3rd Webinar on Climate Risk
5th Dcember 2025

Data You Can Actually Use

Turning Climate Datasets into Actuarial Insights

Piyush Jain
Business Actuary, Kshema General Insurance Limited

Institute of Actuaries of India

The Actuarial Challenge

Institute of Actuaries of India

The Data Void

Actuaries require consistent, long-term, and granular data to price risk effectively. However, Traditional station data is often:

- **Inconsistent:** Missing records.
- **Sparse:** Low density in rural areas.
- **Short-term:** Insufficient tail history.

The Science Barrier

Climate scientists produce terabytes of data, but it is often::

- **Too Coarse:** 100km grids vs. Farm risk.
- **Biased:** Differs from ground truth.
- **Complex:** Hard-to-process formats.

Why Use Public Datasets?

Institute of Actuaries of India

Reliability

Backed by IMD, ECMWF, NASA. Rigorous QC processes.

Granularity

High spatial (10m-30km) and temporal (hourly) resolution.

Cost-Efficiency

Open-access reduces costs, allowing budget focus on modeling.

Session Agenda

- **1. Data Landscape:** Weather vs. Climate vs. Remote Sensing
- **2. Weather Data Deep Dive:** IMD, ERA5, CHIRPS, CMIP6.
- **3. Access Workflows:** Python, GEE & APIs.
- **4. Methodologies:** Downscaling & Bias Correction.
- **5. Applications:** Pricing & Modeling.

Weather Datasets

From Ground Observations to Reanalysis

Understanding Data Types

Institute of Actuaries of India

Observation

Measured directly by instruments (Rain Gauges, Thermometers). **The "Ground Truth".** Limited by spatial coverage.

Reanalysis

Weather models run historically, assimilating observations to fill gaps. **Best for consistency** and gap-free history.

Satellite

Derived from remote sensors or physics simulations. **Crucial for future risk** and remote areas.

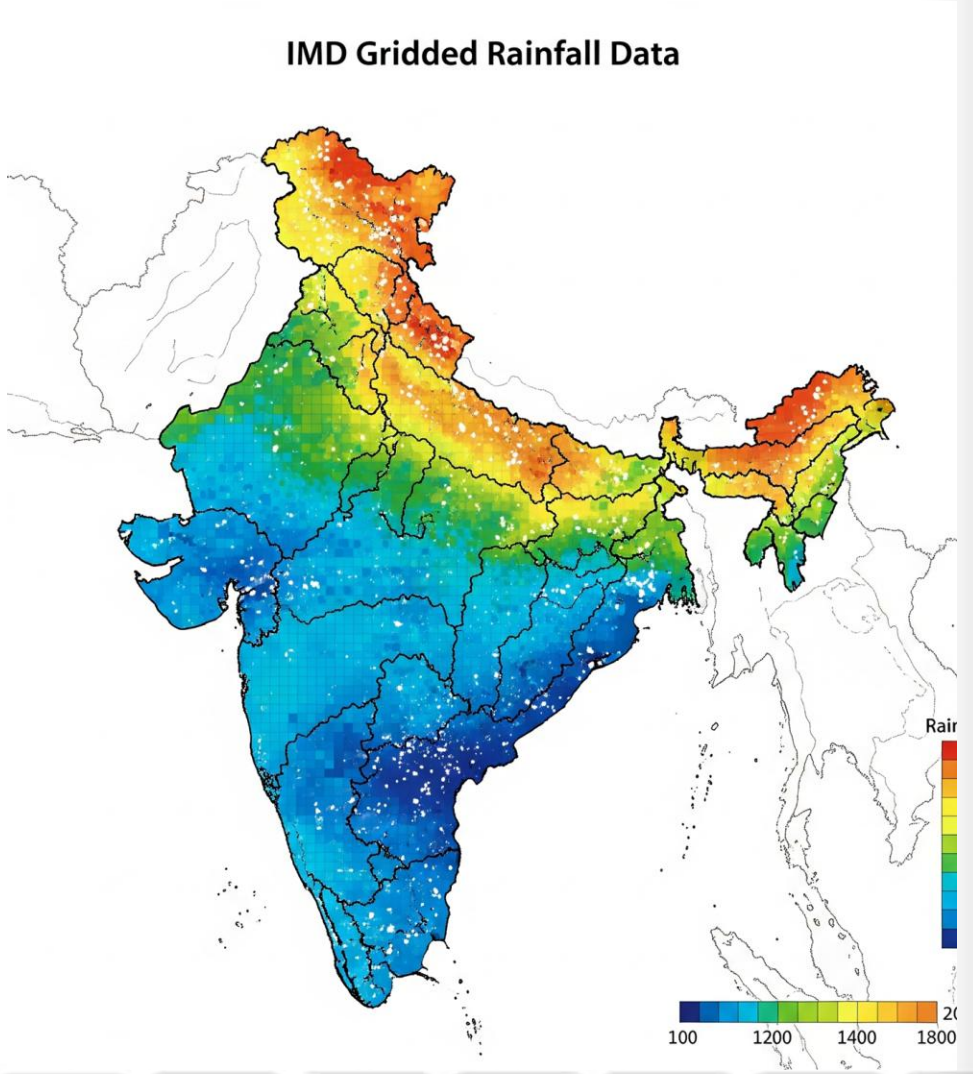
IMD Gridded Data

Institute of Actuaries of India

The Gold Standard

Derived from thousands of rain gauges interpolated across India. Represents "Ground Truth".

