheme is on ensuring higher incomes for farmers by increasing the water availability and productivity of land. The existing method of suitable site identification for these structures has been based on local people's opinion, and taking into consideration factors such as geomorphology, climate, annual rainfall, vegetation cover, distances from farms and so on. However, misidentification of suitable locations for building water assets has resulted in non-operation, non-utilization and inefficient investment. Therefore, there is a need for site suitability assessment in terms of whether the site provides groundwater recharge or surface water storage or has a high runoff.

The India Observatory (IO), the technological branch of the Foundation for Ecological Security (FES), has created Composite Land Assessment & Restoration Tool (CLART) with the goal of empowering rural communities for planning the sites for soil and water conservation measures. The primary objective is to facilitate the restoration of degraded ecosystems and enhance economic opportunities in rural areas. It achieves this by providing access to geospatial datasets such as rock type, slope and landcover in a geographic information system (GIS) based android application. We intend to reproduce CLART in a reprogrammable/reconfigurable manner to allow dynamic scoring based on community requirements for their landscape.

Input Layers

- **Lineaments:** as mentioned in section 7.0.1.
- **Lithology:** as mentioned in section 7.0.2.
- **Drainage density:** Drainage network of the watershed helps in visualizing which areas have a high groundwater recharge. Drainage density is directly related to slope and inversely proportional to permeability. The steeper the slope with low permeability, the higher the drainage density, thus less infiltration and more surface runoff. We used QGIS Line Density tool, which calculates drainage density for each pixel as the sum of lengths of the drainage lines in the search radius weighted by their stream orders. We used elevation raster [34] as input to Line Density tool with 1km and 10m as the search radius and pixel size respectively.
- **Slope percentage:** Slope represents the rate of change of elevation for each DEM pixel. Steep slopes generally reduce recharge as runoff flows very rapidly and would not permit infiltration. Plains, however, enhance groundwater recharge because higher retention time is provided for rainwater to infiltrate the soils. We used SRTM DEM [34] to obtain pixel-wise slope on Earth Engine.

Methodology We processed the input layers above to generate pixel level score for each layer as shown below:

- The Central Ground Water Board (CGWB) document defines rainfall infiltration factor (RIF) for each major aquifer [28]. Each lithologic group in the shapefile was mapped to one of the major aquifers using the string matching algorithm to assign scores of

recharge. The range of RIF was equally binned into three bins to assign scores of 1 (low recharge), 2 (moderate recharge), 3 (high recharge). The higher the RIF, the more the recharge. The shapefiles were then rasterized with the recharge score as the burn in value.

- The observed range of drainage density values was binned into three equal parts to assign scores of 1 (high recharge), 2 (moderate recharge), 3 (low recharge). The lower the drainage density, the more the no. of higher stream orders that enable more groundwater recharge.
- The lineament layer was processed to generate a proximity mask by taking a 2m buffer around the lineament to assign a score of 10 within the buffer and a score of 1 outside the buffer.

Recharge potential is a measure of the ability of an area to recharge the groundwater. The scores of lithology, drainage density and lineaments are multiplied to determine the recharge potential as shown in the equation 18 and table 4.1. Recharge potential is combined with different slope percentages and land cover categories to derive treatment codes as shown in the figure 4.7. The flowchart of methodology is shown in the figure 4.8, where community based score reassignment can be performed to produce CLART output accordingly.

$$Recharge\,potential = Lineament * Lithology * Drainage\,density \tag{4.1}$$

where lineament scores can have values of 1, 10. Lithology and drainage density scores can have values of 1, 2, 3 resulting the table 4.1.

Hosting specifications

- Layer type: raster
- Spatial resolution: 30m
- Temporal resolution: static
- Codebase: Github repository

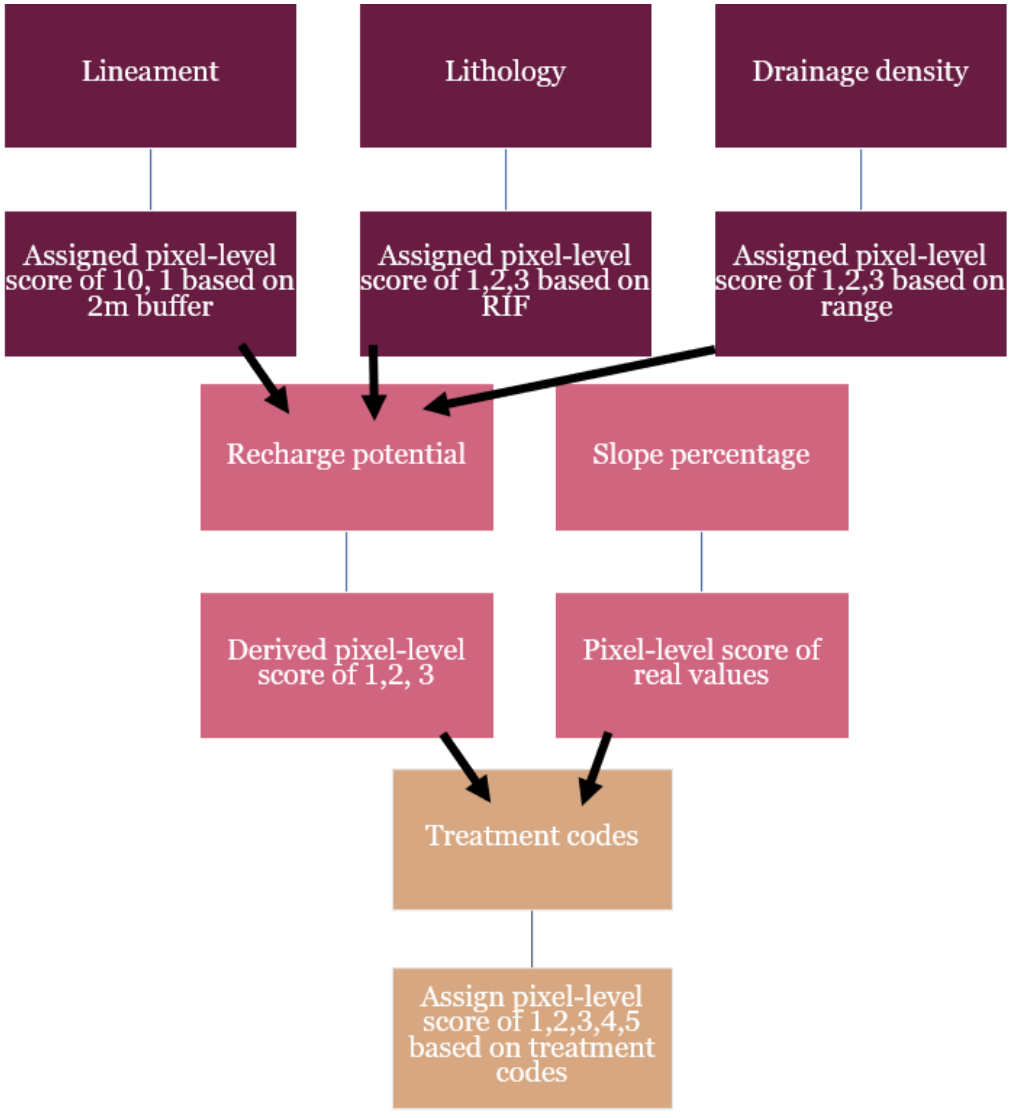
Table 4.1: Recharge potential values and their corresponding scores for recharge

