Flood Assessment in India (Western Ghats)

Group 6:

Neel Shah (23B4244)
Pearsh Gumasta (23B4232)
Akshat Karkar (23B4229)

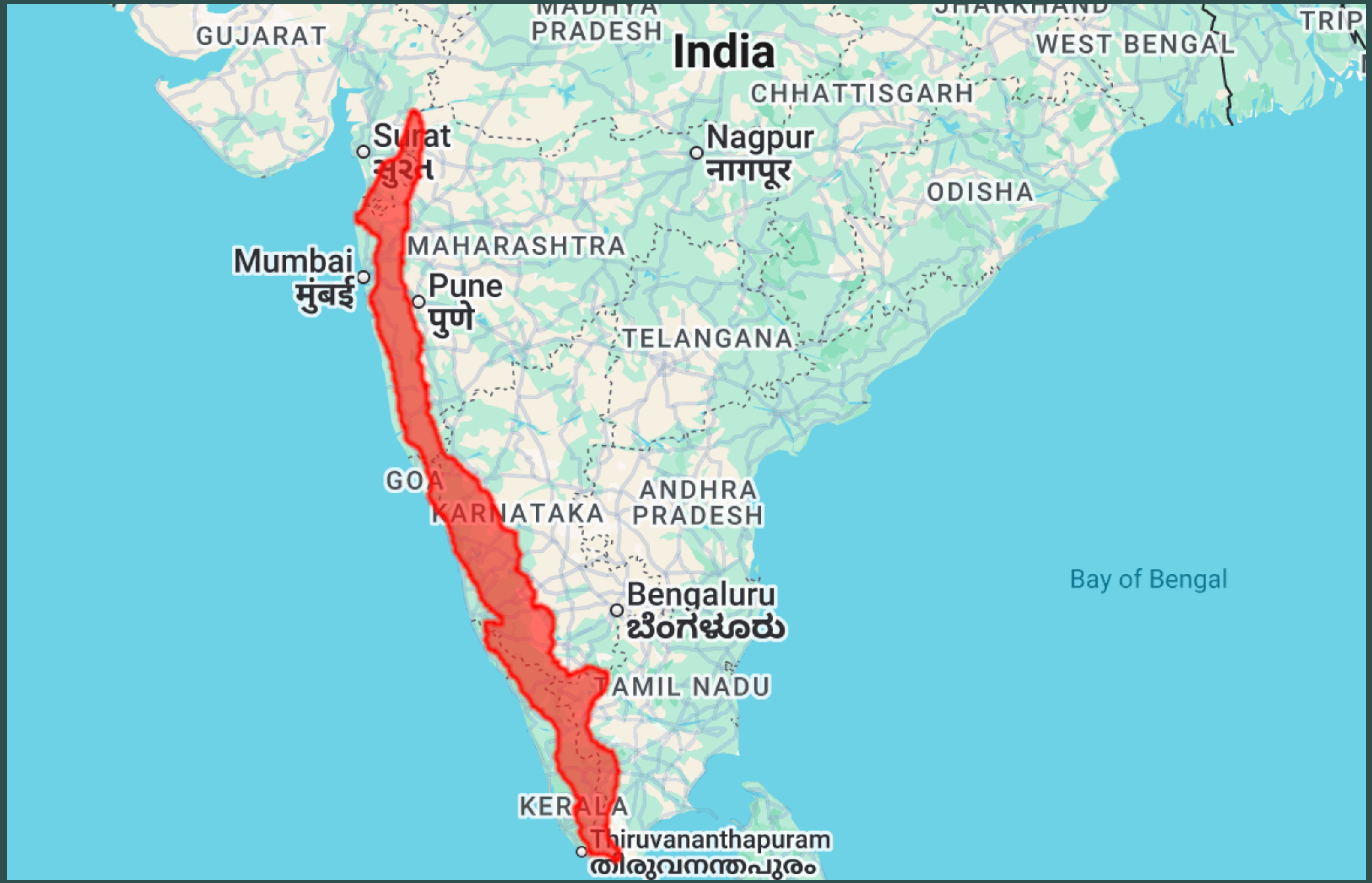
Why Western Ghats?

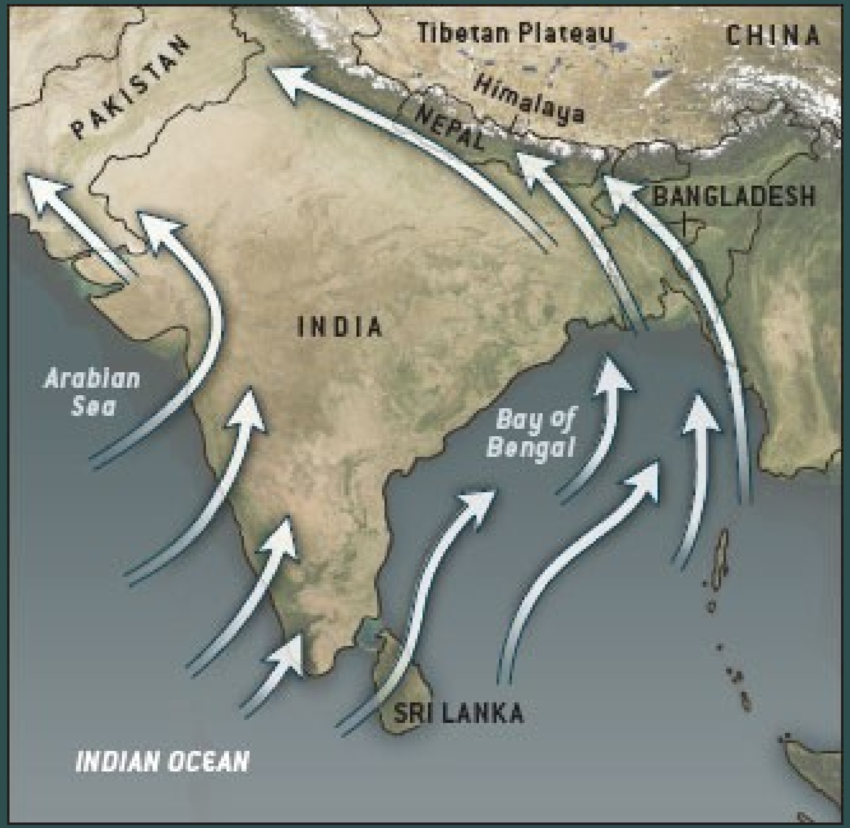
- Unique Topography: A **1,600 km** long mountain barrier. (Area = 125,541.10 sq. km)

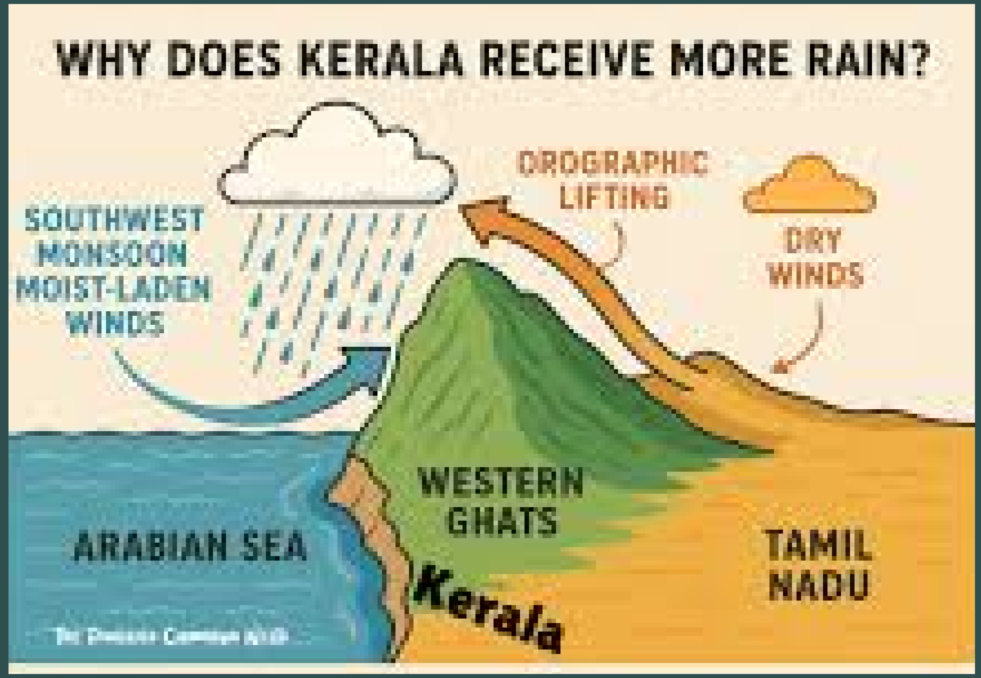
- Monsoon Interaction: The first land barrier to intercept the moisture-laden **Southwest Monsoon**.