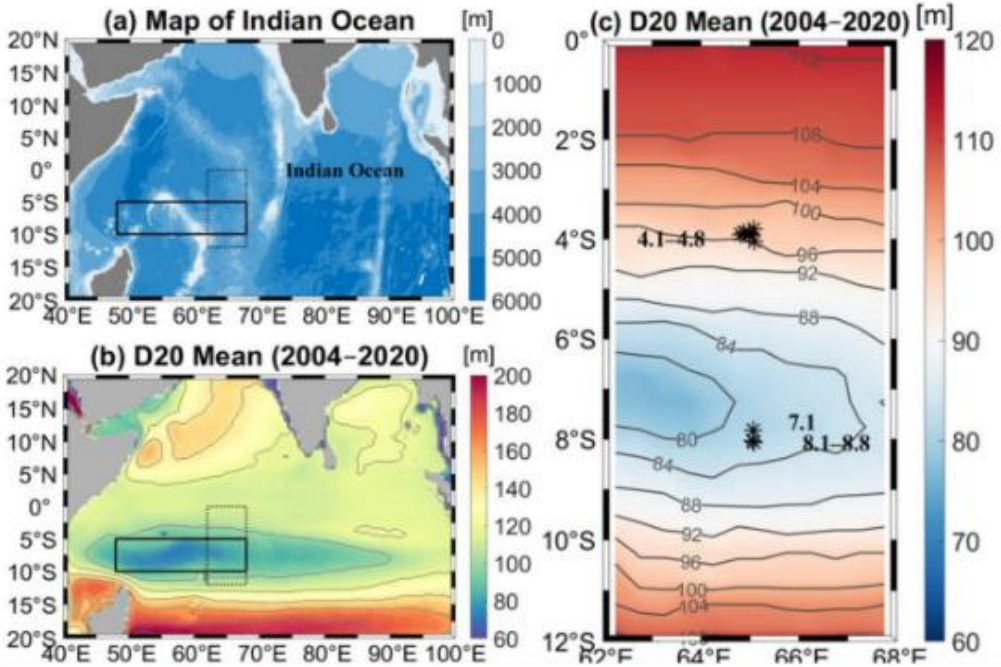
- **Type:** Observation
- **Res:** 0.25° (Rain), 1° (Temp)
- **Freq:** Daily
- **Period:** 1901 – Present



IMD: Available Parameters

Parameter	Variable Name	Unit	Resolution
Precipitation	Daily Rainfall Accumulation	mm	0.25° (~27km)
Temperature	Daily Max Temp (Tmax)	°C	1.0° (~110km)
	Daily Min Temp (Tmin)	°C	1.0° (~110km)
	Daily Mean Temp (Tmean)	°C	1.0° (~110km)

ERA5 Reanalysis (ECMWF)



The Gap Filler

A physical model assimilating global observations.

Provides complete atmospheric data.

- **Type:** Reanalysis
- **Res:** ~31km
- **Freq:** Hourly
- **Period:** 1940 – Present

ERA5: Surface Parameters

- **Temperature:** 2m Air Temp, 2m Dewpoint.
- **Pressure:** Surface Pressure, MSLP.
- **Wind:** 10m U/V Components, Gusts.
- **Precipitation:** Total Precip, Snowfall.
- **Clouds:** Total Cloud Cover, Base Height.
- **Radiation:** Net Solar/Thermal Radiation.
- **Soil:** Soil Temp & Moisture (4 levels).
- **Marine:** Wave Height, Period.

ERA5: Upper Air Parameters

37 Pressure Levels (1000 hPa to 1 hPa). Critical for Aviation & Cyclone Modeling.

Variable	Description	Application
Geopotential Height	Height of pressure surfaces	Cyclone Steering
Temperature	Air temp at altitude	Lapse Rates, Stability
Specific Humidity	Moisture content	Storm Energy
U / V Wind	Zonal / Meridional Wind	Wind Shear, Jet Streams

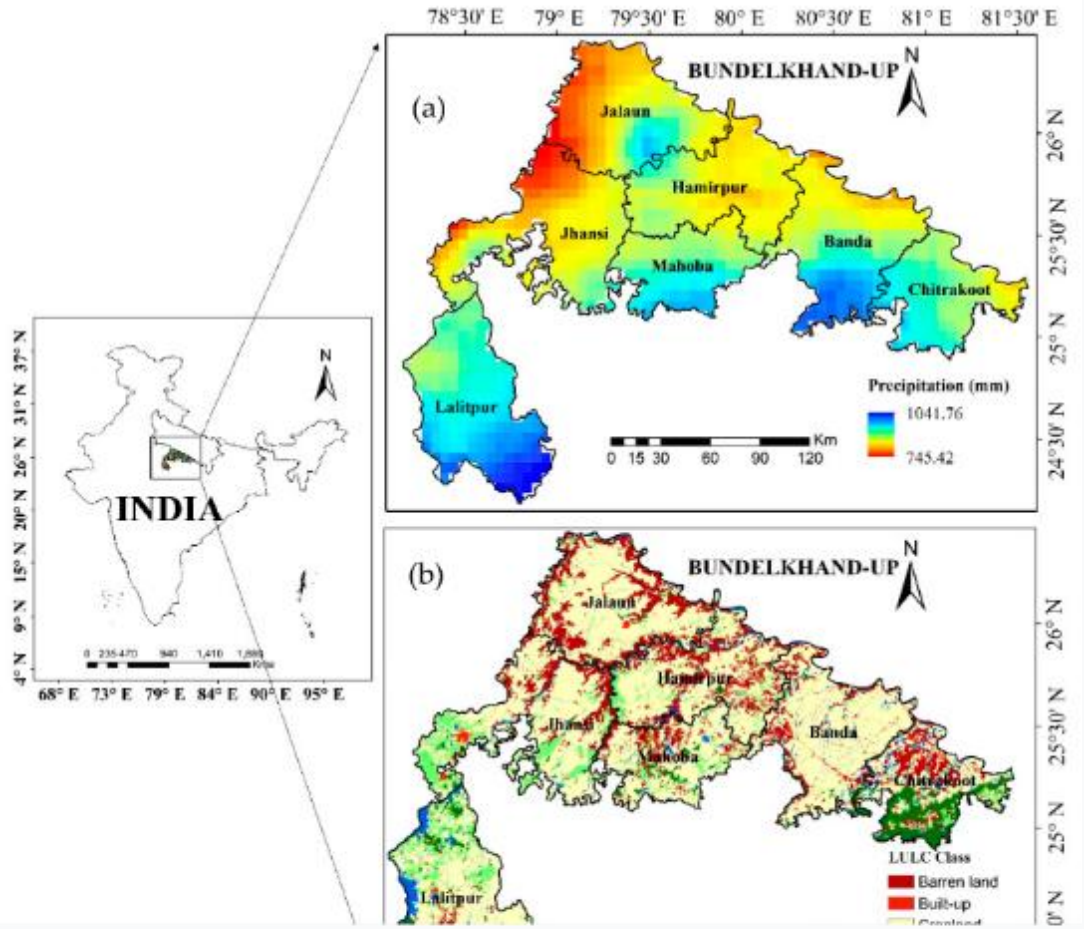
CHIRPS (UCSB/CHC)

