Creating Flash Flood Models of New Jersey Using Random Forest Machine Learning

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

Sean Lamb, AS | EJB School of Planning and Public Policy

Introduction

Background Context

Prioritizing people and property:

Flash floods endanger health, property, and mobility, in particular on households with less institutional, economic and social privilege. In the United States, floods kill more people than tornadoes, hurricanes or lightning (FEMA, 2022). Floods are the most frequent natural disaster (WHO, 2021) which can cause power outages, pollute drinking water, displace people from their homes, damage community infrastructure which can negatively affect a person's life, especially if not aware of the unfortunate circumstances.

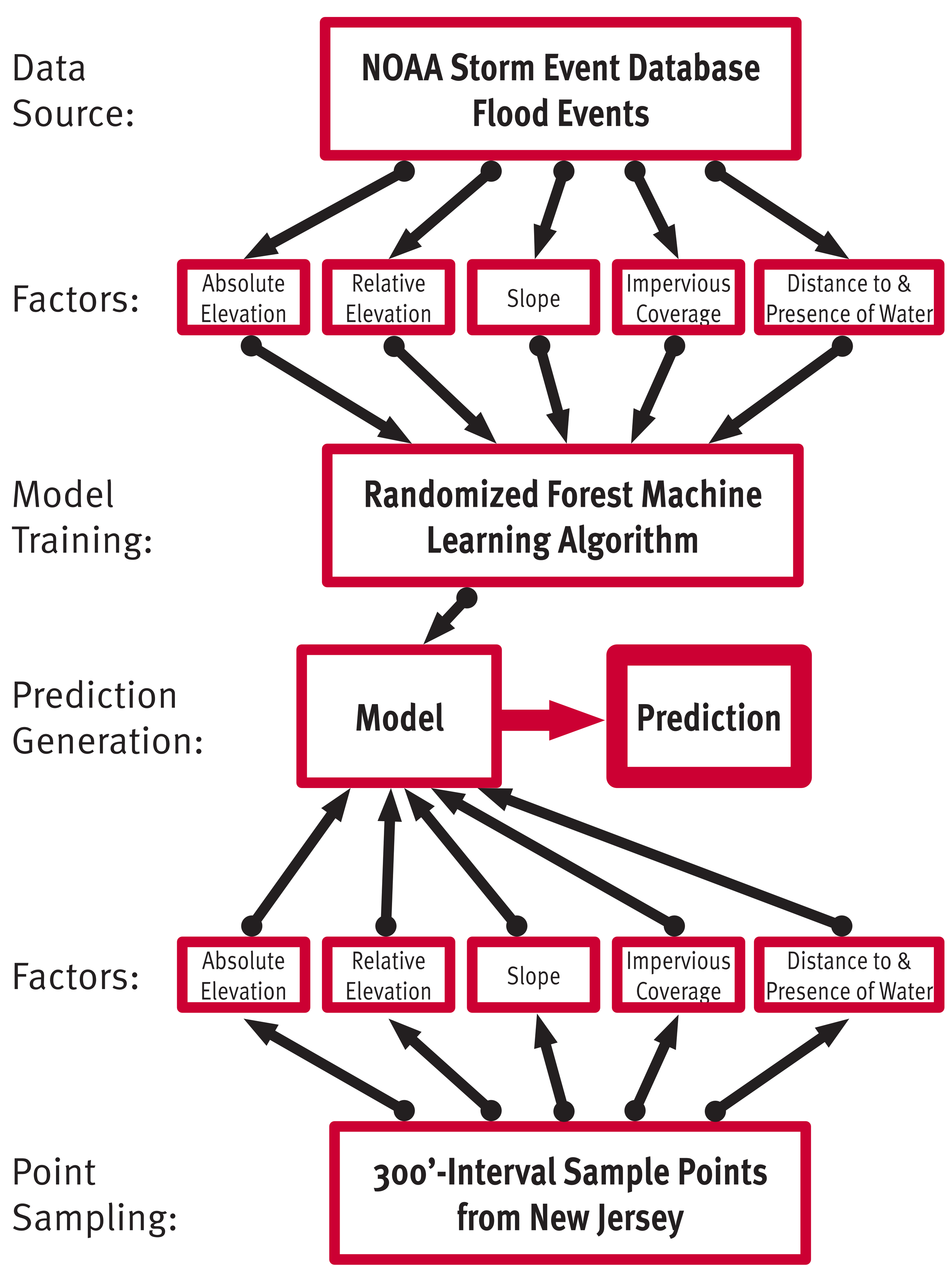
'Bottom-up' modeling with machine learning:

Flood risk modeling typically uses fixed inputs from the physical environment and weather to generate predictions. In this project, we work from the 'ground up' using sources of reported flood incidents. This builds on literature using crowdsourced data on flash flood hotspot and other hazard events, such as from 311 sources (Liu et al, 2024) or waze data (Esparza et al 2023). Others use community and crowdsourced data primarily to check errors and groundtruth flood forecasting data (e.g. Puttinaovarat and Horkaew, 2020).

