W201, Mon 4pm, Group 2, Fall 2021, Carlos Rivera

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

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

Overview

Located below the level of three major river systems, Bangladesh is subject to catastrophic damage during hurricane season when the rivers overflow. Bangladesh has a population of 165 million (more than half the total US population) but has a similar landmass as the state of Michigan. In 2020, heavy monsoons caused nearly 25 percent of the country to be submerged,[1] causing almost 2 million displacements nationwide, destroying thousands of homes, and killing hundreds[2]. The combination of poverty, vulnerabilities to rising sea levels, inadequate infrastructure investment, and increasingly severe weather events make Bangladesh prone to a severe humanitarian crisis that can spill over to neighboring countries. However, proactively managing emergency services and urban infrastructure can mitigate the effects of torrential storms, especially as climate change is expected to worsen it.

According to the Bangladesh Disaster-Related Statistics, in the 5-year periods between 2009-2014, nationally 34.5% of households reported to have been negatively affected by flooding while the reported figure in the national capital region of Dhaka was 52%. Economically, during the same time period, the Bangladesh Bureau of Statistics estimated that floods caused US$ 43bn in economic damages with the crops being most economically impacted with losses of US$ 22bn[3]. The UN forecasts the expected human and economic cost of severe weather events in Bangladesh will exponentially increase from now until 2030[4]. As such, a better flood forecasting system can help offset some of this rising cost.

The intent of this research is to build a highly reliable predictive model that captures the complexities of rain, runoff, and storm to reduce the economic, social and human cost related to

[1] Sengupta, S., & Manik, J. A. (2021, May 19). *A quarter of Bangladesh is flooded ... - The New York Times*. New York Times. Retrieved December 3, 2021, from https://www.nytimes.com/2020/07/30/climate/bangladesh-floods.html.

[2] *Bangladesh*. IDMC. (n.d.). Retrieved December 5, 2021, from https://www.internal-displacement.org/countries/bangladesh#:~:text=The%20storm%20triggered%20around%202.5,the%20end%20of%20the%20year.

[3] BBS. (2015a). Bangladesh Disaster Related Statistics 2015. Dhaka: Bangladesh Bureau of Statistics (BBS).

[4] Khatun, F., & Saadat, S. Y. (2021, September 22). *CLIMATE CHANGE IN BANGLADESH: A SUSTAINABLE DEVELOPMENT PERSPECTIVE*. UN. Retrieved December 5, 2021.

climate change. While having a better weather forecasting model will not reduce the increasing severity of extreme weather for Bangladesh, it can provide a more timely update to Bangladeshi policymakers on how to prepare and plan for natural disasters to minimize the human and economic cost.

The target audience includes Bangladesh government entities that rely on weather forecasts for emergency and infrastructure project planning.

- The current prediction model used by the Bangladesh government:
  - FFWC (Flood Forecasting and Warning Center) currently only offers a five-day deterministic flood forecast with varying levels of accuracy[5].
    - This currently only covers certain regions of Bangladesh.
    - This is not sufficient preparation time as evidenced by the lives lost each year.
- Additionally, there are ongoing attempts to design an improved prediction model:
  - ReqSim is attempting to offer a 10-day water level forecast[6].
  - SERVIR-HKH is attempting to offer a 10-15 day probabilistic forecasts at only certain locations along the Bangladesh border[7]. (Tsering et al., 2021)

Research Question

The main focus of this research is to answer the following questions:

- *Can our iterative approach provide the most accurate predictive ability in forecasting floods from events such as rain, storm surges or river runoff across all parts of Bangladesh?*
- *How far in advance can we functionally predict these events to maximize preparation?*

Data

We plan to use five categories of data for our study:

[5] *Flood forecasting and warning centre, BWDB, Bangladesh*. Flood Forecasting and Warning Centre, BWDB, Bangladesh. (n.d.). Retrieved December 3, 2021, from http://www.ffwc.gov.bd/.

[6] *Reqsim*. ReqSim | Requisitely Simple Flood Forecasts. (2020, July 19). Retrieved December 3, 2021, from https://waterdiplomacy.org/reqsim/.