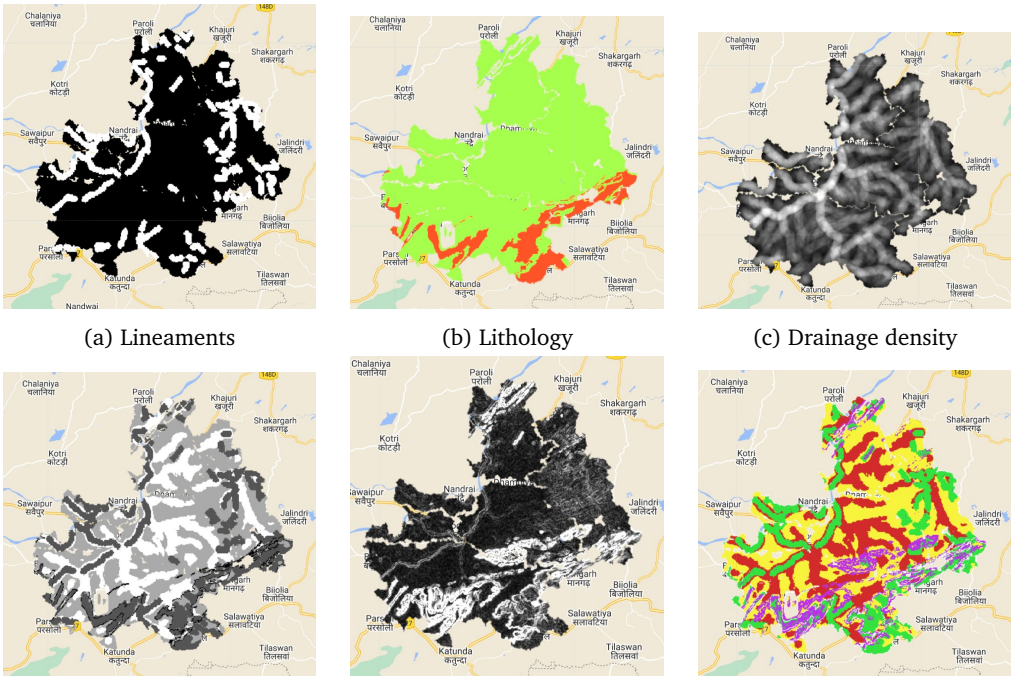
Recharge potential values	**Score**
1,2,10,20,30,40,60,90	1 (high recharge)
3,4	2 (moderate recharge)
6,9	3 (low recharge)

Figure 4.7: The recharge potential scores are combined with different slope percentages and land cover categories to derive treatment codes in CLART.

Recommended Treatment Code	Recommended Treatment Type	Recharge Potentiality	Slope	Land Use/Land Cover
1	Good Recharge structure (Percolation tank, WHS, CCT etc)	Very High (5) High (4)	3-5 % (2) 5-7% (3)	Current fallow (5), Other Waste land (9), Gullied (10), Scrubland (11)
2	Moderate Recharge structure (WAT, GP, LBCD etc)	Moderate (3)	0-3% (1), 3-5 % (2) 5-7% (3)	Current fallow (5), Other Waste land (9), Gullied (10), Scrubland (11)
3	Surface water Harvesting structure (WHS, FP, FB etc)	Low (2) Very Low (1)	0-3% (1), 3-5 % (2)	Current fallow (5), Other Waste land (9), Gullied (10), Scrubland (11)
9	Regeneration (Plantation, Veg Int, Grass seeding, stone bunding, bench terracing etc)	1,2,3	Slope >10%	Current fallow (5), Other Waste land (9), Gullied (10), Scrubland (11) Mixed, degraded forrest, Decidous forest

Figure 4.8: Flowchart of methodology to produce reprogrammable CLART using score assignment





(a) Lineaments (b) Lithology (c) Drainage density

(d) Recharge potential (e) Slope percentage (f) CLART

Figure 4.9: The figure shows the component layers for generating CLART output for Mandalgarh block in Bhilawara district of Rajasthan

4.2.2 Runoff accumulation

Introduction Experienced CSO field staff and volunteers also indicated that it would be useful to judge runoff accumulation capacity for the site while planning water structures under MGNREGA. Runoff accumulation estimates and its sensitivity to drought and non-drought years can help quantification of storage and infiltration to support supplemental irrigation cycles.

