STEP 0 – Prioritising a geographical region

Notebook: country-ranking.piynb

STEP 0 – Prioritising a geographical region. (Notebook: country-ranking.piynb)

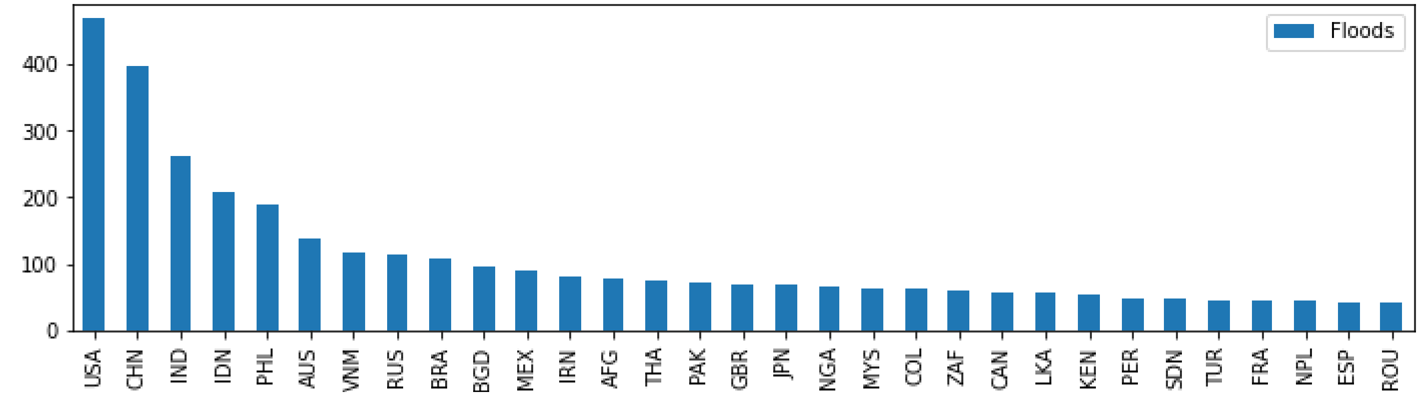
Goal:

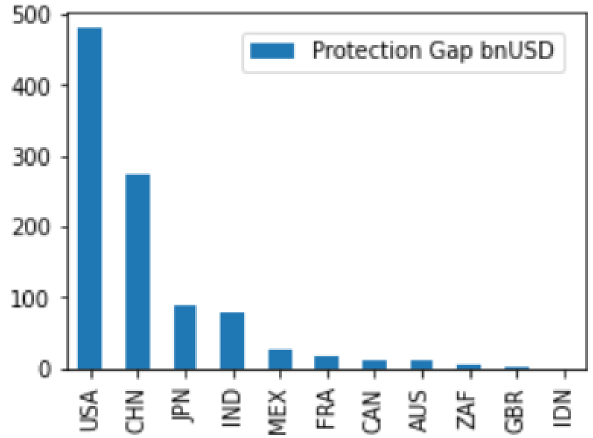
Prioritize a country in terms of

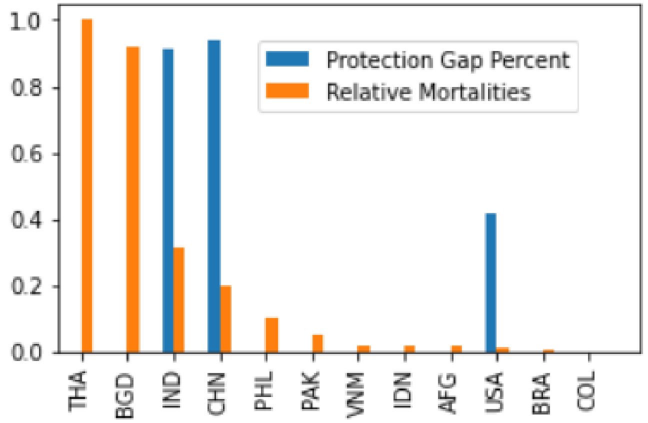
- highest historical flood frequency
- significant insurance gaps.
- high commercial potential for a flood extent product

Method

- Gather data on flood frequency, insurance gaps and commercial potential per country
- Rank countries by each of these in turn
- If the ranking varies for different metrics, form an overall priority score from a weighted sum of each (normalized) metric (← this was deemed unnecessary)
- If not, choose a country that ranks high in most metrics and add any further justifications







Results

Focus on India

Although India has a low GDP per capita, it has a need for support in natural disaster mitigation:

- 3rd highest flood rate in the reporting period
- 4th highest protection gap
- 3rd highest impact of a flood, indicated by protection gap percent and relative mortalities.