Institute of Actuaries of India

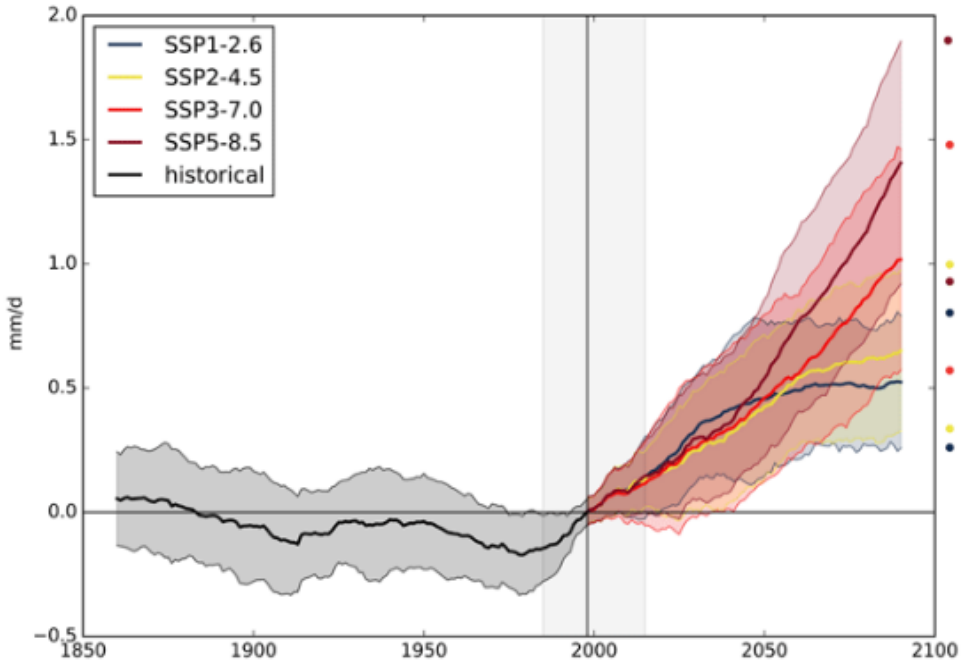
The Agri Specialist

Designed specifically for drought monitoring in data-sparse regions.

- **Type:** Sat + Station Blend
- **Res:** 0.05° (~5.5km)
- **Freq:** Daily
- **Period:** 1981 – Present



CMIP6 Climate Models



The Future Gazer

Essential for understanding long-term risk under different Shared Socioeconomic Pathways (SSPs). Used for climate change stress testing.

- **Type:** GCM Projections
- **Res:** ~100km
- **Period:** 2015 - 2100
- **Use Case:** Stress Testing

CMIP6: Standard Parameters

Code	Standard Name	Description
tas	Near-Surface Air Temp	Mean temp at 2m.
pr	Precipitation Flux	Rainfall + Snowfall rate.
psl	Sea Level Pressure	Pressure at sea level.
sfcWind	Near-Surface Wind	Wind speed at 10m.
rsds	Downwelling Shortwave	Solar radiation.

Secondary Datasets

Institute of Actuaries of India

GPM (IMERG)

Type: Satellite (NASA)

Use Case: Flash Flood Events

Provides 30-minute rainfall data, critical for analyzing recent extreme events like urban flooding where daily data misses the peak intensity.

APHRODITE

Type: Gridded Observation

Use Case: Historical Validation

Excellent for cross-verifying IMD data, especially in Himalayan regions, though the record ends around 2015.

Weather Dataset Comparison

Institute of Actuaries of India

Dataset	Type	Resolution	Period	Best For
IMD Gridded	Observation	0.25° / Daily	1901-Present	Ground Truth, Calibration
ERA5	Reanalysis	31km / Hourly	1940-Present	Cat Modeling, Wind, Cyclones
CHIRPS	Sat + Station	5km / Pentad	1981-Present	Agri-Insurance, Droughts
CMIP6	Model	~100km / Daily	2015-2100	Future Climate Scenarios
GPM	Satellite	10km / 30min	2000-Present	Real-time Flood Events

Key Parameters for Actuaries

- **Precipitation:** Total accumulation, Max daily intensity (Agri/Flood).
- **Temperature:** Min/Max daily, Heatwave duration (Agri/Health).
- **Wind Speed:** 10m gusts, Max sustained wind (Cyclone/Property).
- **Soil Moisture:** Volumetric water content (Drought/Agri).
- **Solar Radiation:** Surface net solar radiation (Renewables).
- **Relative Humidity:** Dew point, Heat index (Health/Livestock).

Accessing the Data

Python & API Workflows

Accessing IMD (Python)

Use **xarray** for NetCDF (.nc) files.

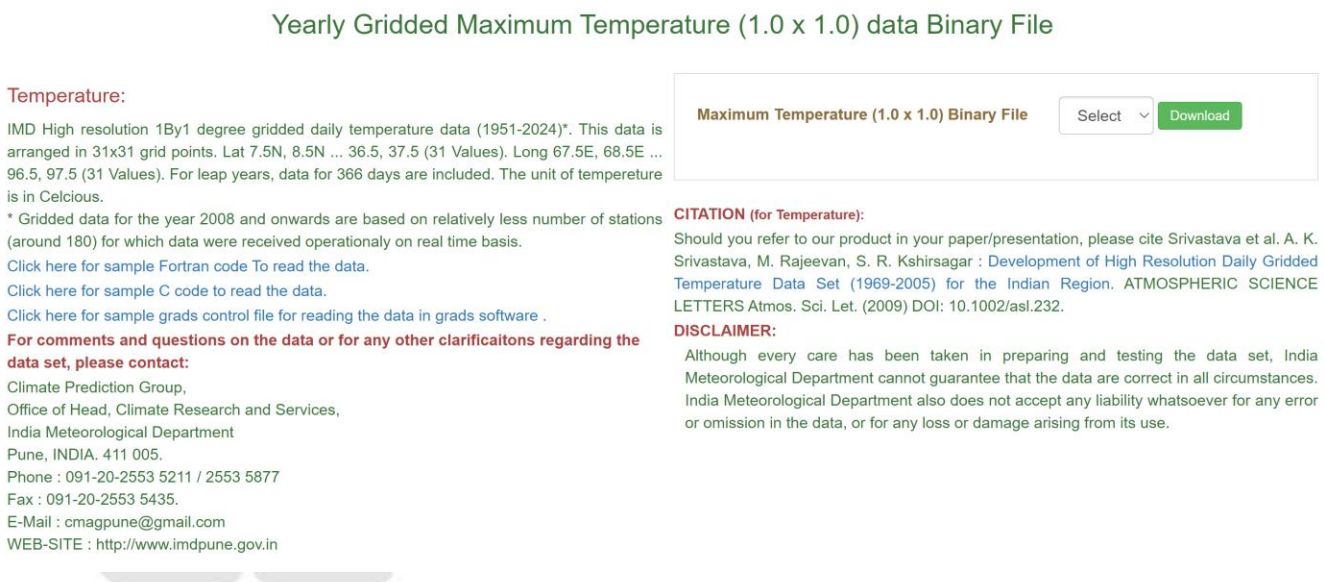
```
import xarray as xr # Load NetCDF file ds = xr.open_dataset('IMD_Rain_2023.nc' )# Extract time series for Mumbai
          # method='nearest' finds closest grid point   data = ds.sel(lat=19.07, lon=72.87, method='nearest' )
# Convert to DataFrame df = data.to_dataframe() print(df.head())
```