Across flash flood modeling with similar datasets and challenges, variations on the random forest model have often been found to be most effective (Band et al 2020; Liu 2024), or at least relatively similar in accuracy to other approaches (Arabameri 2020; Hosseini et al 2020).

Methodology

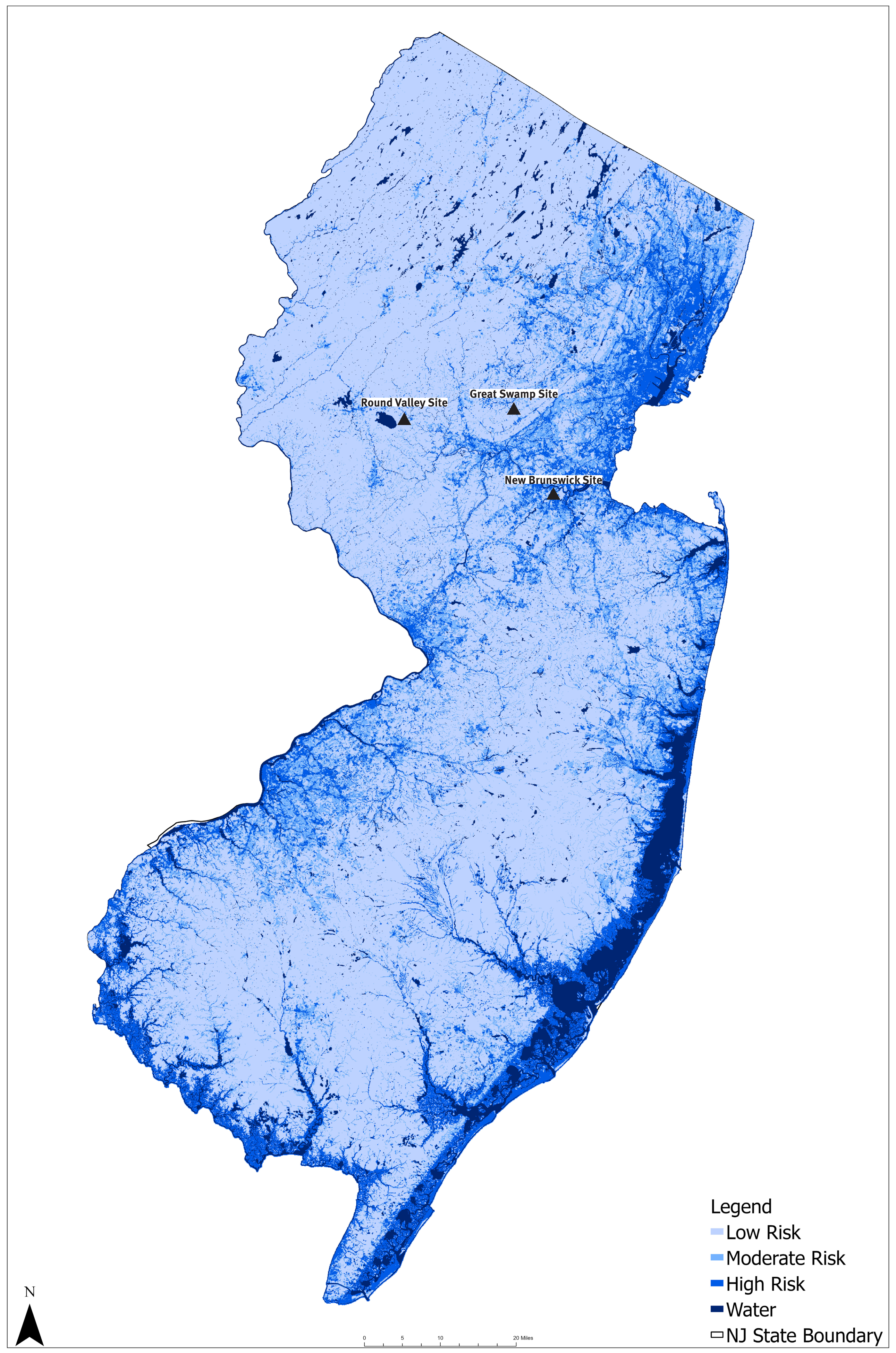
Model Generation



Prediction-Map Generation

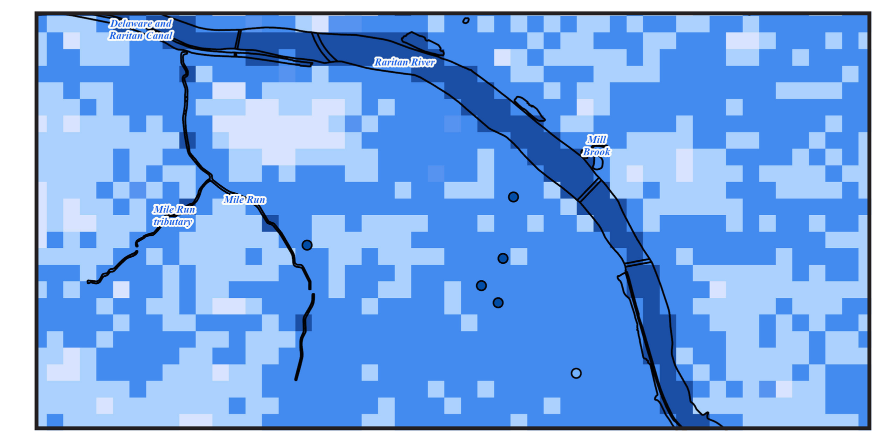
Results

Full Map

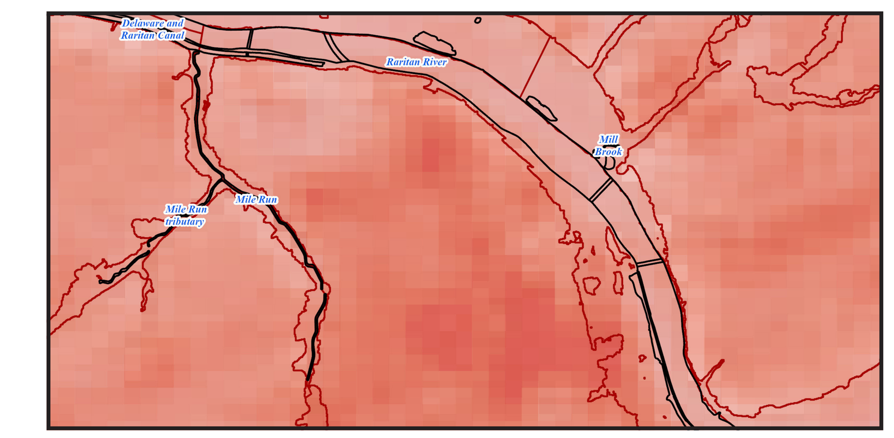


Selected Highlights

Raritan River / New Brunswick

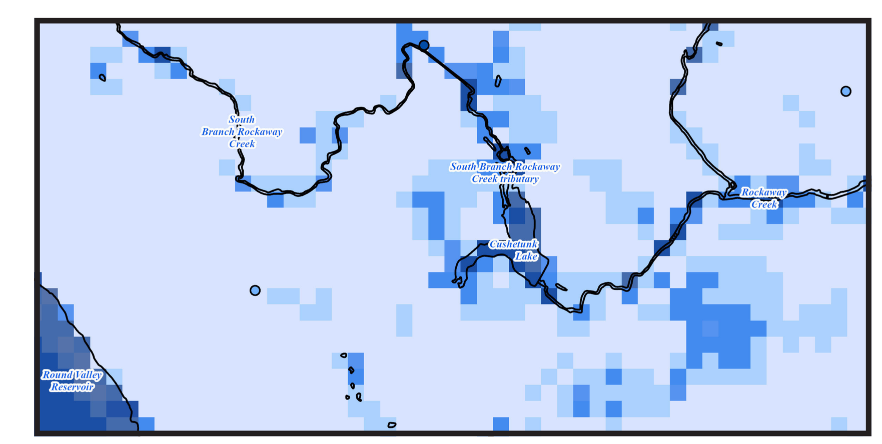


Flood report prediction map with flood report data

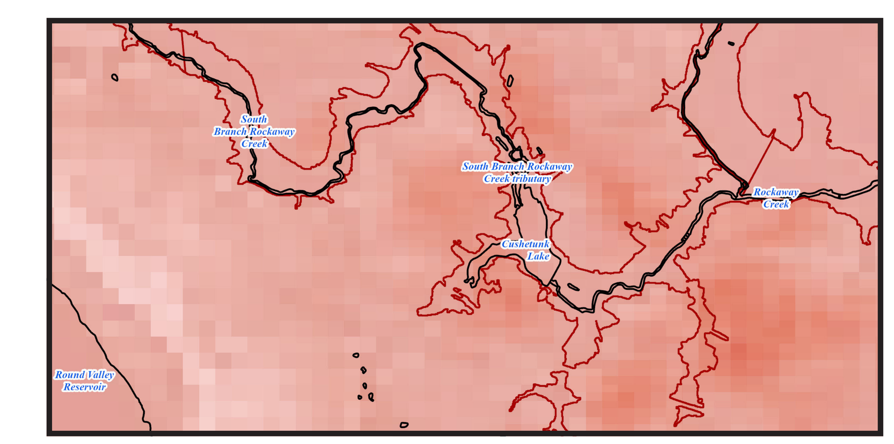


Factors map (combined relative elevation, slope, impervious coverage) and FEMA flood zones

Round Valley Reservoir / Whitehouse Station