[7] Tsering, K., Shrestha, M., Shakya, K., Bajracharya, B., Matin, M., Lozano, J. L., Nelson, J., Wangchuk, T., Parajuli, B., & Bhuyan, M. A. (2021). Verification of two hydrological models for real-time flood forecasting in the Hindu Kush Himalaya (HKH) region. *Natural Hazards*. https://doi.org/10.1007/s11069-021-05014-y

- **Bangladesh Disaster Related Statistics:** First, we will leverage the Bangladesh Bureau of Statistics (BBS) to provide the baseline projections of the social and economic cost of ongoing floods and severe weather events.
- **Primary Hydrological Data Source:** Second, we will use hydrological data from the Bangladesh Water Development Board (BWDB) to collect historical water level data across all of the nation's major rivers.
- **Secondary Hydrological Data Source:** Third, we will use a variety of widely available rainfall and water data from satellite and ground measurements to supplement our data as needed. Daily historical data at 0.25 × 0.25° grid resolution will be obtained from the ERA5 dataset from ECMWF (European Center for Medium Range Weather Forecasts)
- **Primary Meteorological Data Source:** Fourth, we will use data from the Bangladesh Meteorological Department (BMD), which provides 36 years of daily rainfall and temperature data from 1979 to present
- **Secondary Meteorological Data Source:** And fifth, we will use the NOAA Global Integrated Surface Dataset, which will gives us additional meteorological data such as humidity, wind, cloud coverage and historical patterns of extreme weather such as cyclones

Study Design:

We will take the following steps for our study design:

1. First, we will establish the baseline prediction method, which is the five-day deterministic flood forecast provided by the Flood Forecasting and Warning Center. If ReqSim or SERVIR-HKH models are completed at the time of this project and if they have better predictive capabilities, we will use one of them as the baseline method instead.
2. Second, we will monitor actual weather and hydrological conditions to capture live data. Methods of capturing live data include but are not limited to:
    a. Extracting information from local weather stations.
    b. Deploying hydrological sensors across the field.
3. Third, we will develop/train new predictive models against subset of both historical and new (hydrological sensor-extracted) data.
4. Fourth, we will make sure to take into account data-overfitting problems in previous experimental implementations. Even though most regions of the country are at risk of floods, these risks are statistically disproportionate across certain regions. For example, the west is at a higher risk of droughts, while the east is at a higher risk of major floods. As such, we will take into account the regional variance in weather patterns and not over-extrapolate the national result on every region.
5. Fifth, we will compare live data against results of the new model and baseline prediction method.

    a. We will compare the Nash-Sutcliffe efficiency (NSE) and Root Mean Squared Error (RMSE) of the new method versus the baseline to determine whether this new approach is superior.
6. Sixth and, finally, we will determine the predictive ability and accuracy of the new model vs. baseline prediction method.

Sample

- Since floods occur sporadically and at different levels throughout the country, we will stratify based on the flood-prone geographical areas outlined by the Bangladesh Agricultural Research Council.
- Longitude and latitude points are included in the meteorological and hydrological datasets from the Bangladesh Meteorological Department; these data points will be used for grouping into geographical zones.
- We will interpolate for any missing values since the data is uniformly distributed.

Variables and/or Intervention

- For its analysis, the predictive model will take as input historical hydrological/meteorological data such as *precipitation*, *river flows/levels*, and *storm sturges*
- Post-experiment analysis will include a comparison of variables of our predictive model against the current predictive model used by the Bangladesh government. These include: *reliability*, *Nash-Sutcliffe efficiency coefficient*, and *root mean squared error*. The calculation of these variables is described in the statistical methods section below.
- Overweight recent observations versus those in the distant past because climate change may be gradually changing the underlying relationship among the regressor variables.