Sample code for illustration only.

https://imdpune.gov.in/cmpg/Griddata/Rainfall_25_NetCDF.html

https://imdpune.gov.in/cmpg/Griddata/Min_1_Bin.html

https://imdpune.gov.in/cmpg/Griddata/Max_1_Bin.html



Yearly Gridded Maximum Temperature (1.0 x 1.0) data Binary File

Temperature:

IMD High resolution 1By1 degree gridded daily temperature data (1951-2024)*. This data is arranged in 31x31 grid points. Lat 7.5N, 8.5N ... 36.5, 37.5 (31 Values). Long 67.5E, 68.5E ... 96.5, 97.5 (31 Values). For leap years, data for 366 days are included. The unit of temperetureis in Celcious.

* Gridded data for the year 2008 and onwards are based on relatively less number of stations (around 180) for which data were received operationaly on real time basis.

Click here for sample Fortran code To read the data.

Click here for sample C code to read the data.

Click here for sample grads control file for reading the data in grads software .

For comments and questions on the data or for any other clarificaitons regarding the data set, please contact:

Climate Prediction Group,
Office of Head, Climate Research and Services,
India Meteorological Department
Pune, INDIA. 411 005.
Phone : 091-20-2553 5211 / 2553 5877
Fax : 091-20-2553 5435.
E-Mail : cmagpune@gmail.com
WEB-SITE : http://www.imdpune.gov.in

Maximum Temperature (1.0 x 1.0) Binary File Select Download

CITATION (for Temperature):

Should you refer to our product in your paper/presentation, please cite Srivastava et al. A. K. Srivastava, M. Rajeevan, S. R. Kshirsagar : Development of High Resolution Daily Gridded Temperature Data Set (1969-2005) for the Indian Region. ATMOSPHERIC SCIENCE LETTERS Atmos. Sci. Let. (2009) DOI: 10.1002/asl.232.

DISCLAIMER:

Although every care has been taken in preparing and testing the data set, India Meteorological Department cannot guarantee that the data are correct in all circumstances. India Meteorological Department also does not accept any liability whatsoever for any error or omission in the data, or for any loss or damage arising from its use.

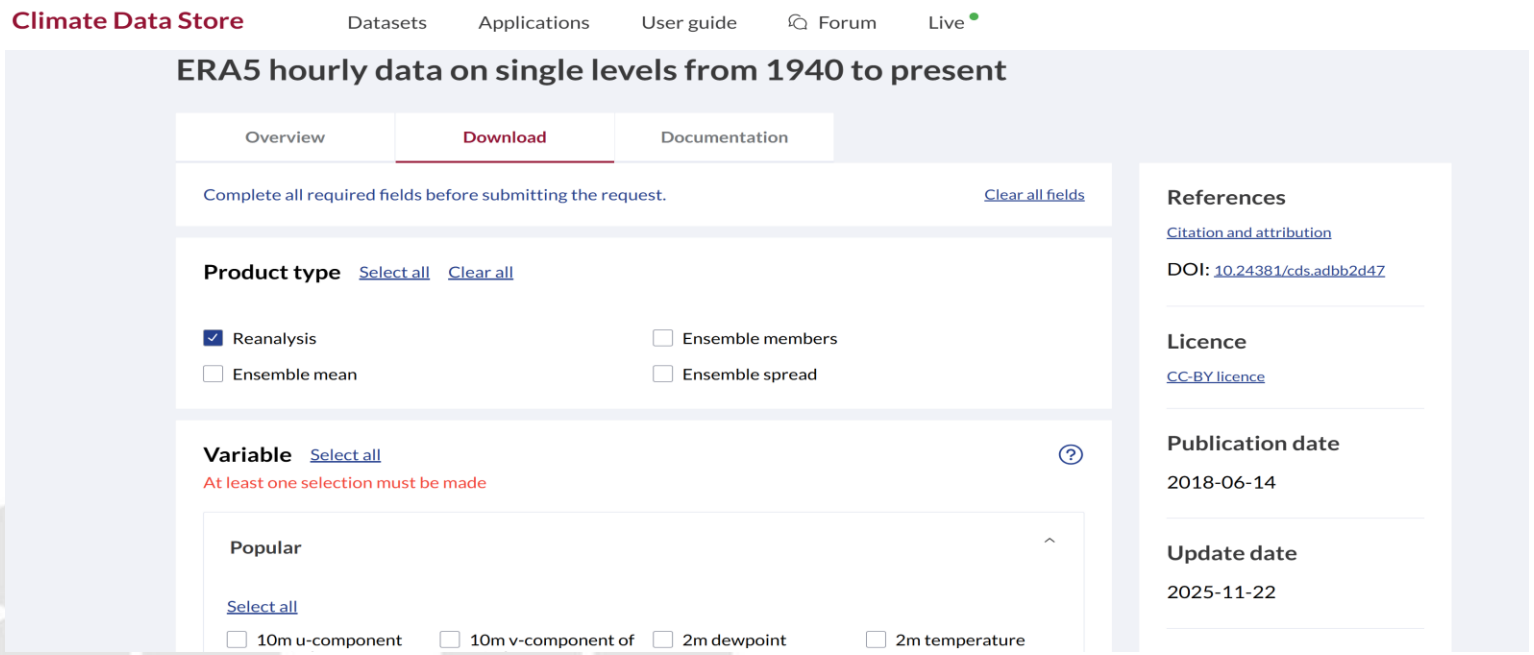
Accessing ERA5 (API)

