Creating historical time series of satellite observed inundation for risk transfer applications in Bangladesh

Jonathan Giezendanner, Mitchell Thomas, Max Mauerman, Upmanu Lall, Arifuzzaman Bhuyan, Mehadi Hasan, A.K.M. Saiful Islam, **Beth Tellman**

jgiezendanner@arizona.edu
Twitter: @JoGiezi

Understanding flood risk in human altered landscapes from cities to farms: inferences from satellites and machine learning

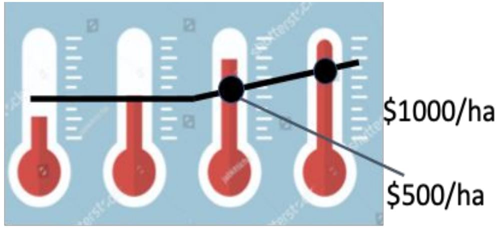
- 15% of flood losses absorbed by agricultural sector (FAO 2015)
- Asia lost 48 billion USD in agricultural production from 1980-2013 (60% due to floods) (FAO 2015)
- Insurance can support farmers’ sustainable development (Benami et al 2021)

<1% insurance penetration in Bangladesh!

Interpress News Service: Mintu Deshwara/ Sheikh Nasir

Bangladesh: world’s first satellite based agricultural flood index insurance

Index Based Insurance



- Payout based on **measurable proxy** for losses
- Payout issued when **pre-defined threshold** is reached
- Interesting in remote areas, generates cheap premiums, less moral hazard

Weather-based

from an environmental variable, such as rainfall

Area Yield

based on yield estimate sampling

Livestock mortality

from livestock mortality estimates

For Floods: based on Return Period vs Fractional Flooded Area estimates

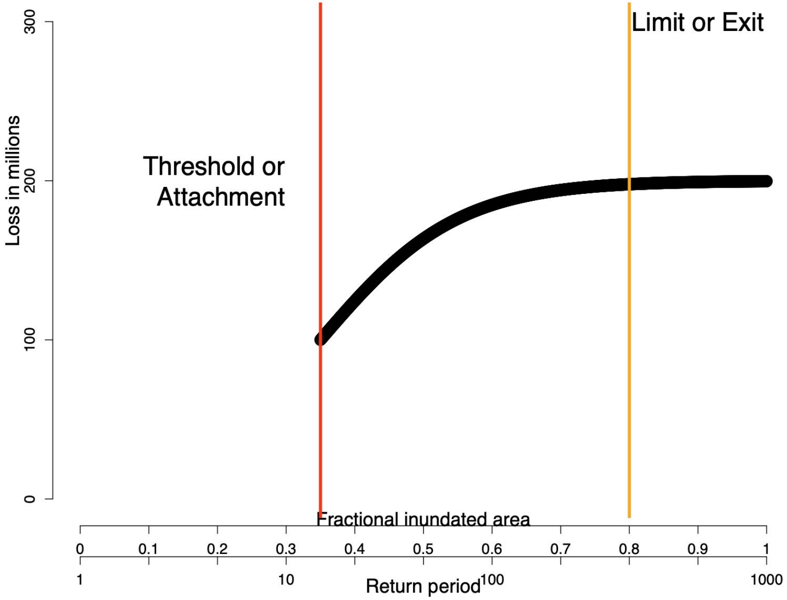
Insurance premium price based on return period estimates

How insurance companies determine premium price

$$R_\theta = Pp_{exc} + kPsp_{exc} + f1 + f2P$$

- **PREMIUM**: R_θ
- **Payout amount ($200M)**: P
- **Trigger: Probability of exceedance**: p_{exc}
- **Risk price factor**: k
- **5|95 spread of probability of exceedance (uncertainty)**: s
- **Transaction cost (15%)**: $f1$
- **Profit margin (6%)**: $f2$