Note USA and China are strong candidates but are less likely to invest in European Satellite business, having their own.

STEP 0 – Prioritising a geographical region. (Notebook: country-ranking.piynb)

Data sources

Flood frequency (and mortality) data taken from the publication

"Spatiotemporal variation in global floods with different affected areas and the contribution of influencing factors to flood-induced mortality (1985–2019)", Natural Hazards (2022) 111:2601–2625

Insurance gap data

Natural catastrophe Insurance Protection Gap = total economic losses (USD Bn) - Insured losses (USD bn)

Data from https://www.swissre.com/risk-knowledge/mitigating-climate-risk/natcat-protection-gap-infographic.html

Data covers last 10 years 2014–2023

Commercial potential

From https://www.linkedin.com/pulse/how-evaluate-international-market-potential-primetarget/

"...Economic and macro-level indicators can allow you to draw some conclusions on a country's market potential for your products and services. Looking at indicators such as GDP, GDP per capita, inflation and median income is important to analyze the size of an economy and the opportunities it may offer..."

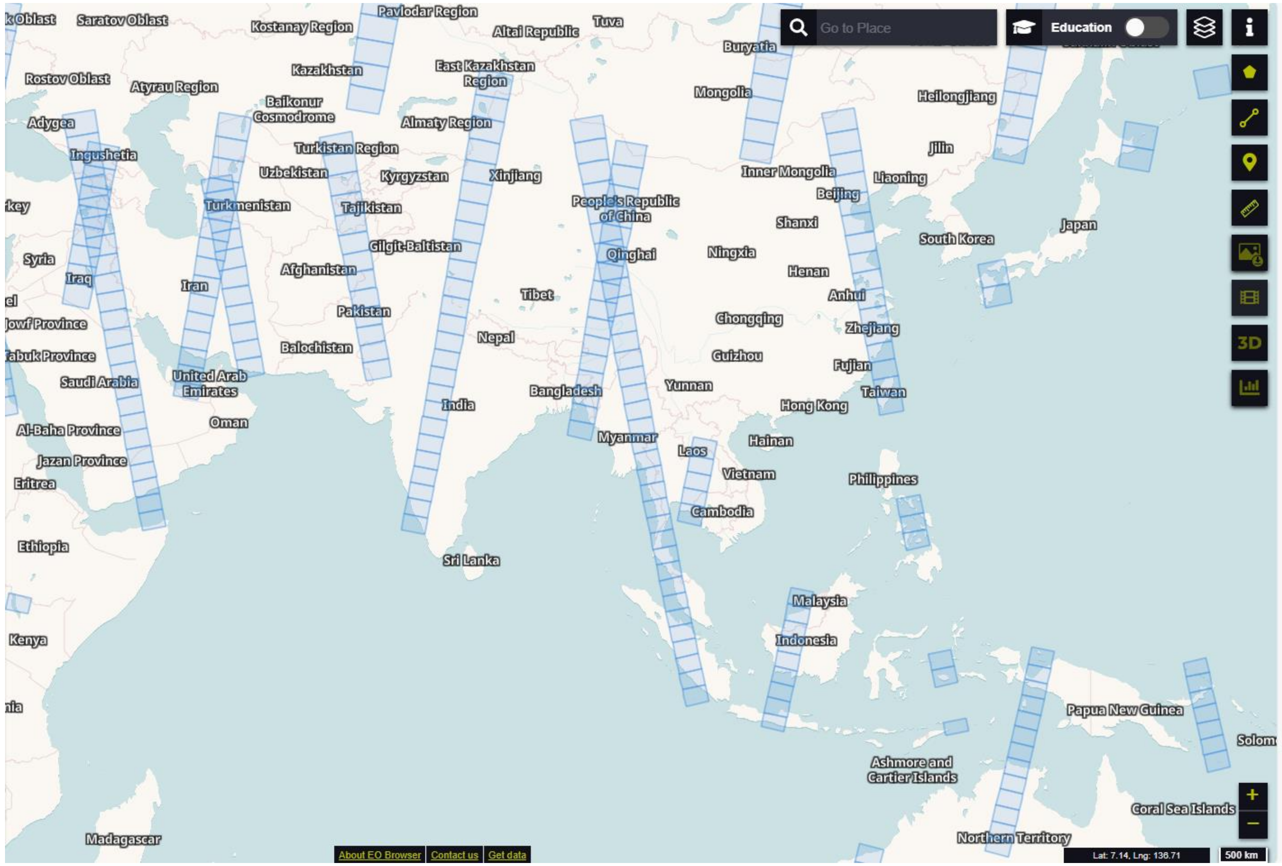
"".. Estimating potential demand requires research and analysis of data that is specific to the company's industry and product segments it targets. ..""

