Machine Learning for Generalizable Prediction of Flood Susceptibility

Dylan J. Fitzpatrick [*1] **Chelsea Sidrane** [*2] **Andrew Annex** [*3] **Diane O'Donoghue** [4] **Yarin Gal** [5] **Piotr Bilinski** [6]

Abstract

Flooding is a destructive and dangerous hazard and climate change appears to be increasing the frequency of catastrophic flooding events around the world. Physics-based flood models are costly to calibrate and are rarely generalizable across different river basins, as model outputs are sensitive to site-specific parameters and human-regulated infrastructure. Statistical models trained primarily from remotely-sensed Earth observation data could reduce the need for extensive in-situ measurements. In this work, we develop generalizable, multi-basin models of river flooding susceptibility using geographically-distributed data from the USGS stream gauge network. Machine learning models are trained in a supervised framework to predict two measures of flood susceptibility from a mix of river basin attributes, impervious surface cover information derived from satellite imagery, and historical records of rainfall and stream height. We report prediction performance of multiple models using precision-recall curves, and compare with performance of naive baselines. This work on multi-basin flood prediction represents a step in the direction of making flood prediction accessible to all at-risk communities.

1. Introduction

Among natural disasters, flooding is one of the most destructive, dangerous, and common hazards. In the U.S., 75% of all Presidential disaster declarations are associated with flooding, and floods cause an average of $6 billion in property damage per year [13, 18]. Additionally, climate change appears to be exacerbating the incidence of catastrophic flooding events [16]. If we can improve the access to flood prediction systems, lives could be saved and damage lessened. In this work, we apply machine learning to the problem of forecasting river flooding hazards.

The physics-based models that the National Oceanic and Atmospheric Administration (NOAA) uses to predict river levels and issue official flood warnings are time-consuming and costly to use. These models rely on field observations and calibration to small geographic areas and consequently their outputs are not generalizable across different river basins. If flood prediction can be done without reliance on these expensive *in-situ* measurements it could reduce costs for existing flood prediction systems as well as expand flood prediction efforts to areas that could not previously afford river gauging infrastructure.

Existing work using machine learning methods to predict flood susceptibility has also been constrained to small geographic areas. Khosravi et al. [9] predict flood susceptibility in a single watershed, and Tayfur et al. [15] predict a hydrograph for a single stretch of river. Assem et al. [1] predict river levels at 3 stations in a single catchment. Work by Kratzert et al. [10] on estimating runoff as a function of rainfall using LSTMs suggests that machine learning models are able to generalize across many catchments in a manner that traditional models cannot, providing promising evidence that statistical flood susceptibility models need not be limited to small geographic areas.

The work presented here explores prediction performance of generalizable, multi-basin models on two measures of flood susceptibility using data from six U.S. states across more than six river basins: South Dakota, Nebraska, South Carolina, Virginia, New York, and New Jersey. The models trained in this work rely on USGS stream gauge data for ground truth and as inputs for a subset of our experiments, but also incorporates remotely-sensed data and is a step in the direction of a flood prediction model that is less dependent on *in-situ* measurements.

2. Data

To predict flood susceptibility, we rely on information sourced from a mix of satellite-derived and in-situ data. The USGS stream gauge network provides river height measurements at 15 minute intervals [17]. Data collected between 2009 and 2019 was used. Flood thresholds at stream gauge

[*]Equal contribution [1]Carnegie Mellon University [2]Stanford University [3]Johns Hopkins University [4]kx [5]University of Oxford [6]University of Warsaw. Correspondence to: Dylan J. Fitzpatrick <djfitzpa@cmu.edu>.

locations are determined by NOAA for four separate flood categories [12]. The ‘minor flood’ threshold was used to binarize the USGS stream gauge readings for experiments that involve predicting flood occurrence. This resulted in a dataset where 5.5% of the data points indicated that flooding occurred.

