Rising Waters: Can We Use Machine Learning to Accurately and Effectively Predict Bangladesh's Floods and Hydrological Data?

Audience: Central and local governments of Bangladesh

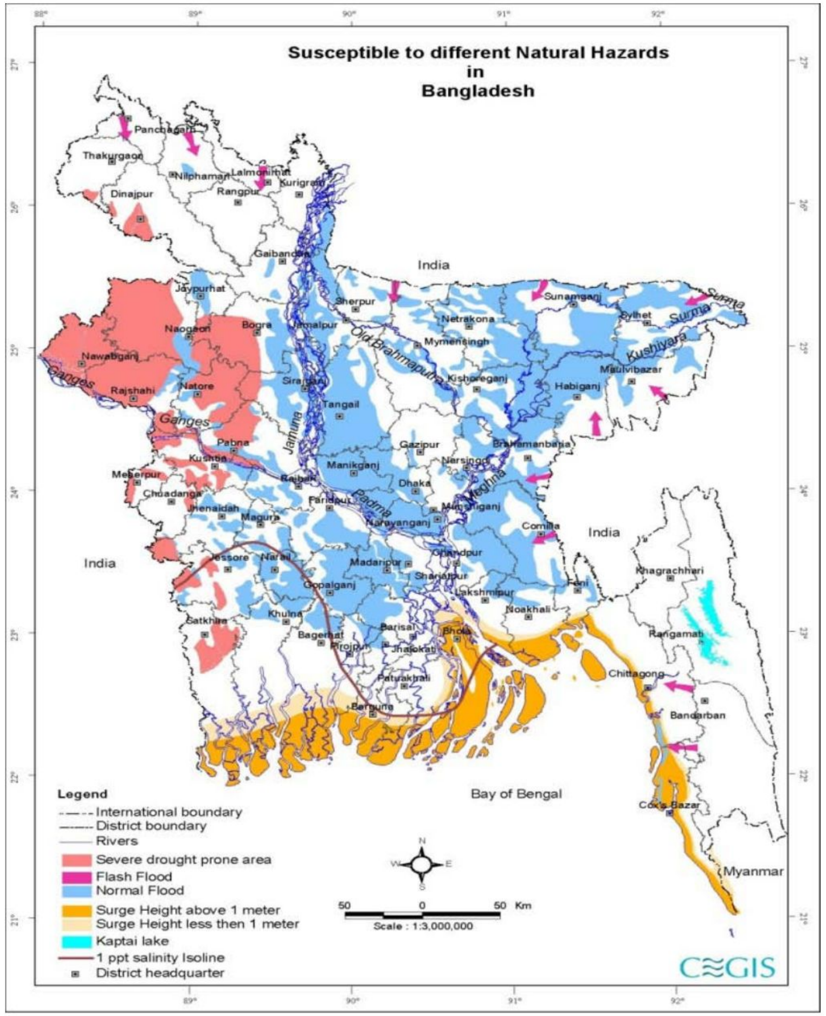
David Campos, Chun Him Cheung, Kolby Devery, Redwan Hussain

UNIVERSITY CALIFORNIA, BERKELEY | Fall 2021

MIDS W201 | MON 4PM PT | GROUP 2

Bangladesh is subject to catastrophe due to floods

- Population of ~165 million people.
- Located at the delta of the Padma and Jamuna rivers in the northeastern part of the Indian subcontinent.



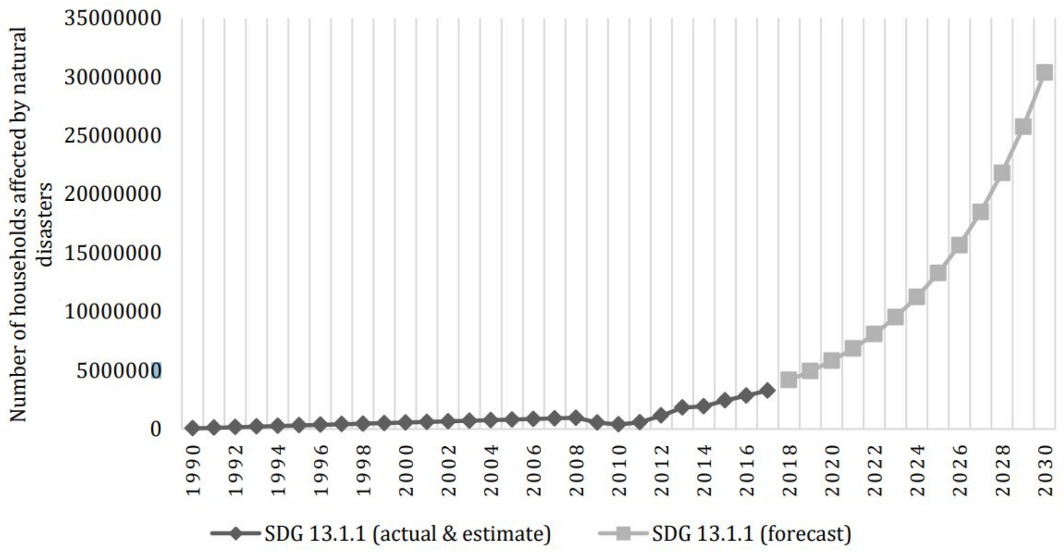
Bangladesh is subject to catastrophe due to floods