- The Result (**Orographic Rainfall**): This interaction creates one of the world's most intense rainfall regimes, leading to a high frequency of flash **floods** and **landslides**.

- Critical Risk Zone: A **UNESCO World Heritage site** with high population density in its coastal plains and foothills.







Scope of Study:

Spatial Reference Frameworks

Spatial Data Models

Data Acquisition Models

→ **Spatial Data Analysis** → **Geo Visualisation**

Kerala 2018 Flood Case study!

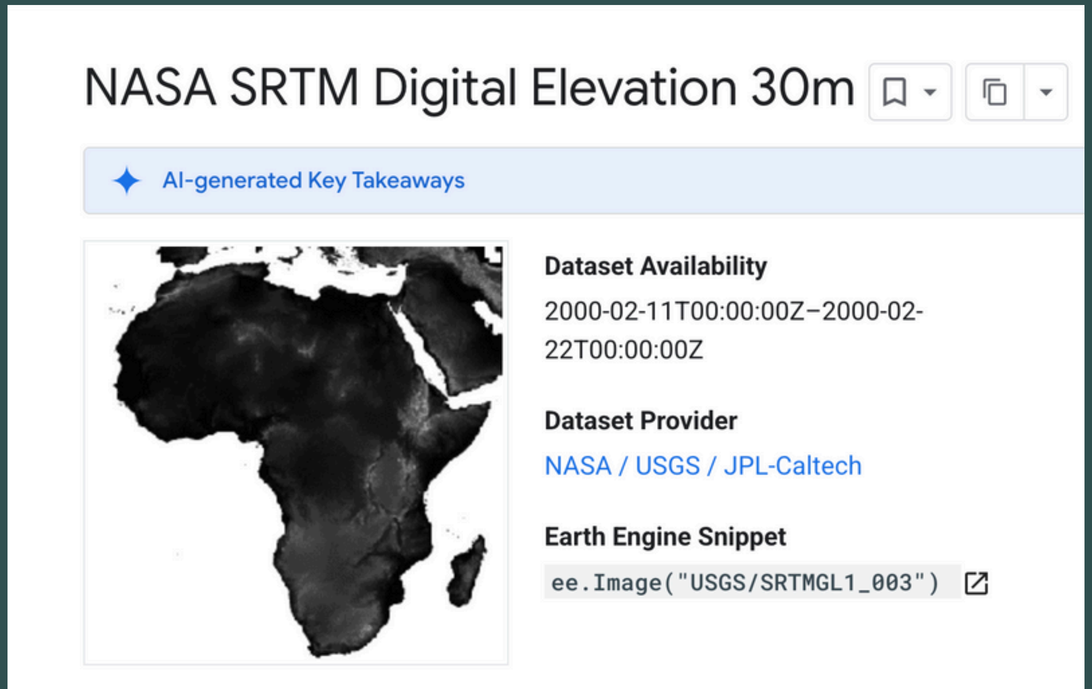
Enabled Via Google Earth Engine

Various Datasets (vector & raster) are used to model the behavior and assess the changes over time (time series) and static datasets are used.

Complex Mathemically operations are performed on the datasets to land to conclusion. Discussed in detail in the slides later.

We plot different layers of map over the western ghats to demonstrate the flood prone areas with hotspot & cluster analysis.

Factors Affecting Floods:
(Part 1/2) The Physical Landscape



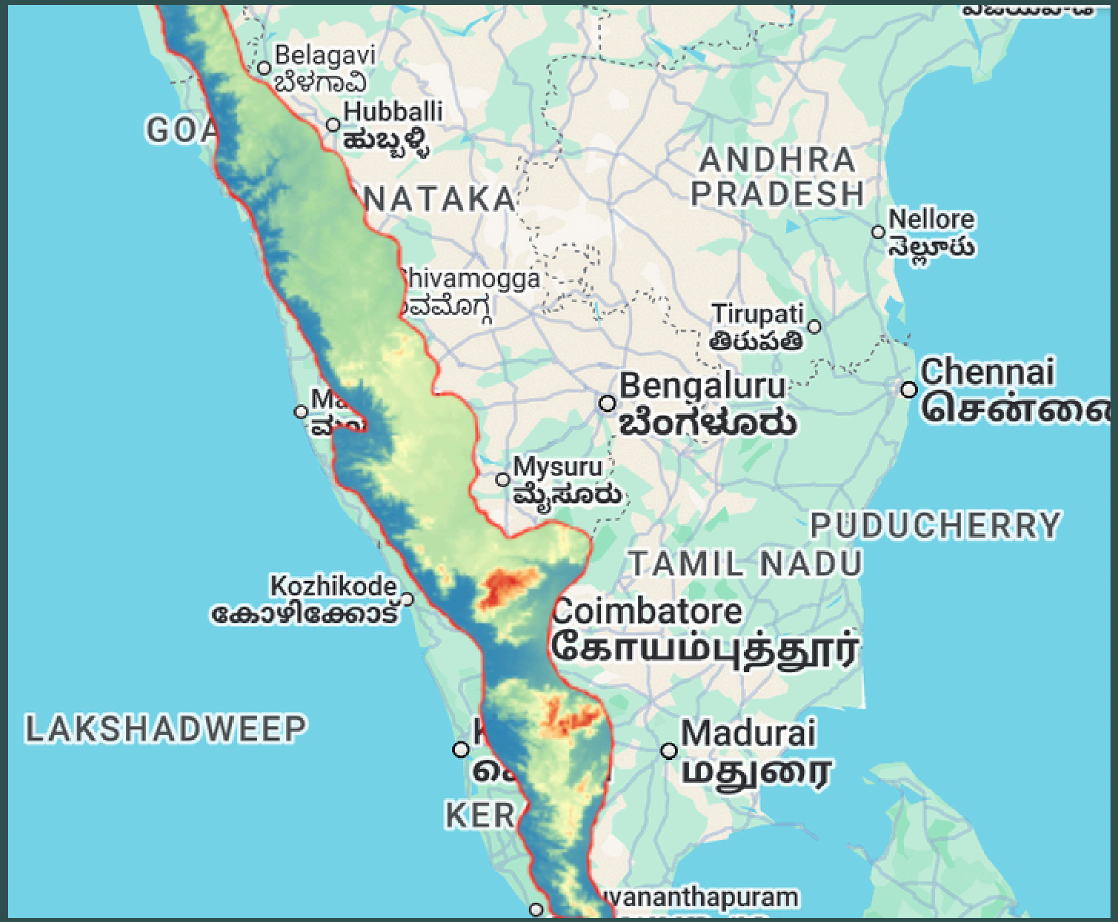
Source Data: 30-meter SRTM Digital Elevation Model (DEM)

Elevation	**Slope**	**Aspect**
Water flows downhill and pools at low elevations.	Flat areas (low slope) cannot drain quickly leading to water accumulation.	The compass direction a slope faces. This determines exposure to the monsoon.