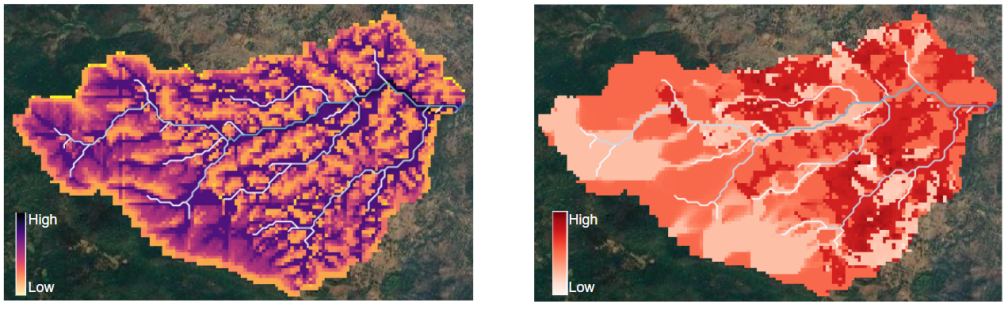
Input layers We use pixel-level elevation data, upstream-downstream microwatershed connectivity mentioned in section 4.2.3, drainage lines and pixel-level runoff data mentioned in section 3.1.2.2 to compute pixel-level runoff accumulation.

Methodology The pixel-level runoff raster is accumulated using flow direction data to compute runoff accumulation at each pixel. The microwatershed connectivity graph is analyzed to identify if there are any upstream micro-watersheds contributing runoff. If no upstream micro-watersheds exist, the process directly outputs the runoff accumulation raster. If any upstream microwatersheds are present, then we add the total runoff from the upstream micro-watersheds to all pixels along the path of the drainage line connecting the micro-watersheds.

Hosting specifications

- Layer type: raster
- Spatial resolution: 30 m
- Temporal resolution: yearly
- Codebase: Github repository

Figure 4.10: The figure in left shows average runoff accumulated in non-drought years and the figure on the right shows sensitivity of runoff accumulation during drought years expressed as percentage deviation.



4.2.3 Upstream-downstream waterbodies

Introduction Water bodies may also influence one another [57]. New upstream water bodies would reduce the runoff flow for downstream water bodies. To determine where and how many such upstream water bodies can be constructed to not severely impact downstream water bodies requires simulations of rainfall and runoff. In ongoing research, we are building this method by first constructing a connectivity graph of water bodies, and then studying how the addition of new nodes on this graph would alter the runoff flow. As an example, figure 4.11 shows the drainage lines and water bodies in a micro-watershed, and the resulting waterbody connectivity graph.

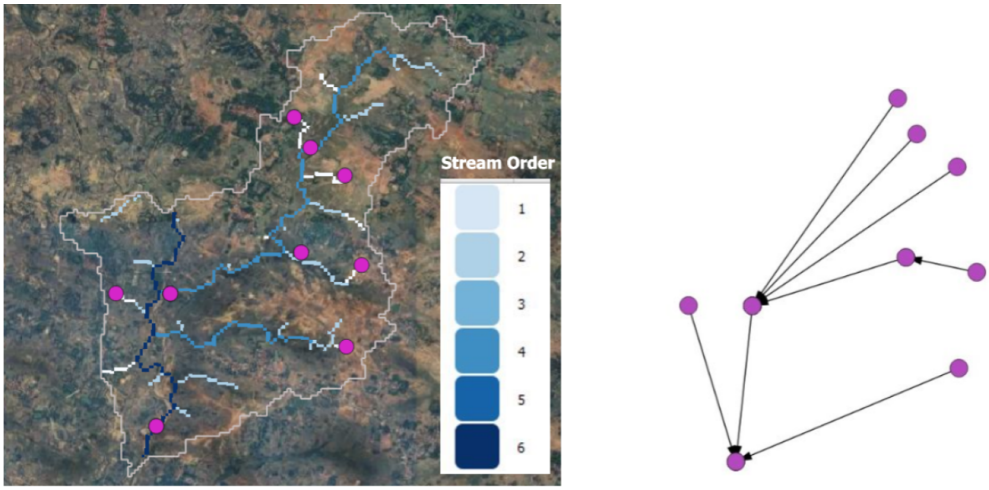
Input layers We use water bodies detected using land use land cover classification (as mentioned in section 3.4.5), drainage network of stream orders and elevation data to determine the connectivity of water bodies.

Methodology We intersect the water bodies with the drainage network. For each non-intersecting waterbody, the centroid is computed, and elevation data is used to determine the downslope direction. The downslope path is iteratively traced until it intersects with a drainage line, and this newly constructed path is integrated into the existing drainage network, ensuring all waterbodies are connected. Once connectivity is established, a waterbody connectivity graph is created by identifying the downstream waterbody within the 1-neighborhood of each waterbody. A directed edge is added from the upstream to the downstream waterbody, forming a waterbody connectivity graph.

Hosting specifications

- Layer type: vector
- Spatial resolution: NA
- Temporal resolution: yearly
- Codebase: Github repository

Figure 4.11: The figure in left shows drainage lines (in shades of blue) with stream order. Using the stream order hierarchy and spatial arrangements of water bodies (in purple circle as shown on top of these drainage lines), waterbody graph is generated (using directional edges in black) to identify upstream-downstream waterbodies as shown in the figure on the right.



4.3 Resource mapping

4.3.1 Farm ponds and wells

Introduction Rural areas face big challenges from climate change, and they need good ways to manage their resources and develop sustainably. Programs like Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) helps by building water storage structures like farm ponds and wells to support farming and village life. But there are problems with how this work is planned and carried out. Sometimes efforts are duplicated, resources aren't used well, and decisions aren't based on good data.

The traditional way of surveying and monitoring these water structures involves people doing manual checks, which takes a lot of time and effort and isn't very accurate. While some researchers have used satellite images to study land use and infrastructure, there's still a need for better ways to spot smaller structures like farm ponds and wells from satellite pictures. Some studies have used advanced techniques to detect changes in these structures over time, but there are still gaps in mapping their exact locations and creating systems that can work across large areas. Getting better at finding and mapping these water structures would help manage water resources more fairly and efficiently across rural communities.

Our work addresses these gaps by introducing an automated framework that integrates deep learning YOLOv11 model with high-resolution satellite imagery to detect and geotag ponds and wells across rural landscapes. The system processes satellite imagery at approximately 1.194 meters per pixel at zoom level 17 for ponds and 0.597 meters per pixel at zoom level 18 for wells. The YOLOv11 model, trained on satellite imagery from Maharashtra, achieves 80% mean average precision for farm ponds and 89.3% mean average precision for wells. To demonstrate scalability, image tiles across diverse geographical regions throughout India are processed through the trained model and predicted bounding boxes are converted to geographic coordinates using affine transformation matrices for precise GIS integration. The model's robustness was verified through human verification studies, which confirmed correct detection rates of 70% for ponds and 80% for wells across different landscapes.

