Automated Detection of Atmospheric Rivers over the Indian Region Using Reanalysis Data

INTERNSHIP REPORT

UNDER THE SUPERVISION OF
Mrs. Shivali Verma
Scientist/Engineer - 'SE',
NRSC, ISRO Hyderabad

SUBMITTED BY
Advaitha Patti
23CSB0F29

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE AND ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY WARANGAL
(Government)
Hanamkonda, Telangana 506004

2025-2026

National Remote Sensing Centre (NRSC)

National Remote Sensing Centre (NRSC) at Hyderabad is responsible for remote sensing satellite data acquisition and processing, data dissemination, aerial remote sensing and decision support for disaster management. NRSC has a data reception station at Shadnagar near Hyderabad for acquiring data from Indian remote sensing satellites as well as others.

NRSC operates through multiple campuses to meet national and regional remote sensing data and applications needs of the country.

- Main Campus at Balanagar, Hyderabad for Administration, Remote Sensing Applications and Aerial Services
- The Campus at Shadnagar for Satellite Data Reception, Data Processing and Dissemination, Earth and Climate Studies and Disaster Management Support
- Five Regional Centres at Sector 9, KBHB in Jodhpur (Regional Centre-West), Sadhiknagar at New Delhi (Regional Centre-North), New Salt Lake City in Kolkata (Regional Centre-East), Amaravathy Road in Nagpur (Regional Centre-Central), Karthik Nagar in Bangalore (Regional Centre-South) for promoting remote sensing applications for various states.
- Outreach facility at Jeedimetla in Hyderabad for providing training for professionals, faculty, and students, and for general outreach.
- Aircraft operations facility at Begumpet Airport, Hyderabad

NRSC Ground station at Shadnagar acquires Earth Observation data from Indian remote-sensing satellites as well as from different foreign satellites. NRSC is also engaged in executing remote sensing application projects in collaboration with the users. The Aerial Services and Digital Mapping (ASDM) Area provides end-to-end Aerial Remote Sensing services and value-added solutions for various large scale applications like aerial photography and digital mapping, infrastructure planning, scanner surveys, aeromagnetic surveys, large scale base map, topographic and cadastral level mapping, etc.

Table of contents

ACKNOWLEDGEMENT

I would like to take this opportunity to express my sincere gratitude and deep respect to everyone who contributed to the successful completion of this work. I am especially thankful to Mrs. Shivali Verma, Scientist/Engineer - SE, NRSC and Dr. M. V. Ramana, Group Head, Atmospheric Sciences Land Surface Process and Climate Group(ASSL&CG), NRSC for their exceptional mentorship and guidance.Their insightful suggestions and steadfast support greatly shaped the course of this project and had a lasting impact on my professional growth.

I am equally grateful to Smt. Savitha Sunkari for offering me the privilege of being a part of such a prestigious organization. Her belief in my potential and continuous encouragement have been instrumental in my personal and academic journey. It has truly been an honour to work at the National Remote Sensing Centre, ISRO, within the Earth and Climate Science Area, and I remain genuinely thankful for the experience.

I extend my heartfelt thanks to my family and friends, whose unwavering love and belief in me have been a constant source of strength and inspiration. Their support kept me motivated throughout this demanding yet fulfilling journey. I am also indebted to my peers and fellow interns at NRSC, whose friendship and collaborative spirit made this experience both enriching and enjoyable.

To conclude, this journey would not have been possible without the countless contributions big and small of the many individuals I've had the good fortune to meet along the way. I am deeply thankful to each one of them for their direct or indirect support. Your contributions have left a lasting impression, and I offer my heartfelt thanks for being a part of this meaningful chapter in my life.

Abstract

Atmospheric Rivers(ARs)are long, narrow plumes of enhanced moisture transport that play a critical role in extreme precipitation and flooding events. While well studied in mid-latitude regions, ARs over the Indian subcontinent have received limited attention despite their increasing relevance to multiday flood events. This study presents a comprehensive AR detection framework for India that integrates multiple data sources to identify and track ARs both retrospectively and in near real-time.

The detection algorithm is built upon Integrated Vapor Transport (IVT) fields derived from *ERA5 reanalysis, GFS forecast data, and satellite-based total precipitable water(TPW)products from INSAT-3D/3R and AMSR2.* The method employs IVT thresholding, object-based segmentation, landfall and size filtering, and AR axis tracing to detect AR events. Satellite data are used to capture the moisture structure in regions where reanalysis or forecast data may have limitations, offering a more detailed picture of AR evolution and intensity.

This multi-source AR detection framework enables both *retrospective analysis and operational forecasting*, enhancing our ability to anticipate AR-driven risks. By incorporating satellite observations with reanalysis and forecasts, the approach provides a robust tool for *early warning systems, disaster preparedness, and climate resilience planning* in the Indian context.

Abbreviations

- *AR:* Atmospheric Rivers
- *ECMWF*: European Centre for Medium-Range Weather Forecasts
- *ERA5:* ECMWF Reanalysis v5
- *GFS:* Global Forecast System (GFS)
- *IVT:* Integrated Vapor Transport
- *IWV:* Integrated Water Vapour
- *INSAT:* Indian National Satellite System
- *AMSR:* Advanced Microwave Scanning Radiometer
- *TPW:* Total Precipitable Water
- *NetCDF:* Network Common Data Form

Introduction

Atmospheric Rivers (ARs) are relatively narrow, elongated corridors of concentrated moisture transport in the atmosphere. These systems play a pivotal role in the global hydrological cycle by moving vast quantities of water vapor from tropical to mid-latitude regions. On average, an atmospheric river transports an amount of water vapor roughly equivalent to the flow of the Mississippi River at its mouth. However, exceptionally strong ARs can transport *up to 15 times* that volume, making them potent drivers of hydrometeorological extremes.

When ARs make landfall, they often release the accumulated moisture in the form of intense rainfall or snowfall. These precipitation events can lead to a range of high-impact hazards, including *flooding, landslides, infrastructure damage, and disruptions to transportation and livelihoods*. ARs that stall over watersheds are especially dangerous, as they can sustain heavy precipitation over a prolonged period. A well-documented example is the *Pineapple Express*, an AR system that transports moisture from the vicinity of Hawaii to the western coast of the United States, often leading to significant flood events.

Despite their potential for destruction, not all atmospheric rivers result in damage. Many ARs are relatively weak and instead provide *beneficial rainfall and snow accumulation*, contributing significantly to regional water resources. Thus, ARs are not only a key source of hydrological extremes but also *essential components of regional water supply systems.* Recent studies have highlighted ARs as *the dominant mechanism for long-range moisture transport* in several regions around the world, such as the western United States, South America, and parts of Europe. Emerging evidence suggests that *the Indian subcontinent* also shows increasing incidents of ARs.

ARs can be characterized based on several factors, including the magnitude of their moisture transport, duration, and their hydrological impacts. Their persistence is governed by large-scale atmospheric pressure gradients and oceanic moisture availability, and some ARs can endure for several days. The *magnitude of an AR typically increases with its duration,* which is often linked to the extent and continuity of moisture uptake over oceanic regions. For instance, long-duration ARs affecting the western United States may source moisture from much of the northern Pacific Ocean.

Climate records indicate an *increasing frequency and intensity of ARs* globally, a trend that is projected to intensify further in a warming climate. This amplification has critical implications for both flood risk and water availability, especially in regions vulnerable to multiday extreme precipitation.

In India, *prolonged heavy rainfall events have increased markedly in recent decades,* contributing to a growing frequency of urban and riverine floods. These changes are expected to worsen with continued climate change. Yet, the contribution of atmospheric rivers to these events in India remains largely unquantified. Given the scale and consequences of recent flood events in regions such as Kerala, Uttarakhand, and Gujarat, there is a pressing need to investigate the occurrence, structure, and driving mechanisms of ARs over the Indian subcontinent.

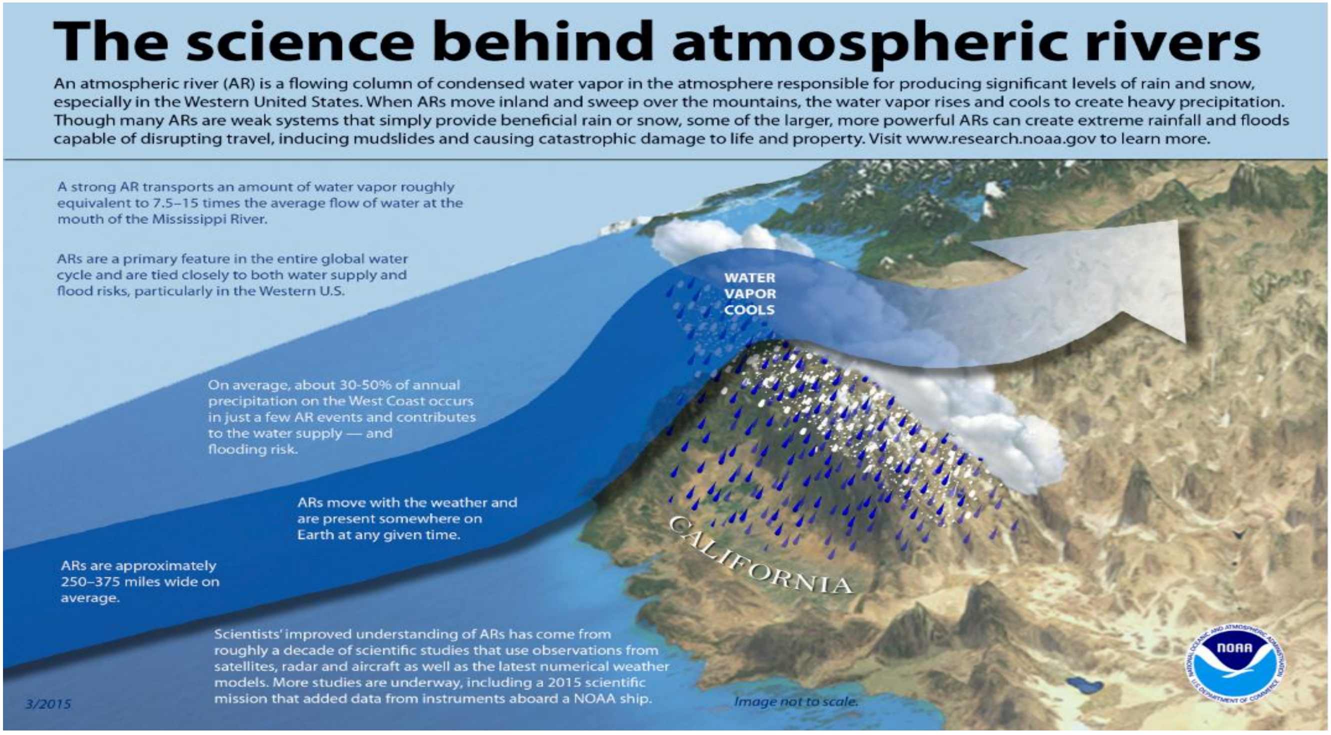


Figure 1: Illustration explaining the science behind atmospheric rivers, highlighting their moisture transport mechanisms and impacts on the U.S. West Coast (Source: NOAA, 2015).

Common detection criteria

According to the American Meteorological Society (AMS) Glossary, an AR is "a long, narrow, and transient corridor of strong horizontal water vapor transport." Most ARs are shallow (confined below ~3 km altitude), transient (lasting 1–5 days), and oriented poleward or along storm tracks.

Most AR detection methods are based on moisture thresholds and geometric characteristics. Widely used criteria include:

a) *Integrated Vapor Transport (IVT) threshold:*
ARs are typically identified where IVT ≥ *150* $kg \cdot m^{-1} \cdot s^{-1}$ over a contiguous region *(Guan and Waliser, 2015)*.

b) *Length threshold:*
The moisture corridor must be at least *1000 km long.*

c) *Length-Width ratio:*
The length-width ratio must be greater than *2 (Guan and Waliser, 2015).*

d) *Poleward orientation:*
The IVT vector should have a dominant northward (or poleward) component *(greater than 50* $kg\ m^{-1}s^{-1}$*)*

(Guan and Waliser, 2015).