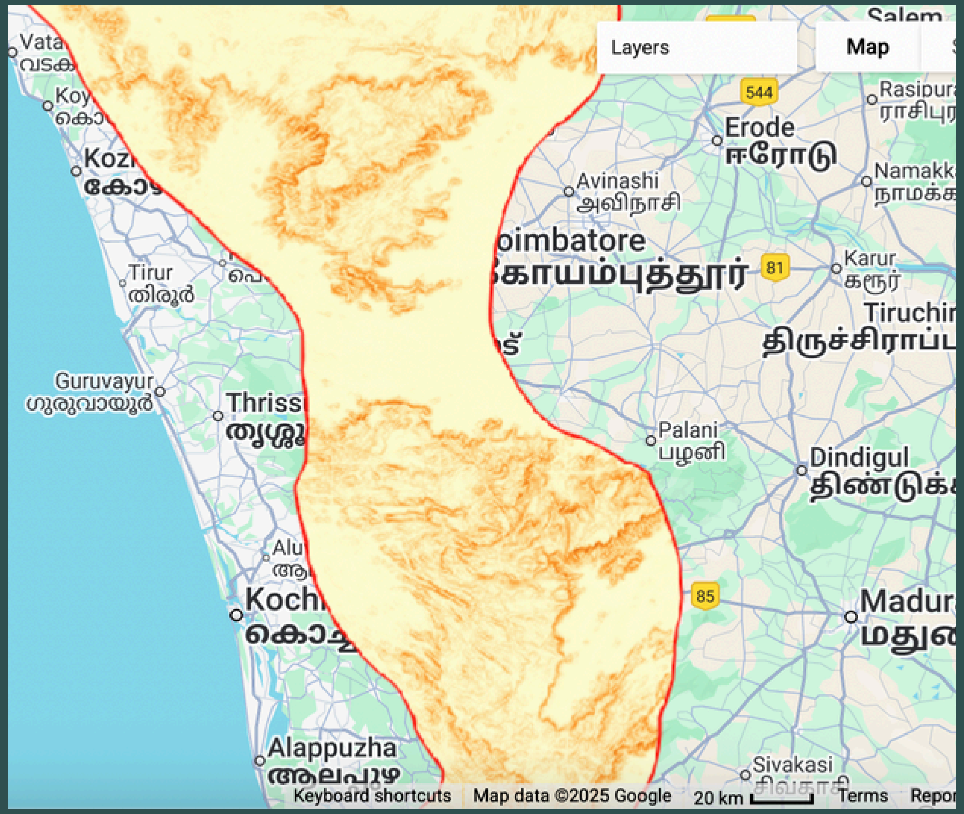
Why this particular Resolution?

- **90m is too coarse! We would miss the small details on narrow valleys which are typical causes of floods.**
- **Ultra-High-Resolution 1-meter (LiDAR) data would be computationally impossible to process for a 1600km region.**

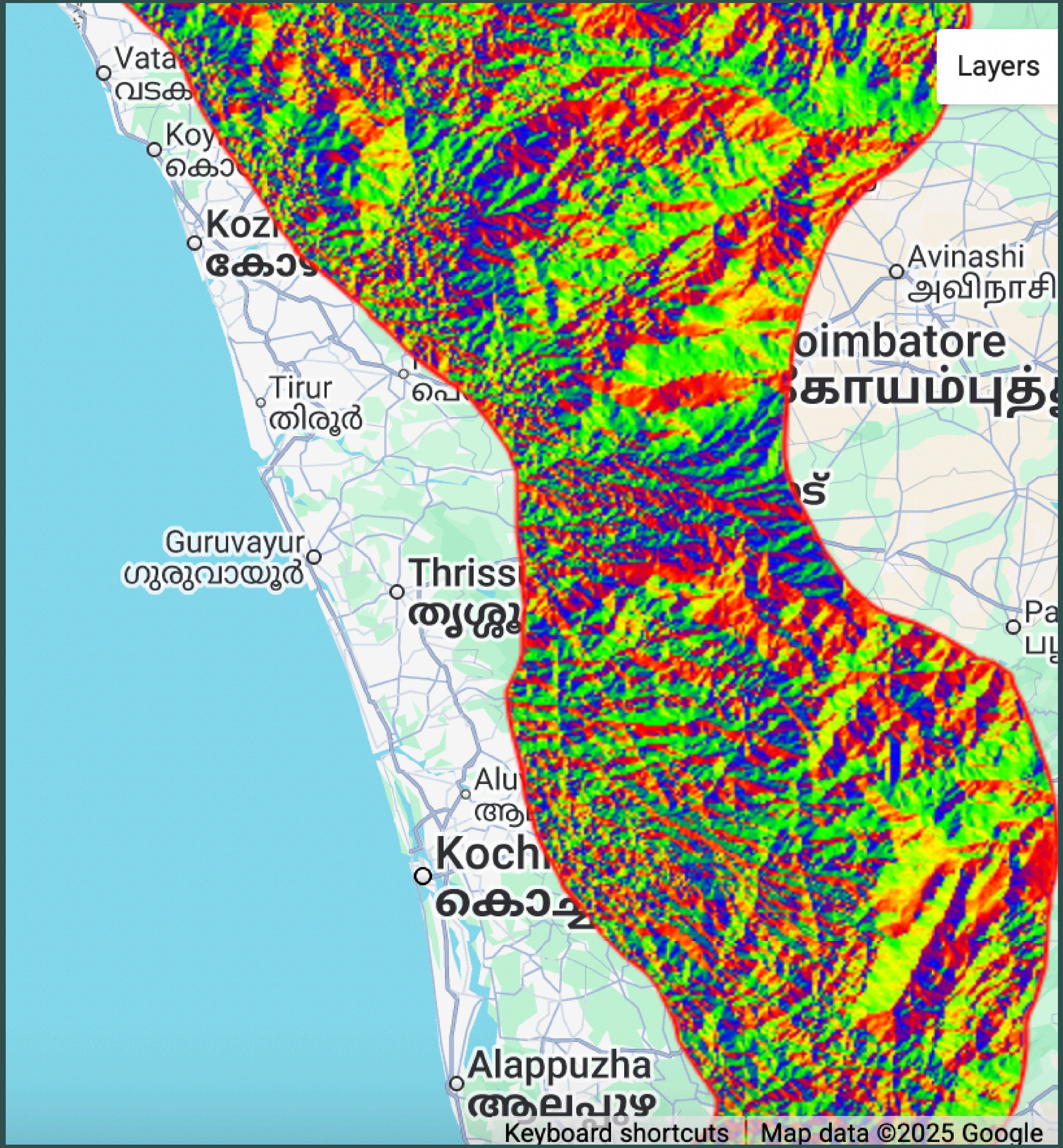
Attribute 1/6 :**Elevation:**



Attribute 2/6 :**Slope:**



Attribute 3/6 :**Aspect:**



Legend

Elevation (m)
- 0 (Low)
- 1250 (Mid)
- 2500 (High)

Slope (°)
- 0 (Flat)
- 30 (Mid)
- 60 (Steep)

