Data Description Colombia Temperature Precipitation Patterns (1990-2023)

All spatial data (including vector, raster, and CSV files; original and processed data) are in the EPSG Code 4326: Datum D_WGS_1984 coordinate system. For this project, given Colombia's most common type of precipitation, this project refers to precipitation and rainfall as the same phenomenon.

The folder structure of the data is described in Figure 1 as a tree structure and the specific description of each folder is the respective document section.

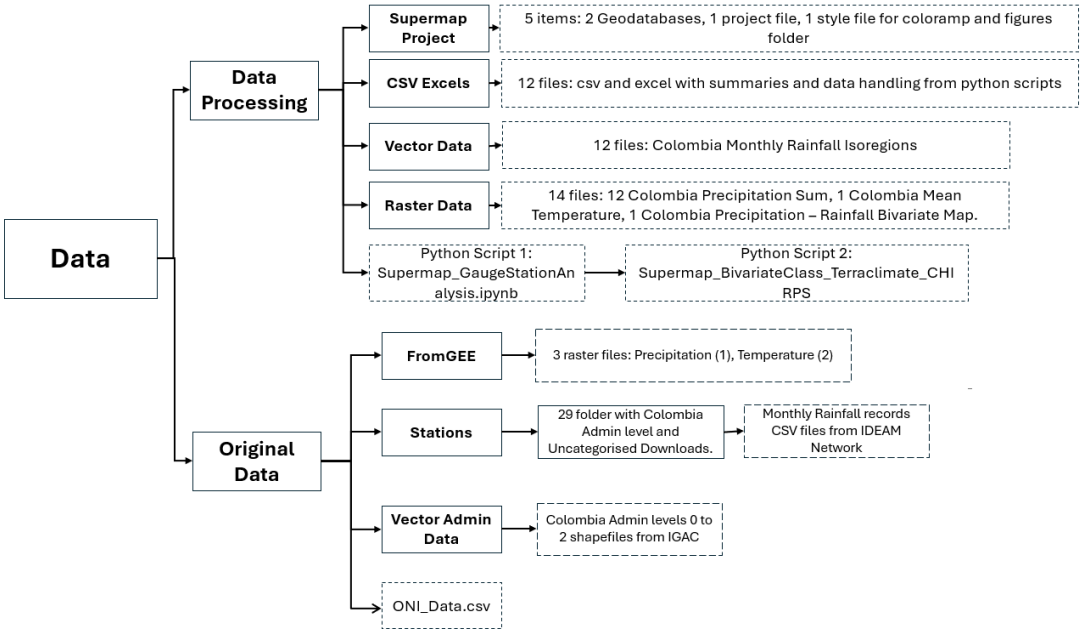

Figure 1 Tree structure for the folder data and subfolders of original and processing data.

1 ORIGINAL DATA (RAW DATA DESCRIPTION)

The original data comes from different sources, which are:

- Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia (IDEAM), the main institute of the Colombian government in charge of recording different environmental variables, such as rainfall, temperature, river discharge, water level, etc. [1]
- Instituto Geografico Agustin Codazzi (IGAC) oversees cadastral data. In this project, we used their database to download Colombia Administration Levels Shapefiles. [2]
- National Oceanic and Atmospheric Administration (NOAA) for the Oceanic Niño Index (ONI). [3]
- Terraclimate Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho accesses through Google Earth Engine (GEE) with geemap python package. See [4] (For Python scripts on data processing folder for reference). For dataset reference please visit [5], [6]
- Climate Hazards Center InfraRed Precipitation (CHIRPS) With Station Data (Version 2.0 Final) also accessed through geemap represents 5-day composite

of mean rainfall, also called *pentad*. This worked for calculating the monthly isoregions composite (isohyets). See [7] for more information about this dataset

Find in Table 1 the data sources' relationship with each folder in the original data project subfolder.

Table 1 Original Data Folder Sources relationship

Project Original data subfolder name	Data Source	Description
Stations	IDEAM	Set of CSV files with monthly rainfall data records, grouped by Colombia Admin Level and Last-minute downloads in the subfolders.
From GEE	Terraclimate, CHIRPS GEE	Raster data with 5km/pixel resolution from GEE. Mean for rainfall and Min, Max for temperature across 1990 to 2023 for Terracliomate. Multiannual Monttlhy mean of precipitation from CHIRPS data catalog.
Vector Admin Data	IGAC	Colombia admin level 0, 1 and 2 shapefiles, as boundaries references.
ONI_Data.csv	NOAA	Csv file for ONI moving window 3-months references.