Datasets and Region of Study

This study utilizes a combination of reanalysis, satellite-based remote sensing, and forecast model outputs to detect and analyze Atmospheric River (AR) events over the Indian subcontinent. The datasets were chosen based on their temporal resolution, spatial coverage, and suitability for moisture transport diagnostics such as Integrated Vapor Transport (IVT) and Integrated Water Vapor (IWV).

a) **ERA5 Reanalysis data (hourly data on pressure levels from 1940 to present) for IVT calculation**:

Used for calculating Integrated Vapor Transport (IVT), which is a core variable in AR detection algorithms. ERA5 provides hourly atmospheric data on pressure levels globally at a horizontal resolution of ~0.25°. Specific humidity and wind components at multiple pressure levels are used to compute the vertically integrated moisture flux.

Source: European Centre for Medium-Range Weather Forecasts (ECMWF) through the *Copernicus Climate Data Store (CDS).*

Link: **https://cds.climate.copernicus.eu**

b) ERA5 Reanalysis data (hourly data on single levels from 1940 to present) for IWV calculation:

Used for calculating Integrated Water Vapor (IWV) to assess the total column moisture content.Hourly surface-level fields such as total column water vapor are available, which help in validating AR moisture signatures.

Source: European Centre for Medium-Range Weather Forecasts (ECMWF) through the *Copernicus Climate Data Store (CDS).*

Link: **https://cds.climate.copernicus.eu**

c) INSAT 3D/3R satellite data:

Provides high-frequency, high-resolution observations of Total Precipitable Water(TPW)over the Indian region.
The INSAT-3D/3DR satellites operated by ISRO provide TPW and other atmospheric parameters at a spatial resolution of ~8 km and temporal frequency of 30 minutes to 1 hour.

Source: India Meteorological Department(IMD) and Indian Space Research Organisation (ISRO).

Link **: https://www.mosdac.gov.in**

d) AMSR satellite data:

Used for daily global retrievals of TPW over ocean regions. The Advanced Microwave Scanning Radiometer 2 (AMSR2) onboard GCOM-W1 provides TPW data at a spatial resolution of ~25 km, particularly useful for observing over oceans where ARs typically form.

Source: NASA's Earthdata portal.

Link: **https://earthdata.nasa.gov**

e) GFS data (Forecast data):

Enables daily detection and short-term prediction of ARs for operational and early warning applications. The Global Forecast System (GFS) provides weather forecasts at a 0.25° resolution with hourly outputs up to 120 hours ahead and 3 hourly outputs from 120 to 384 hours ahead.

Source: National Centers for Environmental Prediction (NCEP), accessed via *NOAA*.

Link: **https://nomads.ncep.noaa.gov**

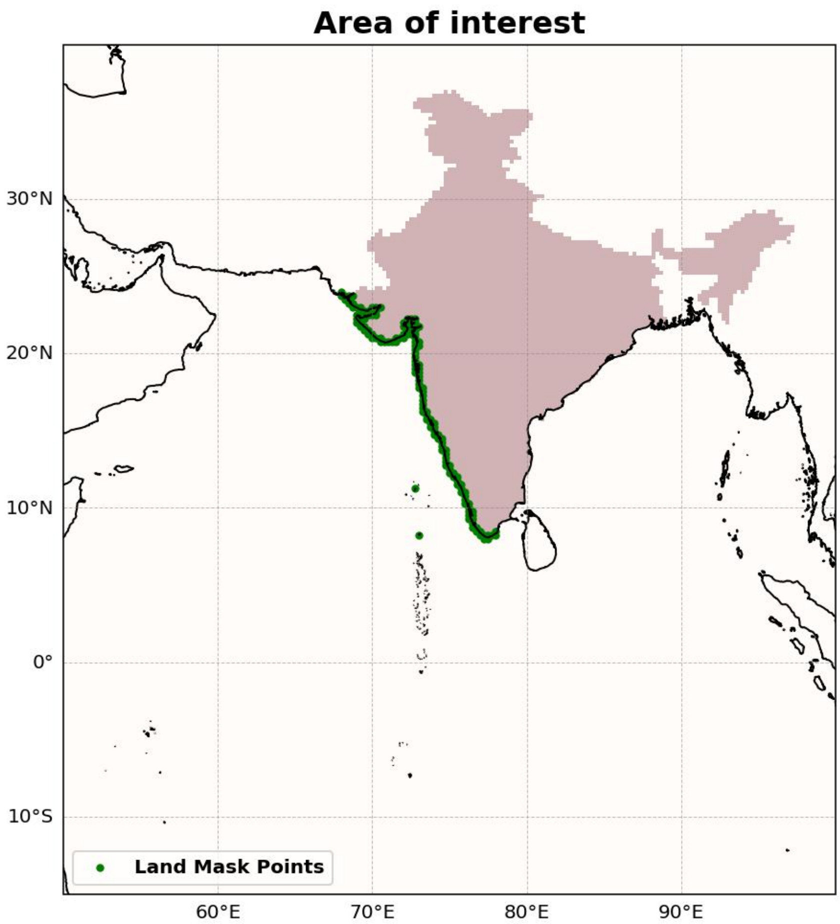


Figure 2: Area of Interest and LandFall
The red-shaded region represents the landfall used in the analysis, while the green dots represent coastal grid cells along the Indian landmass used for landfall detection logic

Co-ordinates:
North: 40 degrees
South: -15 degrees
West: 50 degrees
East: 100 degrees

The study focuses on the Indian subcontinent, particularly the western and southwestern coastal regions (Gujarat and Kerala), which are frequently affected by intense precipitation events.

Tools used

a) Development Environment:

- **Language used:**

A powerful, high-level programming language widely used in scientific research and data analysis. Python was the core language for implementing the Atmospheric River (AR) detection pipeline. It offers excellent readability, rapid development capability, and strong support for working with large scientific datasets. Python's modularity allowed for building a flexible AR detection workflow that could be adapted to different datasets and regions, making it ideal for climate data processing and geospatial analysis tasks.

- **Anaconda Distribution:**
  Provides an isolated and managed environment for scientific computing with pre-installed libraries (e.g., NumPy, SciPy, matplotlib) and tools like Spyder.
- **Spyder IDE:**
  Used for writing, debugging, and executing Python scripts for AR detection and analysis.
- **Command Line / Terminal (Anaconda Prompt or system terminal):**
  For running scripts, installing packages with conda or pip, and batch-processing AR detection routines.

b) Visualization Tools:

- **Panoply:**
  Used to view and quickly inspect NetCDF and HDF5 datasets, especially for checking variable structures and spatial patterns before processing in Python.

- **QGIS (Quantum Geographic Information System):**

Used to visualize *GeoTIFF* files generated from AR detection results. QGIS helps overlay AR-related data (e.g., IVT fields or AR masks) on geographic maps, inspect spatial alignment with regional boundaries (like Kerala or Gujarat), and validate output visually with high-resolution geographic context.

c) File Transfer:

- **FileZilla:** Used for secure SFTP-based data transfer between local systems and remote servers or repositories.

d) Python Libraries used:

- **Numerical and Scientific computing:**

✔ numpy – Used for efficient array operations and general numerical analysis.

✔ scipy – Specifically scipy.ndimage for image processing tasks such as object labeling (label, sum, mean) during AR detection.

- **File Handling & Data Formats:**

✔ netCDF4 – For reading and manipulating NetCDF-formatted climate datasets (e.g., ERA5).

✔ h5py – For accessing HDF5-formatted satellite data (e.g., AMSR2 Level-2 TPW).

✔ pygrib – For reading and extracting variables from GRIB-formatted data (e.g., GFS model outputs for IVT components).

✔ glob, os – For file searching, path handling, and directory management.

- **Date & Time Handling:**

✔ datetime – For working with timestamps, time intervals, and organizing data by date.

- **Geospatial Calculations:**

✔ geopy.distance – Used to calculate distances on Earth's surface (e.g., AR axis length and width) based on latitude and longitude.

- **Plotting & Visualization:**

✔ matplotlib.pyplot – For plotting time series, IVT/TPW fields, and AR axes.

✔ cartopy.crs, cartopy.mpl.ticker – For map projections and formatting geospatial plots.

✔ matplotlib.ticker – For customizing plot axes and tick marks

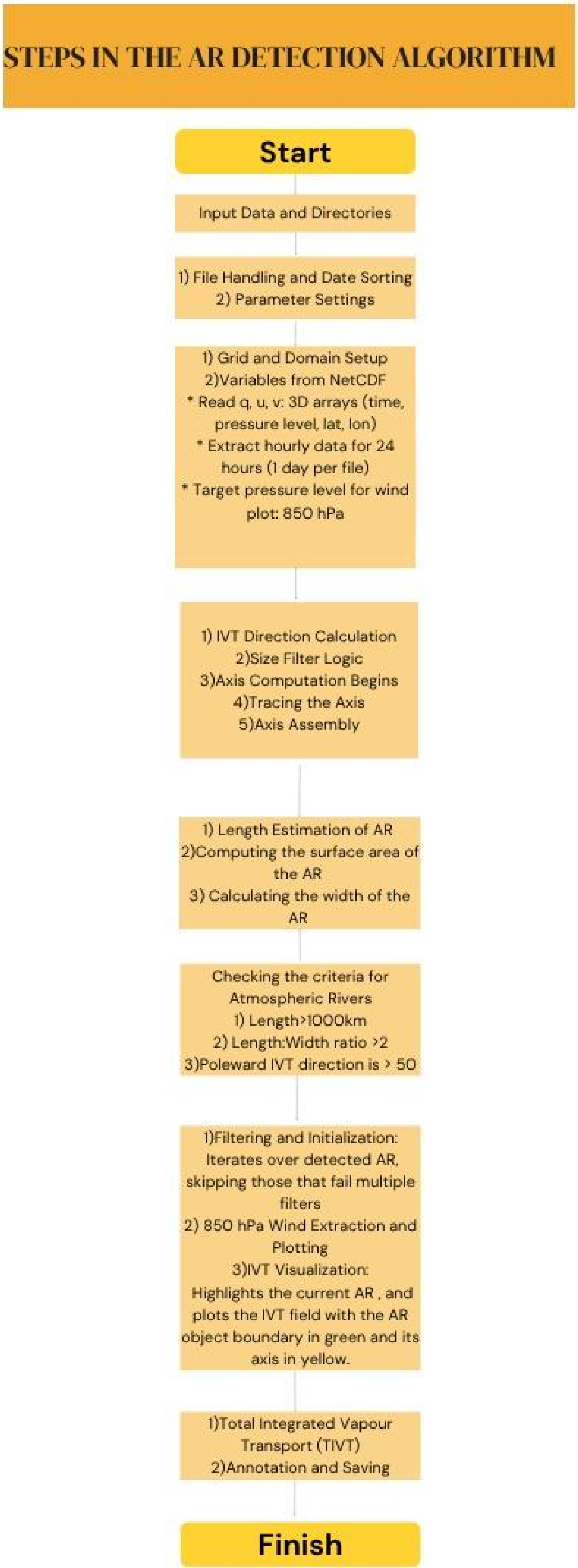


Figure 3: Flowchart illustrating the steps involved in the Atmospheric River (AR) detection algorithm.