Based on return period vs fractional inundated area



Tellman et al, 2022

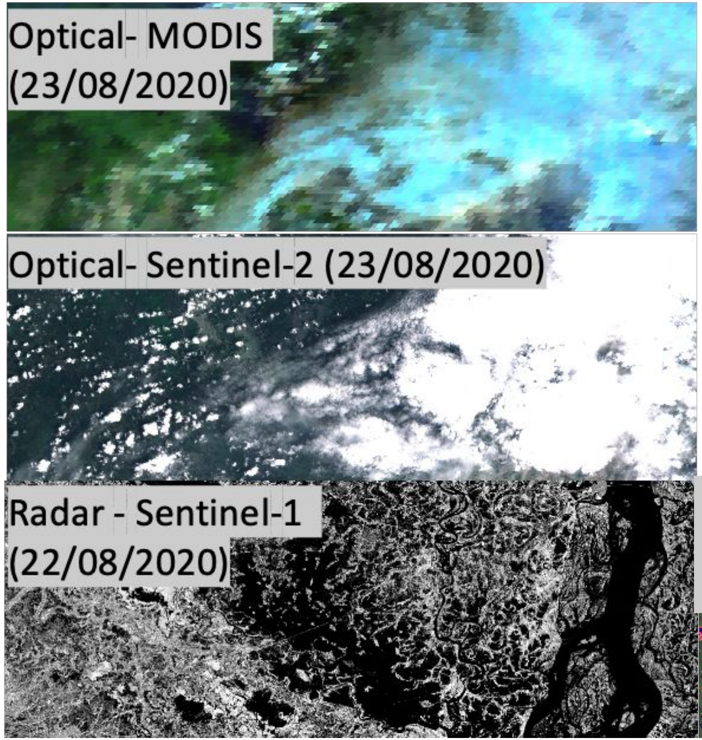
Requires accurate historical estimate of yearly maximum flood extend (capture peaks)

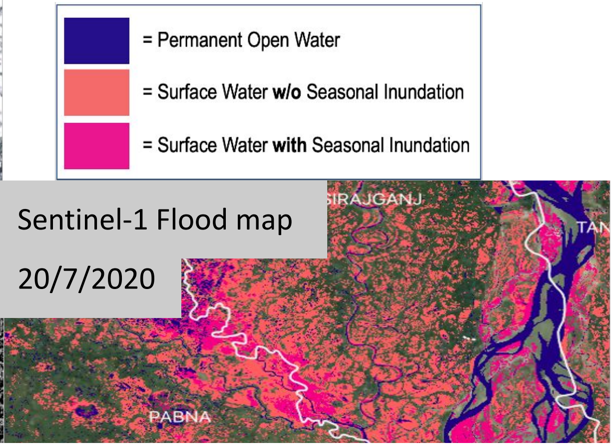
Best satellites for flood mapping start ~2017 (Sentinel-1) but insurance requires a >15 year time series to establish contracts

Sentinel-1= higher spatial accuracy + temporal consistency, and more correlated to damage (Thomas et al, submitted)

MODIS Terra Satellite image showing flood affected areas in Bangladesh (17 July 2019). Photo: NASA MODIS

GREEN DELTA INSURANCE





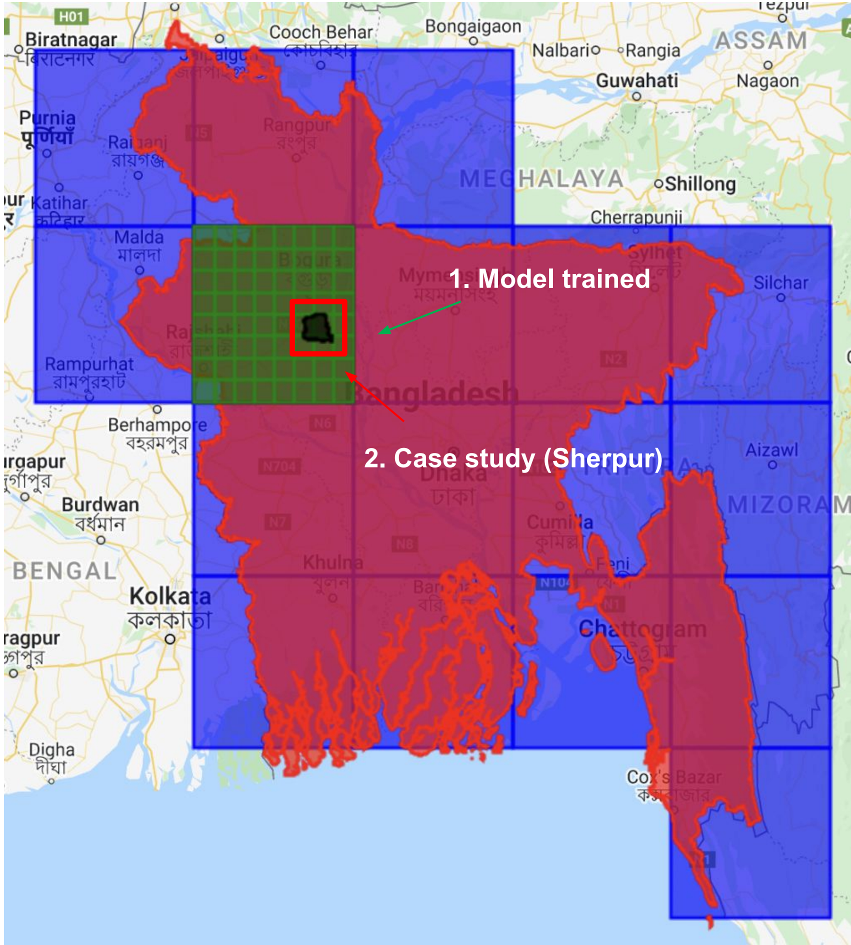
Goal: create historical (20+ years) time series of flooded areas over Bangladesh for return period estimates