Aspect (Direction)
- North
- East
- South
- West

(Part 2/3) Remote Sensing for Vegetation: **NDVI**

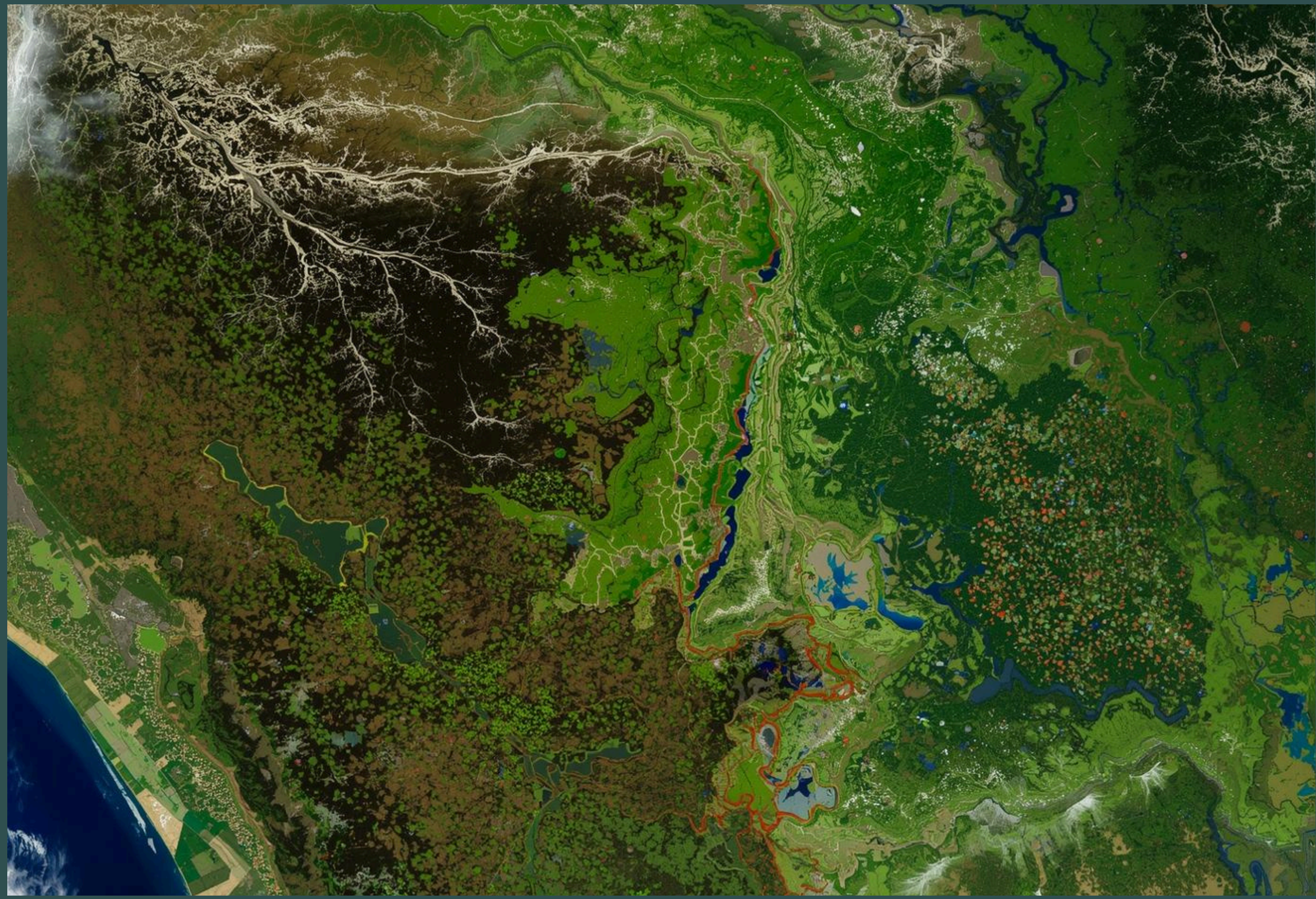
Source: Sentinel-2 Satellite (10m Resolution)

Why?

- **High NDVI (Dense Forest): Acts like a "sponge." Intercepts rainfall and slows down surface runoff. (Lowers flood risk)**
- **Low NDVI (Bare Soil / Urban): Has very low absorption. Leads to rapid and high-volume runoff. (Increases flood risk)**

Why 10m resolution?

- **Detail is Critical: Flood risk is often hyper-local. A 30m sensor might "average" a small, bare field (high risk) with the dense forest around it, making the risk invisible.**
- **10m "Sees" the Risk: Our 10m data from Sentinel-2 can "see" these small, critical areas like a bare patch of ground or a new clearing that produce rapid runoff. This creates a much more accurate model of the "sponge" effect.**



$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Method: We processed a cloud-free composite of monsoon-season imagery to get a representative map of vegetation health.

(Part 3/3) Hydrological Parameters:

The final two factors model the "event" itself “The water”

Max 24-Hour Rainfall

Source: CHIRPS Daily (a 10-year satellite & station dataset).

- What it Measures: The worst-case single-day rainfall for each pixel.

- Why?

Floods are threshold-based events. They are caused by extreme intensity (e.g., 300mm in one day), not by high averages. This metric is a direct proxy for a flood-triggering downpour.

- How:

We applied a .max() reducer to a 10-year time-series of daily rainfall images.

Distance to Rivers

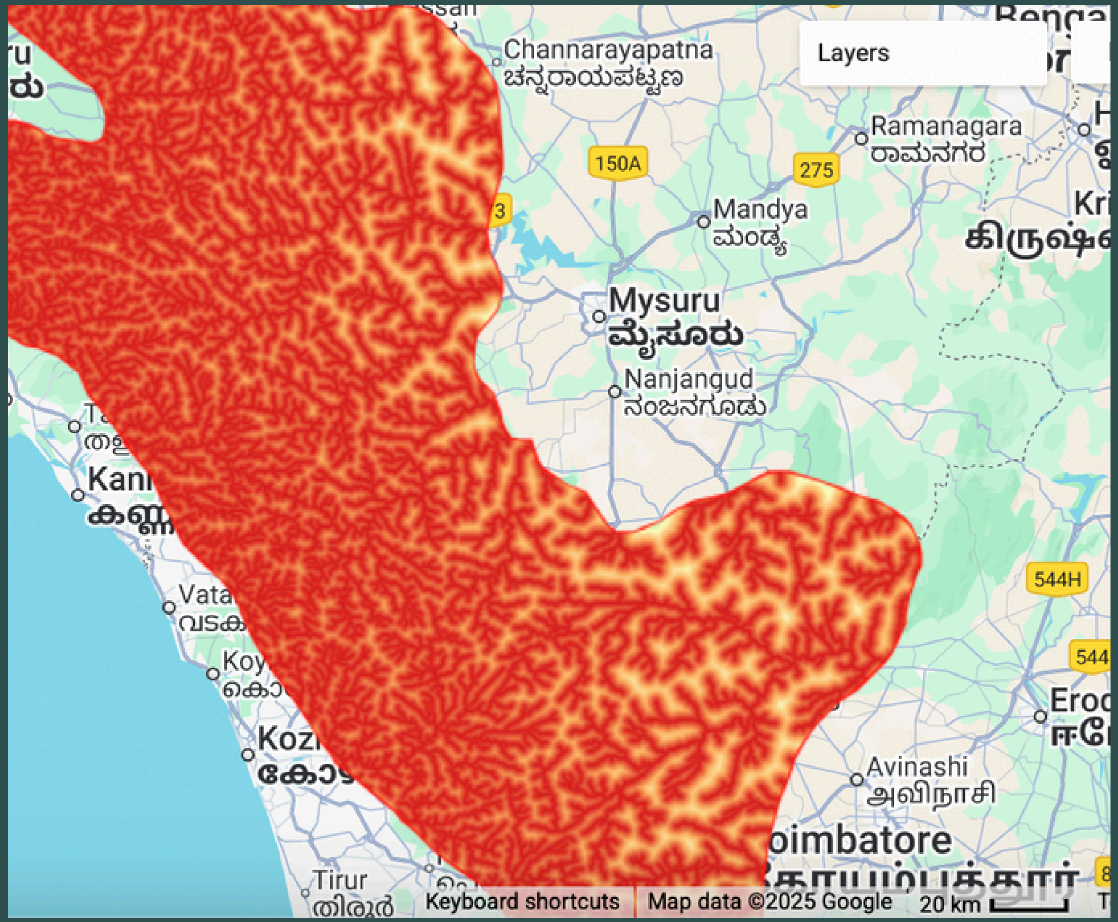
Source: WWF HydroSHEDS (a global vector dataset of river lines).

- What it Measures: A continuous raster map of the distance (in meters) from any pixel to the nearest river.
- Why?