Formulae used in the code

1) IVT Calculation:

$$IVT_x = 1/g \int_1^N \quad q_i \times u_i \times \Delta p_i \times 100$$
$$IVT_y = 1/g \int_1^N \quad q_i \times v_i \times \Delta p_i \times 100$$

IVT

$$= 1/g \sqrt{\left(\left(\int_1^N \mathrm{q_i} \times \mathrm{u_i} \times \Delta\mathrm{p_i} \times 100 \quad\right)^2 + \left(\int_1^N \mathrm{q_i} \times \mathrm{v_i} \times \Delta\mathrm{p_i} \times 100\right)^2\right.}$$

where:

$$q_i = specific\ humidity\ at\ the\ i^{th}\ pressure\ layer$$
$$u_i = Zonal\ wind\ at\ the\ i^{th}\ pressure\ layer$$
$$v_i = Meridonial\ wind\ at\ the\ i^{th}\ pressure\ layer$$
$$\Delta p_i = Pressure\ difference\ at\ the\ i^{th} layer$$
$$g = 9.8\ m/s^2 : gravitational\ acceleration$$
$$N = Number\ of\ pressure\ layers$$

2) IVT sums and means:

$$IVT_{sum} = sum(IVT)$$

$$IVTx_{sum} = sum(IVT_x)$$

$$IVTy_{sum} = sum(IVT_y)$$

The IVT values along the pressure dimension are summed up.
IVT sums are generated for each hour

The sum generated at each hour is stored in variables as below:

hourly_sum : stores IVT_{sum} at each hour
hourlyx_sum : stores $IVTx_{sum}$ at each hour
hourlyy_sum : stores $IVTy_{sum}$ at each hour

Now calculating the average of IVT across all hours

Mean of hourly_sum , hourlyy_sum and hourlyx_sum is calculated

3) IVT_Direction:

$$IVT\_dir = \{Eastward\_IVT/Northward\_IVT) \times (180/\pi) + 180\,\}\,\%360$$

IVT_dir stores the direction (in degrees) from North (0^o)

Multiplying by 180/π gives the direction in degrees

%360 gives the remainder when divided with 360. This is essential to ensure that the direction does not go out of bounds [0-360]

4) IVT_Threshold:

Calculated based on the 85th percentile of IVT from the previous 1 day.

$$IVT_{min} = 150\ kg.m^{-1}.s^{-1}$$

$IVT_{threshold}$ is calculated as the

$$maximum\ between\ IVT_{min}\ and\ IVT_{85th\ percentile}$$

5) Total precipitable water:

Total precipitable water is calculated as :

(L1_PREC_WATER) + (L2_PREC_WATER) + (L3_PREC_WATER)

where:
L1_PREC_WATER (Pressure levels: 1000-900mb)
L2_PREC_WATER (Pressure levels: 900-700mb)
L3_PREC_WATER (Pressure levels: 700-300mb)

Methodology:AR Detection Algorithm

1) Initialization and Parameter Definition:

This section sets up the working environment, loads required libraries, defines the working directories, and specifies the key parameters required for Atmospheric River (AR) detection.

i. Library Imports and Setup:

- numpy, scipy, and matplotlib for numerical processing and plotting,
- cartopy for geospatial projections and map plotting,
- netCDF4 for handling NetCDF climate datasets,
- datetime and geopy for working with dates and geographic distances,
- warnings is used to suppress warning messages for clean execution.

ii. Directory Paths and File Handling:

Two directories are defined:
datadir: The path containing daily IVT-related NetCDF data files.
datadir2: The directory where AR detection outputs are stored.
The script filters for .nc files in datadir and sorts them based on the date extracted from their filenames using the extract_date() function. The filenames are assumed to follow a pattern ending in a date string like "DD-MM-YYYY.nc"

iii. Parameter Definitions:

Several key detection parameters are defined to guide the AR detection logic:

min_ivt: The minimum fixed threshold of IVT magnitude to be considered (150 kg m^{-1} s^{-1}).
threshold_percentage: Percentile threshold for adaptive detection (e.g., 85th percentile).
min_length: Minimum length of a potential AR (1000 km)
min_size: Minimum number of connected grid cells to be considered a valid AR object (60).

iv. Landfall Mask Loading:

A .mat file containing a binary landfall mask (e.g., highlighting Gujarat to Kerala coastline) is loaded. This mask is used later to check whether an IVT object intersects the coastal region of interest.

v. Grid and Spatial Extent:

The grid resolution is set to 0.25° to match ERA5 data.

The geographical bounds of the domain are defined as:

- Latitude: -15° to 40° N
- Longitude: 50° to 100° E

2) Daily IVT Calculation and Thresholding :

This section focuses on extracting meteorological variables from ERA5 NetCDF files, computing Integrated Vapor Transport (IVT), and applying a dynamic threshold to identify regions with high moisture transport.

i. Data Extraction and Preprocessing:

For each NetCDF file (representing one day), the script:

- Opens the dataset and extracts:
  - longitude and latitude grids,
  - Specific humidity and horizontal wind components across pressure levels,

Constructs a list of hourly timestamps and prepares meshgrids for lat-lon spatial calculations.

Calculates pressure differences for IVT integration.

ii. Hourly IVT Computation:

For each hour (24 total per day):

- Calculates layer-wise average values of specific humidity and wind components across adjacent pressure levels.
- Computes the IVT components:
  - Eastward IVT and northward IVT
  - Magnitude of IVT by combining the components.
- Stores hourly sums of IVT and individual components.

iii. Daily Averaging:

After looping through all hours of a day:

- Averages the hourly IVT values to get a daily mean IVT field, and its eastward and northward components.
- A rolling window of the previous 1 day's Integrated Vapor Transport data is maintained to compute the 85th percentile threshold.

iv. IVT Direction:

Computes the IVT direction (in degrees) from the eastward and northward components and adjusts to lie within 0° to 360°.

v. Seasonal Masking and Threshold Application:

If the month is between October and January (i.e., during the Indian winter season), additional masking is applied:

High IVT values (>600 kg m^{-1} s^{-1}) are removed to suppress outliers,

The latitudes below 8° N are masked.

Computes a threshold as the greater of the percentile value or a fixed min_ivt limit.

Creates a binary object_mask indicating regions that exceed the IVT threshold.

vi. Component Filtering for Physical Relevance:

To ensure physical plausibility, regions where either eastward or northward IVT components are positive are retained

3) Object Identification, Landfall Detection, and Size Filtering:

This section of the algorithm identifies individual IVT structures (potential Atmospheric Rivers), filters them based on landfall and size, and prepares the data for further analysis, such as AR axis tracing.

i. Connected Object Labeling:

Using the binary object_mask (where grid cells with IVT > threshold are True), the script applies a connected-component labeling algorithm (scipy.ndimage.label). This assigns a unique integer label to each spatially contiguous region of high IVT, treating each as a potential AR object.

ii. Landfall Detection:

A pre-defined **land mask** is multiplied by the binary object_mask to retain only grid cells over land.

Then, using the labeled object mask (labelled_mask), the script finds which labeled objects intersect land by checking for any non-zero overlap with the land mask.

The IDs of land-intersecting ARs are stored in the landfall_filter list.

iii. Size Filtering:

To reduce false positives and computational load:

The number of grid cells in each object is computed.

Any object that is *smaller than* the minimum size (in terms of grid cells) is added to the size_filter list and subsequently excluded from further processing.

4) AR Axis and Landfall Detection:

For each detected Atmospheric River(AR)object, we compute the axis by tracing from the maximum IVT point in the direction of IVT flow using the trace_axis() function. Tracing is done forward and backward to capture the full extent of the AR. If the IVT maximum lies over land, its location, magnitude, and direction are recorded as the landfall point. The resulting axis mask and coordinates are stored for further analysis, with all data organized into structured lists and arrays. Small objects are excluded using a precomputed size filter.

5) AR Object Length Estimation:

To estimate the length of each Atmospheric River(AR)object, we first skeletonize the object's mask to reduce it to a one-pixel-wide structure. This skeleton is converted into a graph, where nodes represent grid cells and edges connect adjacent pixels. The longest path within the graph is then identified, and its length is computed in kilometers using geodesic distances between pixel coordinates. Objects filtered out by size are skipped.

6) Surface Area and Width Calculation:

For each AR object, the surface area is computed by summing the area of all grid cells it covers, using a spherical Earth approximation based on grid resolution and latitude. Width is then estimated by dividing the object's surface area by its length (previously computed). This provides an effective width in kilometers for each AR.

7) AR Object filtering criteria:

i. Length Criterion:

Each object is evaluated for its axis length, computed by skeletonizing the object and determining the longest continuous path. Objects with axis lengths less than 1000 km are excluded.

ii. Narrowness Criterion:

Next, a *length-to-width ratio* is calculated for each object. The width is derived by dividing the surface area of the object (in km^2) by its axis length.

iii. Meridional IVT Criterion:
For all remaining objects, the *mean meridional (northward) component* of Integrated Vapor Transport (IVT) is evaluated. An object is discarded if its poleward IVT is less than or equal to $50\ kg\ m^{-1}s^{-1}$

8) Atmospheric River (AR) Snapshot Generation:

i. AR Object Isolation:
Each AR object is isolated from the larger IVT field using its unique label. The isolated region is then:

Displayed as a filled contour over a map with coastlines and gridlines.
Outlined in *green* for the object boundary and *yellow* for the AR axis (computed earlier).

ii. Spatial and Temporal Metadata:
The date of the snapshot is extracted from the file's metadata.
The geographic coordinates (latitude, longitude) of the core IVT point are calculated.
The distance to the nearest land from the core is estimated using geodesic distance to land points.

iii. Physical Attribute Calculation:
For each AR object, the following key characteristics are computed:

- *Length and Width* (in km): Based on the skeletonized axis and object area.
- *Mean IVT Magnitude and Direction*: Representing average moisture transport intensity and orientation.
- *Landfall IVT Magnitude and Direction*: Based on the object's intersection with the land mask.
- *TIVT (Total IVT)*: Calculated by summing IVT values along a line perpendicular to the AR axis at the core point.

iv. Visual Layout and Annotation:
A multi-element plot is generated for each AR:

- IVT field displayed with color shading.
- Green and yellow contours for AR boundary and axis.
- Annotations include:
  - ✔ Date
  - ✔ Object length and width

- ✔ IVT magnitude and direction
- ✔ Landfall metrics
- ✔ TIVT value
- ✔ Distance from land

Text annotations are positioned clearly beneath the plot, and axis labels are manually added for clarity.

v. Export to GeoTIFF:

To preserve the spatial fidelity of the IVT field:

- ➢ The IVT array is georeferenced using its known latitude/longitude extent and resolution.
- ➢ The object is saved as a GeoTIFF with accurate metadata, suitable for loading into GIS software.

vi. Seasonal Condition Handling:

- ➢ For *October to January* (winter months), all AR objects are plotted regardless of physical filtering.
- ➢ For other months, only AR objects that pass all filtering criteria (size, length, narrowness, and poleward IVT) are considered for plotting. This ensures that the snapshots represent meteorologically robust ARs.

Methodology: Downloading and Visualizing Total Precipitable Water (TPW) Data using Satellite imagery

This project automates the process of downloading, processing, and visualizing recent *Total Precipitable Water* (TPW) data using the INSAT 3D/3DR satellite and the AMSR2 satellite.

1. Methodology for INSAT 3D/3DR Satellite:

i. Connect to MOSDAC SFTP and list the available files

This section of the code establishes a secure connection to the SFTP server hosted at *mosdac.gov.in* using the paramiko library. It authenticates with the provided username and password, and then navigates to the remote folder named StandingOrder. The script retrieves a list of all files in this directory and filters them to find those that start with the prefix 3RSND_, which identifies the relevant satellite data files.For each of these filtered files, check whether the file already exists locally in the specified downloaded directory. If a file is not present locally, the script downloads it from the SFTP server. This ensures only new data is transferred, optimizing bandwidth and storage efficiency. Once the download is complete, the SFTP connection is properly closed.

Library Used: paramiko (for secure file transfers using SFTP).

ii. Reading and Processing the HDF5 Data

Once the latest data files are downloaded, the next step is to extract and process the Total Precipitable Water (TPW) data. Search the local directory for HDF5 (.h5) files that match the naming pattern *3RSND_*_L2B_SA1_V01R00.h5*, which indicates they contain the required satellite data. Sort these files by their modification time and select the three most recent files. This approach ensures that the most current atmospheric data is used for analysis.

Libraries Used: h5py, numpy, glob.

For each selected file, extract the three specific datasets using the h5py library.
i. L1_PREC_WATER (Pressure levels: 1000-900mb)
ii. L2_PREC_WATER (Pressure levels: 900-700mb)
iii. L3_PREC_WATER (Pressure levels: 700-300mb)

These represent precipitable water content derived from different algorithmic processing levels. To maintain data quality, any invalid values (represented by -999.0) are replaced with NaN (Not a Number), ensuring they are excluded from calculations.

iv. Accumulate and average the TPW from the files

Read the files and sum the three layers to compute the total TPW for each file. These totals are accumulated across all three files, and finally, an average is calculated. Compute the *average TPW* by dividing the total by the number of files. This average TPW represents the recent atmospheric water vapor content over the observed region.

v. Extract the latitude and longitude grids
i. Extract the latitude and longitude from the first .h5 file.
ii. These are used to place the TPW values correctly on the map.
iii. Values are divided by 100 to convert to actual coordinates.

vi. Visualizing the TPW Data
- Libraries Used: Matplotlib, cartopy.
- Projection Used: Plate Carrée (simple geographic projection).