Use the **CDS API** to request specific subsets.

```
import cdsapi c = cdsapi.Client() c.retrieve( 'reanalysis-era5-single-levels' , { 'variable' :
['total_precipitation' , '10m_u_component_of_wind' ], 'year': '2023', 'month': '07', 'day': '01', 'time': '12:00',
'format': 'netcdf', }, 'era5_sample.nc' )
```

Sample code for illustration only.

https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download



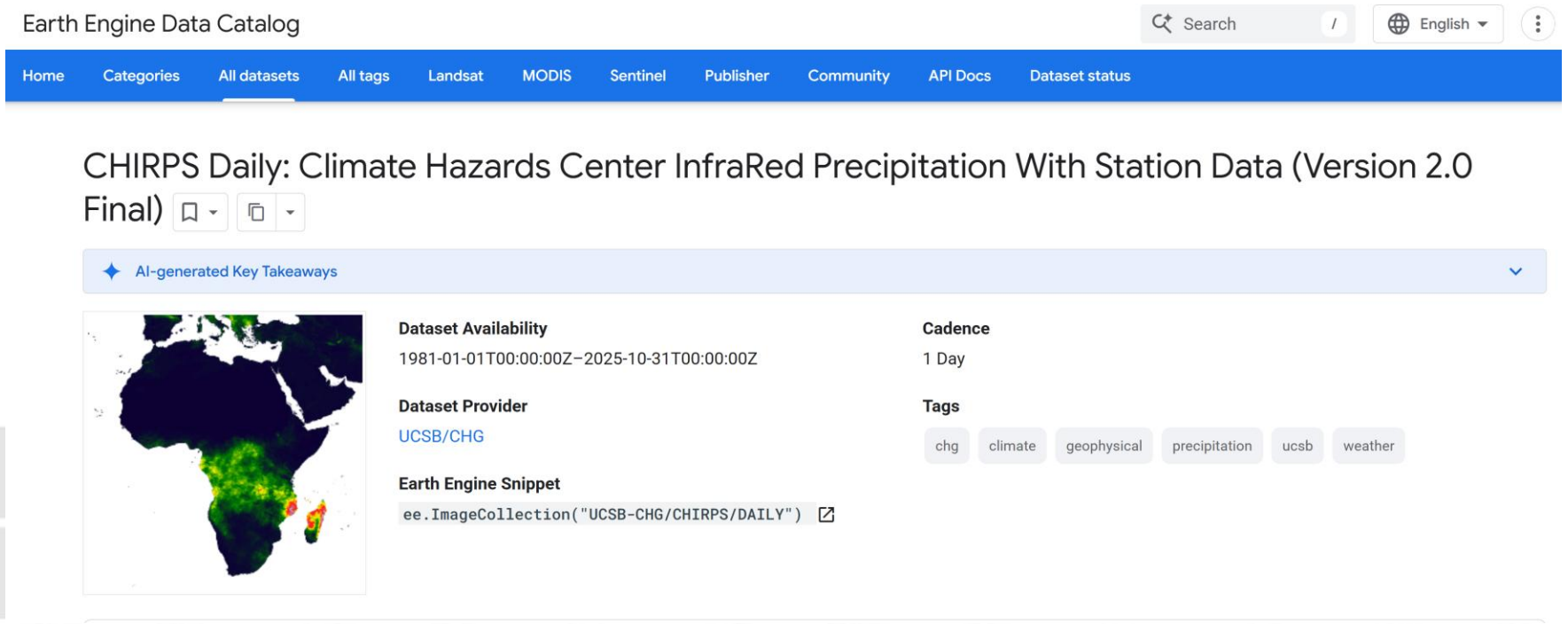
Accessing CHIRPS (GEE)

Google Earth Engine processes heavy GeoTIFFs in the cloud.

```
var chirps = ee.ImageCollection("UCSB-CHG/CHIRPS/DAILY" ); // Filter Date  var rain = chirps.filterDate('2023-06-01'
    , '2023-09-30' ); // Calculate Mean over District  var meanRain = rain.mean().reduceRegion({ reducer:
ee.Reducer.mean(), geometry: district_geom, scale: 5500 });
```

Sample code for illustration only.

https://developers.google.com/earth-engine/datasets/catalog/UCSB-CHG_CHIRPS_DAILY#colab-python



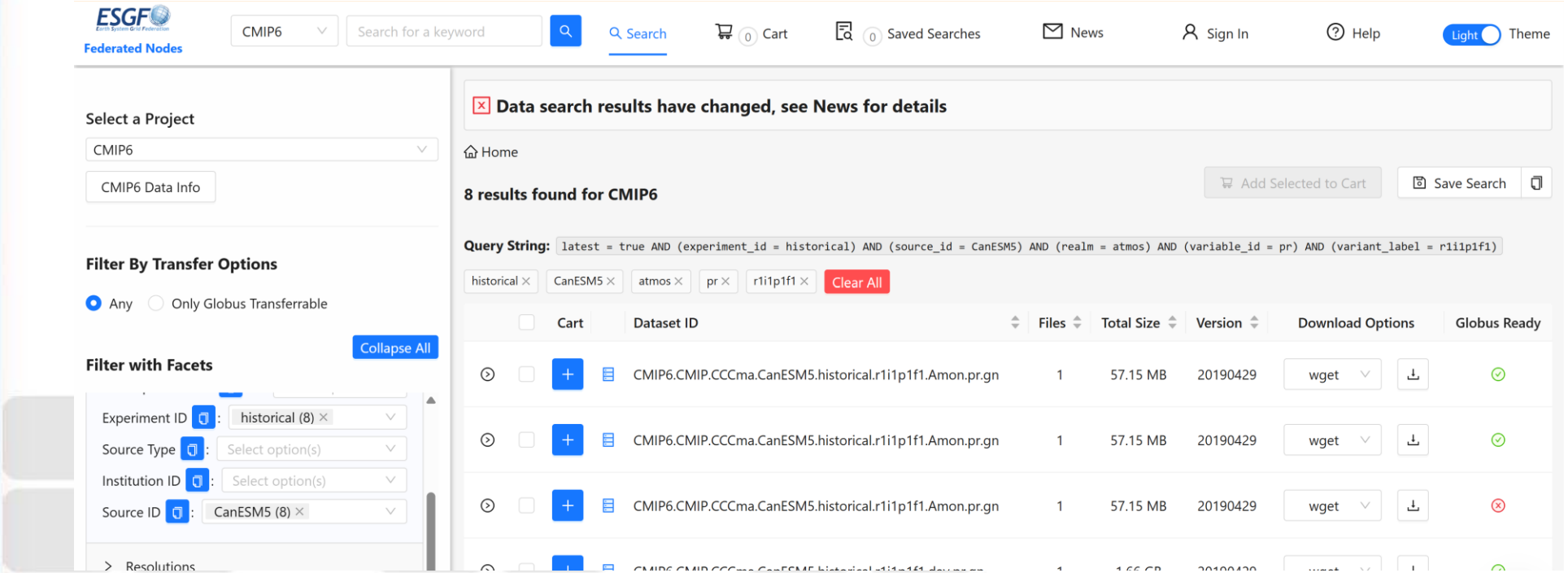
Accessing CMIP6

Handle non-standard calendars using **cftime**.

```
import xarray as xr import cftime # Load Dataset ds = xr.open_dataset('pr_day_MRI-ESM2-0.nc' )# Decode time
 automaticallyds = xr.decode_cf(ds) # Extract future rain future_rain = ds['pr'].sel(lat=28.6, lon=77.2,
method='nearest')
```