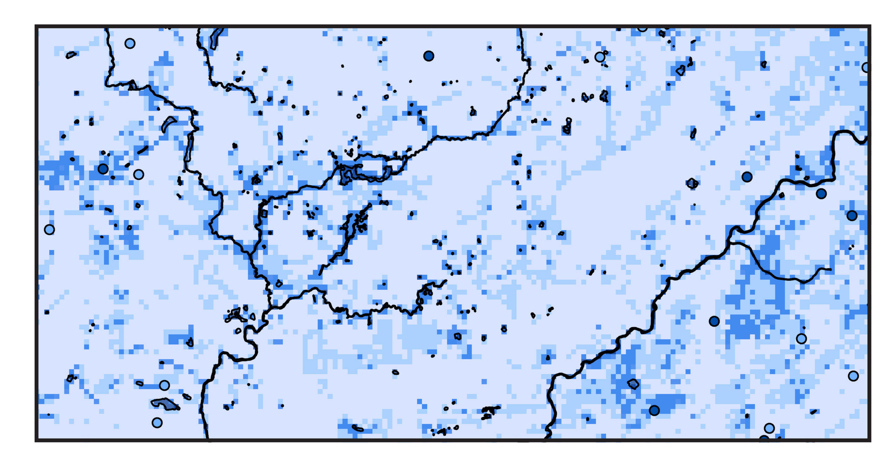


Flood report prediction map with flood report data

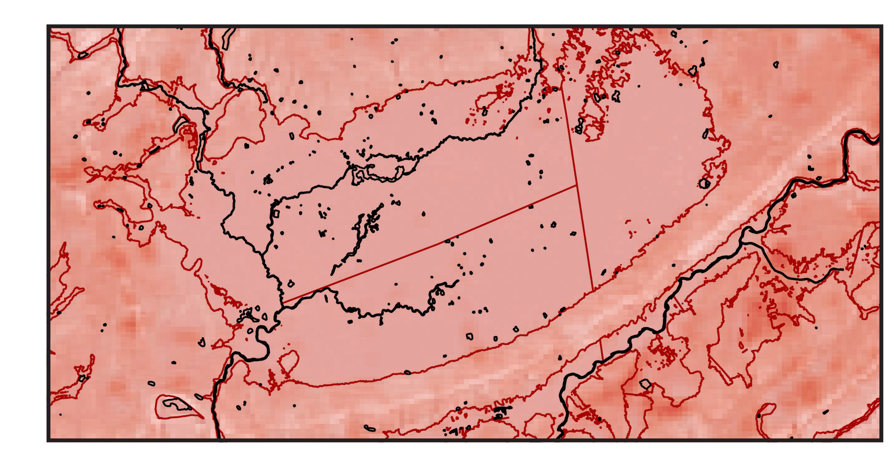


Factors map (same as above) and FEMA flood zones

Great Swamp National Wildlife Refuge



Flood report prediction map with flood report data



Factors map (same as above) and FEMA flood zones

Legend

Flood Report Data
- Non-Flood Point
- Flood Point
- Low Risk
- Moderate Risk
- High Risk
- Water

- NJ State Boundary
- Waterbody 2015 (NHD)
- FEMA Flood Zone +3 FT (Rutgers)

Elevation: 1790.6 – -116.352

Relative Elevation: 167.528 – -166.121

Slope: 728.092 – 0

Impervious Coverage Percentage: 100 – 0

Final Notes

Discussion

Potential policy and advocacy uses
- Guide investments for stormwater infrastructure; DCIA (impervious area with drainage) may matter more than simply the impervious ratio in high-rainfall events (Sohn et al 2020).
- Help community organizations in underserved areas advocate for other forms of assistance
- Assist public health targeting of programs (e.g. mold abatement)
- Help organizations or local governments 'flag' dangerous areas that have recurrent flooding

Limitations
- The model depends on reported flood events, therefore it is biased towards dense areas
- The non-flood points are also dependent on the existence, or lack of, flood reports
- There are additional factors which were not considered for this model, but could be incorporated into one in the future, such as soil absorbtion and drainage systems

Future Research
- The accuracy of the model can be tested by comparing flood events to the prediction map.