The average TPW values, calculated from the three most recent HDF5 files, are plotted over a geographic coordinate grid using the Plate Carrée projection, which is a straightforward latitude-longitude layout suitable for global and regional maps.This visualization provides a clear and intuitive understanding of atmospheric moisture distribution over the observed region.

2. Methodology for AMSR2 Satellite:

i. Data Acquisition from NASA Earthdata

To obtain Total Precipitable Water (TPW) data from the AMSR2 (Advanced Microwave Scanning Radiometer 2) Level 2B Ocean product, the script leverages NASA's Common Metadata Repository (CMR) API. The script searches for data files (granules) starting from the current UTC date and moving backwards one day at a time for up to 30 days. If any granules are found for that date, their URLs are extracted, specifically filtering for .he5 file links. The script then initiates a download for each file, using authenticated HTTP requests through the Earthdata Login credentials.

ii. User Authentication and Secure Download

Access to NASA's Earthdata services requires registered user authentication. The script handles this using basic HTTP authentication. The download process is conducted using the requests library with streaming enabled to handle large files efficiently. The tqdm module is used to provide a visual progress bar for each download.
Libraries Used: tqdm

iii. Data Extraction and Preprocessing

After downloading, the script changes the working directory to the data folder and lists all available .he5 files. Using the h5py library, it opens each HDF-EOS5 file to extract three primary datasets:

- *Total Precipitable Water (TPW)*
- *Latitude*
- *Longitude*

The TPW values are checked for validity, and all values less than zero are masked as NaN, as they indicate missing or invalid data.
A geographical mask is applied to extract only the region of interest, defined by user-specified latitude and longitude bounds (50°E to 100°E longitude and -15°N to 40°N latitude).
Libraries Used: h5py

iv. Visualizing the TPW data

The visualization step uses matplotlib and cartopy for map-based plotting. A global Plate Carrée projection is selected to maintain spatial fidelity over the region of interest.Each granule (file) is plotted in sequence, and overlapping regions are visualized transparently. This helps to capture the satellite's multiple swaths and their collective coverage across the day.
Libraries Used: Matplotlib, cartopy.

Sequential Flow of the Algorithm

1) Monsoon date: 20th June 2024:

i. IVT plot

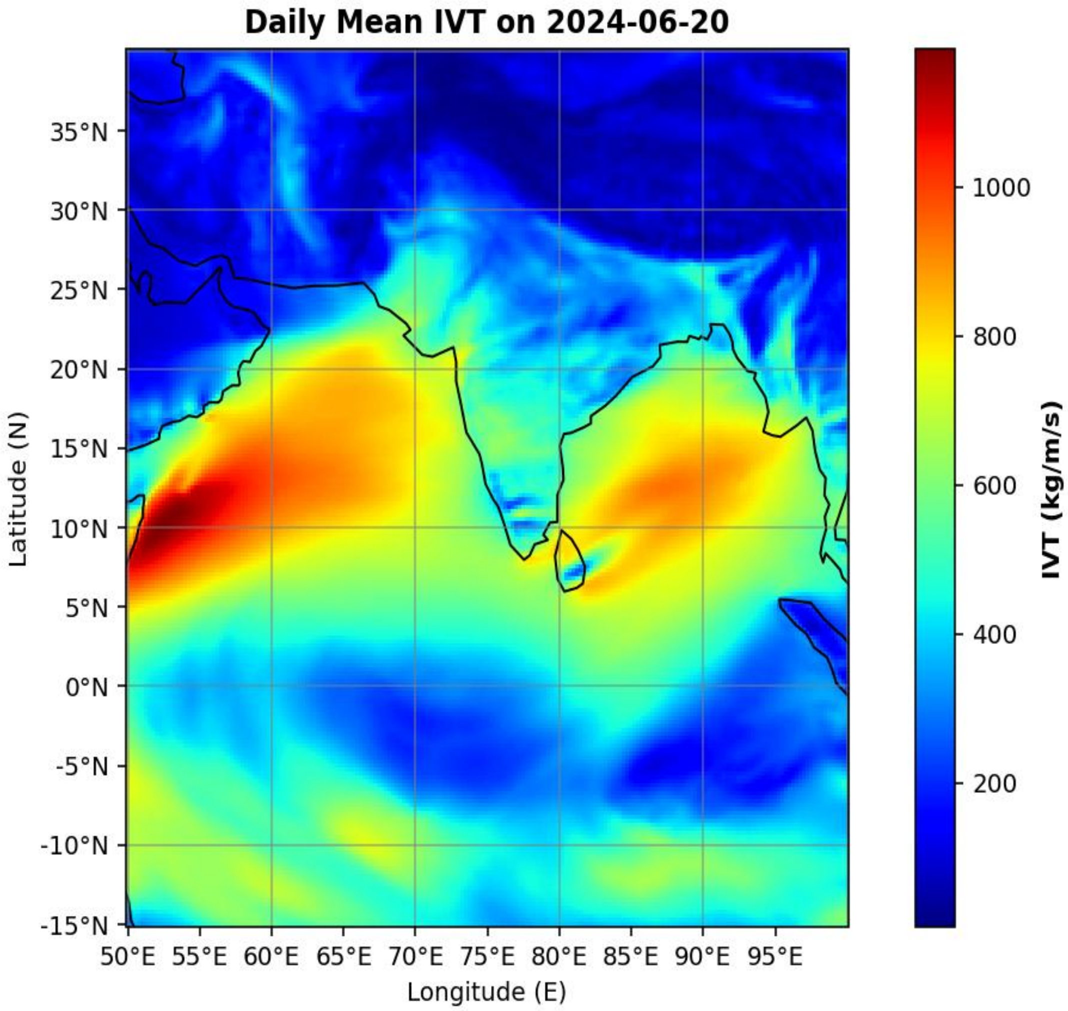


Figure 4 : Daily Mean Integrated Vapor Transport (IVT) over the Indian Subcontinent on 20 June 2024

ii. Northward IVT plot:

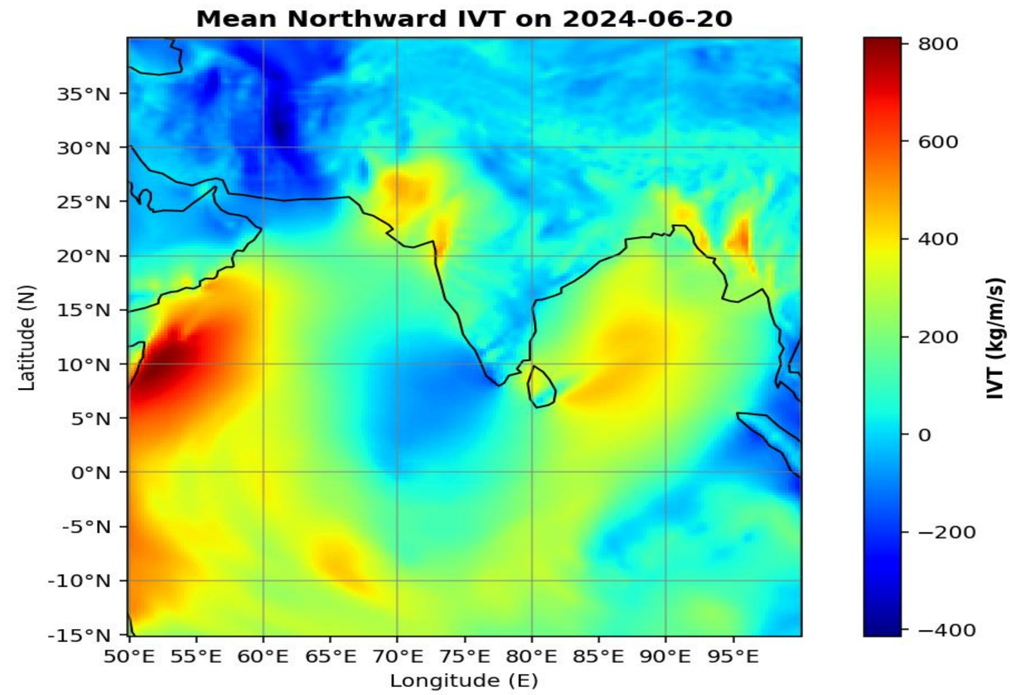


Figure 5: Daily Mean Northward Integrated Vapor Transport (IVT) over the Indian Subcontinent on 20 June 2024

iii. Eastward IVT plot:

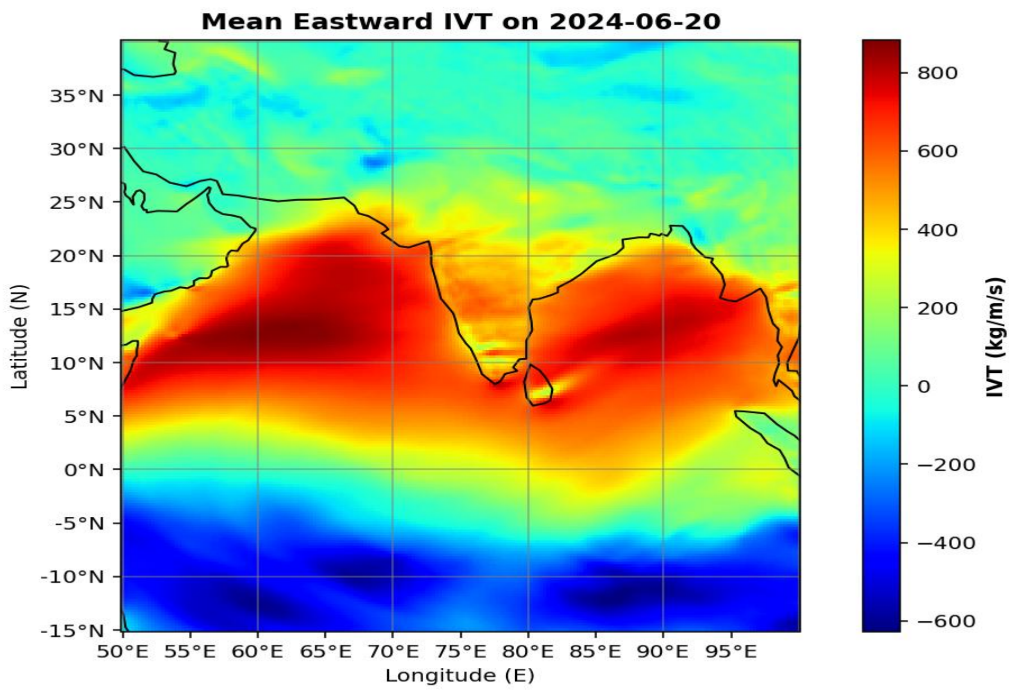


Figure 6: Daily Mean Eastward Integrated Vapor Transport (IVT) over the Indian Subcontinent on 20 June 2024

iv. Labelled mask:

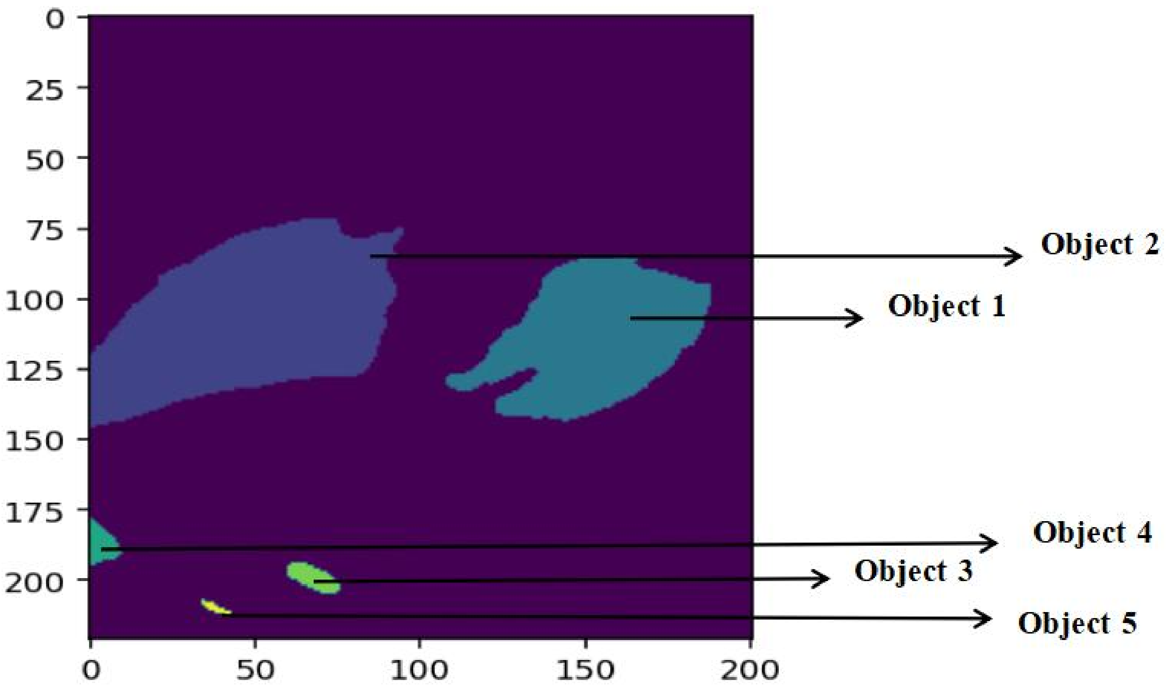


Figure 7: Labeled Atmospheric River Candidate Objects Identified from IVT Thresholding

v. AR Snapshots:

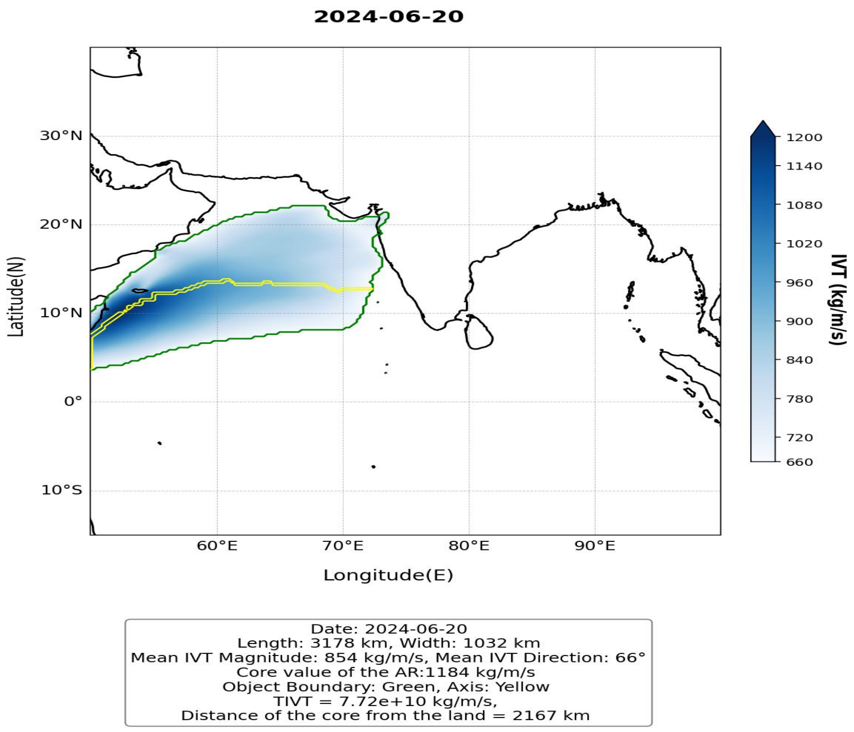


Figure 8(a)

vi. Wind plot:

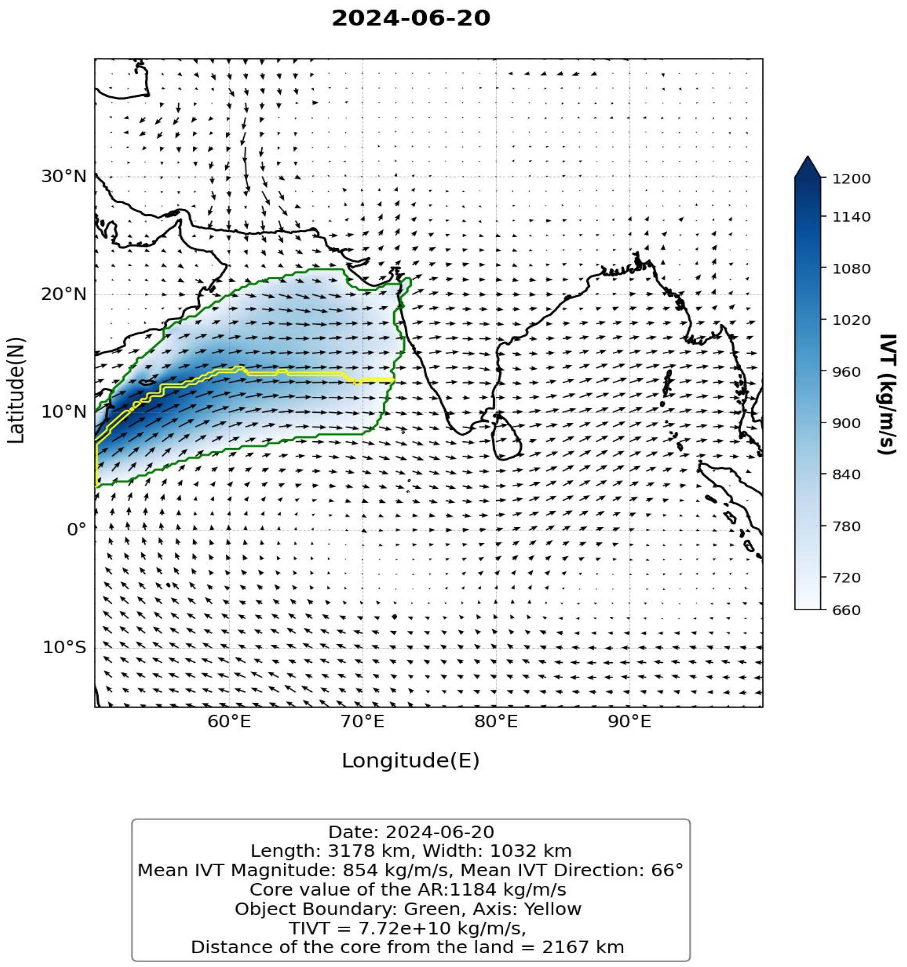


Figure 8(b)

Figure 8(a,b): Detected Atmospheric River over the Arabian Sea on 2024-06-20 Showing IVT Magnitude (shaded), Wind Vectors, Object Boundary (green), and Core Axis (yellow)

vii. INSAT data plot for Total Precipitable Water(TPW) :

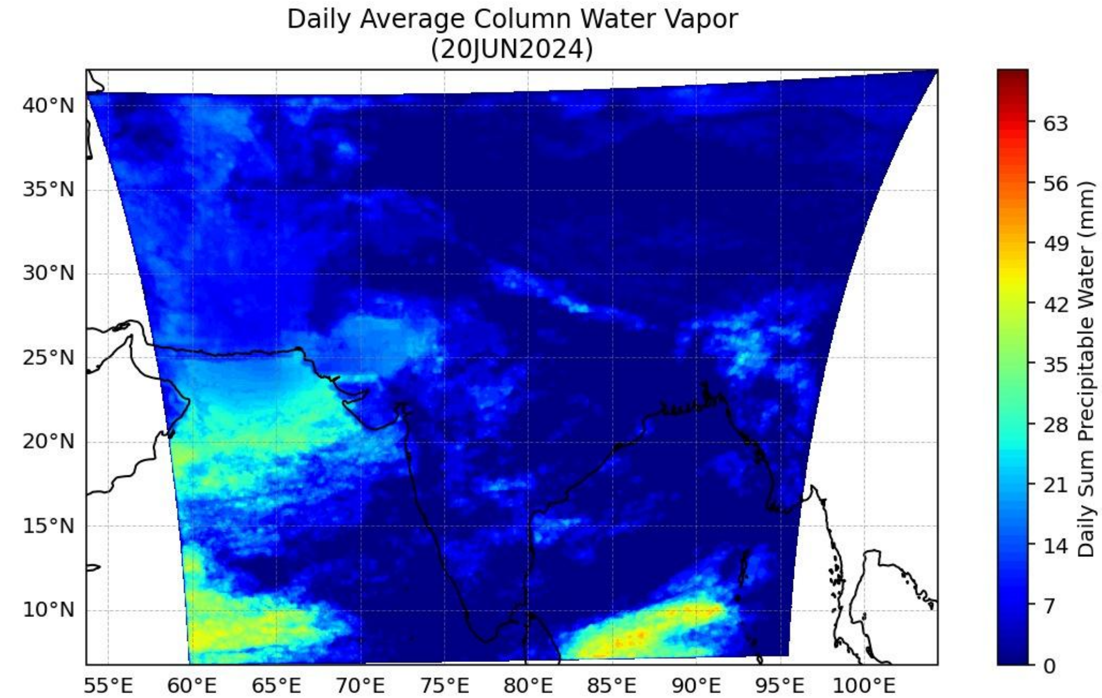


Figure 9: Daily Average Column Water Vapor over the Indian Subcontinent on 20 June 2024 from INSAT 3D Observations

viii. AMSR data plot for Total Precipitable Water(TPW) :

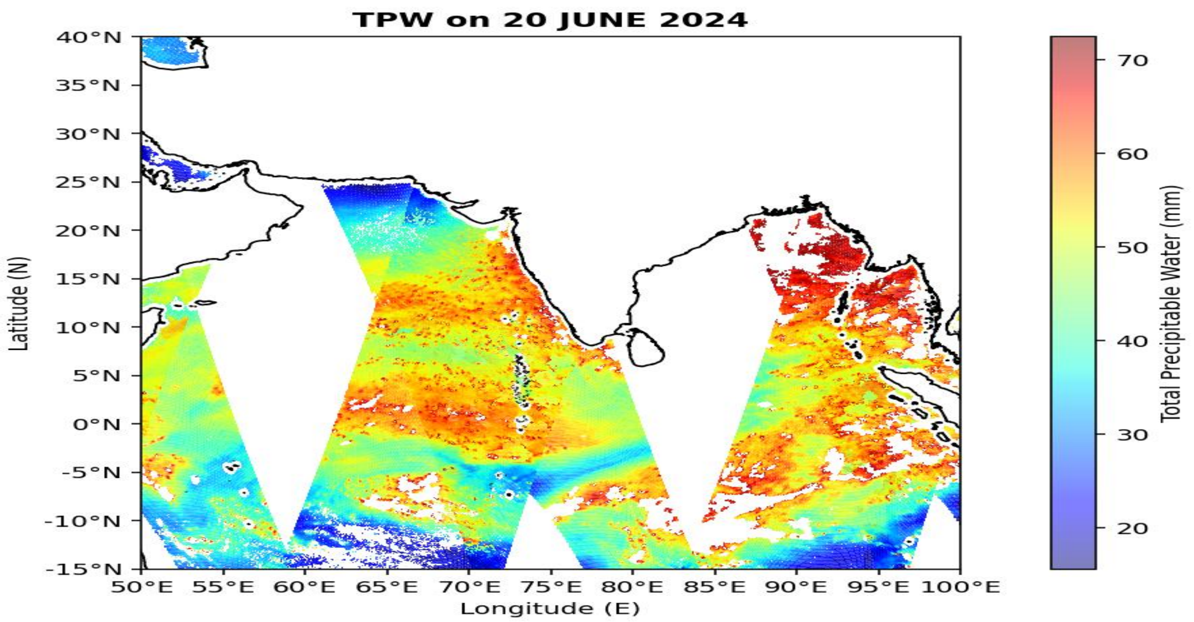
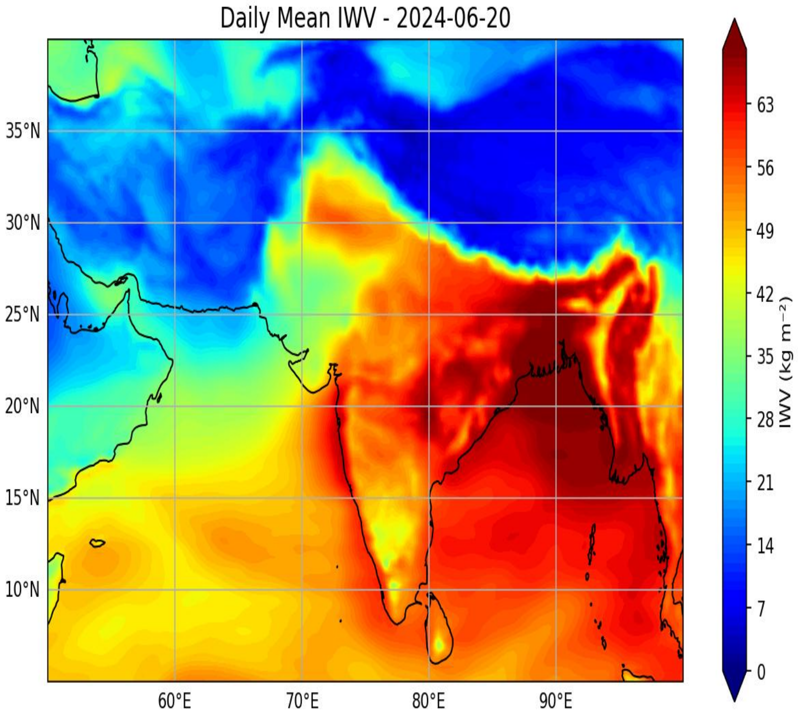