→ Insufficient time for market demand analysis or affect of socio-political factors.

Make do with GDP per capita available from https://data.worldbank.org/indicator/NY.GDP.PCAP.CD

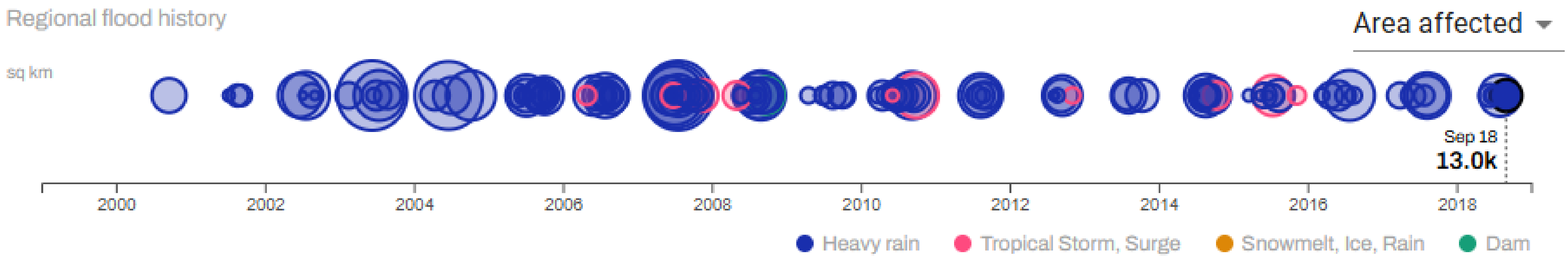
→ data as of 28.07.2024

STEP 0 – Prioritising a geographical region. (Notebook: country-ranking.piynb)



LEFT: Sentinel-1 coverage over & around India during the 24-hour period 15.06.2018 (source: Copernicus Data Workspace)

BOTTOM: Timeline of flood seasons and key causes over the last couple of decades in the region of India (source: Global Flood Database.)



Goal: Find out the probability of Sentinel-1 capturing floods in India and investigate the role of flood characteristics on that probability.

Approach:
IF we know when the floods occur in India, and where (which we do from the Global Flood Database),

THEN whether or not we will be able to capture all of those floods is completely determined by whether or not Sentinel-1 is in the right place.

SO:

- determine the capture proportion (spatial proportion of flood extent in Sentinel's ground-coverage area) per day in historical data and use that as a proxy for the probability of capturing full flood extent on any given day
- determine the binary "captured or not" value (whether any of a flood's extent was in Sentinel's ground-coverage area) per flood per day, and use that as a proxy for the probability of capturing full flood extent on any given day
- calculate features of each flood extent on each day, to see how well those serve as capture predictors in 2 Machine Learning models

    1) regression, to predict capture proportion

    2) binary classification, to predict "captured or not"

Hunch: flood characteristics (other than their location) have no significant impact on capture probability, which is rather governed by the prescribed orbit of Sentinel-1

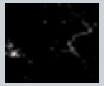
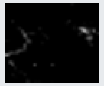
STEPS 1-4: Gathering and combining data

Python Sources:

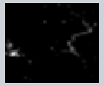
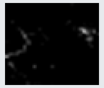
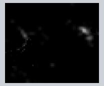
- **STEP 1 – Data Acquisition/Global Flood Database/various.ipynb files**
- **STEP 2 – add daily snapshot images.ipynb**
- **STEP 3 – add GFD metadata and Sentinel WGS info.ipynb**
- **STEP 4 – add capture metrics.ipynb**

STEPS 1-4: Gathering and combining data. Overview of steps

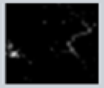
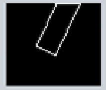
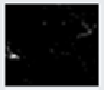
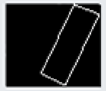
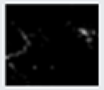
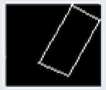
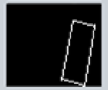
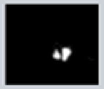
1) Store ground truth images in a dataframe along with metadata from GFD

meta	meta	meta	img (GFD)
A	..	..	
B	..	..	
C	..	..	