Input layers The input to our model consists of high-resolution satellite images, obtained from Google Maps at specific zoom levels - level 17 for farm ponds and level 18 for wells.

Methodology The study employed YOLOv11, the latest iteration in the YOLO series, to detect farm ponds and wells from satellite imagery. The model, which was pre-trained on COCO and ImageNet datasets, was fine-tuned using custom datasets. For ponds, the dataset included 752 training and 77 test images containing 1,194 annotations, while the wells dataset comprised 544 training and 60 test images with 841 annotations. The satellite imagery was collected at zoom level 17 for ponds and zoom level 18 for wells, providing an optimal balance between area coverage and detail for precise identification. The data to train ponds can be found here, and the data to train wells can be found here.

The training process involved 100 epochs with specific data augmentation techniques to improve the model's generalization capabilities. These augmentations included a 30% probability for vertical and horizontal flipping, rotation within a range of $\pm 10°$, and translation up to $\pm 10\%$. To optimize the training process, the OneCycleLR learning rate scheduler was employed with an initial learning rate of 0.01 and a final learning rate of 0.2. Farm ponds were categorized into two classes - Dry Farm Ponds and Wet Farm Ponds - to account for visual variability, while wells were treated as a single class.

For inference, the model was tested on data from diverse geographical locations including Masalia (Jharkhand), Mohanpur (Bihar), Pindwara (Rajasthan), and Paithan (Maha-

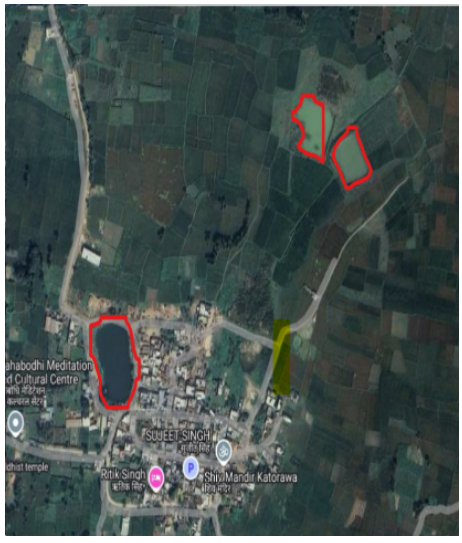
rashtra). The model predicted mask coordinates as (x_1, y_1) points representing polygon vertices for each detected asset. These pixel-based coordinates were then converted to geographic locations using formulas that considered the tile coordinates and zoom level. The conversion process involved calculating the longitude and latitude of tile corners and then mapping the predicted polygon coordinates to real-world geographical coordinates.

To improve accuracy in pond detection, an entropy-based filtering solution was implemented to reduce false positives. This approach calculated entropy values for predicted masks of wet farm ponds, with a threshold of 5.2 used to distinguish true positives from false positives. Higher entropy values indicated greater uncertainty in predictions, allowing the system to filter out likely false detections. For dry farm ponds and wells, the entropy calculation step was bypassed as these didn't exhibit the same characteristic entropy behavior as wet farm ponds.

The final stage involved converting the processed data into a standardized format for visualization and analysis. The detection results, including image paths, predicted classes, and geographic coordinates of polygon vertices, were compiled into CSV files. This data was then converted into GeoJSON format, where each detected polygon was represented as a geographical feature with associated class labels and geometry. Finally, the GeoJSON features were transformed into GeoDataFrames using the Web Mercator coordinate reference system (EPSG:3857), enabling visualization and analysis in GIS tools. The accuracy of the detection was validated through human evaluation, with wells detection achieving higher accuracy (0.81), precision (0.82), and recall (0.95) compared to pond detection (accuracy: 0.69, precision: 0.77, recall: 0.80).

Hosting specifications

- Layer type: Vector
- Spatial resolution:
  - Zoom level 17 for farm ponds (approximately 1.19 meters/pixel)
  - Zoom level 18 for wells (approximately 0.60 meters/pixel)
  - 256 x 256 pixels per tile (standard Google Maps tile size)
- Temporal resolution: static
- Codebase: Github repository



(a) Visualization of farm ponds geolocations in QGIS



(b) Visualization of wells geolocations in QGIS

Figure 4.12: Screenshots showing geolocations of predicted assets visualized in QGIS: (a) farm ponds and (b) wells.

4.3.2 Farm boundaries and plantations

Introduction Since agriculture in most parts of India is rain-fed, global products like Dynamic World struggle to differentiate between agricultural land and scrubland. The error is evident in the areas of Surgana where a hilly but scrubby area comprising of a lot fields in it is classified mostly as scrubland. These classification errors hinder downstream tasks such as predicting cropping intensity on agricultural land. To address this, we developed a local compute pipeline that downloads high-resolution images at 1.19 m pixel resolution and applies computer vision models to segment the images into farmland, plantations, and scrubland.

Next, we designed an algorithm to classify these segments based on the insight that, unlike the irregularly shaped scrublands, agricultural fields exhibit either lower randomness—quantified by entropy measures—or structured planting patterns, which can be detected using techniques like the Hough transform. The resulting farmland, plantation, and scrubland boundaries are then converted into vector assets and uploaded to Google Earth Engine. Within our LULC pipeline, these vectors are used to refine downstream tasks, such as predicting cropping patterns exclusively within farm areas.

Methodology First, the region of interest (ROI) is divided into blocks, each containing a 16×16 grid of 256×256-pixel images, with each pixel representing a resolution of 1.19 m as shown in the figure 4.15. The entire ROI must be encompassed within these blocks. The pipeline consists of three sequential stages, executed serially for each block as shown in the figure 4.13.

In the first stage, high-resolution images for the block are downloaded and merged into a GeoTIFF, preserving spatial information. This spatial data is later used to restore boundary details detected in the RGB image. Additionally, individual 256×256 RGB images are stored separately, as the computer vision model processes images of this size. The Wang et al. model [76] is applied to the block, generating predictions for all 16×16 images. This is a FracTAL-ResUNet model pre-trained on France field boundaries and uses transfer learning and weak supervision to fine-tune the model on 10,000 farm boundaries from all over the India. The model uses Airbus SPOT imagery as input. The output of the model consists of a field extent prediction, a field boundary prediction, and a distance to boundary prediction.