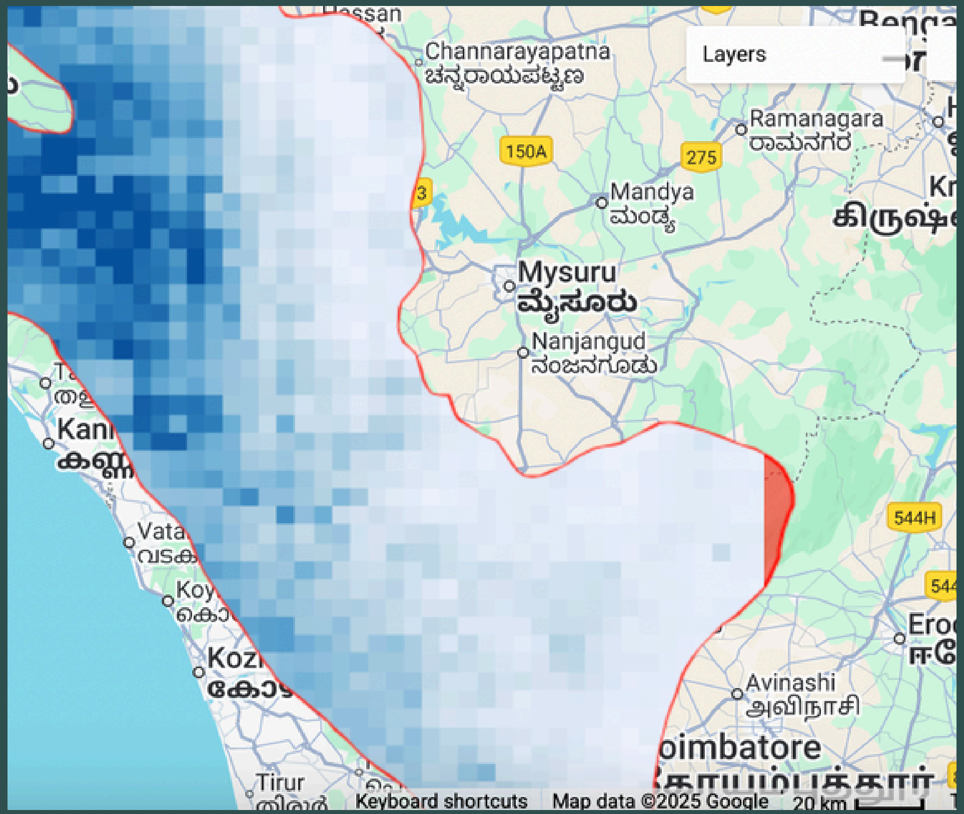
This models a specific and major risk: fluvial (river) flooding. Areas in the immediate floodplain (low distance) are the first to be inundated when a river overflows its banks.

How: A classic Proximity Analysis—a key GIS operation that converts vector lines into a risk based raster surface.

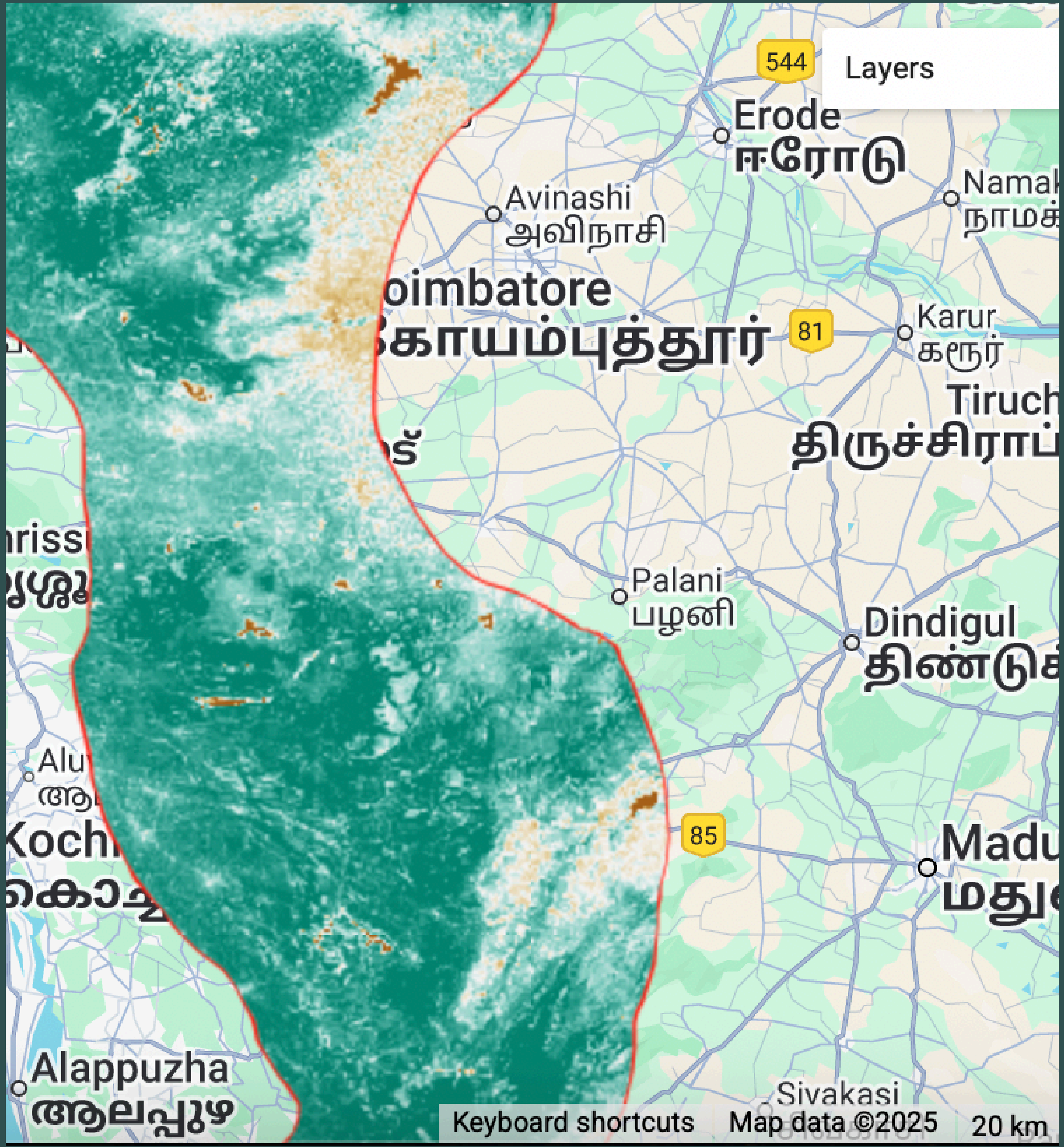
Attribute 4/6 **Distance from river:**



Attribute 5/6 **:Max 24 hr rainfall:**



Attribute 6/6 **:NDVI:**



Legend (Dynamic Layers)

NDVI (Vegetation)

- 0.0 (Bare)
- 0.4 (Sparse)
- 0.8 (Dense)

Max 24h Rainfall (mm)

- 50 (Low)
- 225 (Mid)
- 400 (High)

Distance to River (m)

- 0 (Close)
- 5,000 (Mid)
- 10,000 (Far)

Methodology - Combining the Factors

The **Problem**: We have 6 factors in different units (meters, degrees, mm, etc.). How do we combine them to create a model to assess floods and map them?

Step 1: **Normalization**

All 6 layers are mathematically normalised between 0 to 1, into a “Vulnerability Index”

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Direct Relation:

Rainfall

Normalise values equally between 0 and 1 from minimum to maximum respectively

Indirect Relation:

Slope, Distance from river, Elevation, NDVI

Normalise values equally between 0 and 1 from minimum to maximum respectively

1 - Normalised value

Complex Relation:

Aspect

1.0 if ghats face west

0.5 if ghats face north or south

0 for ghats facing east

remaining values distributed equally in degrees between 0 and 1

Methodology - Principal Component Analysis

Why **PCA**?
PCA is the best unsupervised learning technique for finding the axis of maximum variation.