2) Split the 'duration' layer of each tiff by day number, which expands the dataset to daily snapshots of flood extent

meta	meta	meta	img (GFD)
A	..	day 0	
A	..	day 1	
A	..	day 2	
B	..	day 0	
B	..	day 1	
C	..	day 0	
C	..	day 1	

3) Acquire ground-coverage (bounding boxes) from Sentinel Hub API within the same field of view and 24 hour period as each daily snapshot

meta	meta	meta	img (GFD)	bbox (Sent-1)
A	..	day 0		
A	..	day 1		
A	..	day 2		
B	..	day 0		
B	..	day 1		
C	..	day 0		
C	..	day 1		

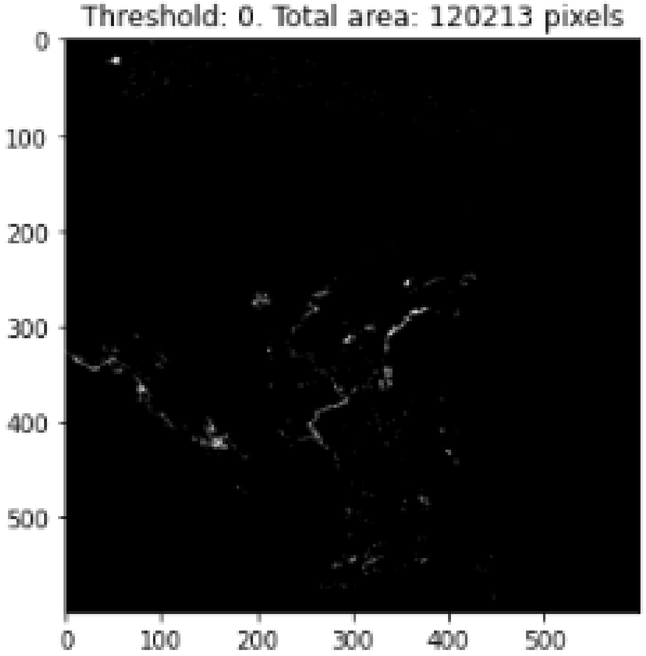
4) Add a column for capture metrics, derived from overlap between the 2 geometries: (i) WSG co-ordinates inside the flood extent and (ii) WSG co-ordinates inside the Sentinel ground-coverage.

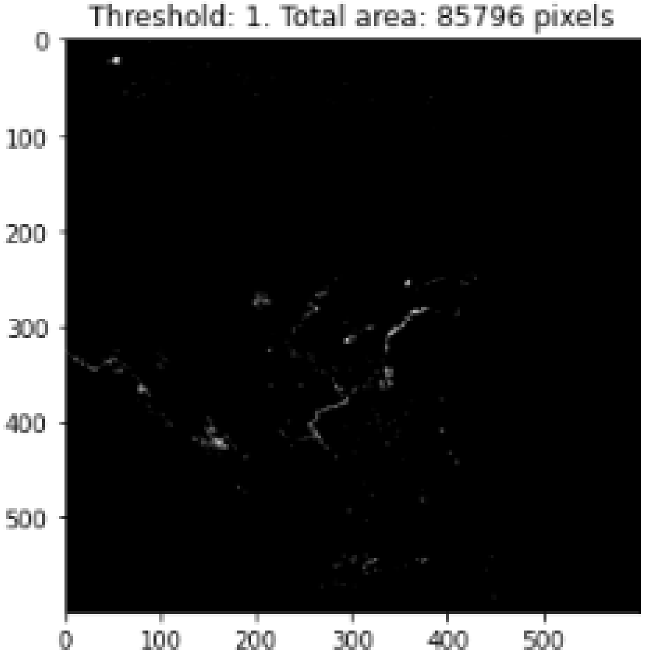
STEPS 1-4: Gathering and combining data. – The data in numbers