Statistical Methods

We will use the following statistical methods to test the efficacy of our predictive model:

- Examine the full joint frequency distributions of our model's flood forecasts vs. the corresponding actual observations to build prediction reliability diagrams at various lead times. This will be done through analysis of historical data as well as data collected in real time over the course of the research.
- Calculate the **Nash-Sutcliffe efficiency (NSE)** coefficient which is commonly used to assess the predictive ability of hydrological models. It is calculated as one minus the

ratio of the error variance of the model divided by the variance of the observed. Full equation in appendix.
- Calculate the **Root Mean Squared Error (RMSE)** of our modeled predictions. This is the quadratic mean of the differences between the observed and predicted values.

Potential Risks

- **Over-fitting:** We run the risk of over-fitting our data too closely to historical, adding an unforeseen bias to the data. We may need to develop a set of robust model validation tests mixed with field observations.
- **Parameter instability:** This model may not fully capture factors that will have a more profound effect on flooding in Bangladesh in the future. For example, melting of glaciers as climate patterns worsen in 10+ years. According to Germanwatch, a non-profit organization specializing in environmental policy research, Bangladesh is ranked 7th amongst countries most affected by climate change from 2000 through 2019[8]. Climate change may profoundly change the relationship between our input variables, and thus, make the out-of-sample forecast less accurate relative to our in-sample data. The approach we use for our model may need continual updates and added complexity to maintain a high level of predictive accuracy.
- **Communication with stakeholders and local community:**
    - Efforts must be made to communicate the findings and for the model to gain legitimacy in the eyes of policymakers and stakeholders. The results of the model will have an impact on the lives of those living in the areas our model identifies as high-risk for incoming weather events. This may result in relocation, calls for evacuation and other government action that can affect the lives of local residents. As such confidence and trust between the model's designer, local government officials and local residents are crucial.
    - Secondly, the government and local community must be made aware of **false positives.** No model is perfect and the public should not lose faith in the model after the first few false positives. The emphasis to the public is not this model we develop is perfect, but is an improvement on existing methodology.
- **Fair and equitable access to results:** We must also ensure the data needed for this model is widely available and usable for the public over time. It is important for our model not to be viewed as partisan to any particular interest group. Our intention is to better inform the Bangladeshi community on incoming severe weather events where all stakeholders have

[8] Eckstein, D., Künzel, V., & Schäfer, L. (2021). *Global climate risk index*. Germanwatch e.V. Retrieved December 3, 2021, from https://germanwatch.org/en/19777.

equal access to the data. We intend to continually publish the data, methodology and result of the findings in an open-source manner to maintain transparency and credibility of the project. If any field observations are needed long term, this may add overhead or require resources which have to be pre-committed .

- **Regional variance:** It may not be worth examining every region within Bangladesh as severe rainfall and floods are not a major concern in the western region. This model is meant to detect severe water-related events and is most applicable to central areas of the country where floods are most common.

Deliverables

After 3 years of data collection we will provide:

- A report of our approach and how well the training data performed in forecasting new flood data
- If the approach does not provide improved hydrological forecasting or lead time compared to baseline, we will provide lessons learned and recommendations on how to address moving forward
- If we succeed:
  - The accuracy of prediction achieved and the maximum number of days that can be projected while maintaining accuracy
  - Best practices for new data acquisition for continued accuracy of model
- Live testing of the model with new data over a span of the next three years. The point is to demonstrate that this new approach is superior with data the model has not yet seen. If this is the case, then the findings can be presented to policymakers and stakeholders to convince why decisions should be made based on this new model.
- A summary of opportunities for continual research