- 2020: heavy monsoons caused nearly 25 percent of the country to be submerged
  - ~2 million displacements
  - Hundreds of deaths
- Going forward, the UN estimates that the number of households to be impacted by natural disaster will grow exponentially.

Data

- Bangladesh Disaster Related Statistics
- Bangladesh Water Development Board (BWDB) Data:
    - Historical water level data across all of the nation's major rivers
- Bangladesh Meteorological Department (BMD) Data
    - Daily rainfall and temperature data
- NOAA Global Integrated Surface Dataset:
    - Meteorological data such as humidity, wind, cloud coverage, and historical patterns of extreme weather such as cyclones

Bangladeshi households impacted by weather events: p

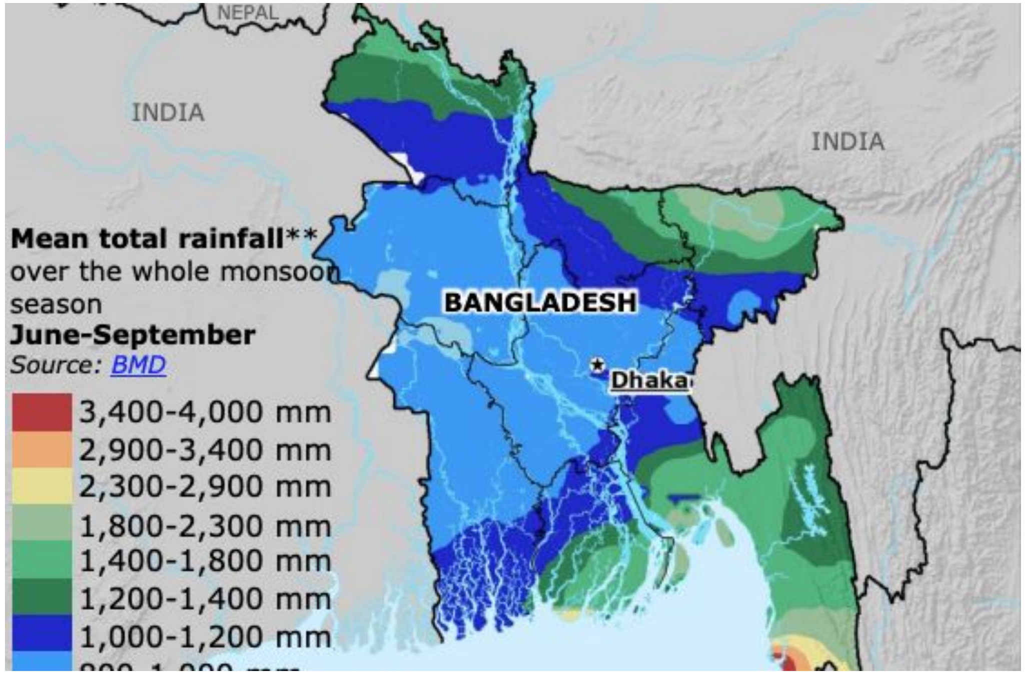


Study design

- Quantitative study
- Establish baseline predictive method used by government
- Capture mix of historical and live data
- Feed captured data to train model
- Calculate predictive ability of our model against baseline prediction

Sample

- Stratification based on flood-prone geographical areas
- Use latitude and longitude points from dataset for grouping
- Interpolate for missing values



Source: United Nations Office for the Coordination of Humanitarian Affairs

Variables and/or Intervention

- Hydrological/meteorological data:
  - Precipitation, river flows/levels, storm surges

Statistical Methods

$$NSE = 1 - \frac{\sum_{t=1}^{T} (Q_o^t - Q_m^t)^2}{\sum_{t=1}^{T} \left(Q_o^t - \overline{Q}_o\right)^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

- **Nash-Sutcliffe efficiency coefficient (NSE)**
  - Assesses predictive ability of hydrological models
  - One minus ratio of the error variance of model divided by the variance of the observed
- **Root Mean Squared Error (RMSE)**
  - Square root of predicted minus actual data divided by

Potential Risks

- Overfitting
- Parameter instability of model
- Communication with stakeholders and local community
- Fair and equitable access to results
- Regional variance

Deliverables

After 3 years, we will provide:

- Technical details of predictive model (e.g. accuracy, # of projection days)
- Live test results of new model
- Summary report with recommendations

Thank You