Sample code for illustration only.

https://aims2.llnl.gov/search/cmip6/



Remote Sensing

Exposure, Damage, and Index Design

Institute of Actuaries of India

Optical vs. Radar

Institute of Actuaries of India

Optical (Sentinel-2)

Tech: Reflected Sunlight.

Use: Vegetation Health (NDVI).

Con: Blocked by clouds.

Radar (Sentinel-1)

Tech: Active Microwave.

Use: Flood Mapping.

Pro: Sees through clouds/rain.

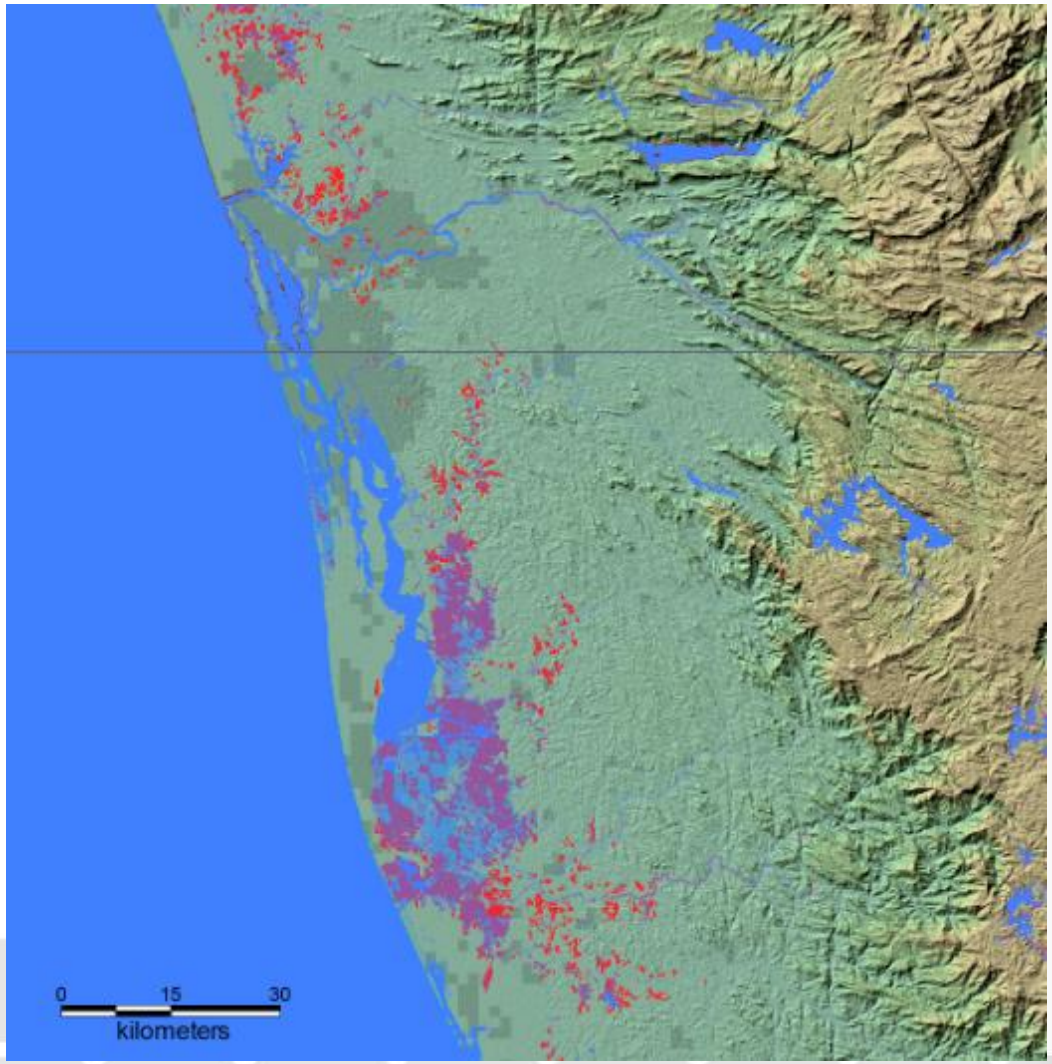
Sentinel-1 (SAR)

Institute of Actuaries of India

The Flood Watcher

Essential for monsoon flood mapping when optical satellites are blind.

- **Res:** 10m
- **Revisit:** 6-12 Days
- **Params:** Backscatter (dB), Polarization (VV/VH)



Workflow: Flood Extent Mapping

Institute of Actuaries of India

1. Acquire

Get Sentinel-1 GRD image during the flood event.