Records of time to peak river levels after precipitation events were obtained from the Flooded Locations & Simulated Hydrographs (FLASH) Project [5]. The time-to-peak data was split into 4 bins with roughly equal frequency, a time to peak of less than 3.12 hours, between 3.12 and 7.44 hours, between 7.44 and 18 hours, and greater than 18 hours, and assigned categorical labels indicating bin number. For historical rainfall data, daily total precipitation from the PRISM climate data set is used [14]. At each gauge location, river basin attributes affecting regional surface runoff and groundwater drainage are obtained from the EPA StreamCat dataset [8]. Average upstream impervious surface cover at gauge locations is calculated from the satellite-derived National Land Cover Dataset (NLCD) [11]. We include elevation at each river gauge location by overlaying Shuttle Radar Topography Mission (SRTM) 30-meter resolution data [19], and include a characteristic length parameter describing the scaling relationship between channel slope and drainage area [6]. All input data including rainfall, basin characteristics, elevation information, and impervious surface cover is normalized by feature to fall approximately in the range $[0, 1]$ using the training set statistics. Each stream gauge had up to 10 years of data at 15 minute intervals which was summarized into monthly statistics. The entire record of each such location was randomly assigned to either the training set, validation set, or the test set to achieve a 60-20-20 training, validation, and test split.

3. Methodology

We train statistical models to predict two separate measures of flood susceptibility: (1) binary flood occurrence at gauge locations within a given month (approx. 50,000 gauge-months), and (2) time to peak river level after precipitation events (approx. 3000 events). These two measures of flood susceptibility are indicated on a plot of river height over time for a single location in Figure 1, where flood occurrence indicator is triggered once the river height crosses above a fixed flood threshold. Taken together, these two prediction targets provide critical information for both long-term and regional-scale flood planning, and short-term planning for localized flash flooding following extreme rainfall events. In Section 4, we report results for a single time-to-peak bin (greater than 18 hours from start of precipitation to peak river level).

A key objective of this work is to explore the relative importance of different features for making accurate predictions. In particular, how critical are recent historical records of river height and accurate rainfall forecasts for predicting future flood susceptibility? These questions can help us understand the relative feasibility of statistical flood prediction for an ungauged river basin and for an area where accurate rainfall forecasts aren’t available. To answer these questions, we construct three experiments that include different groups of features for the statistical models. Experiment 1 addresses the problem of predicting flood susceptibility in ungauged locations along a river network. For these experiments, no information on prior river levels is provided as input the predictive models. Experiment 2 assumes a well-gauged river basin, and forecasts flood susceptibility using historical data on river height at prediction locations. Finally, Experiment 3 provides an informative upper bound on prediction performance in the presence of accurate rainfall forecasts by including true rainfall observations as an input to predictive models.

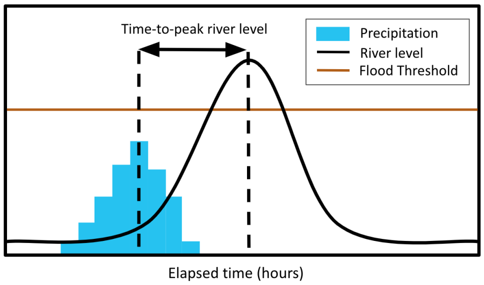
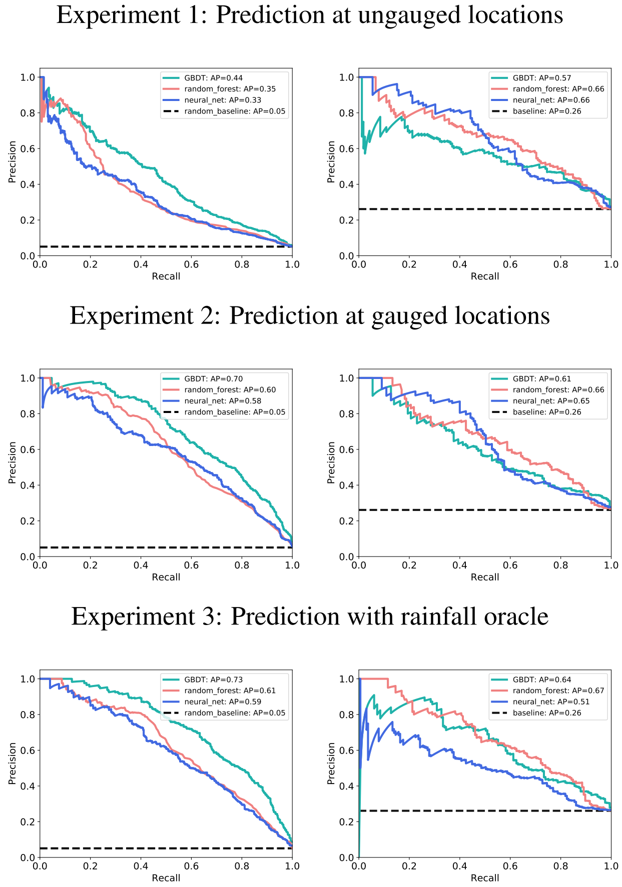