- The **process**:

- Resampled **5000** random samples from the region selected and ran PCA to analyse the variations in the dataset.
- Tried with 100,000; 50,000 and 10,000 samples but GEE was rejecting the code due to high computation

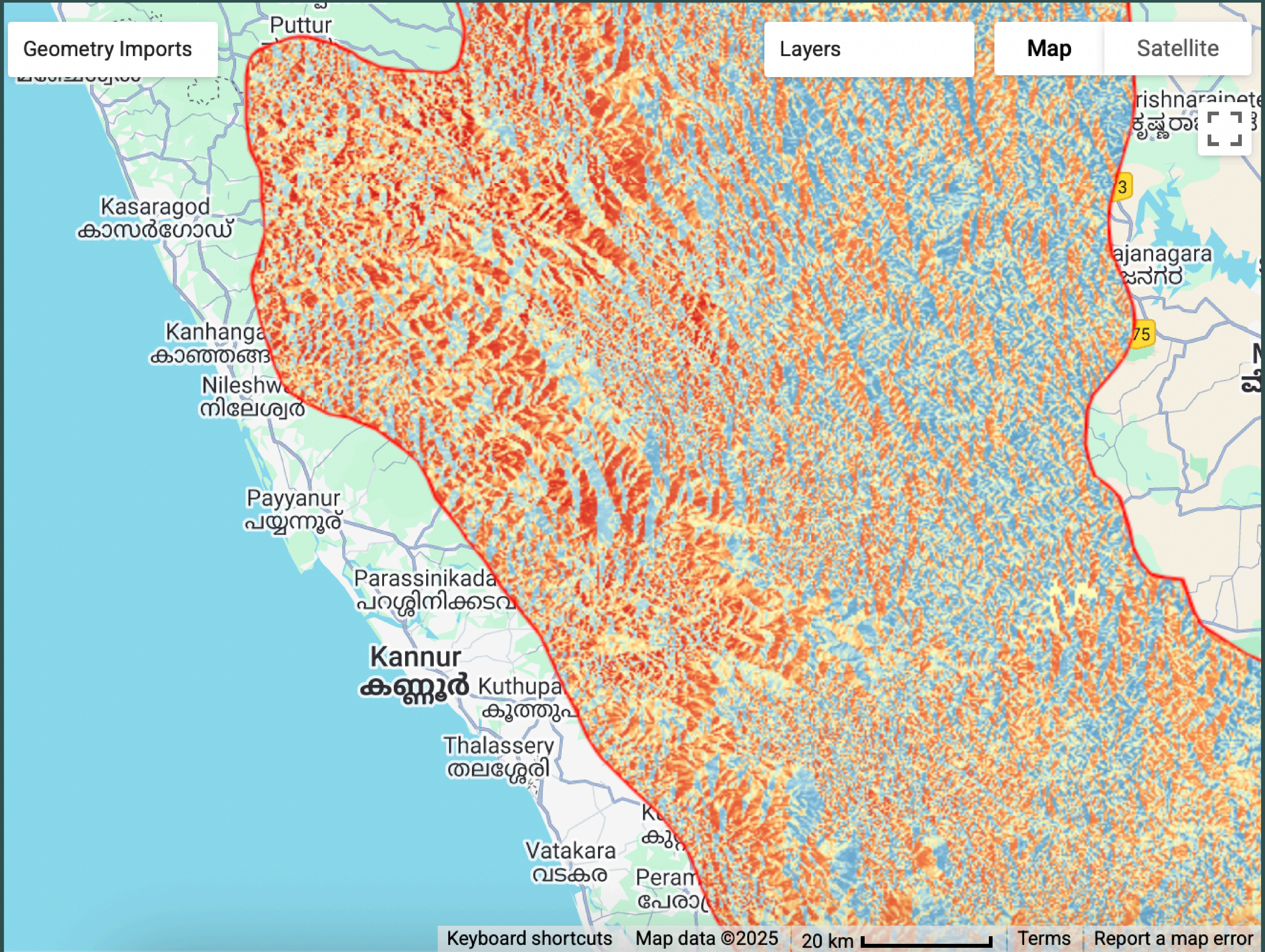

```
PCA-Derived Weights (FINAL):                                JSON
▾Object (6 properties)                                      JSON
   aspect: 0.7236845299402813
   distance: 0.0448051719622337
   elevation: 0.06621879694712195
   ndvi: 0.023674261632087978
   rainfall: 0.13960628722944346
   slope: 0.002010952288831581
```

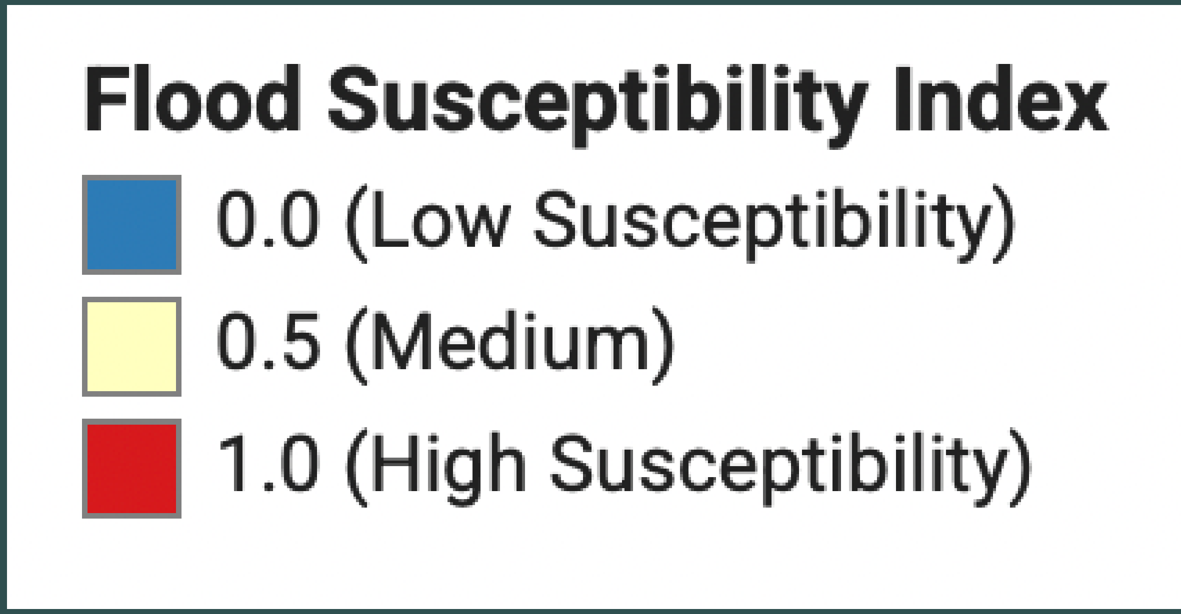

Aspect showed the highest 72.37% variation in the data

We will use these coefficients for modelling "Flood Susceptibility Index (FSI)"

FSI= (0.72(aspect) + 0.045(distance)+0.06(elevation) +0.024(ndvi)+0.14(rainfall+ 0.002(slope))/2

Flood Susceptibility Index (FSI)





Validation (2018 Kerala Floods)

Source:Sentinel-1 SAR GRD, 10m resolution

Why SAR Specifically? Traditional validation approaches fail in monsoon regions:

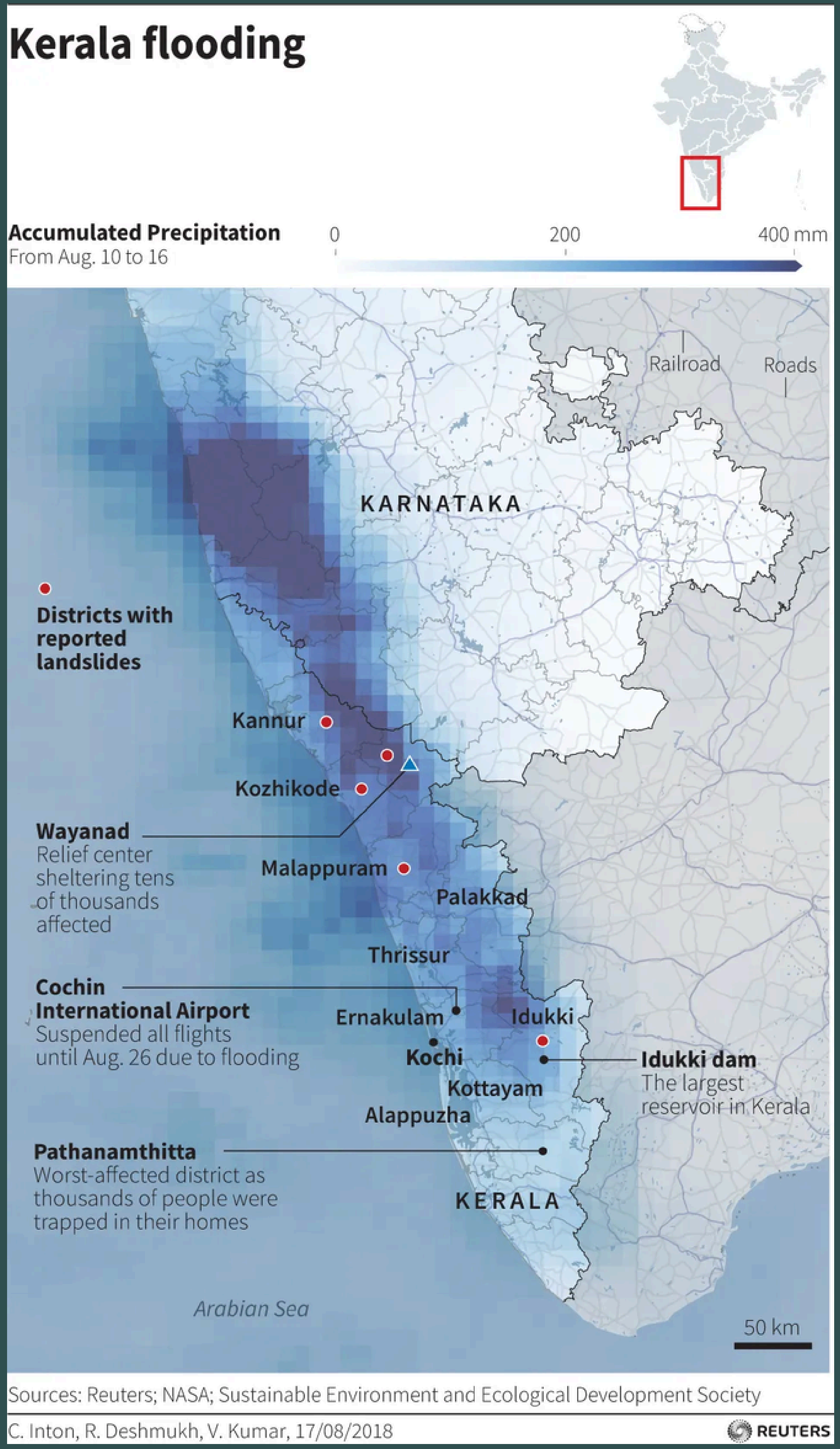
- **Optical satellites (Landsat, Sentinel-2) can't see through clouds during floods**
- **Ground surveys are dangerous, expensive, and spatially limited**
- **Historical flood records are often incomplete or poorly georeferenced**

Synthetic Aperture Radar (SAR) solves this because:

- **All-weather capability: Microwave signals penetrate clouds and rain**
- **Day/night operation: Active sensor with its own energy source**
- **Water sensitivity: Smooth water surfaces create distinctive radar signatures**

Timeperiod chosen:

Before Period: May 1-31, 2018 (Dry season) After Period: August 18-25, 2018 (Peak Kerala floods)



Methodology - Ratio-Based Change Detection

How?

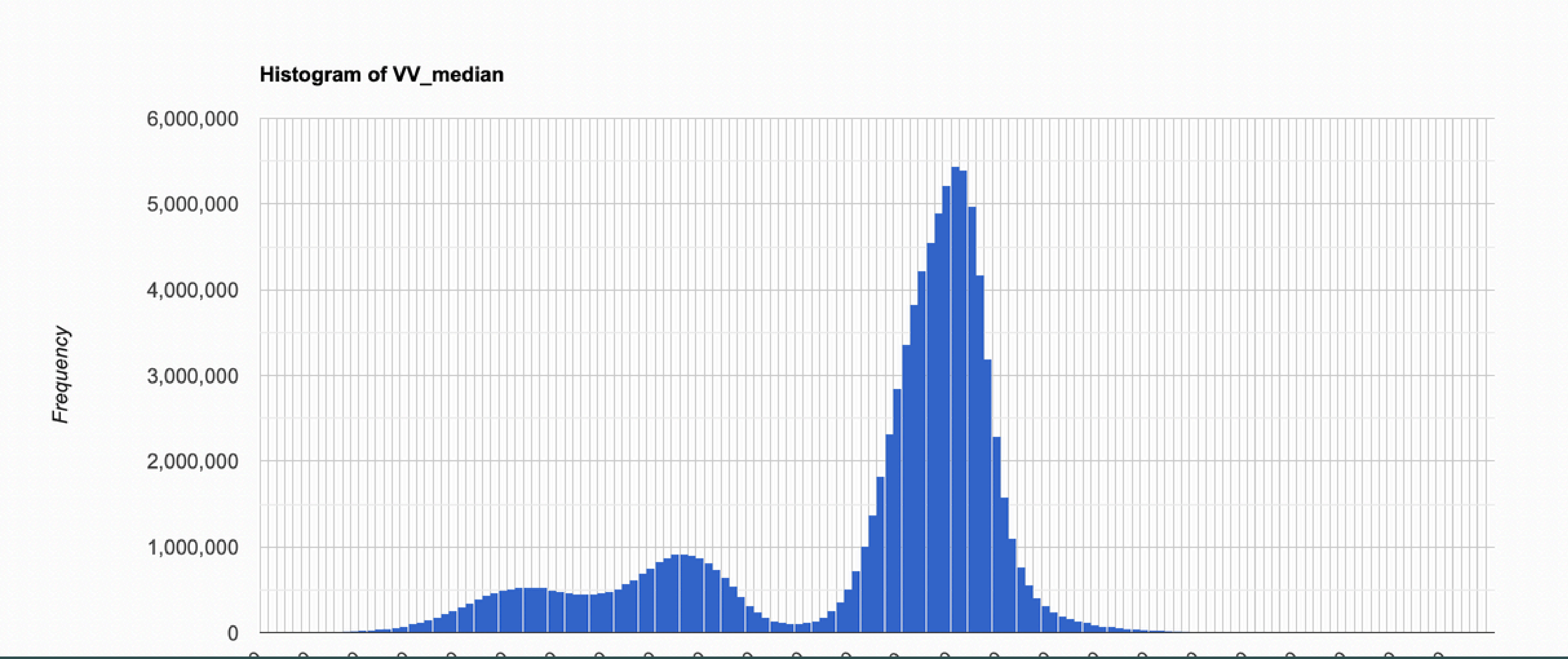
We took snapshots from May 1-31 and other in August 18-25 (peak floods),2018, took their median values.

Land generally reflect high amount of light back so it has higher intensity (-10 - -20 db) and Flood prone areas reflect less so less intensity(-20 - -25db)

We mapped the ratio of **before to after** for **SAR Backscatter** values of spatial resolution.

Interpretation:

- Ratio > 1.0 = Pixel got darker
- Ratio > 1.5 = Pixel got significantly darker
- Ratio < 1.0 = Pixel got brighter (impossible for floods)

