Capstone 3 Project Proposal

By Kealie Pretzlav

Problem Statement

Floods occur every year all over the United States, impacting human life, local economics, and infrastructure. Climate change projections indicate that floods will likely increase in frequency and magnitude in the coming years. Management of flood infrastructure, early warning systems, and dam spillway operations could all greatly benefit from flood likelihood predictions. Long-term flood probabilities over the next several decades can help prioritize flood infrastructure investment and aid in the design of robust mitigation strategies. I will combine historical weather predictions, and precipitation records to predict flood in the San Lorenzo River and to answer the following questions:

1. Can machine learning be used to accurately predict flooding in the near future on the San Lorenzo River? Can it predict whether the river will flood or not (ie categorical model)? Or, can I accurately predict the numerical flow rate?
2. Can climate projections be used to design a series of “what if” flooding scenarios to be evaluated against the model(s)? For example, moderate winter storm, large spring storm, under climate change projections for RCP 4.5 and 8.5.

Datasets

To complete this project, I will use three primary data sources, USGS flow records of the San Lorenzo River, historical precipitation forecast, historical rainfall data, and long-term projections of precipitation.

San Lorenzo River Flow

The USGS has an extensive network of historical stream gages all over the country. Data collection techniques are highly standardized and generally produce very high quality data. USGS Flow records are the national standard for collecting, calibrating, and cleaning historical flow data. The San Lorenzo River has one of the longest records in the state of California, beginning in 1936, which will provide an excellent long-term historical record for characterizing climate change.

15-minute data (1987 to present) and daily data (1936 to present) at two locations:
In Santa Cruz:
https://waterdata.usgs.gov/monitoring-location/11161000/#parameterCode=00065&period=P7D&showMedian=false

At Big Trees:
https://waterdata.usgs.gov/monitoring-location/11160500/#parameterCode=00065&period=P7D&showMedian=false

Historical Precipitation Record

There are two precipitation gages with long-term records in the San Lorenzo River watershed:
Ben Lomond:
https://www.ncei.noaa.gov/access/past-weather/USC00040673/data.json

And Santa Cruz:
https://www.ncei.noaa.gov/access/past-weather/USC00047916/data.json

Short-term precipitation forecasts

I will predict flooding in the San Lorenzo River using the short term weather prediction. For each day of the model simulation, I will use the prediction for up to three days in the future to predict whether the San Lorenzo River will flood in the next three days. Two potential sources of this data include the National weather service and Weather Underground.

https://www.weather.gov/documentation/services-web-api
https://www.wunderground.com/history

Climate change projections

To simulate changing precipitation patterns, I will use RCP 4.5 and RCP 8.5 as “best” and “worst” case scenarios of changing climate, sourced from Cal-adapt

https://analytics.cal-adapt.org/

Other Data

Other relevant data may be added to the feature set as needed. For example, the number of years since a wildfire occurred in the watershed may be a significant feature for flood prediction.

Modeling Methods and Final Products

I will first develop and time series model to predict flooding in the San Lorenzo River. The target variable will be flow at one of the stream gage locations. I may begin with a simplified categorical model, predicting flood or not flood at the gage location, converting the flow record

into a categorical target variable using a flow threshold for flooding. Next, a similar model will be used to predict flow rate. The timeseries model will take into account the previous flow rate, the precipitation forecast, and previous actual precipitation perhaps over a short period of time or as a cumulative annual total. Finally, model results will be interpreted in the context of the climate change projection data for RCP 4.5 and 8.5 as “best” and “worst” case scenarios. The final products for this project will be a report and slide presentation.

Possible Constraints

Dimensionality of the model could get overly complicated, especially when using both the historical precipitation record, and the precipitation forecast out several days. One major constraint for understanding projections of flooding into the future is that we must assume that the historical model is representative of the conditions we expect to see.