Figure 10: Daily Average Column Water Vapor over the Indian Subcontinent on 20 June 2024 from AMSR2 Observations

ix. Comparison between TPW plot from ERA5 data and INSAT data :



Figur 11(a)

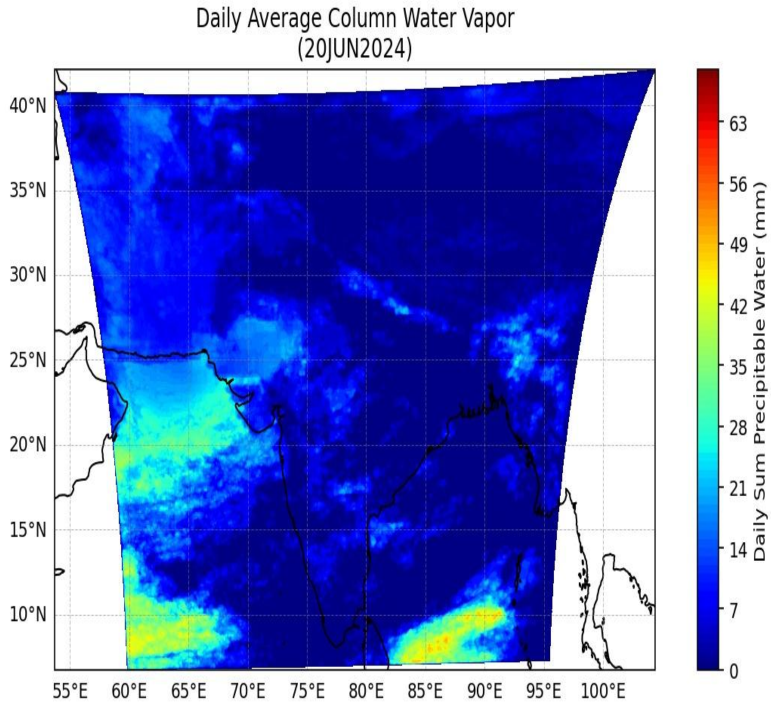


Figure 11(b)

Figure 11 :

(a): Daily mean IWV from ERA5 reanalysis data, expressed in kg/m². It shows moisture accumulation over Indian subcontinent.

(b) : Daily mean TPW retrieved from a satellite-based product (INSAT-3D), expressed in mm. This plot reveals localized zones of elevated water vapor.

x. GeoTiff Images of the detected ARs :

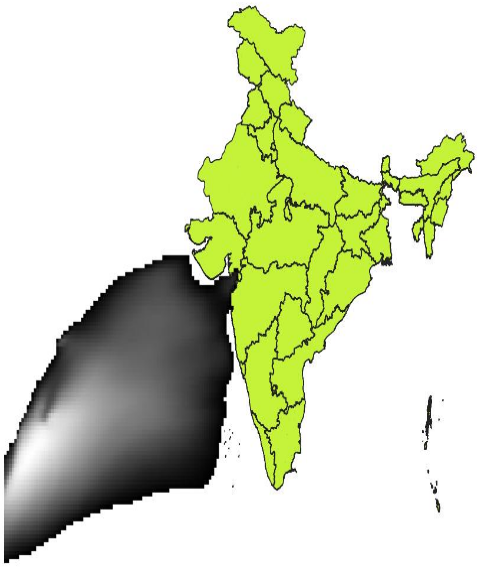


Figure 12:
GeoTIFF image generated using QGIS software, depicting an AR object approaching the western coast of India. The AR is represented in grayscale, where lighter shades indicate stronger IVT. The state boundaries of India are overlaid in green to provide clear geographic context.

2) Winter date: 12th January 2020:

i. IVT plot:

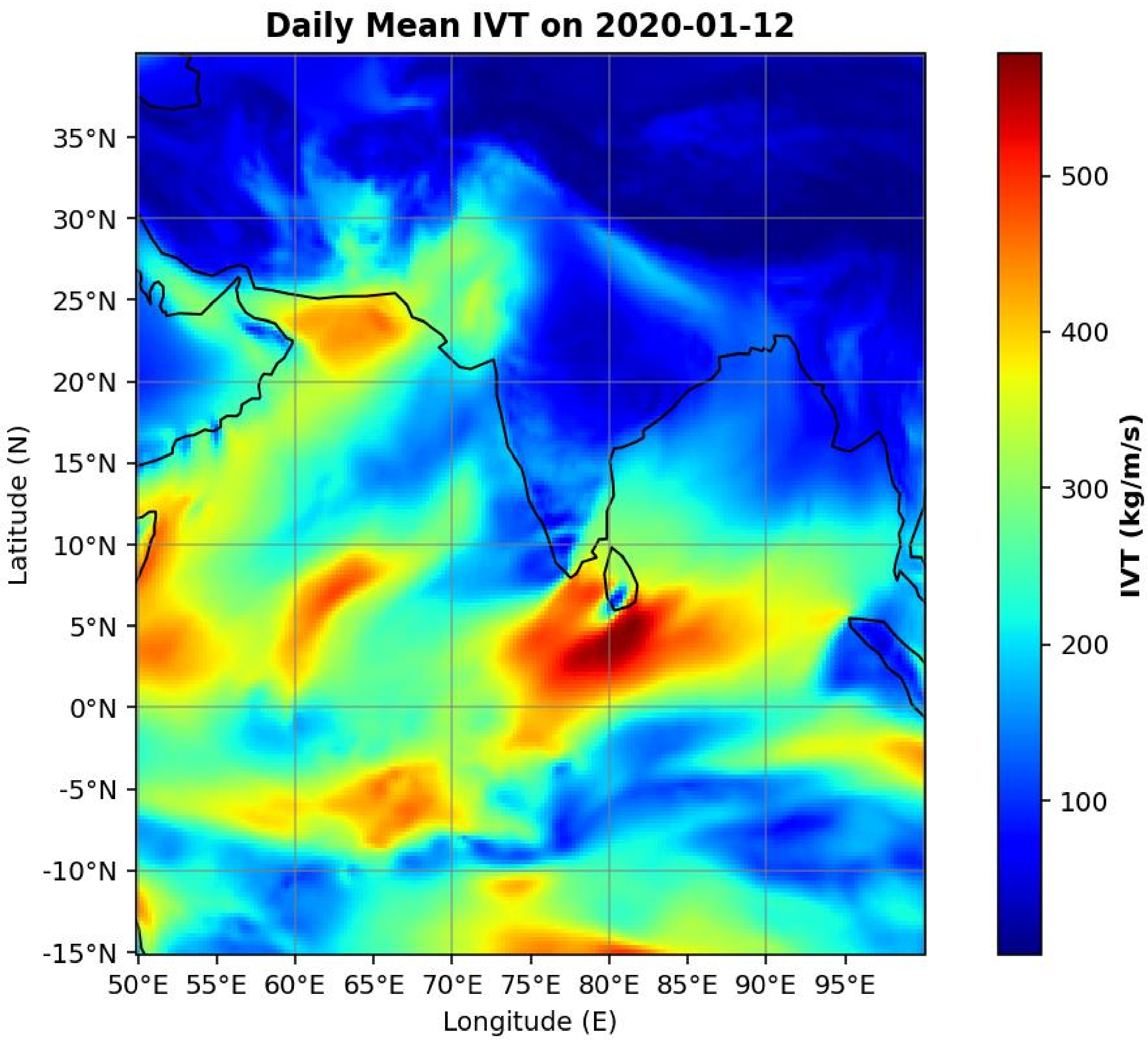


Figure 13: Daily Mean IVT over the Indian Subcontinent on 12 January 2020

ii. Northward IVT plot:

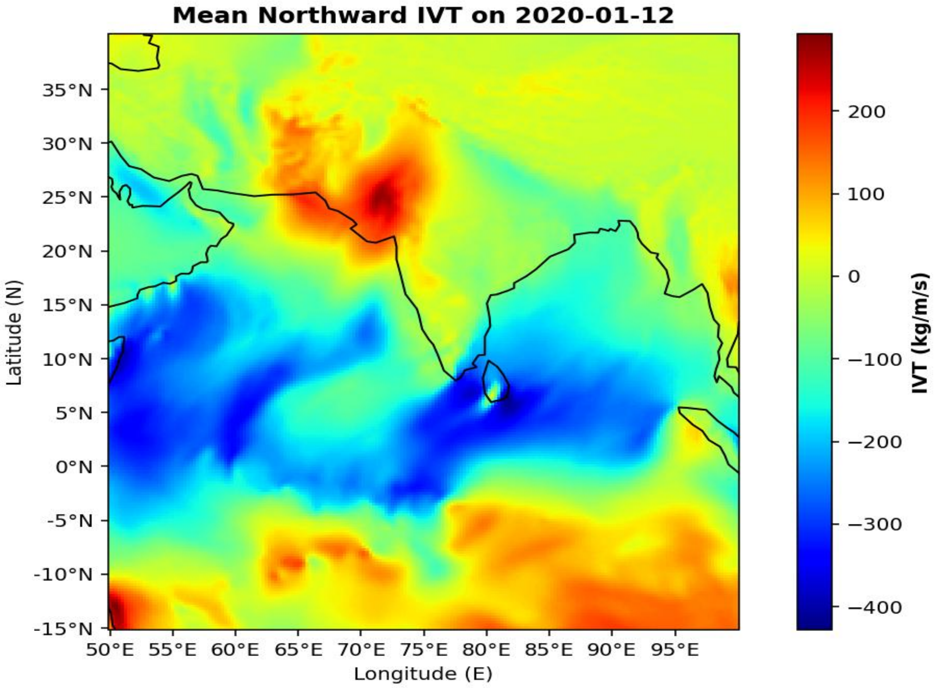


Figure 14: Daily Mean Northward IVT over the Indian Subcontinent on 12 January 2020

iii. Eastward IVT plot:

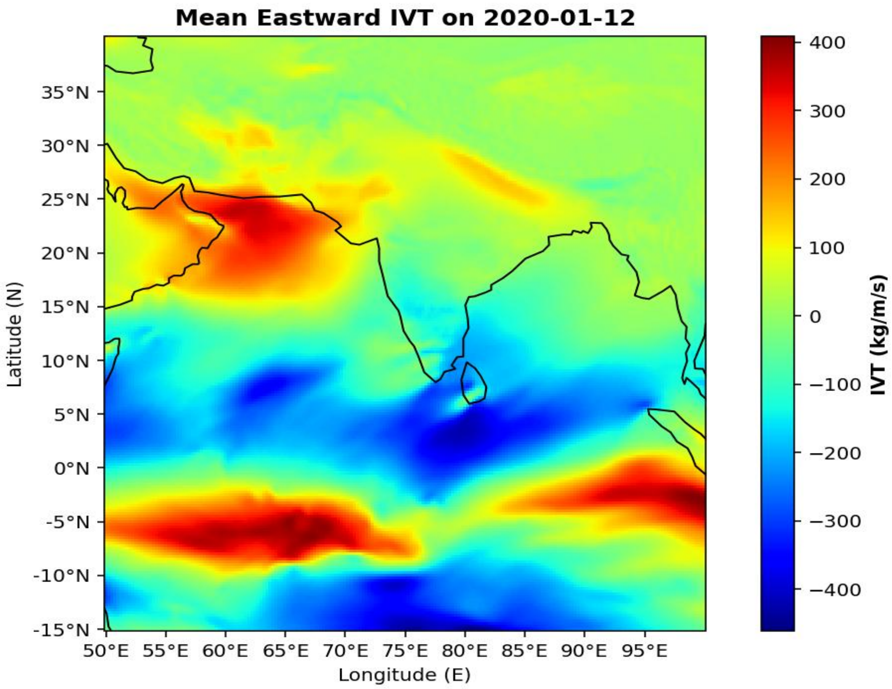


Figure 15: Daily Mean Eastward IVT over the Indian Subcontinent on 12 January 2020

iv. Labelled mask:

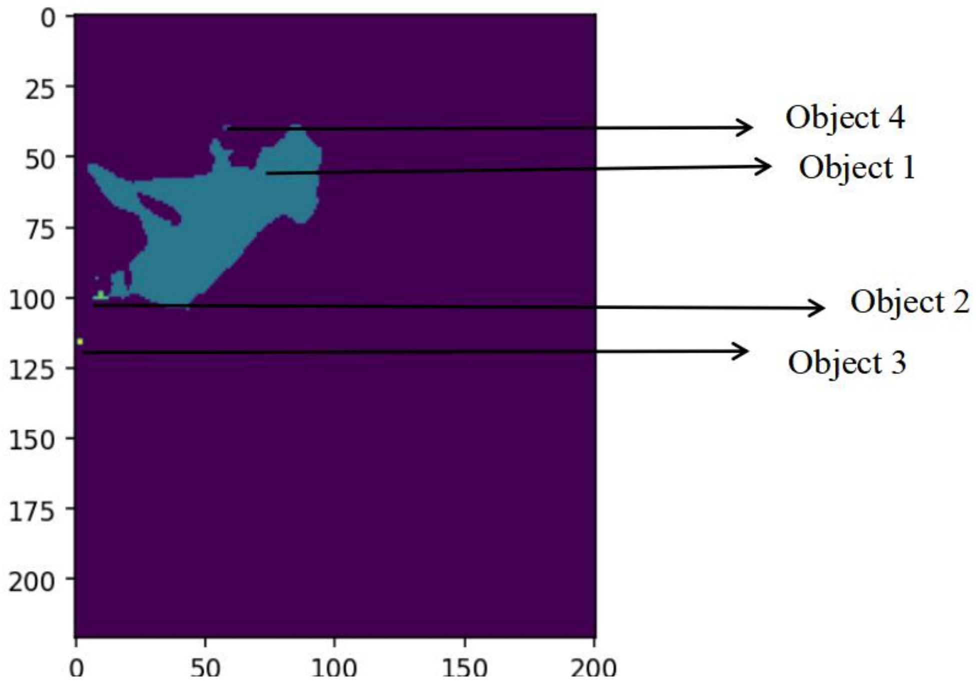


Figure 16: Labeled Atmospheric River Candidate Objects Identified from IVT Thresholding

v. AR Snapshots:

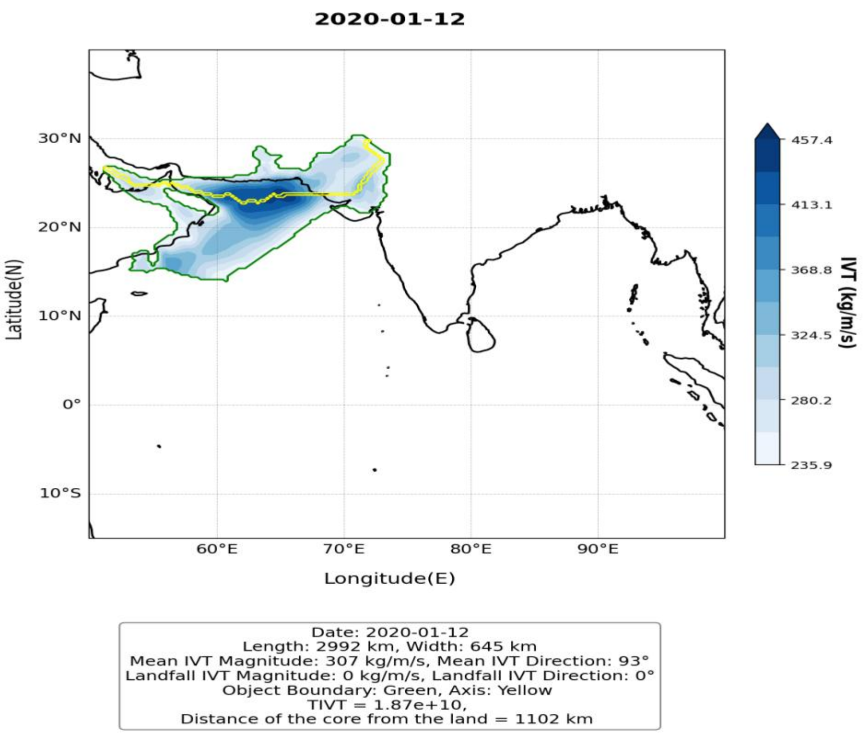


Figure 17(a)

vi. Wind plot:

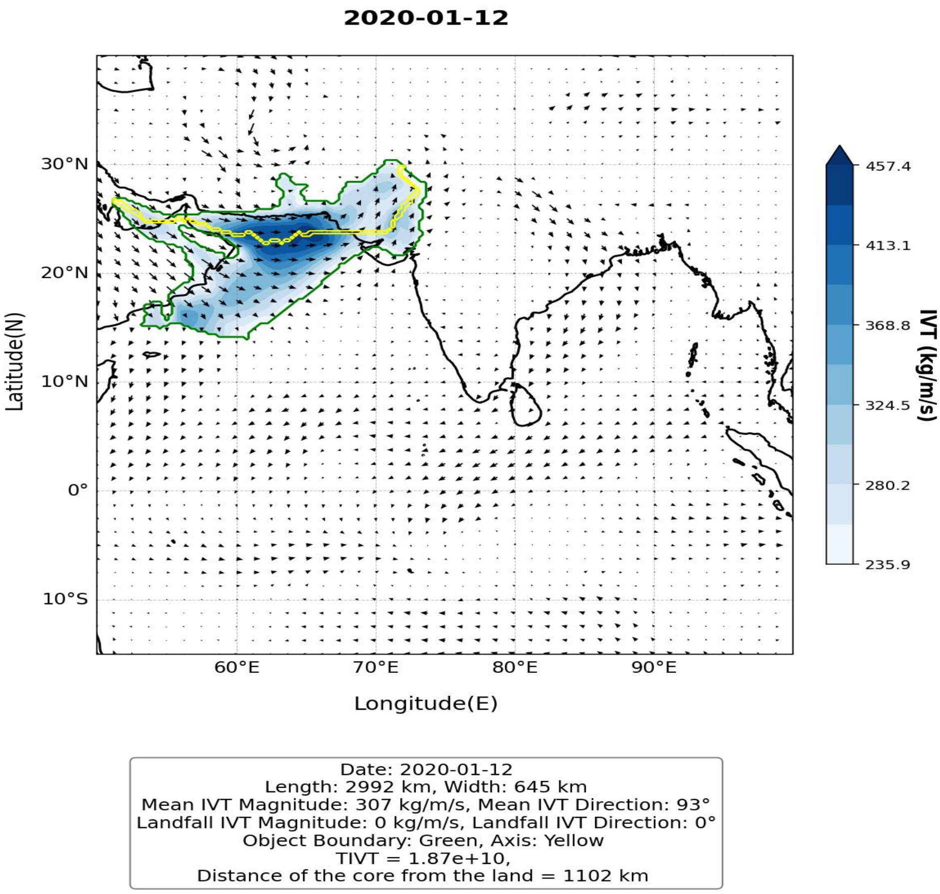


Figure 17(b)

Figure 17(a,b): Detected Atmospheric River over the Arabian Sea on 2024-06-20 Showing IVT Magnitude (shaded), Wind Vectors, Object Boundary (green), and Core Axis (yellow)

vii. INSAT data plot for TPW(Total Precipitable Water) :

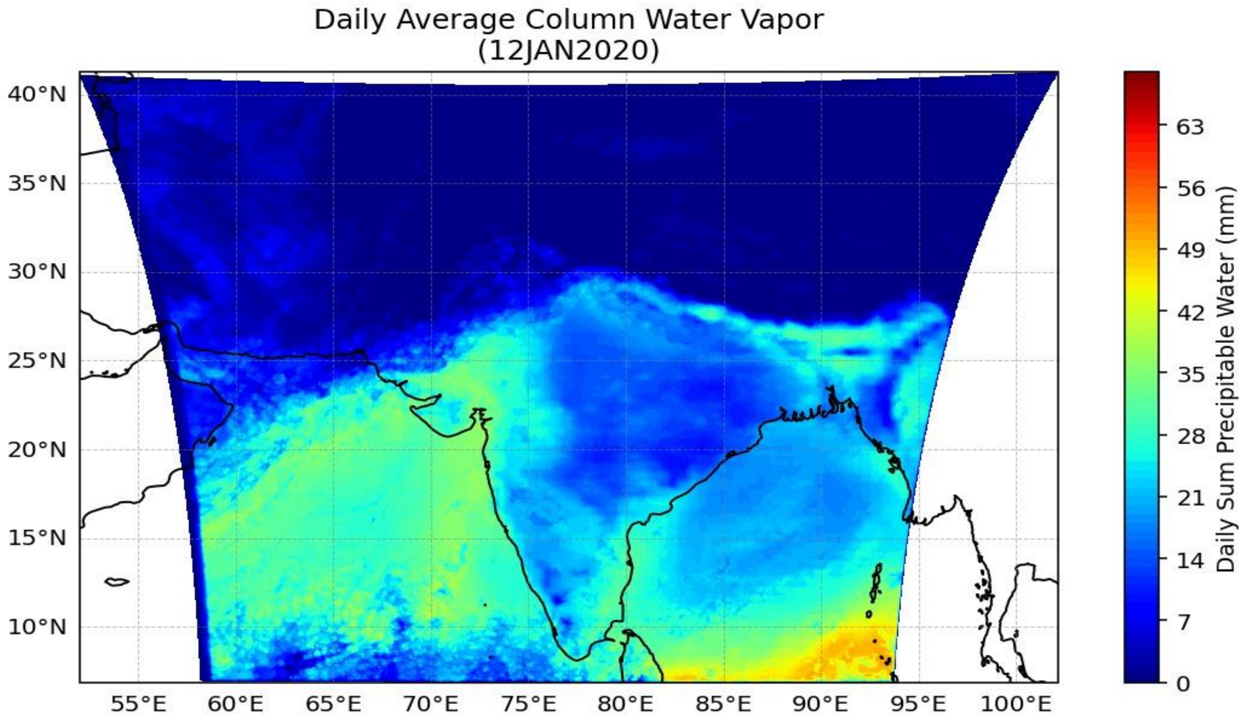


Figure 18: Daily Average Column Water Vapor over the Indian Subcontinent on 12 January 2020 from INSAT 3D Observations

viii. AMSR data plot for TPW(Total Precipitable Water) :