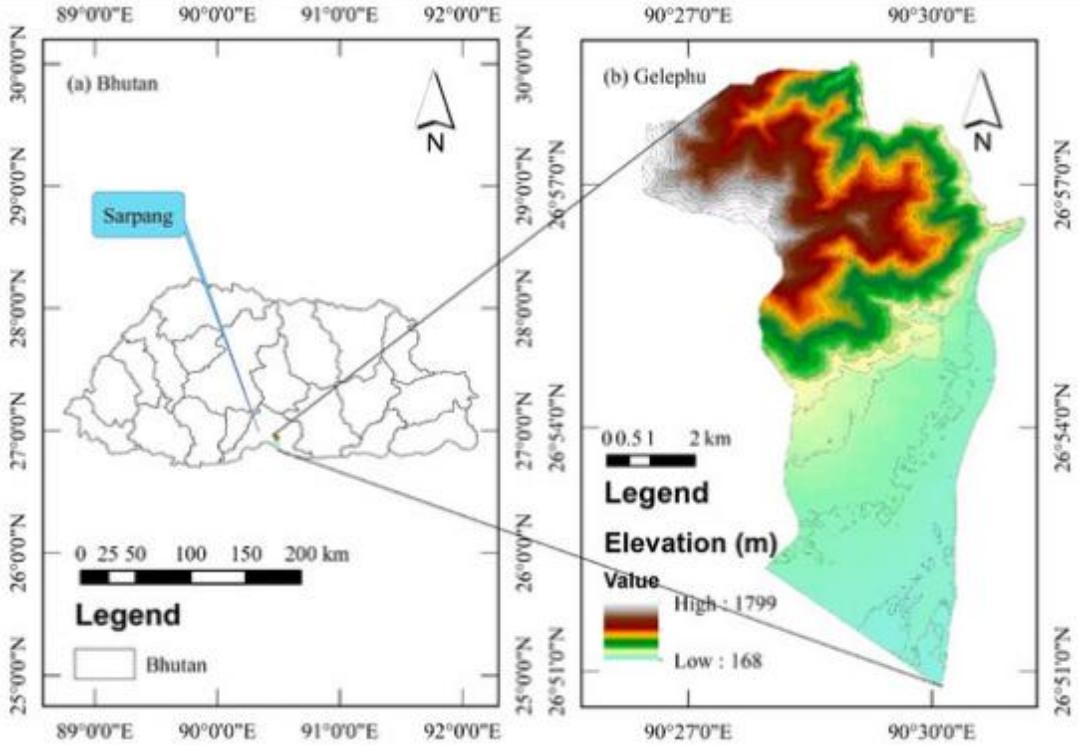
2. Threshold

Water reflects radar away (dark). Threshold backscatter < -15dB.

3. Overlay

Combine flood mask with asset locations to check exposure.

Sentinel-2 (Optical)



The Crop Monitor

High-res multispectral imaging for agriculture.

Band	Use
Red (B4)	Chlorophyll absorption
NIR (B8)	Biomass reflection
SWIR	Moisture content

Calculating NDVI (GEE)

Normalized Difference Vegetation Index = (NIR - Red) / (NIR + Red)

```
import ee ee.Initialize() # Load & Filter Sentinel-2 s2 = ee.ImageCollection("COPERNICUS/S2_SR") \
.filterBounds(point).first() # Calculate NDVI ndvi = s2.normalizedDifference(['B8', 'B4'])
print(ndvi.reduceRegion(ee.Reducer.mean(), point, 10))
```

Topography & Exposure

Dataset	Usage
NASADEM / SRTM	**Flood Accumulation:** Water flows downhill. **Slope:** Landslide risk.
ESA WorldCover	**Exposure Mapping:** Distinguish 'Built-up' (Urban) from 'Cropland' (Rural).

Satellite Sources Comparison

Satellite	Type	Resolution	Revisit	Key Use Case
Sentinel-2	Optical	10m	~5 Days	Crop Health, NDVI
Sentinel-1	Radar (SAR)	10m	6-12 Days	Flood Mapping (Cloud-free)
MODIS	Optical	250m	Daily	Regional Drought, LST
Landsat	Optical	30m	16 Days	Historical Baseline (>40 yrs)

Data Formats Explained

Institute of Actuaries of India

NetCDF (.nc)

Multi-dimensional arrays (Lat, Lon, Time). The standard for climate data. Handle with **xarray**.

GeoTIFF (.tif)

Raster images with georeferencing. Standard for satellite/CHIRPS. Handle with **rasterio** or **rioxarray**.

GRIB (.grib)

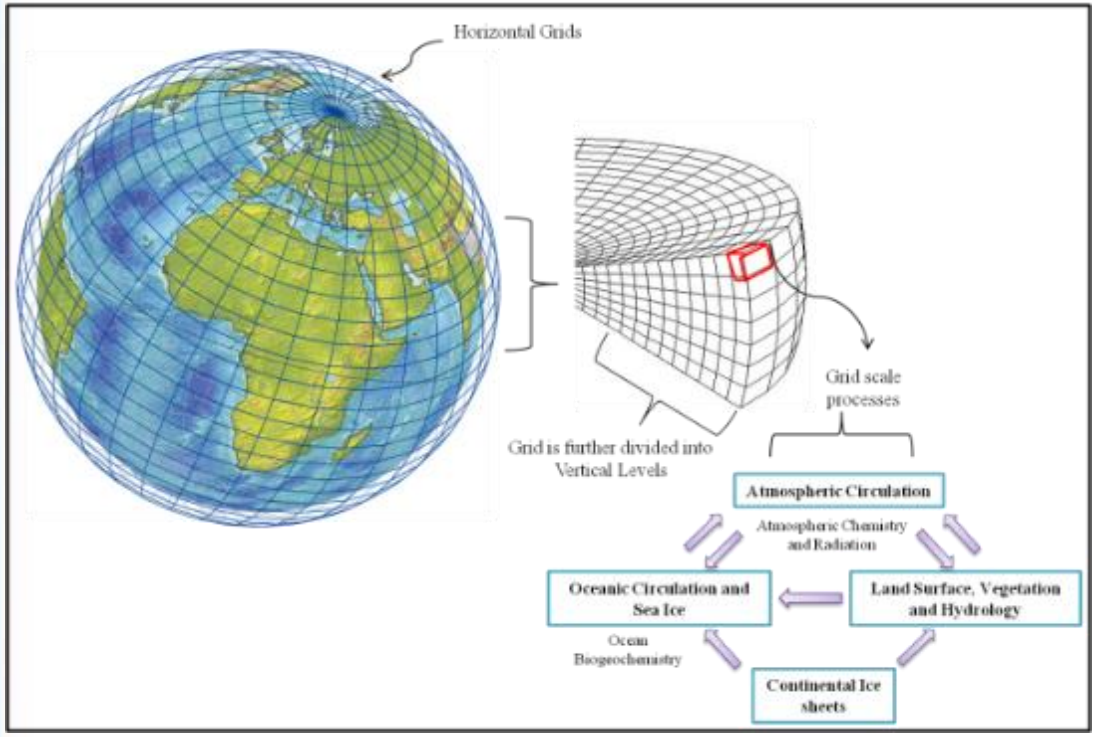
Binary format used by meteorologists (ERA5 raw). Compact but hard to read. Convert to NetCDF first.

Data Processing

Statistical Downscaling & Bias Correction

Institute of Actuaries of India

The Scale Problem



Global vs. Local

Global: Models average weather over 100km blocks.