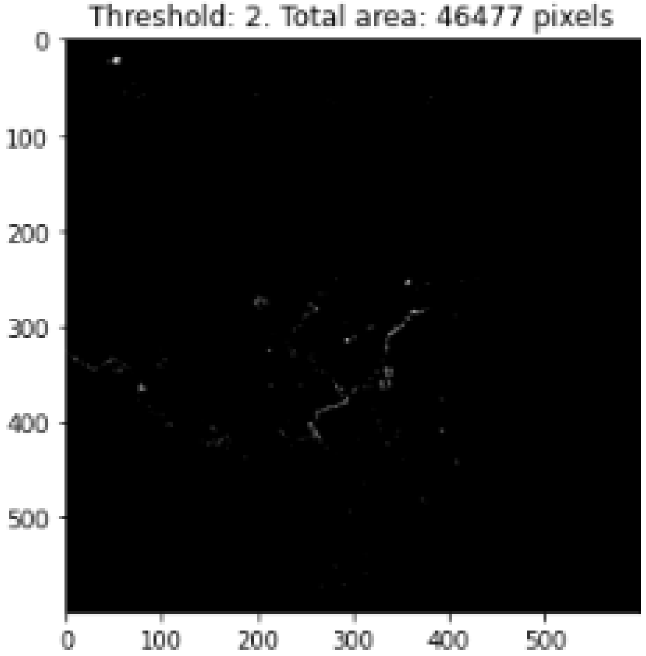
<u>From the Global Flood Database:</u>

- 6 flood events in India occurring in year 2018

- Total of 61 distinct flood ’snapshots’, where
  - one flood shapshot = one unique image of that flood’s extent
  - on one particular day
  - → derived from the day numbers provided in the “duration” layer of tif images (example below)