Conclusion

This project highlights the critical importance of prioritizing people and property in flood risk assessment, particularly for vulnerable communities. Floods are a frequent and deadly natural disaster, causing widespread harm to health, property, and mobility. By adopting a 'bottom-up' approach to flood modeling, this project leverages crowdsourced data to improve prediction accuracy and capture the localized impact of flash floods. Using reported flood incidents and applying machine learning techniques, especially those rooted in random forest methodologies, this approach offers an enhanced, data-driven method for assessing flood risk and informing protective actions.

References

Arabameri, Alireza, Sunil Saha, Wei Chen, Jagabandhu Roy, Biswajeet Pradhan, and Dieu Tien Bui. "Flash flood susceptibility modelling using functional tree and hybrid ensemble techniques." Journal of Hydrology 587 (2020): 125007.

Band, Shahab S., Saeid Janizadeh, Subodh Chandra Pal, Asish Saha, Rabin Chakrabortty, Assefa M. Melesse, and Amirhosein Mosavi. "Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms." Remote Sensing 12, no. 21 (2020): 3568

Brown, V., Fauver, S., Geppi, D., Haynes, A., Sprague-Hilderbrand, J., Jacobson, R., Klockow, K., Nagele, D., Rouleau, T., & Wowk, K. (2016, July). Risk Communication and Behavior: Best Practices and Research Findings. Repository. library.noaa.gov. https://repository.library.noaa.gov/view/noaa/29484

Esparza, Miguel, Hamed Farahmand, Samuel Brody, and Ali Mostafavi. "Examining data imbalance in crowdsourced reports for improving flash flood situational awareness." International journal of disaster risk reduction 95 (2023): 103825.

FEMA. (2022). Protective Actions Research. Fema.gov. https://community.fema.gov/ProtectiveActions/s/article/Flood-Impact

Hussain, Z., & Zetkulic, A. (2021, August 11). Protecting and Empowering Communities during Disasters. RMI. https://rmi.org/community-resilience-hubs/

Hosseini, Farzaneh Sajedi, Bahram Choubin, Amir Mosavi, Narjes Nabipour, Shahaboddin Shamshirband, Hamid Darabi, and Ali Torabi Haghighi. "Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method." Science of the total environment 711 (2020): 135161.

Li, Q., & Lin, Y. (2023). How Can Community-Based Organizations Improve Flood Risk Communication? A Case Study of China Based on Grounded Theory. Systems, 11(2), 53. https://doi.org/10.3390/systems11020053

Liu, Z., Felton, T., & Mostafavi, A. (2024). Interpretable machine learning for predicting urban flash flood hotspots using intertwined land and built-environment features. Computers, Environment and Urban Systems, 110, 102096.

Mackinnon, J., Heldsinger, N., & Peddle, S. (2018). A COMMUNITY GUIDE TO EFFECTIVE FLOOD RISK COMMUNICATION Promoting Personal Preparedness. https://uwaterloo.ca/partners-for-action/sites/default/files/uploads/files/a_community_guide_to_effective_flood_risk_communication_24jan18_1.pdf

Mcdermott, R. (n.d.). Knowing in Community: 10 Critical Success Factors in Building Communities of Practice The Limits of Knowledge Management. https://crowdsynergy.wdfiles.com/local--files/community-of-practice/KnowingIn-Community.pdf

Parker, D. J. (2004). Designing flood forecasting, warning and response systems from a societal perspective. Meteorologische Zeitschrift, 13(1), 5–11. https://doi.org/10.1127/0941-2948/2004/0013-0005

Puttinaovarat, Supattra, and Paramate Horkaew. "Flood forecasting system based on integrated big and crowdsource data by using machine learning techniques." IEEE Access 8 (2020): 5885-5905.

World Health Organization. (2021). Floods. Www.who.int; World Health Organization. https://www.who.int/health-topics/floods#tab=tab_1

Acknowledgements

This poster was created as part of a larger research project with a team, consisting of Amara Thompkins, Eric Yang, and Carrie Ward, under the guidance of Bernadette Baird-Zars, Ph.D.