Methods:

- Create a Fusion algorithm (Random Forest) to estimate fraction of flooded area for each MODIS pixel
- Sentinel-1 data (2017 - 2021) to generate weak labels (Thomas et al., submitted)
- Infer time series based on MODIS historical data (2001 – 2021).

Regions of interest:

1. Train on large region
2. Discuss results on Sherpur



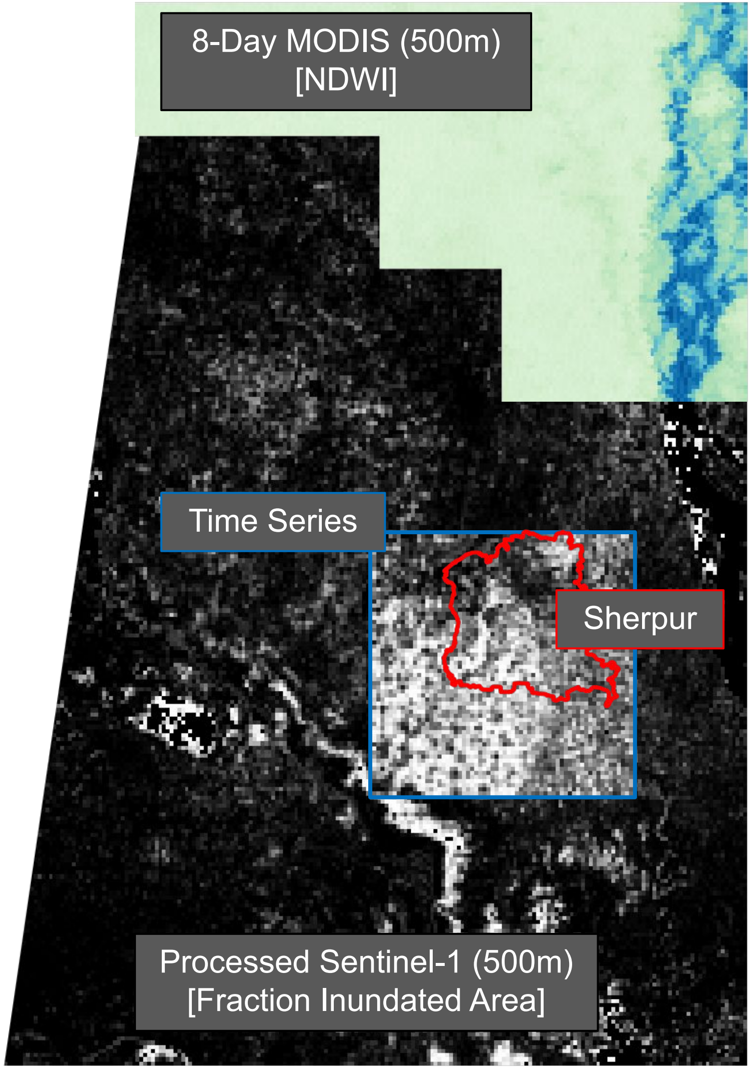
Data

Fraction Inundated Area Estimate:

- Sentinel-1 (2017-2021, 10m.)
- Thresholding algorithm (Thomas et al., submitted), binary map at 10m resolution
- Calculate fraction of inundated area ([0,1]) for each MODIS pixel at 500m resolution

Input features:

- 8-Day MODIS composite image at 500m resolution
- Elevation (FABDEM)