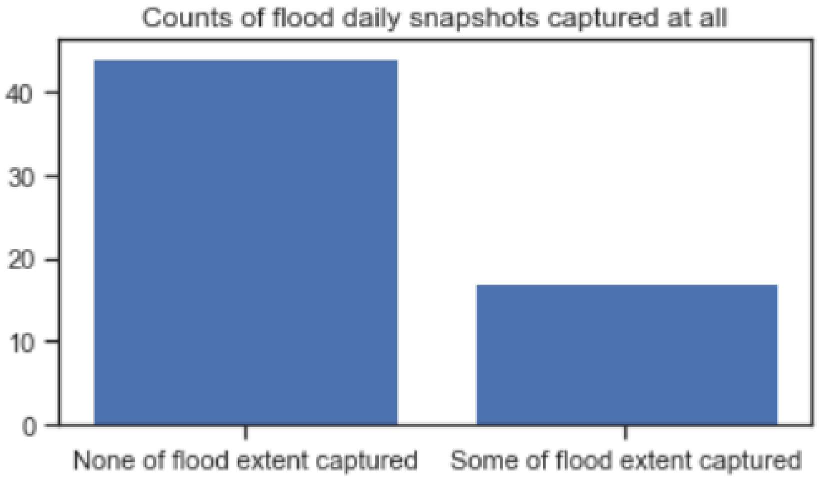


STEPS 1-4: Gathering and combining data. – The data in numbers

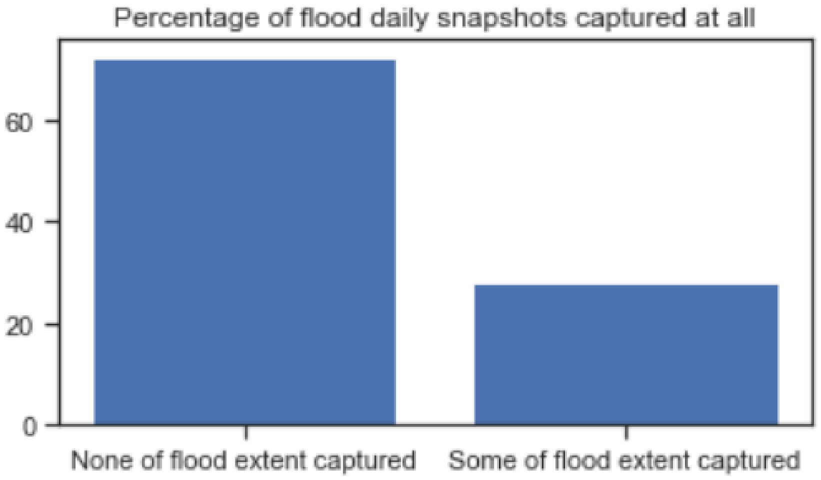
<u>Target variables</u>

We have 2 proxies for capture probability

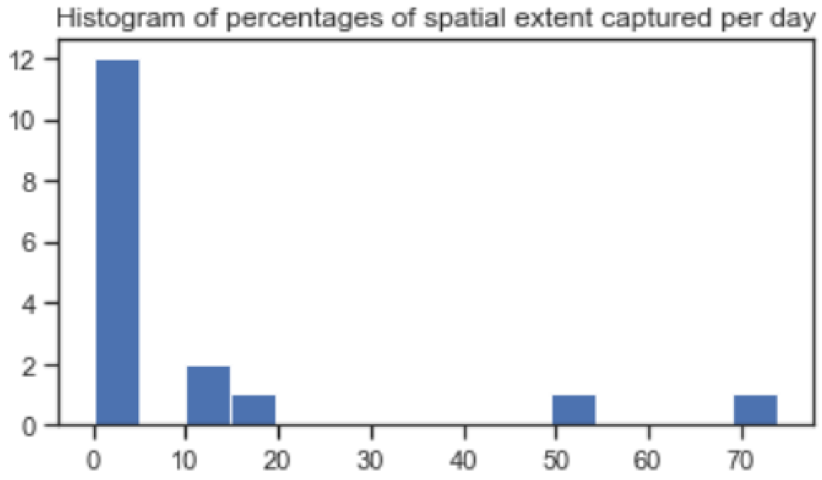
- Captured at all (binary variable) indicated by 1 or more ground truth pixel falling inside the Sentinel ground-coverage bounding box during the same day
- Capture probability (continuous variable) given by the percentage of full extent captured) on any given day



17 out of 61 daily flood snapshots are captured at all …



… this represents a “capture probability” of 28% on a per-day basis.



For any flood event captured at all, the capture proportion ranges from 0.02% to 74%.

This means that no daily snapshots have their full extent captured (100%)

STEP 5 – Analysis

- Report the observed proxies for capture probability
- Enrich with flood characteristics and analyse the relationship between those and the capture metric

Notebook:Step 5 - Analysis.piynb

STEP 5: Analysis

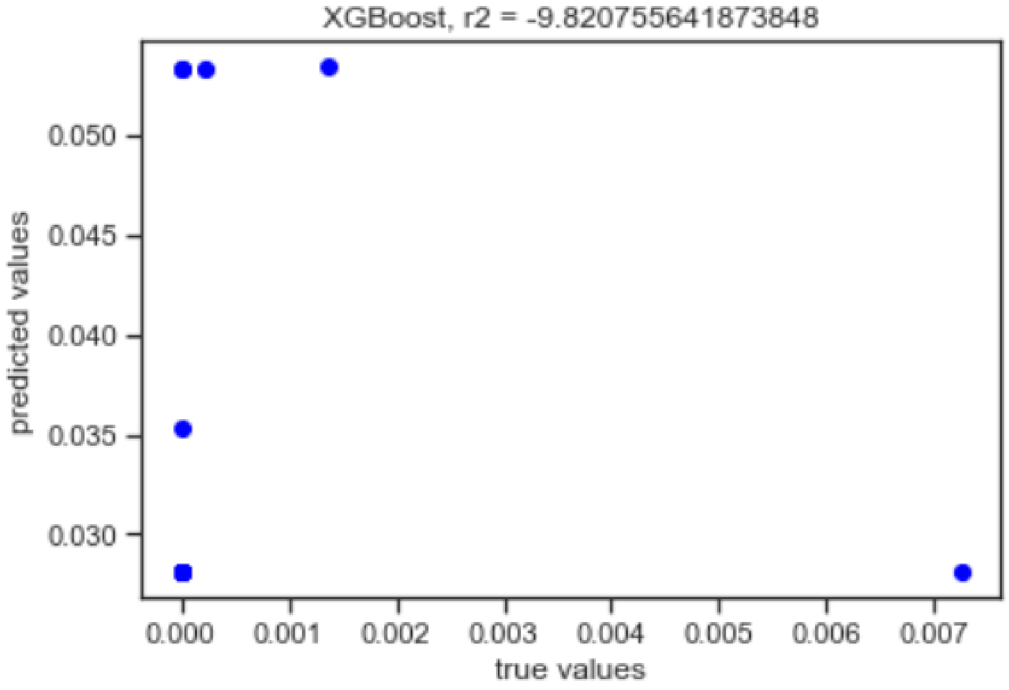
Spatial features from daily snapshot (binary) images of flood extent

Feature name	Description
perimeter	The number of pixels around the boundaris of (one or more) discrete shapes in a binary image of flood extent
shape_irregularity	complexity indicated by the ratio of the perimeter (in pixels) to the total area (in No of pixels)
n_discrete_patches	The number of disconnected patches or 'islands' in a binary image of flood extent
shape_euler	For 2D objects, the Euler number is the number of disconnected patches or 'islands' minus the number of holes in those.