2 PROCESSED DATA

There are 2 jupyter notebooks for data processing:

- Supermap_GaugeStationAnalysis.ipynb
- Supermap_BivariateClass_Terraclimate_CHIRPS.ipynb

The first script focuses on cleaning and joining CSV gauge data stations records and macroclimatic data (from NOAA), while the second is dedicated to downloading and processing raster data from the two GEE dataset.

2.1 GAUGE AND MACROCLIMATIC FILE PROCESSING

From Supermap_GaugeStationAnalysis.ipynb

The attribute table for the gauge stations contains several columns. The main ones used for processing are outlined below. In the table, you will find the description of the key attributes for the IDEAM gauge files:

- **CodigoEstacion**: Unique identifier for each gauge station. Works as the key join columns between spatial and non-spatial data.
- **Latitude/Longitude**: Geographical coordinates of the station.

- **Elevation**: Elevation of the station in meters.
- **FechaInstalacion**: Multiannual mean temperature at the station.
- **Fecha**: Date where each value of rainfall representative for each month was recorded (or analysis by IDEAM)
- **Valor**: monthly mena rainfall value.
- **Departamento:** Colombia admin 1 level whre the station is located.
- **Municipio:** Colombia admin 2 level where the station is located.

2.2 MACROCLIMATIC DATA JOIN

From Supermap_GaugeStationAnalysis.ipynb

The macroclimatic data is joined with the gauge station data based on their temporal index, creating a merged dataset that allows for analysis of regional climate variability in Colombia. From this analysis the anomalies where the phase of the El Niño-South Oscillation (ENSO) phase were joined to each gauge station record.

2.3 RASTER DATA DOWNLOAD AND PROCESSING

From Supermap_BivariateClass_Terraclimate_CHIRPS.ipynb

The second script handles the downloading and processing of raster data, particularly focusing on temperature and precipitation layers.

2.3.1 Temperature and Rainfall Mean, Bivariate Raster Generation

The raster for mean temperature was generated by stacking the minimum and maximum temperature data obtained from the TerraClimate dataset. The resulting temperature raster reflects the multiannual mean temperature, which is critical for understanding long-term climatic trends in the region.

Using the mean temperature raster along with the mean precipitation data from TerraClimate, we employed the rasterio library to generate 16 classes of combined precipitation-temperature levels. These levels are represented in a histogram (See Figure 2 and final map, middle left diagram) and provide insight into the relationship between temperature and rainfall across Colombia. The output, named *raster_data_with_coordinates_biclass*, inside the subfolder CSV Excel files, consists of a set of points classified according to a specific bivariate color ramp, ensuring clear visualization of the patterns. The corresponding bivariate color ramp used in this analysis can be found in the processes folder at: *DataProcessing/SupermapProject/Bivar_new_Style.scs*

The bi_class.csv main product file has the following attributes table:

- **User_temp**: User-provided temperature value for each data point (°C).
- **ppt**: Precipitation value associated with each data point (mm).
- **temp_class**: Classification of temperature based on predefined temperature ranges.

- **ppt_class**: Classification of precipitation based on predefined precipitation ranges.
- **bi_class**: Combined classification of temperature and precipitation.
- **color**: Hex code representing the bivariate color for the corresponding temperature-precipitation class.
- **class_int**: Integer code corresponding to the bivariate class, from 1 to 16 based on bi_class. This is the field that was classified inside Supermap
- **geometry**: Geospatial coordinates in a geometry format (e.g., point, polygon).
- **X**: X-coordinate (longitude) of the data point.
- **Y**: Y-coordinate (latitude) of the data point.