Appendix

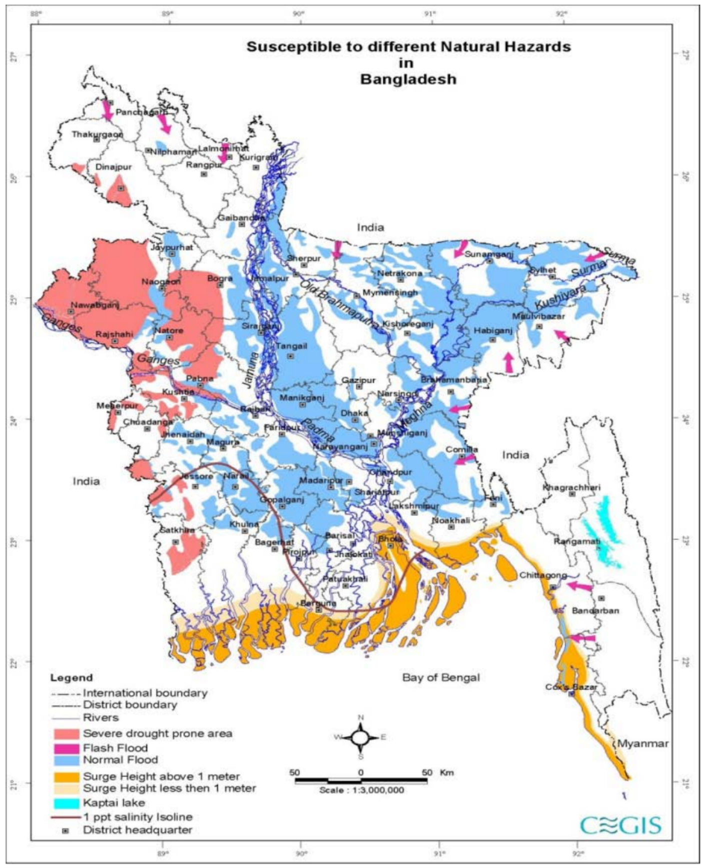
Figure 1: Geographical location of Bangladesh

[9]

---

[9] Encyclopædia Britannica, inc. (n.d.). *Bangladesh*. Encyclopædia Britannica. Retrieved December 5, 2021, from https://www.britannica.com/place/Bangladesh.

Figure 2: Vulnerability to Different Natural Hazards



[10]

Nash-Cutliffe efficiency coefficient:

$$NSE = 1 - \frac{\sum_{t=1}^{T} \left(Q_o^t - Q_m^t\right)^2}{\sum_{t=1}^{T} \left(Q_o^t - \overline{Q}_o\right)^2}$$ [11]

---

[10] Dastagir, M. R. (2015). Modeling recent climate change induced extreme events in Bangladesh: A Review. *Weather and Climate Extremes*, *7*, 49–60. https://doi.org/10.1016/j.wace.2014.10.003

[11] Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — a discussion of Principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Where: $\overline{Q_0}$ $is\ the\ mean\ of\ observed\ data,\ Q_m\ is\ modeled\ data,\ and\ Q_0^t\ is\ observed\ data\ at\ time\ t$

Root Mean Squared Error:

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

[12]

Statements of Contribution:

- **David Campos:** Responsible for developing the research question as well as revising/editing the final report. Added visuals to the slide deck and helped to compile audio recording. If we were to do this project again, we would ensure the structure of our project is sound and have group alignment before moving forward.
- **Kolby Devery:** Helped to flesh out the initial concept and literature review. Also worked on the data and study design of the project. So far, we have been able to work pretty efficiently and delegate out tasks and deliverables. The most challenging part of the design has been arriving at a proper scope and a well-defined research question. We have been at risk of scope creep and constantly need to be cognizant of going beyond the scope of the paper. If we were to do it again, it would be helpful to ensure we have a clear and narrow scope from the outset.
- **Redwan Hussain:** I helped conduct an initial literature review to finalize the scope and target geography for the project. Additionally, I contributed to developing the study design. I attribute our success to being effective at conducting sprint-like meetings. Given more time, I would have opted for a narrower scope and focused on only one of the infrastructure projects.
- **Chun Him Cheung:** Wrote the overview, helped develop concepts for analysis and review the history of multilateral institutions and their investment strategies in Bangladesh. Given more time, we would have started with a narrower project scope.

---

[12] Willmott, C. J., & Matsuura, K. (2006). On the use of dimensioned measures of error to evaluate the performance of Spatial Interpolators. *International Journal of Geographical Information Science*, *20*(1), 89–102. https://doi.org/10.1080/13658810500286976