In the next stage, we combine the predictions from all images within a block to form a single, larger, and complete image. The watershed algorithm is then applied to this composite image which interpret it as a topographic map, where pixel intensity corresponds to elevation. The algorithm identifies objects by initializing seeds at local minima and allowing them to grow outward as the intensity increases, resembling water spreading through a basin. The expansion ceases when regions encounter the boundaries of neighboring objects. This process utilizes field extent and boundary predictions to segment the image into multiple regions, each of which can represent a farm, plantation, or scrubland.

In the next stage, we need to classify these segments into their respective categories. To achieve this, we formulated rules based on entropy, Hough transform, segment size, fractal dimension, and rectangularity to classify the simpler segments into farm, plantation, and scrubland categories. These easily classified segments serve as a foundation, leveraging their properties to incorporate the more challenging ones. An example of a difficult segment to classify solely using these rules is a farm containing several trees, which increases overall entropy. Since the trees are not arranged in a grid pattern, the Hough transform fails to detect perpendicular lines. However, the segment's rectangularity and fractal dimension prevent it from being classified as scrubland, as scrubland segments tend to be irregular. These segments then fall into the gray area as described in 4.14. To incorporate these challenging segments, we analyze their mean NDVI time series and compare them with the

easy positives. We identify the five most similar segments within the easy positives and assign the most frequent (mode) label among them to the ambiguous segment. Through this approach, all segments are ultimately classified into the desired categories.

The predicted farm, plantation and scrubland boundaries are then converted to vector format using the spatial information previously stored and uploaded to Google Earth Engine as an asset.

Hosting specifications

- Layer type: Vector
- Input spatial resolution: 1.19 m pixel
- Temporal resolution: static
- Codebase: Github Repository

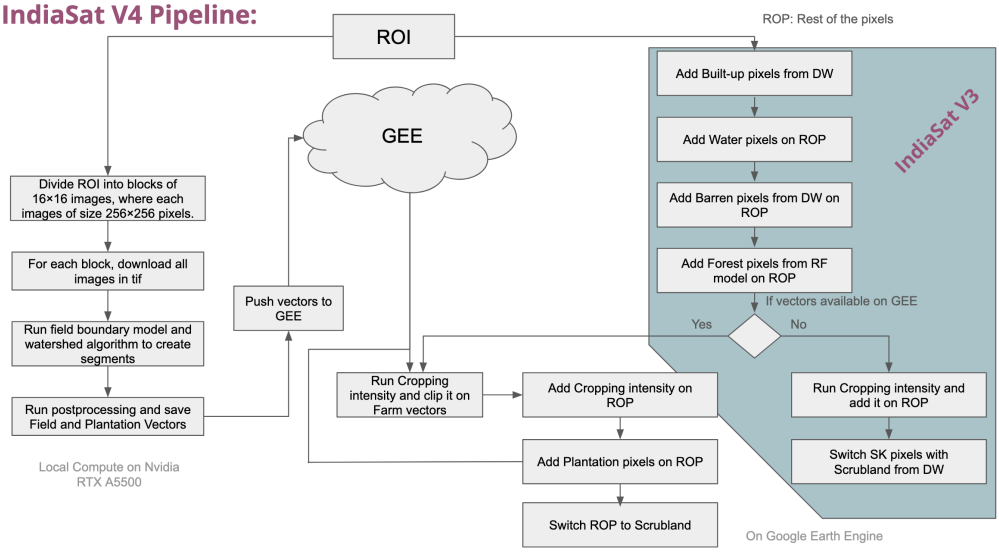


Figure 4.13: Farm-boundary pipeline integrated with LULC pipeline to create IndiaSAT v4

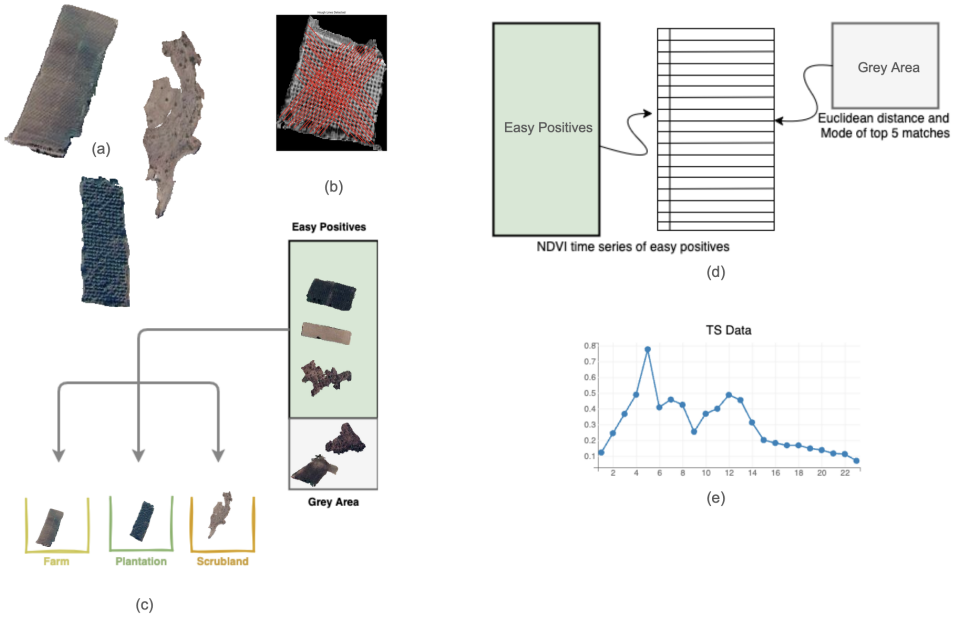


Figure 4.14: (a) Example of farm, plantation and scrubland segment, (b) Result of Hough transform when applied on a gridded plantation, (c) Illustration showing easier segments being separated into different buckets of farm, plantation and scrubland using rules (d) Using mean NDVI timeseries to bring in gray area (e) Example of a NDVI time series