Local: A single block contains mountains, valleys, and cities.

Need: Downscaling bridges this gap to estimate farm-level risk.

Statistical Downscaling

"Predicting local-scale weather (predictands) from large-scale atmospheric states (predictors) using historical statistical relationships."

Computationally efficient for point-based pricing compared to dynamical models.

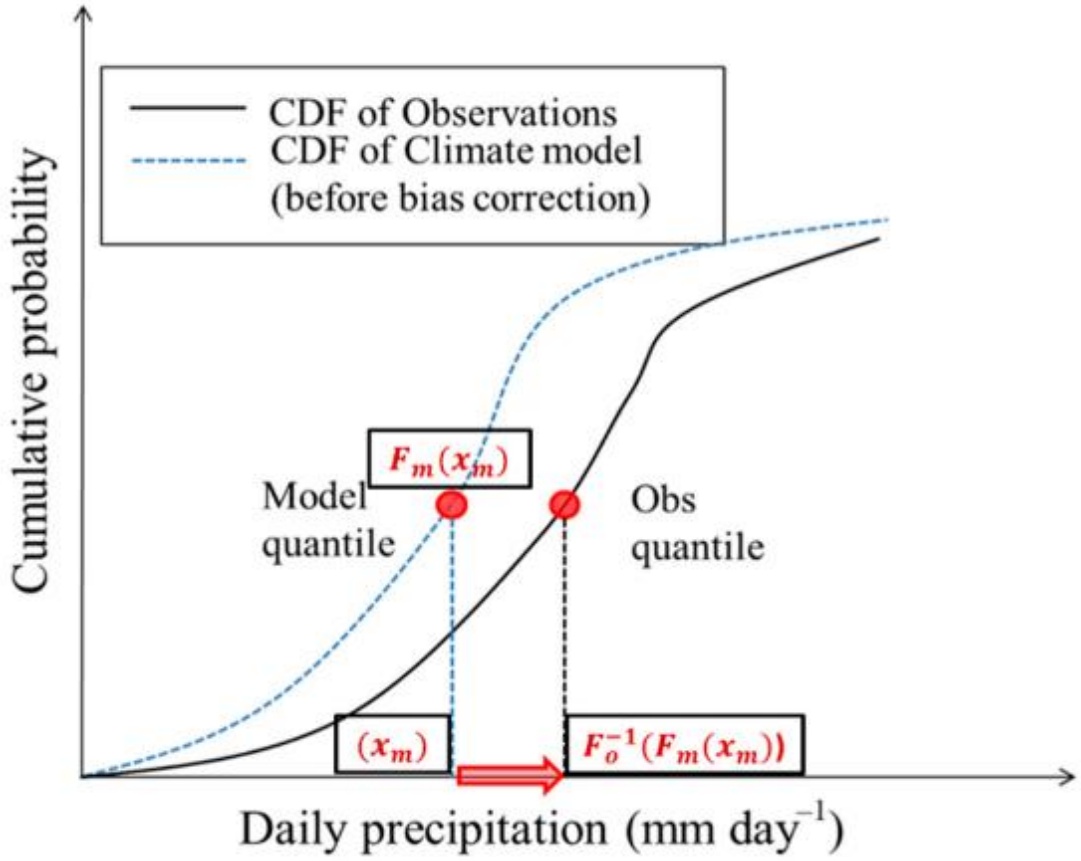
Bias Correction

Institute of Actuaries of India

Why Correct?

Models are imperfect simulations. Common biases:

- **Warm Bias:** Consistently too hot.
- **Drizzle Bias:** Too many light rain days.
- **Variance Bias:** Underestimating extremes.



Real-World Use Cases

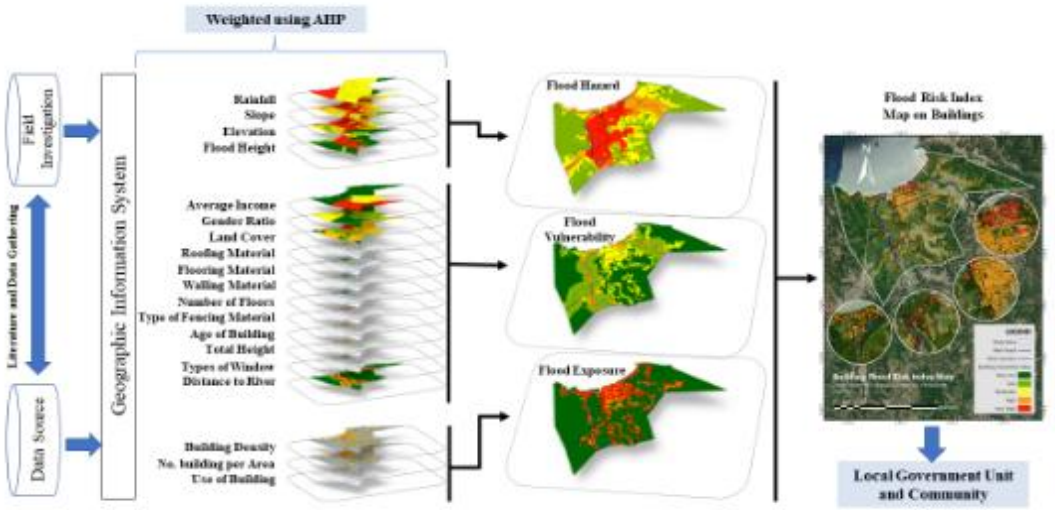
Case 1: Insurance Product Design

- **Issue:** No yield data = Basis Risk.
- **Proxy:** CHIRPS Rainfall.
- **Structure:**
  - History: 30yr CHIRPS.
  - Trigger: Rain < 70% Avg.
  - Payout: Immediate.

Benefit

Transparent settlement, zero claims handling cost, rapid farmer support.

Case 2: Underwriting: Flood Risk Scoring



Portfolio Analysis

Input: Sentinel-1 (Flood Mask) + NASADEM.

Method: Overlay historic flood footprints on insured assets.

Outcome: Granular risk score per property (Safe vs. Inundated).

Case 3: Catastrophe Model: Cyclone Tail Risk

The Data

ERA5: Max Wind Gusts (1940-2023).

Provides 80+ years of history to capture rare tail events.

The Math

GEV Distribution: Fit Generalized Extreme Value to annual maxima.

Result: Calculate 1-in-100 year wind speed for Reinsurance pricing.