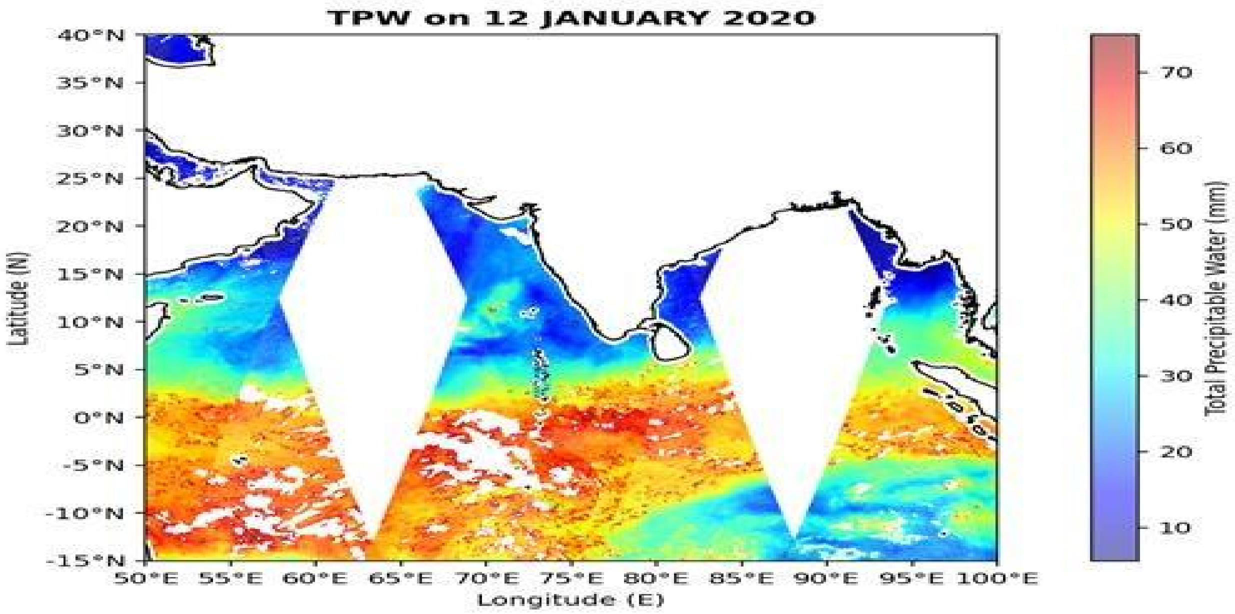


Figure 19: Daily Average Column Water Vapor over the Indian Subcontinent on 12 January 2020 from AMSR2 Observations

ix. Comparison between Total Precipitable Water (TPW)plot from ERA5 data and INSAT data :

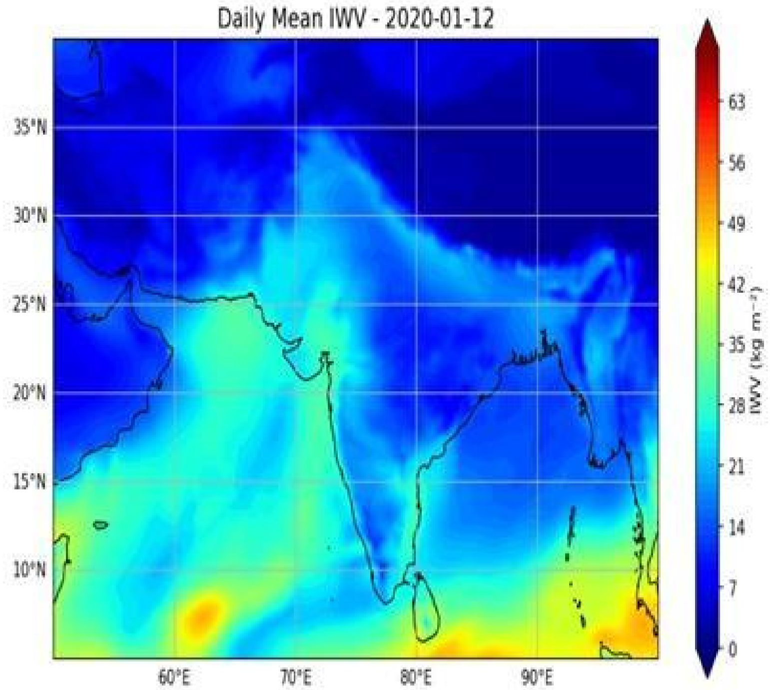


Fig 20(a)

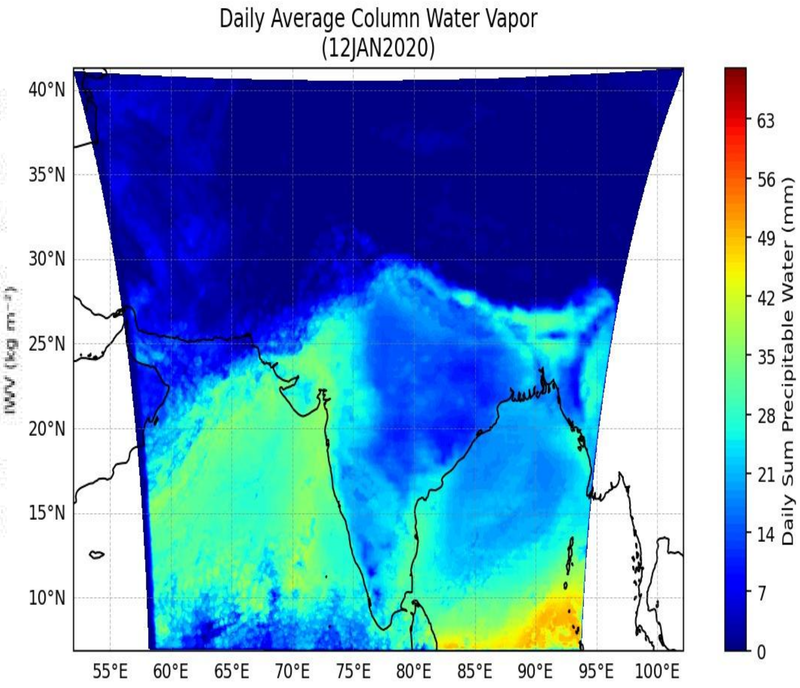


Fig 20(b)

Figure 20 Comparison of Daily Mean IWV from ERA5 and Satellite-Derived Column Water Vapor on 12 January 2020. (a) IWV from ERA5 reanalysis data, expressed in kg/m². It shows moisture accumulation over Indian subcontinent. (b) TPW retrieved from a satellite-based product (INSAT-3D), expressed in mm. This plot reveals localized zones of elevated water vapor.

x . GeoTiff Image of the detected AR:

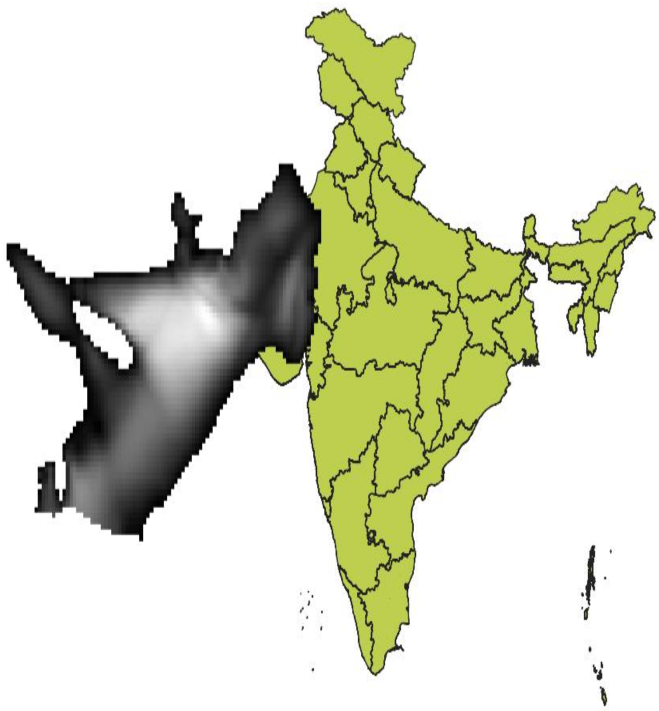


Figure 21:
GeoTIFF image generated using QGIS software, depicting an AR approaching the western coast of India. The AR is represented in grayscale, where lighter shades indicate stronger IVT. The state boundaries of India are overlaid in green to provide clear geographic context.

Conclusion

This study presents a comprehensive framework for detecting Atmospheric River(ARs)over the Indian subcontinent, spanning both the summer monsoon and winter seasons. ARs are shown to be critical drivers of large-scale moisture transport with wide-ranging impacts—from intense precipitation and flooding during the monsoon to persistent fog and low visibility during the winter months, particularly over the Indo-Gangetic plains.

To address the gap in regional AR detection, we developed a custom algorithm that integrates multiple data sources—including ERA5 reanalysis, GFS forecasts, and satellite-based TPW products from INSAT-3D and AMSR2. The algorithm incorporates key elements such as IVT calculation, adaptive thresholding, object-based segmentation, landfall and size filtering, and AR axis tracing. It is designed to be modular, scalable, and applicable in both retrospective and near-real-time settings.

The flexibility of the algorithm allows for detection across different seasons and synoptic conditions, capturing the full range of AR behavior over India. It effectively identifies high-impact AR events, including those contributing to floods during the monsoon and low-visibility conditions during winter. By using informed thresholds and geospatial filtering, the algorithm ensures both physical relevance and robustness across diverse meteorological scenarios.

As climate change continues to amplify moisture availability and atmospheric instability, ARs are expected to become more frequent and intense. This makes accurate detection and characterization of ARs vital for disaster preparedness, hydrological forecasting, and long-term climate planning. The algorithm developed in this study offers a valuable tool for improving early warning systems, enhancing situational awareness, and informing policy decisions related to water management and climate resilience in India.

Challenges and Limitations

Despite its overall effectiveness, the developed Atmospheric River (AR) detection algorithm has several limitations that may influence its performance under specific conditions. These limitations are outlined below:

1. Satellite Data Inconsistency:

Although satellite datasets(e.g., INSAT-3D/3DR and AMSR2)are used primarily to support and enhance the detection process, particularly in regions where reanalysis or forecast data may lack detail, they can present inconsistencies. Issues such as cloud cover, missing swaths, and instrument-related errors may affect data quality, especially over coastal or topographically complex regions, potentially reducing confidence in supplementary TPW-based validation or visualization.

2. Sensitivity to Threshold Selection:

The algorithm applies fixed IVT thresholds or percentile-based criteria (e.g., the 85th percentile) to identify high moisture transport regions. However, these thresholds are not universally applicable across different climatic conditions, regions, or seasons. For instance, a static threshold may lead to under-detection in drier months or over-detection during the monsoon peak, affecting the consistency and accuracy of AR identification.

Future Work and Improvement

1. Data Fusion Using ML:
Develop ML-based data fusion models that can intelligently combine satellite data (INSAT-3D/3DR, AMSR2) with reanalysis and forecast datasets. These models can learn to handle data gaps, cloud contamination, and instrument errors to produce more consistent and reliable moisture transport estimates, especially over complex terrain and coastal regions.

2. Adaptive Thresholding via ML Classification:
Replace static IVT thresholding with supervised ML classifiers (e.g., Random Forest, Neural Networks) trained on labeled AR events across different regions and seasons. This approach can adaptively adjust detection criteria, improving sensitivity in drier months and specificity during monsoon peaks.

References

1. Guan, B. & Waliser, D. E. Tracking atmospheric rivers globally: spatial distributions and temporal evolution of life cycle characteristics. *J. Geophys. Res. Atmos.* 124, 12523–12552 (2019).

2. Banerji, U. S., Arulbalaji, P., & Padmalal, D. Holocene climate variability and Indian Summer Monsoon: an overview. *SAGE J.* https://doi.org/10.1177/0959683619895577 (2020).

3. Dettinger, M. D., Ralph, F. M., Das, T., Neiman, P. J. & Cayan, D. R. Atmospheric rivers, floods and the water resources of California. *Water* **3**, 445–478 (2011).

4. https://www.noaa.gov/stories/what-are-atmospheric-rivers

5. Dhana Lakshmi, D. & Satyanarayana, A. N. V. Influence of atmospheric rivers in the occurrence of devastating flood associated with extreme precipitation events over Chennai using different reanalysis data sets. *Atmos. Res.* **215**, 12–36 (2019).

6. https://www.nature.com/articles/s43247-023-00775-9

7. Pan, M. & Lu, M. A novel atmospheric river identification algorithm. *Water Resour. Res.* **55**, 6069–6087 (2019).

8. Guan, B. & Waliser, D. E. Detection of atmospheric rivers: evaluation and application of an algorithm for global studies. *J. Geophys. Res. Atmos.* **120**, 12514–12535 (2015).

9. https://github.com/daemonkennett/ar_detection/blob/main/example_output.png