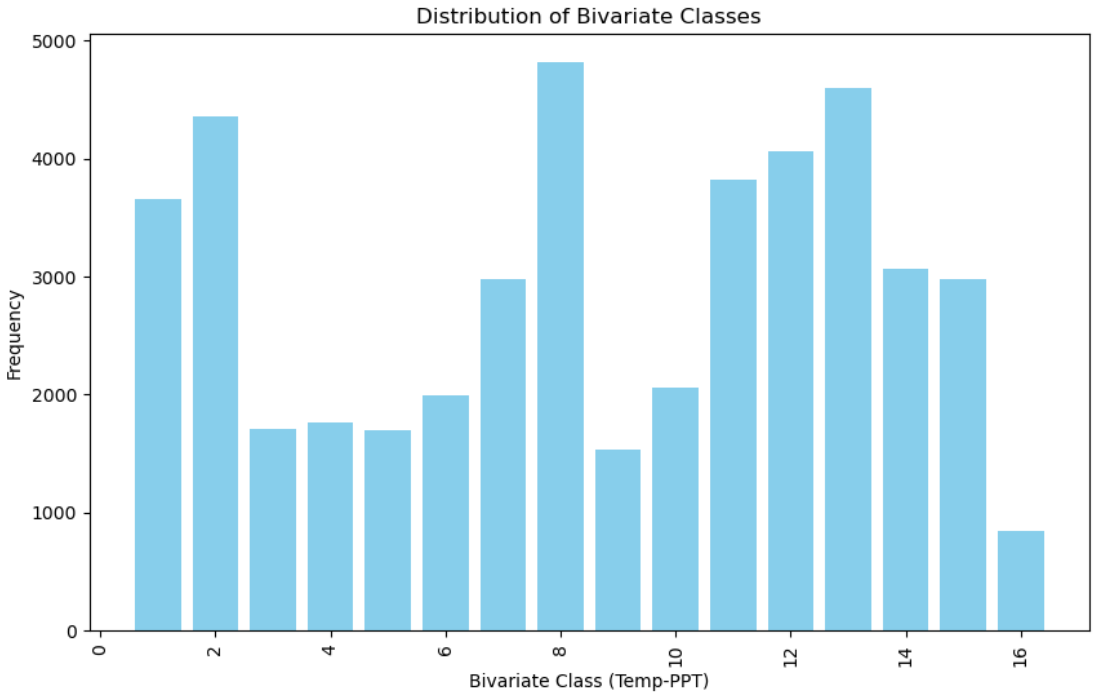

Figure 2 Bivariate classes based on raster analysis of precipitation and temperature

2.3.2 Multiannual Monthly Isoregion Analysis

We utilized CHIRPS data in Google Earth Engine (GEE) and Python's geemap library to calculate the multiannual monthly mean precipitation. The workflow involved stacking pentad data to create monthly composite images. Specifically, we filtered the CHIRPS ImageCollection for the desired years and months, grouped the pentads to represent a full month, and summed their values to generate monthly precipitation totals. Afterward, a vectorization process was performed using SuperMap, followed by the calculation of isoregions. These isoregions were then classified based on the minimum rainfall value for each month. To streamline and automate this workflow, a model builder was developed in SuperMap, which is included within the project and explained in an accompanying video. A result of the mosaic in each month can be seen in Figure 3.