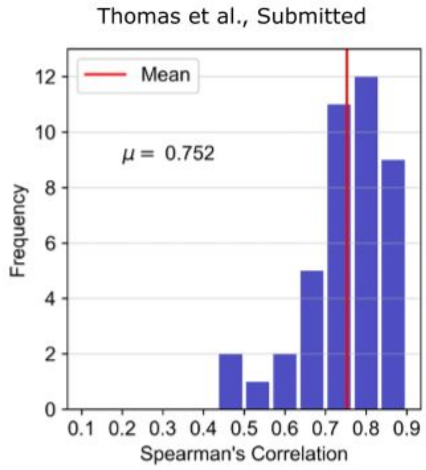


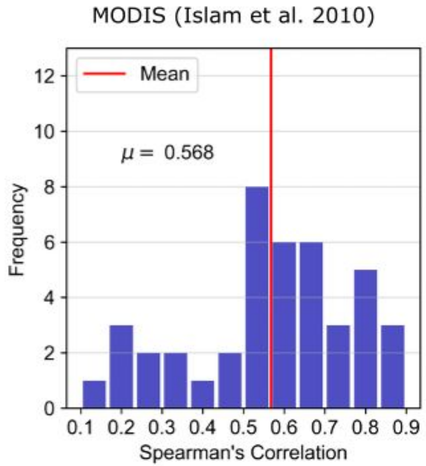
Fraction Inundated Area Estimates

Sentinel-1 rapid and automated flood detection algorithm (Thomas et al, submitted)

- Measure Z-score statistical in radar backscatter during a flood compared to a historical dry baseline
- Computes individual thresholds for VV and VH bands in IW mode
- Processes Ascending and Descending orbit passes separately
- Apply additional backscatter threshold and spatial smoothing

Algorithm outperforms MODIS based algorithm (Islam et al. 2010)





From Thomas et al., Submitted

MODIS Data – 8 Days Composite

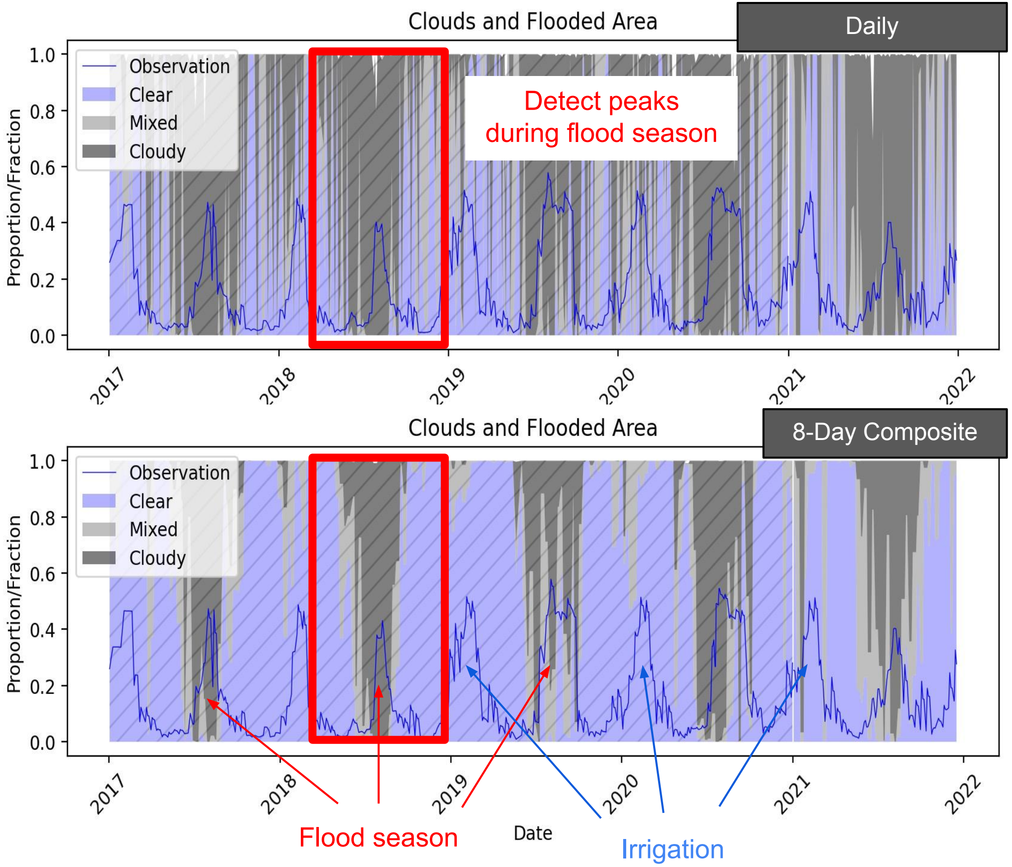
MODIS Cloud Cover compared to Sentinel-1 Weak Labels