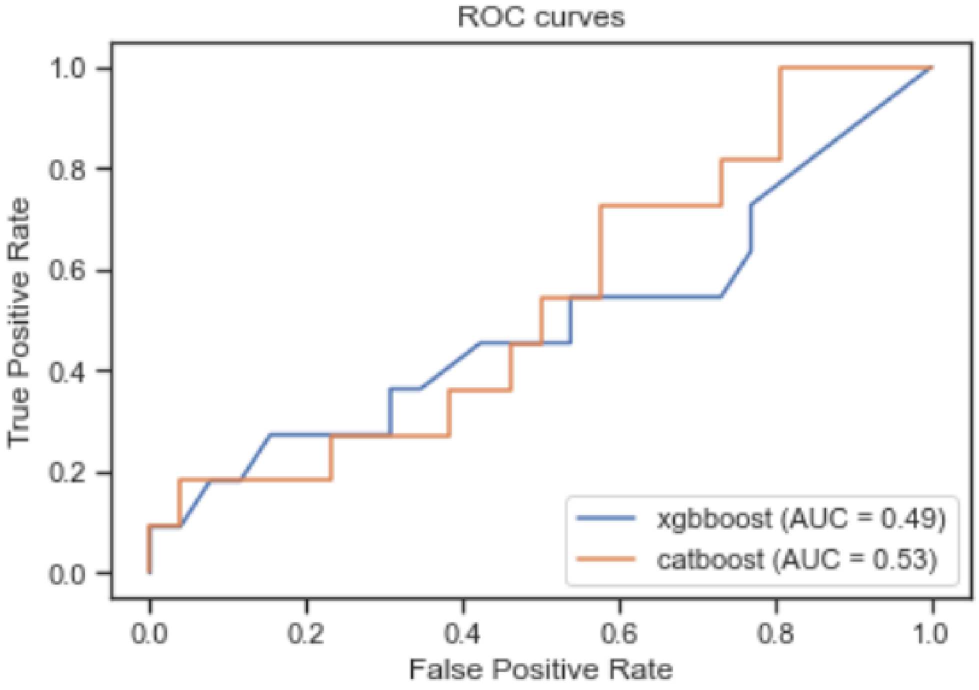
STEP 5: Analysis

Results:

- None of the basic or novel spatial characteristics of floods are predictors of capture probability on their own (see Notebook)
- Introducing multi-dimensional/nonlinear models does not reveal any predictive power of the features either.
- **This confirms the hunch stated at the beginning**, that flood characteristics (other than their location) have no significant impact on capture probability, which is rather governed by the prescribed orbit of Sentinel-1



Regression by XGBoost model fails to predict the capture percentage for any given snapshot extent, i.e. the proportion of a flood falling within Sentinel's ground-coverage box on any given day



Binary classifiers fail to predict the label "captured_any", i.e. whether or not any of a flood is within Sentinel's ground-coverage box on any given day

Recommendations for future work

Rather than predicting whether a flood is in a location that Sentinel-1 can capture, build a global model to predict the success of Sentinel in correctly labelling a flood pixel as a flood pixel, and apply it in India.

The model would be built using Sentinel images not just orbit information as follows:

- Use the current data to identify all Sentinel ground-level bounding boxes that enclose a large proportion of any day's flood extent, anywhere in the world
- For each of those, get the Sentinel image to compare with the ground truth image from the Global Flood Database (GFD)
- Train a pixel-wise classifier to predict whether Sentinel will assign a "flood" label to each pixel, knowing the truth from GFD
- Features at the pixel level can be both pixel-level (e.g. distance from a city centre, elevation etc) and flood-event level
- Apply the model to predict pixel-wise classification success in India

Outcome of the recommended model: ability to say to Indian stakeholders:
"Sentinel-1 will observe flooded locations (250mx250m square areas) with > X% probability

1. in locations having properties x,y,z (which are common in parts of your country), and
2. for floods of type a,b,c, which are historically predominant in your country.

"