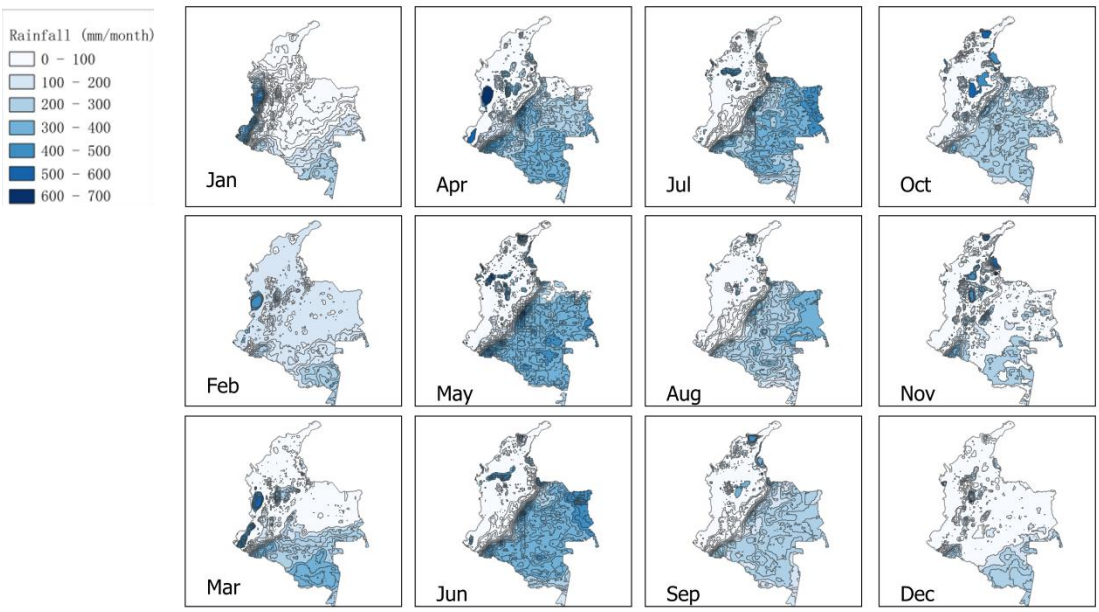

Figure 3 Multiannual monthly isohyets (rainfall isoregions) accros Colombia, generated from CHIRPS pentad data sources, GEE and Supermap. From April to Jun are the peak season in the southeast region of Colombia, Amazon and Orinoquia Basins.

3 FINAL PRODUCTS SUMMARY

In Table 2 you can find the description of each dataset file that were obtained from the different processes mentioned above in this section.

Table 2 Summary table of the final products generated (Data processing subfolder)

Subfolder	Name(s)	Description
SupermapProject	Bivar_new_Style.scs	Bivariate Color Ramp Style
	ColombiaRainfall_Temperature.smwu	Supermap project
	Isoregion_Monthly.udb	Database with just Isoregion, result of Model Builder
	MainDataSoruces.udbx	Rest of proccesses products (see Other raws of this table)
Raster Data	Biclass_Colombia	Final Main raster product
	Colombia_Temperature_MEAN_1990_2023_Scale01	Mean temperature, resutl from the stack of min and max raw raster files
	Pentad_Sum_MONTH_mean	12 raster files, represent the mutiannual montly mean rainfall from the CHIRpS dataset

Vector Data	Pentad_Sum_MONTH_mean	Same as above but in shapefile format
CSV EXCELS	anomalies_df	anomalies, based on ENSO phase, taken from the mean monthly value for each station
	anomalies_df_elnino	same above, but filtered for have just El Niño's phase records
	anomalies_df_lanina	same above, but filtered for have just La Niña's phase records
	descriptive_stats_by_estacion	Percentiles of each gauge station (rainfall)
	descriptive_stats_by_monthestacion	Monthly Percentiles of each gauge station (rainfall)
	descriptive_stats_by_yearestacion_	Yearly Percentiile of each gauge station (rainfall)
	final_combined_df	all data records concatenated
	just_coordinates_df	Each Gauge stations with just Longitud, Latiude, Name and Code
	missing_records_df	Each data station missing percentage records
	missing_Recordsbydpto	Same as above but grouped by admin level 1
	ONI_Data_Cleaned_longformat	ONI data from NOAA in a long format (original one on wide format). This is for easy join with gauge stations records
	raster_data_with_coordinates_biclass	Final Raster with each bivariate class, represented as point cloud

4 REFERENCES

[1] Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), «Plataforma para la descarga de datos hidrometeorológicos». Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), Colombia, 2024. [En línea]. Disponible en: http://dhime.ideam.gov.co/atencionciudadano/

[2] Instituto Geográfico Agustin Codazzi (IGAC), «Geoportal: Datos abiertos subdireccion de cartografia y Geografia». 2024. [En línea]. Disponible en: https://geoportal.igac.gov.co/contenido/datos-abiertos-cartografia-y-geografia

[3] National Oceanic and Atmospheric Administration (NOAA), «Cold & Warm Episodes by Season of the Oceanic Niño Index dataset». 2023. [En línea]. Disponible en: https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

[4] J. T. Abatzoglou, S. Z. Dobrowski, S. A. Parks, y K. C. Hegewisch, «TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015», *Sci. Data*, vol. 5, n.º 1, p. 170191, ene. 2018, doi: 10.1038/sdata.2017.191.

[5] Q. Wu, «geemap: A Python package for interactive mapping with Google Earth Engine», *J. Open Source Softw.*, vol. 5, n.º 51, p. 2305, jul. 2020, doi: 10.21105/joss.02305.

[6] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, y R. Moore, «Google Earth Engine: Planetary-scale geospatial analysis for everyone», *Remote Sens. Environ.*, vol. 202, pp. 18-27, dic. 2017, doi: 10.1016/j.rse.2017.06.031.

[7] C. Funk *et al.*, «The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes», *Sci. Data*, vol. 2, n.º 1, p. 150066, dic. 2015, doi: 10.1038/sdata.2015.66.