Overarching goal:

- Detect inundation peaks during flood season
- Ideally use daily MODIS data to detect flood dynamics as close as possible
- In practice: cloud cover during flood season too strong

Decision:

Use MODIS 8-Days Composite



Random Forest Model

Feature	Importance
MODIS Band 5	0.35
NDWI	0.17
MODIS Band 6	0.16
NDVI	0.12
B1 / B2	0.11
Elevation	0.07

Data:

- 12’072’405 total valid (Sentinel-1 overlap with MODIS) data points, spread over 65’536 spatial pixels and 5 years

Feature Selection:

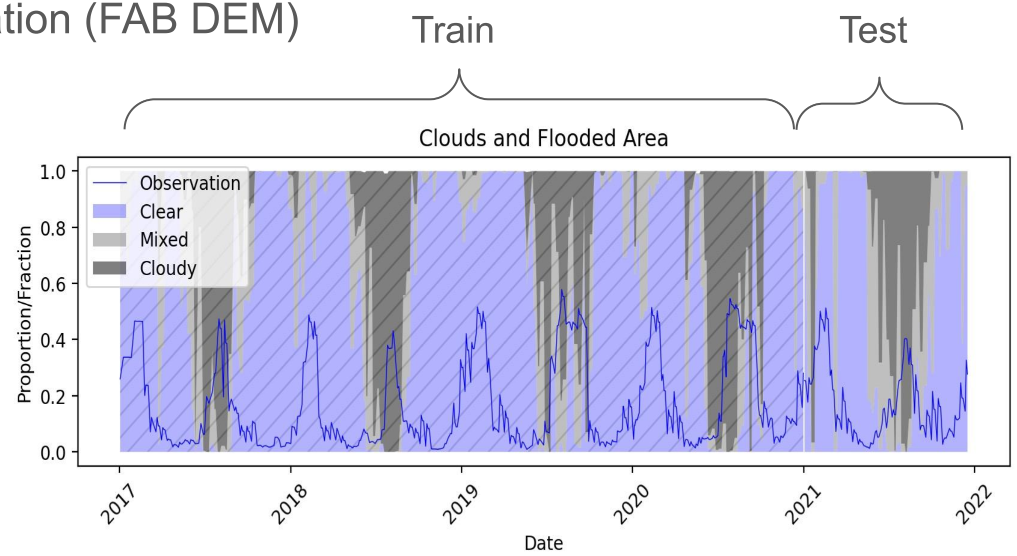
- MODIS bands 5 and 6
- Ratio of MODIS band 1 over 2
- MODIS based NDWI and NDVI
- Elevation (FAB DEM)

Training:

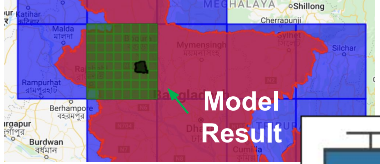
- Trained on whole dataset (not only flood season), to improve water detection

Cross validation:

- Train on 4 years, leave one year out
- Iterate over all years



Results: Aggregated over Grid



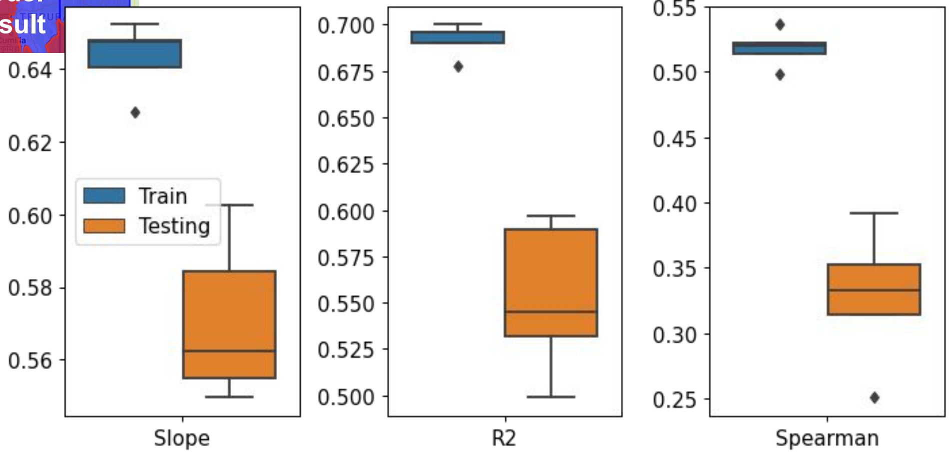
Cross-validation:

- Statistics consistent throughout years

Fit:

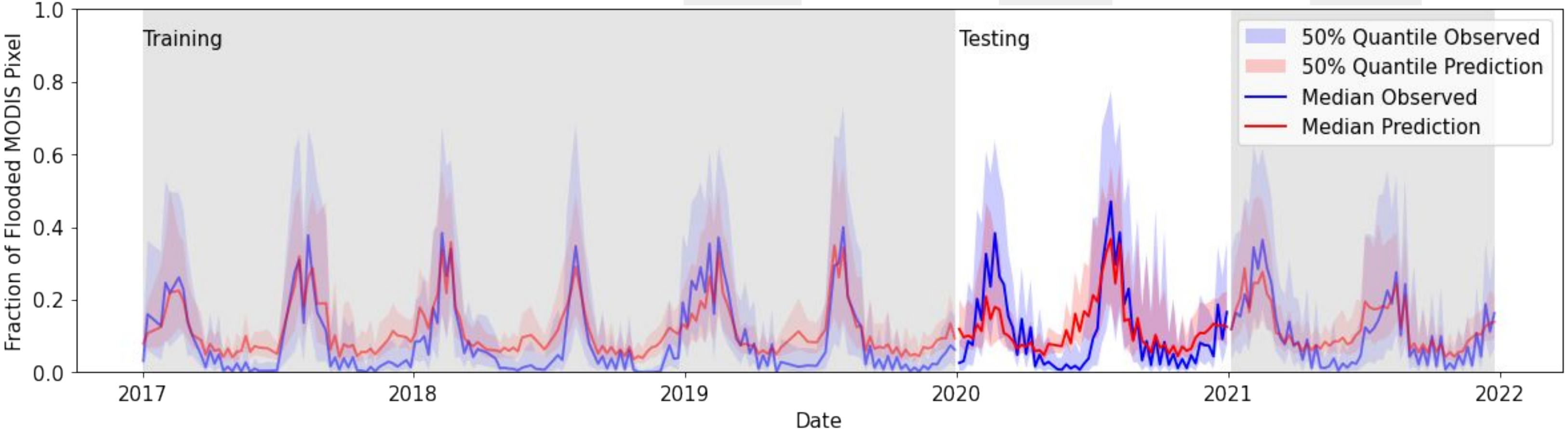
- Overestimation of valleys (dry season)
- Not all peaks captured for test data
- Flood season peak captured, but underestimated

Cross-Validation (distribution over years)