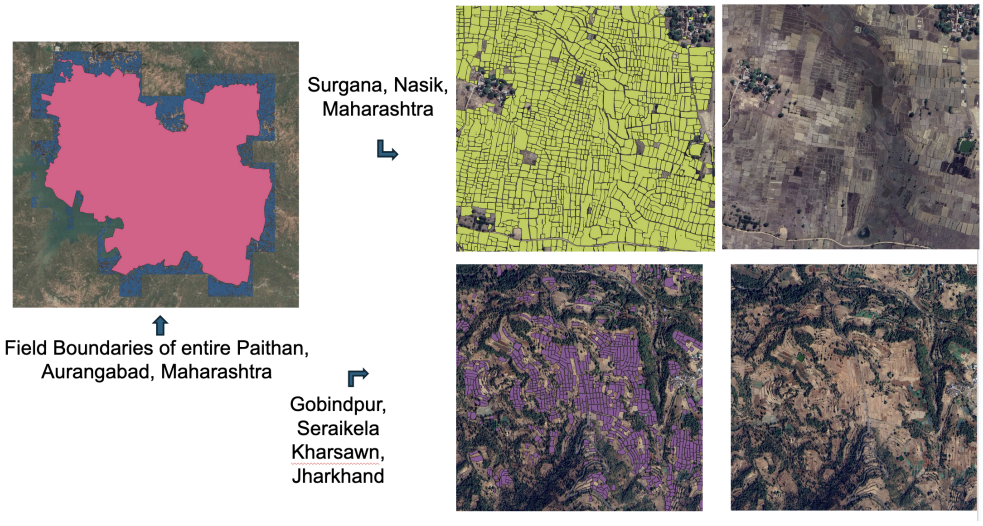


Figure 4.15: Dividing Roi into blocks and results of farm boundaries on several areas

Bibliography

[1] NSIDC DAAC's Data Access and Service API. https://github.com/nsidc/NSIDC-Data-Access-Notebook.

[2] Administrative Boundaries of India. https://drive.google.com/drive/folders/1HJef1T8wJg8eoz4kUU8u1HHNoH3510Zv?usp=drive_link.

[3] Annual Report 2008, INDIA METEOROLOGICAL DEPARTMENT, Ministry of Earth Sciences, Govt. of India. https://metnet.imd.gov.in/docs/imdnews/ANNUAL_REPORT2008English.pdf.

[4] Aquifer properties and decription. https://csciitd-my.sharepoint.com/:x:/g/personal/anz208490_iitd_ac_in/EYzjXM0tf1BCjPH5x5IwtDgBEPYuGu-VTNi655cHIpWZsg?e=Sk643b.

[5] Aquifer Systems of India, Central Ground Water Board, Ministry of Water Resources, Government of India. https://www.cgwb.gov.in/cgwbpnm/public/uploads/documents/16874195126800234 37file.pdf.

[6] Atal Bhujal Yojana (Atal Jal), Department of Water Resources, River Development and Ganga Rejuvenation, Ministry of Jal Shakti, Government of India. https://ataljal.mowr.gov.in/.

[7] Bhukosh, Geological Survey of India. https://bhukosh.gsi.gov.in/Bhukosh/Public.

[8] Bhuvan GEO-MGNREGA. https://bhuvan-app2.nrsc.gov.in/mgnrega/nrega_dashboard_phase2/.

[9] Bhuvan, Indian Geo-Platform of ISRO, NRSC. https://bhuvan-app1.nrsc.gov.in/thematic/thematic/index.php.

[10] CGWB Ground water resources, WRIS India. https://indiawris.gov.in/wiki/doku.php?id=cgwb_ground_water_resources.

[11] CGWB wells and groundwater level data, WRIS India. https://indiawris.gov.in/wris/#/groundWater.

[12] Chapter 2 - FAO Penman-Monteith equation. https://www.fao.org/3/x0490e/x0490e06.htm.

[13] Dynamic World V1. https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_DYNAMICWORLD_V1.

[14] Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System FLDAS Noah Land Surface Model L4 Central Asia Daily 0.01 x 0.01 degree (FLDAS_NOAH001_G_CA_D). https://disc.gsfc.nasa.gov/datasets/FLDAS_NOAH001_G_CA_D_001/summary.

[15] First Census of Water Bodies Data, Open Government Data (OGD) platform of India. https://data.gov.in/catalog/first-census-water-bodies-data.

[16] Global Hydrologic Soil Groups (HYSOGs250m) for Curve Number-Based Runoff Modeling, ORNL DAAC. https://daac.ornl.gov/SOILS/guides/Global_Hydrologic_Soil_Group.html.

[17] Global Satellite Mapping of Precipitation (GSMaP), JAXA Earth Observation Research Center. https://developers.google.com/earth-engine/datasets/catalog/JAXA_GPM_L3_GSMaP_v6_operational.

[18] Ground Water Level Data under Atal Bhujal Yojana. https://www.data.gov.in/catalog/ground-water-level-data-under-atal-bhujal-yojana.

[19] Guidelines/Criteria for evaluation of proposals/requests for ground water abstraction, Central Ground Water Authority. https://www.cgwb.gov.in/sites/default/files/2022-10/revised_guidelines_12112015.pdf.

[20] Indian River Systems, WRIS. https://indiawris.gov.in/wiki/doku.php?id=river_info.

[21] Manual for Drought Management 2016, Ministry of Agriculture and Farmers Welfare, Government of India. https://sdma.cg.gov.in/ManualDrought2016.pdf.

[22] MASTER CIRCULAR A GUIDE FOR PROGRAMME IMPLEMENTATION FY 2018-2019. https://nregs.kerala.gov.in/mgnregs/wp-content/uploads/2021/01/AMC_18-19.pdf.

[23] MGNREGA MIS. https://mnregaweb4.nic.in/netnrega/error.html?aspxerrorpath=/netnrega/MISreport4.aspx.

[24] MGNREGA work category. https://csciitd-my.sharepoint.com/:x:/g/personal/ramneekk_iitd_ac_in/EfM9Wta6TVpAnxbMADWb5ZIBg8fIVp9NhHmfH0LjmgMnlA?rtime=DBETvY_V20g.

[25] MGNREGS Permissible Work List. https://megsres.nic.in/sites/default/files/mgnrega-permissible-work-list.pdf.