Figure 1. River height for a single stream gauge location in the period during and following a precipitation event.

Three classification models are trained in a supervised framework to predict both measures of flood susceptibility: a random forest classifier [2], gradient boosted decision trees (GBDT) [3], and a multilayer perceptron (MLP) with ReLU activations [7]. For all models considered, hyper-parameters are tuned using a held-out validation set and models are evaluated on a held-out test set.

4. Results

The results of all three experiments are shown with precision-recall (PR) curves in Figure 2. PR curves allow a decision-maker to choose where to threshold the probability produced by a classifier in order to produce a desired trade-off between false positives (predicted floods that do not occur) and false negatives (actual floods that were not predicted). The machine learning models are compared to a random baseline classifier that would assign a random probability of flooding to each example, giving a PR curve that is a horizontal line at a precision corresponding to the fraction of positive examples in the test data.


Experiment 1: Prediction at ungauged locations

Experiment 2: Prediction at gauged locations

Experiment 3: Prediction with rainfall oracle

Figure 2. Precision and recall curves for four model classes predicting monthly flood occurrence (left) and time to peak river level (right). Results are reported for three experiments: prediction at ungauged locations (top), prediction at gauged locations (middle), and prediction at gauged locations with a rainfall oracle (bottom). A gradient boosted decision tree (GBDT), random forest classifier, and multi-layer perception are compared to a random baseline.

Figure 2 shows that all models considered outperform the random baseline. Across all three gauge-month experiments, the GBDT is the best classifier in terms of Average Precision (AP). This is likely due to the gradient boosting on difficult-to-classify examples, which makes the GBDT well-suited for an imbalanced data set of rare flooding events. For classifying time-to-peak bins, all models (excluding the random baseline) demonstrate roughly similar performance. This may be because the time-to-peak task uses a more balanced dataset than the gauge-month experiments – 25% positive class labels compared to 5.5% positive class labels.

We also compare predictive performance to the operational flood predictions released by NOAA and collected in the Iowa Environmental Mesonet (IEM) database [4]. We evaluate approximately 4,100 gauge-month predictions by the Northeast River Forecast Center across 54 gauging locations and aggregate three-day-lookahead river level forecasts to produce monthly flood occurrence statistics. We find that NOAA's forecasts have precision of 0.5 and recall of 0.245. This is not directly comparable to our PR curves as a different subset of the data is evaluated, but if this trend is consistent, our best models represents a more than twofold improvement in the proportion of monthly flood events that are predicted ahead of time.