Median Score	0.65	0.56	0.7	0.54	0.52	0.33

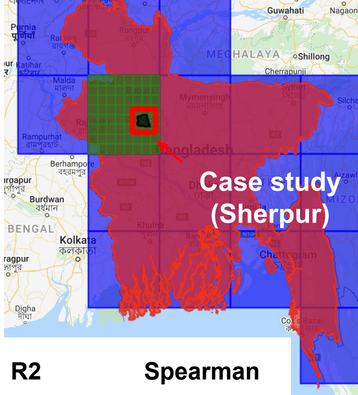
Single Year Test on 2020



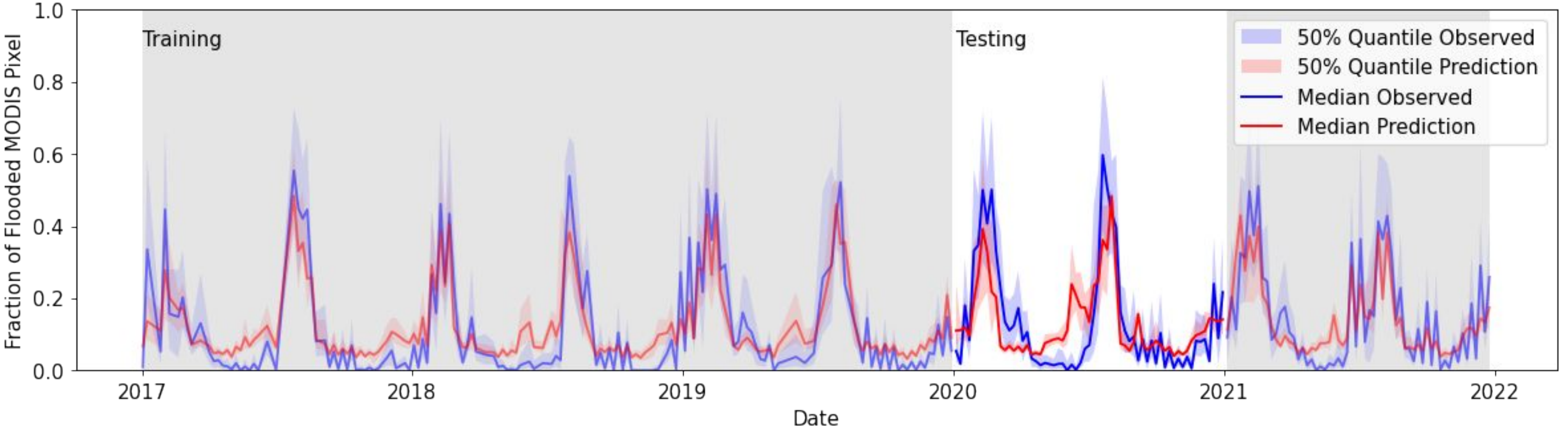
Results: case study

Use case: aggregate data at administrative level (Upazila)

- Focus on Sherpur
- Flood season peak captured, but underestimated

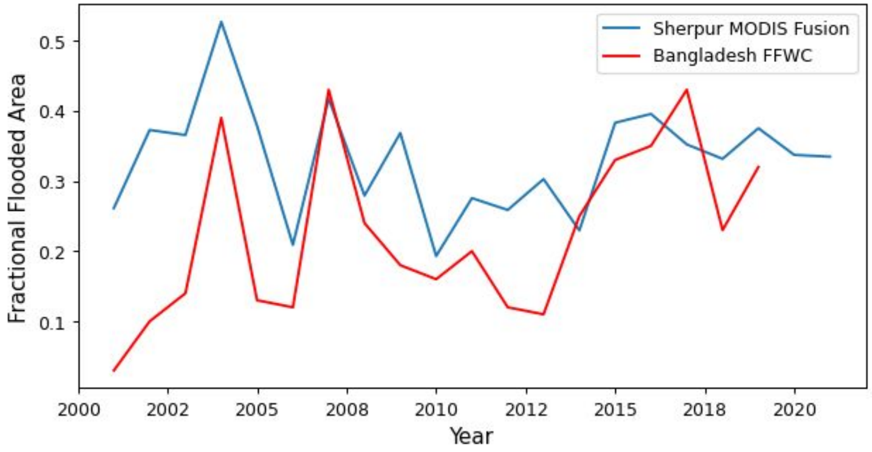


Set	Slope	R2	Spearman
Train	0.52	0.61	0.47
Test	0.42	0.44	0.21

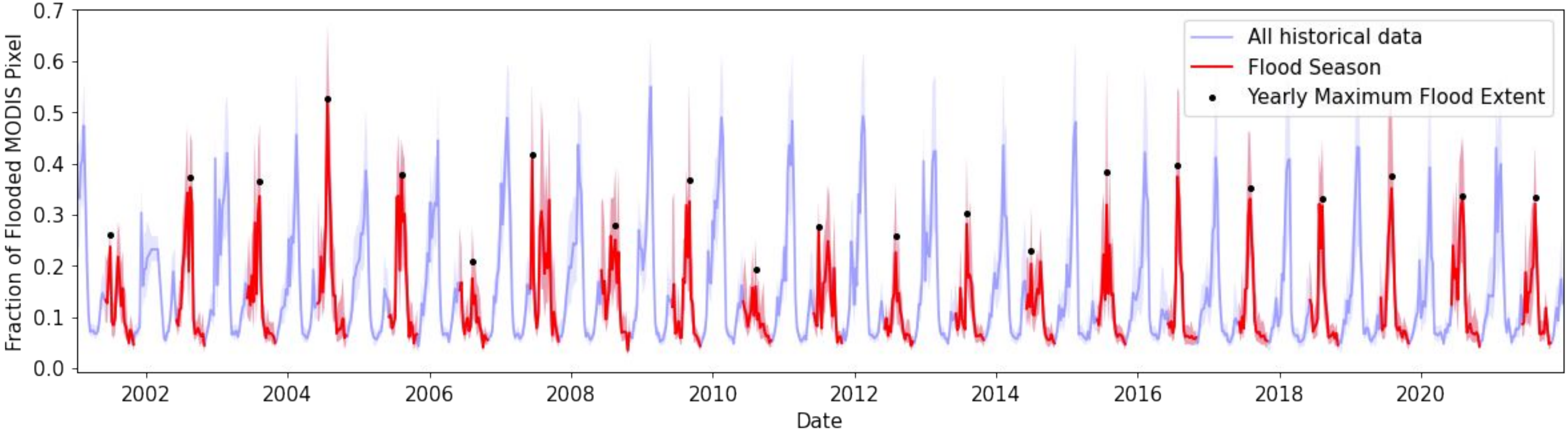


Historical Inference (Sherpur)