[26] MODIS Collections in Earth Engine. https://developers.google.com/earth-engine/datasets/catalog/modis.

[27] Module 5- Introduction to Canal System, Definitions and Unit, National Water Academy, Pune. https://www.nwapune.gov.in/Module%205/English%20Material/5.1-Eng-Narrative-Intro%20to%20canals-RKA.pdf.

[28] National Compilation on Dynamic Ground Water Resources of India, 2022. https://static.pib.gov.in/WriteReadData/userfiles/file/GWRA2022(1)HIDO.pdf.

[29] National Compilation on Dynamic Ground Water Resources of India, 2022. https://static.pib.gov.in/WriteReadData/userfiles/file/GWRA2022(1)HIDO.pdf.

[30] Open Street Maps. https://www.openstreetmap.org/.

[31] Procedure to download and process NREGA geotagged assets. https://drive.google.com/file/d/1vSi8c-orzVuZH9Zd3-_X8Z--vfukrIDE/view?usp=sharing.

[32] Sentinel-1 SAR GRD. https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD.

[33] Sentinel-2 MSI. https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR.

[34] SRTM Digital Elevation Data Version 4. https://developers.google.com/earth-engine/datasets/catalog/CGIAR_SRTM90_V4.

[35] USGS Landsat 7 Level 2. https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LE07_C02_T1_L2.

[36] USGS Landsat 8 Level 2. https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2.

[37] Water Bodies, First Census Report, Ministry of Jal Shakti, Government of India. https://cdnbbsr.s3waas.gov.in/s3a70dc40477bc2adceef4d2c90f47eb82/uploads/2023/04/2023040672.pdf.

[38] Watershed Development for Rainfed Areas under MGNREGS MGNREG, Designed and Developed by WASSAN as part of GIZ Project “Environmental benefits of the Mahatma Gandhi National Rural Employment Guarantee Act". https://e-saksham.nic.in/wp-content/uploads/2018/07/Watershed-development-for-rainfed-areas-under-MGNREGS-A-resource-booklet-fo
pdf.

[39] Peter C. Austin. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46:399–424, 2011.

[40] Chahat Bansal, Arpit Jain, Phaneesh Barwaria, Anuj Choudhary, Anupam Singh, Ayush Gupta, and Aaditeshwar Seth. Temporal prediction of socio-economic indicators using satellite imagery. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, pages 73–81. 2020.

[41] Soumendra N Bhanja, Abhijit Mukherjee, Matthew Rodell, Yoshihide Wada, Siddhartha Chattopadhyay, Isabella Velicogna, Kishore Pangaluru, and James S Famiglietti. Groundwater rejuvenation in parts of india influenced by water-policy change implementation. *Scientific reports*, 7(1):7453, 2017.

[42] Ritu Bharadwaj, Simon Addison, and Mohan Reddy. Climate resilience information system and planning tool for mahatma gandhi national rural employment guarantee scheme: the crisp-m tool, 2021.

[43] Kantha Rao Bhimala, GK Patra, and Sheshakumar Goroshi. Annual and seasonal trends in actual evapotranspiration over different meteorological sub-divisions in india using satellite-based data. *Theoretical and Applied Climatology*, 152(3-4):999–1017, 2023.

[44] Grace Carswell and Geert De Neve. Women at the crossroads: Implementation of employment guarantee scheme in rural tamil nadu. *Economic and Political Weekly*, 48(52):82–93, 2013.

[45] Chao Chen, Xinyue He, Zhisong Liu, Weiwei Sun, Heng Dong, and Yanli Chu. Analysis of regional economic development based on land use and land cover change information derived from landsat imagery. *Scientific Reports*, 10(1):12721, 2020.

[46] Argade S Dadabhau and Wadkar S Kisan. Role of socio-psychological status of beneficiaries in their effective participation in mgnrega. *International Journal of Scientific Research*, 2(6):535–537, 2013.

[47] Sumit Das and Sudhakar D Pardeshi. Comparative analysis of lineaments extracted from cartosat, srtm and aster dem: a study based on four watersheds in konkan region, india. *Spatial Information Research*, 26:47–57, 2018.

[48] Arjun B Doke, Rajendra B Zolekar, Hemlata Patel, and Sumit Das. Geospatial mapping of groundwater potential zones using multi-criteria decision-making ahp approach in a hardrock basaltic terrain in india. *Ecological Indicators*, 127:107685, 2021.

[49] Sheshakumar Goroshi, Rohit Pradhan, Raghavendra P Singh, KK Singh, and Jai Singh Parihar. Trend analysis of evapotranspiration over india: Observed from long-term satellite measurements. *Journal of Earth System Science*, 126:1–21, 2017.

[50] Anant Gulgulia, Aman Gupta, Akshay P Sarashetti, Aaditya Sinha, and Aaditeshwar Seth. Tracking socio-economic development in rural india over two decades using satellite imagery. *ACM J. Comput. Sustain. Soc.*, 1(2), dec 2023.

[51] P K Gupta, Sam Chauhan, and M P Oza. Modelling surface run-off and trends analysis over india. *Journal of Earth System Science*, 125:1089–1102, 2016.

[52] Himanshu, Abhiroop Mukhopadhyay, and M R Sharan. *The National Rural Employment Guarantee Scheme in Rajasthan: Rationed funds and their allocation across villages.* Number 35 in ESID Working Paper. Effective States and Inclusive Development Research Center, 2015.

[53] Guido W. Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, 86:4–29, 2004.

[54] Shriram Kadiya, Sapna Parashar, and Sanket Vatavwala. Work demand pattern analysis for mgnrega: With special reference to 18 indian states. *Scientific Papers Series "Management, Economic Engineering in Agriculture and Rural Development"*, 16(4), 2016.

[55] Stephanie K Kampf, Stephen J Burges, John C Hammond, Aditi Bhaskar, Tim P Covino, Abby Eurich, Hannah Harrison, Michael Lefsky, Caroline Martin, Daniel McGrath, et al. The case for an open water balance: Re-envisioning network design and data analysis for a complex, uncertain world. *Water Resources Research*, 56(6):e2019WR026699, 2020.