5. Conclusions & Future Work

These preliminary results provide promising evidence that multi-basin flood prediction with statistical models is possible and is deserving of additional research despite the fact that such an approach has not yet been used in prior work to the best of our knowledge. We believe that the incorporation of additional remotely-sensed data streams and more sophisticated machine learning techniques has the potential to produce even higher quality multi-basin flood prediction models. Further, while this work focuses on six geographically-distributed states, adding data from stream gauge locations throughout the conterminous U.S. can potentially improve generalizability of the trained models. In summary, our work provides a proof-of-concept that multi-basin flood prediction using statistical models and remotely-sensed data has considerable value, and can overcome many of the shortcomings of physics-based flood prediction models such as reliance on time-consuming calibration to small geographic areas. Regional-scale flood susceptibility models increase the accessibility of accurate and cost-effective disaster planning for populations at risk of experiencing flooding anywhere in the world.

Future work includes training and testing our model using data from the entire U.S., and comparing this to flood predictions from NOAA for the entire U.S. We also plan to closely analyze the geographic areas and climatological and topographical conditions where our predictions perform significantly differently from NOAA's predictions. A flood prediction model that is fast to deploy over large geographic areas and trained using remotely-sensed data would be extremely valuable for helping people around the globe prepare for flooding events, and would only become more useful as climate change increases the frequency and impact of severe flooding events.

6. Acknowledgements

This research was conducted at the NASA Frontier Development Lab (FDL). We gratefully acknowledge the support from NASA, Google Cloud, Intel, USGS and Kx. We also gratefully acknowledge the support of Baishali Chaudhury, Jonathan Stock, Theo Barnhart, Jack Eggleston, and Guy Schumann.

References

[1] Assem, H., Ghariba, S., Makrai, G., Johnston, P., Gill, L., and Pilla, F. Urban water flow and water level prediction based on deep learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 317–329. Springer, 2017.

[2] Breiman, L. Random forests. *Machine Learning*, 45 (1):5–32, 2001.

[3] Chen, T. and Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 785–794, 2016.

[4] Department of Agronomy, I. S. U. Iowa environmental mesonet. URL https://mesonet.agron.iastate.edu/wx/afos/list.phtml.

[5] FLASH. FLASH Flood Observation Data, USGS Event Data, 2016. URL https://blog.nssl.noaa.gov/flash/database/.

[6] Giachetta, E. and Willett, S. D. A global dataset of river network geometry. *Scientific Data*, 5, 2018. URL https://doi.org/10.1038/sdata.2018.127.

[7] He, K., Zhang, X., Ren, S., and Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1026–1034, 2015.

[8] Hill, R. A., Weber, M. H., Leibowitz, S. G., Olsen, A. R., and Thornbrugh, D. J. The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States, 2016.

[9] Khosravi, K., Pham, B. T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., Prakash, I., and Bui, D. T. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at haraz watershed, northern iran. *Science of the Total Environment*, 627:744–755, 2018.

[10] Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M. Rainfall–runoff modelling using long short-term memory (lstm) networks. *Hydrology and Earth System Sciences*, 22(11):6005–6022, 2018.

[11] Multi-Resolution Land Characteristics Consortium. National land cover dataset, percent developed imperviousness, 2016. URL https://www.mrlc.gov/data.

[12] National Weather Service. Advanced hydrologic prediction service, 2019. URL https://water.weather.gov/ahps/.

[13] National Weather Service. Flood Related Hazards, 2019. URL https://www.weather.gov/safety/flood-hazards.

[14] PRISM Climate Group, O. S. U. Prism climate data, 2019. URL http://www.prism.oregonstate.edu/.

[15] Tayfur, G., Singh, V., Moramarco, T., and Barbetta, S. Flood hydrograph prediction using machine learning methods. *Water*, 10(8):968, 2018.

[16] Universite Catholique de Louvain - Centre for Research on the Epidemiology of Disasters. EM-DAT: The International Disaster Database, 2019. URL https://www.emdat.be/.

[17] USGS. National water information system, 2019. URL https://waterdata.usgs.gov/nwis/rt.

[18] USGS. Flood Hazards - A National Threat, 2019. URL https://pubs.usgs.gov/fs/2006/3026/2006-3026.pdf.

[19] USGS Earth Resources Observation and Science (EROS) Center. Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global (30 meters), 2014.