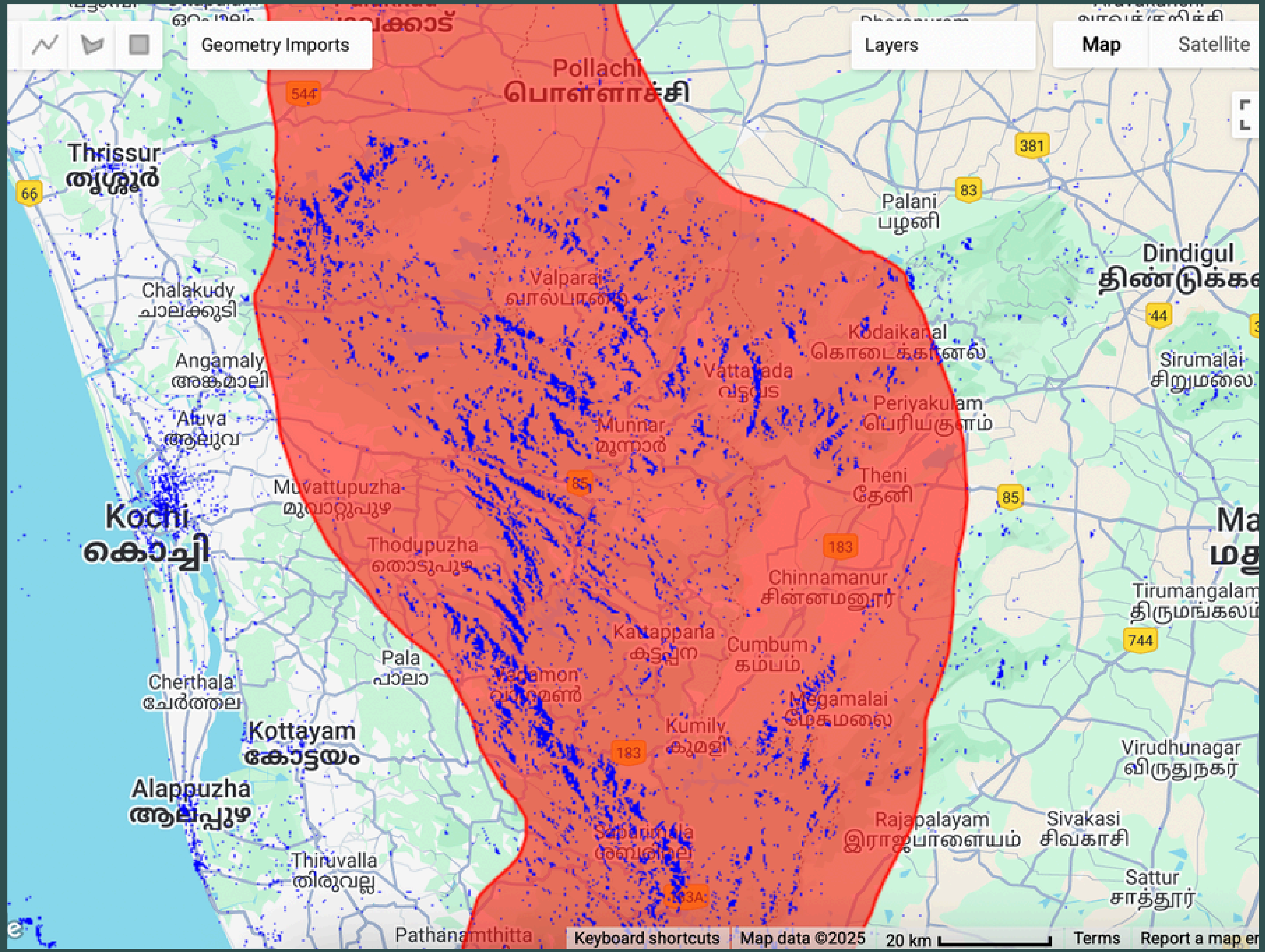


The data from SAR looks like a bimodal distribution of frequency vs intensity. This is for May 2018.

Left peak indicates water bodies(-21db) and right peak indicates land surfaces(-16db)

$$\sigma^0_{linear} = 10^{(dB/10)}$$

Likely flooded areas




```
Object (1 property)
  groups: List (2 elements)
    0: Object (2 properties)
      flood_class: 0
      mean: 0.47569514478575775
    1: Object (2 properties)
      flood_class: 1
      mean: 0.7309519974067753
```


Blue dots represent ratio >1.5 so it is likely to be flooded after the monsoon season in august

Statistical Validation:

Before : FSI (30m) After :FSI (300m)	Flooded areas (10m) Flooded Areas(300m)
Average of all rasters	Mode of all rasters
For every flooded area we computed the average of FSI values	Average FSI in Likely Flooded areas was **55%** higher than in Likely non flooded areas

Environmental Planning: Hotspot Analysis

To find statistically significant clusters of high-risk pixels, we applied the Local Z statistic to the FSI values.

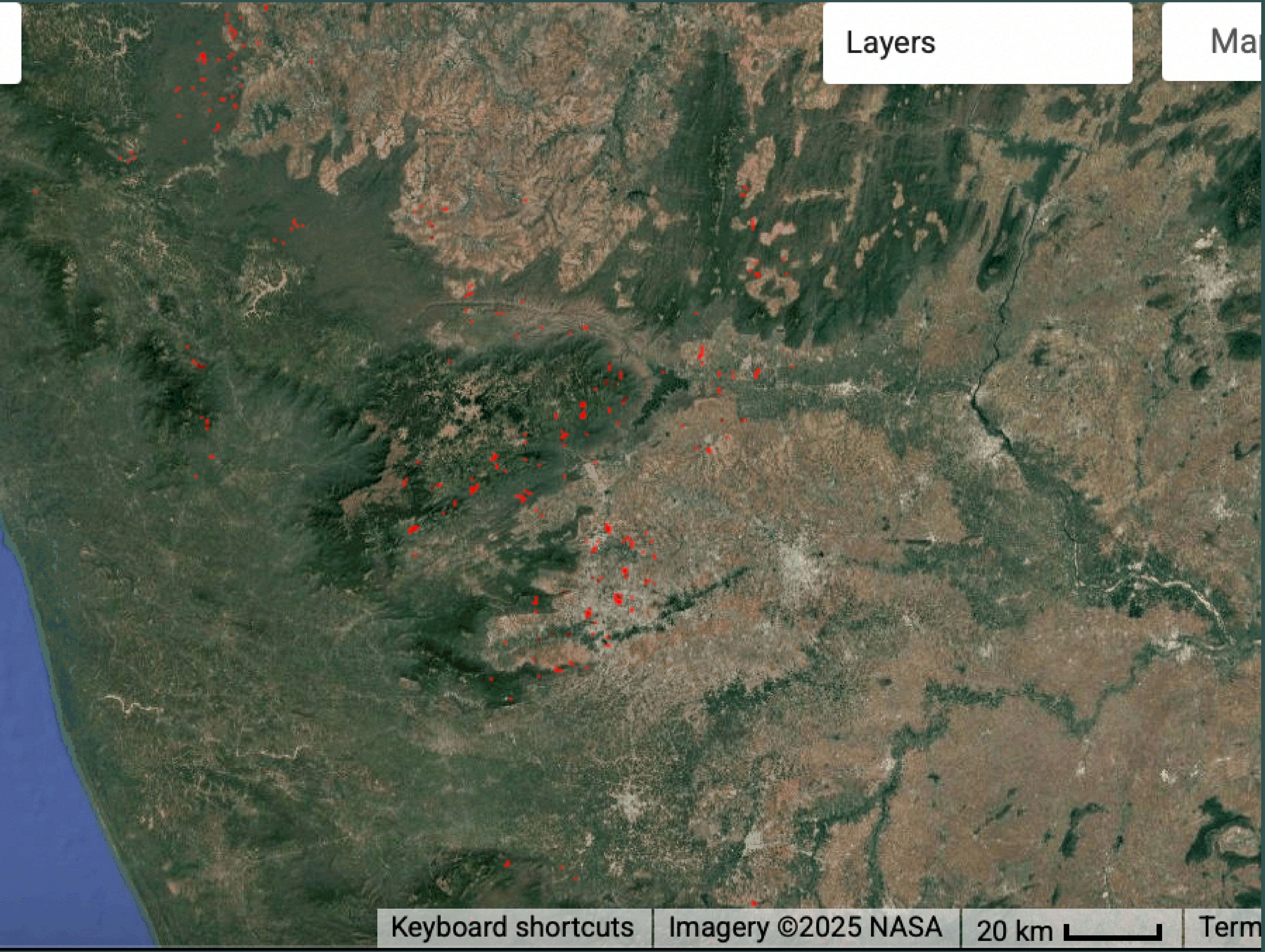
Methodology:

- Taking rasters in a 5km radius neighbourhood.
- That is 5000/300= 16.6 rasters in every direction around the focus.
- For every central pixel, GEE looks at all FSI values within that circle a local neighborhood of around 800 - 900 pixels.

Compute the Local Z statistic:

For every raster A:

$$Z_A = \frac{FSI_A - \mu_{local}}{\sigma_{local}}$$



These red clusters show hotspots with >99% significance level

Conclusion:

The first principles of data acquisition on primary attributes like slope, elevation, aspect played a crucial & significant role in determining FSI.
The FSI was validated against Kerala 2018 floods, and it proved to be more accurate in determining Flood Likely areas.
Hotspot analysis helped us determing highly prone flood areas.

References:

Begged for datasets (most relevant ones) with justifcation for resolutions & complete understanding of SAR from GPT Tools.

Borrow the idea of mapping Primary attributes from following research paper:
https://www.mdpi.com/2072-4292/11/13/1589

Stole the codes written in JavaScript by Gemini 2.5 Pro and Claude Sonnet 4.5