[56] Amy McNally, Kristi Arsenault, Sujay Kumar, Shraddhanand Shukla, Pete Peterson, Shugong Wang, Chris Funk, Christa D Peters-Lidard, and James P Verdin. A land data assimilation system for sub-saharan africa food and water security applications. *Scientific data*, 4(1):1–19, 2017.

[57] Shivani A Mehta, Aila Dutt, Ajay Tannirkulam, Akshay Pratap Singh, Aman Verma, Anamitra Singha, Ananda Sreenidh, Ananjan Nandi, Ankit, Ankit Kumar, Atharv Dabli, Athira P, Balakumaran Ramachandran, Chahat Bansal, Chintan Sanjaybhai Sheth, Craig Dsouza, Deepak Kumar, Dharmisha Sharma, Harshita, Jaskaran Singh, Kapil Dadheech, Ksheetiz Agrahari, Om Krishna, Pooja Prasad, Priyadarshini Radhakrishnan, Ramita Sardana, Rittwick Bhabak, Ruptirumal Sai Bodavula, Saketh Vishnubhatla, Samitha Haldar, Sanjali Agrawal, Shiv Prakash Maurya, Shruti Kumari, Siddharth S, Sukriti Kumari, Vishnu S, and Aaditeshwar Seth. Initial observations from field testing of a digital participatory tool to improve water security in rural india. To appear in ICTD 2024, December 09–11, 2024, Nairobi, Kenya, 2024.

[58] Prasanth Meiyappan, Parth S Roy, Yeshu Sharma, Reshma M Ramachandran, Pawan K Joshi, Ruth S DeFries, and Atul K Jain. Dynamics and determinants of land change in india: integrating satellite data with village socioeconomics. *Regional Environmental Change*, 17:753–766, 2017.

[59] Prasanth Meiyappan, Parth S Roy, Aiman Soliman, Ting Li, Pinki Mondal, Shaowen Wang, and Atul K Jain. India village-level geospatial socio-economic data set: 1991, 2001. *Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC)*, 10:H4CN71ZJ, 2018.

[60] Surendra Kumar Mishra, Manoj Kumar Jain, RP Pandey, and VP Singh. Catchment area-based evaluation of the amc-dependent scs-cn-based rainfall–runoff models. *Hydrological Processes: An International Journal*, 19(14):2701–2718, 2005.

[61] J.L. Monteith. Evaporation and environment. *Symposia of the Society for Experimental Biology*, 19:205 – 234, 1965. Cited by: 4759.

[62] Preetilata Murmu, Mukesh Kumar, Deepak Lal, Irjesh Sonker, and Sudhir Kumar Singh. Delineation of groundwater potential zones using geospatial techniques and analytical hierarchy process in dumka district, jharkhand, india. *Groundwater for Sustainable Development*, 9:100239, 2019.

[63] Rajendran Narayanan, Sakina Dorajiwala, and Sailasri Kambhatla. Length of the last mile delays and hurdles in nrega wage payments. Technical report, Lib Tech India and Azim Premji University, 2020.

[64] Rajendran Narayanan, Laavanya Tamang, and Parul Saboo. Wage payment delays in nrega by the central government across caste and pay-ment type from april heavy wait. Technical report, Lib Tech India, 2021.

[65] Sudha Narayanan, Christian Oldiges, and Shree Saha. Does workfare work? india's employment guarantee during covid-19. *Journal of International Development*, 2021.

[66] Jean-François Pekel, Andrew Cottam, Noel Gorelick, and Alan S Belward. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633):418–422, 2016.

[67] Charles Henry Brian Priestley and Robert Joseph Taylor. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly weather review*, 100(2):81–92, 1972.

[68] Paul R. Rosenbaum and Donal B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[69] ISRO SAC. Desertification and land degradation atlas of india (based on irs awifs data of 2011-13 and 2003-05). *Ahmedabad: Space Applications Centre, ISRO, Ahmedabad, India*, 219, 2016.

[70] Diego Salazar-Martínez, Friso Holwerda, Thomas RH Holmes, Enrico A Yépez, Christopher R Hain, Susana Alvarado-Barrientos, Gregorio Ángeles-Pérez, Tulio Arredondo-Moreno, Josué Delgado-Balbuena, Bernardo Figueroa-Espinoza, et al. Evaluation of remote sensing-based evapotranspiration products at low-latitude eddy covariance sites. *Journal of Hydrology*, 610:127786, 2022.

[71] U Scs. Urban hydrology for small watersheds, technical release no. 55 (tr-55). *US Department of Agriculture, US Government Printing Office, Washington, DC*, 1986.

[72] Aaditeshwar Seth, Aarushi Gupta, Mira Johri, Aishwarya Narayan, Nishanth Kumar, Anupama Kumar, Sultan Ahmad, Arshiya Bhutani, Matiur Rahman, Lamuel Enoch, Deepak Kumar, Ashok Sharma, Amarjeet Kumar, and Lal Ranjan Pappu. Delivery of social protection entitlements in india: Unpacking exclusion, grievance redress, and the relevance of citizen-assistance mechanisms. Technical report, Dvara Research, 2021.

[73] Andrew N Sharpley. Epic-erosion/productivity impact calculator: 1, model documentation. *USDA Techn. Bull. 1759*, 235, 1990.

[74] Prafull Singh, Jay Krishna Thakur, and UC Singh. Morphometric analysis of morar river basin, madhya pradesh, india, using remote sensing and gis techniques. *Environmental earth sciences*, 68(7):1967–1977, 2013.

[75] Kaicun Wang and Robert E Dickinson. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Reviews of Geophysics*, 50(2), 2012.

[76] Sherrie Wang, François Waldner, and David B Lobell. Unlocking large-scale crop field delineation in smallholder farming systems with transfer learning and weak supervision. *Remote Sensing*, 14(22):5738, 2022.

[77] Andrew D. Weiss. Topographic position and landforms analysis. 2001.

[78] Miao Zhang, Bingfang Wu, Hongwei Zeng, Guojin He, Chong Liu, Shiqi Tao, Qi Zhang, Mohsen Nabil, Fuyou Tian, José Bofana, et al. Gci30: a global dataset of 30 m cropping intensity using multisource remote sensing imagery. *Earth System Science Data*, 13(10):4799–4817, 2021.