- Infer time series of fraction of flooded area based on MODIS Fusion algorithm (20 years)
- Extract yearly maximum extent
  - Compare to Bangladesh Flood Forecasting and Warning Center (FFWC) model (coupled hydrologic - rainfall Mike 11 model) (Tellman et al. 2022), 0.27 Spearman



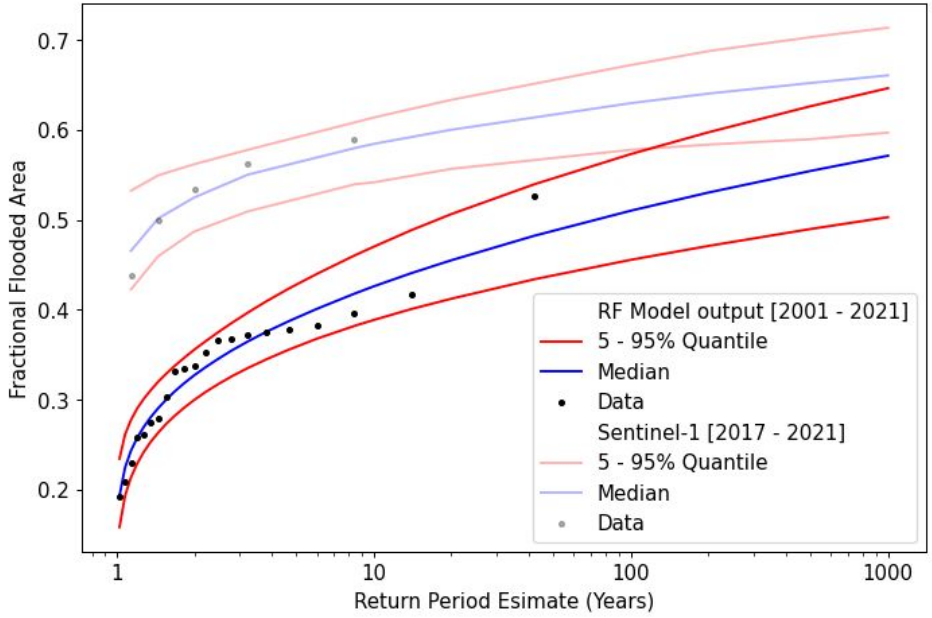
Extract Yearly Maximal Extent



Return Period Estimates (Sherpur)

Return period estimates for Fractional Flooded Area using Beta-2 distribution (Tellman et al., 2022):

- Strong difference for short return periods between 4-years Sentinel-1 estimates and MODIS historical data
- Sentinel-1 series biased by short time series
- MODIS fusion algorithm, more realistic estimate, but underestimates flood peaks
- Real estimate in-between both distributions



Conclusions and Outlook

Project is at its very beginning!

- Promising results!
- Results show a more realistic return period estimate using this historical time series compared to the short term Sentinel-1 observation
- Greatly impacts triggers for parametric insurance and computation of premiums
- Difficulty to predict peaks exacerbated by heavy cloud cover

Data:
- Given the selected MODIS product (8-Days composite), goal is to generate return period estimates with model, but insurance trigger should be done with Sentinel-1 (better model quality)
- If single day MODIS were to be usable, and high model quality, trigger on Fusion algorithm could be considered given higher temporal resolution and probability of capturing flood peak.

Next steps:
- Local validation of flood extends
- Spatial Cross-validation
- Country-scale model
- Deep Learning algorithm to increase performance with presence of clouds

Thank you for your attention!

jgiezendanner@arizona.edu Twitter: @JoGiezi

Reference: **Tellman and et al**, *Earth's Future,* Regional Index Insurance using Satellite-based Fractional Flooded Area, 2022

This work is undertaken as part of the NASA New (Early Career) Investigators (NIP) Program (80NSSC21K1044)