Creating Global Flow Maps using Machine Learning and Free Remote Sensing Data: A Novel Approach and Application in Unmapped Areas

Héctor Leopoldo Venegas-Quiñones[1], Pablo García-Chevesich[2], Rodrigo Valdés-Pineda[1], Ty P.A. Ferré[1], Hoshin Gupta[1], Derek Groenendyk[1], and Laura Bakkensen[1].

[1]University of Arizona, Hydrology and Water Resources Department, 1133 E James E. Rogers Way, Tucson, AZ, The United States

[2] Colorado School of Mines, Department of Civil and Environmental Engineering, Denver, CO, The United States

Abstract

This study developed a Random Forest model to estimate flood risk in Arizona, New Mexico, and Utah by integrating topographic and weather data. The study demonstrated the advantages of this approach, including its ability to estimate flood risk in areas lacking NFHL classifications and provide a comprehensive picture of the overall impact of weather on flood risk in each area. The methodology presented in this study provides a useful framework for assessing flood risk in other regions and highlights the potential for integrating different types of data to provide a comprehensive picture of flood risk.

Introduction

Floods have become increasingly frequent and severe over time, causing significant impacts on the environment, infrastructure, and human life worldwide (Tellman et al., 2021). The primary drivers of this trend include climate change, urbanization, and deforestation, as well as changes in precipitation patterns and land use (Burn & Whitfield, 2023). Flood events can result in severe consequences, including economic losses, property damage, loss of life, and displacement of communities. To effectively manage the risks associated with flooding, it is crucial to have accurate and up-to-date flood maps that can inform decision-making at all levels (Chen, 2022).

Flood maps provide critical information on the location and extent of flood-prone areas, as well as the potential severity of flood events, enabling informed decisions on land use planning, infrastructure design, and emergency response (Mudashiru et al., 2021). Moreover, flood maps can help to raise awareness among the public and encourage proactive measures to mitigate flood damage. To create increasingly detailed and precise flood maps, advances in technology and data collection have enabled the collection of data on factors such as land topography, precipitation patterns, and soil permeability (Chaminé et al., 2021). Therefore, the development of flood maps should be a priority for governments and stakeholders worldwide to provide accurate and up-to-date information on flood risk, especially in the light of increasing frequency and severity of flood events (Kam et al., 2021).

In the United States, the Federal Emergency Management Agency (FEMA) is responsible for creating flood maps to manage flood risks (Mudashiru et al., 2021). FEMA employs a variety of methodologies and classifications, including the Flood Insurance Rate Map (FIRM) to identify areas at risk of flooding based on elevation, land use, and proximity to waterways (National Research et al., 2009). FIRMs are based on various sources of data, such as topographical maps, aerial photography, and hydrologic models. The cost of creating and maintaining these maps is influenced by factors such as the level of detail required, the size and complexity of the area being mapped, and the availability of data and resources. Acquisition and analysis of data on hydrology, topography, and precipitation patterns are significant portions of the cost of creating these maps. FEMA employs experts in fields such as geology, hydrology, and engineering to analyze the data, which requires significant staff time and resources (National Research et al., 2009). It is estimated that creating national Flood Insurance Rate Maps costs FEMA as much as $11.8 billion, with annual maintenance costs ranging from $107 million to $480 million (Managers, 2020). Without accurate and up-to-date FIRMs, individuals and communities may not have the necessary information to make informed decisions about flood insurance, development, and emergency preparedness.

Accurate and up-to-date flood maps are essential for effective flood risk management, and stakeholders worldwide should prioritize their development and maintenance. In the United States, FEMA plays a crucial role in creating FIRMs, and the cost of creating and maintaining these maps is substantial. However, without accurate and up-to-date FIRMs, individuals and communities may not have the necessary information to make informed decisions about flood insurance, development, and emergency preparedness. Therefore, investment in the creation and maintenance of accurate and up-to-date FIRMs is crucial for protecting lives and property from flood damage. Figure 1 shows the geographic areas in the United States where FEMA has and has not yet developed Flood Insurance Rate Maps.

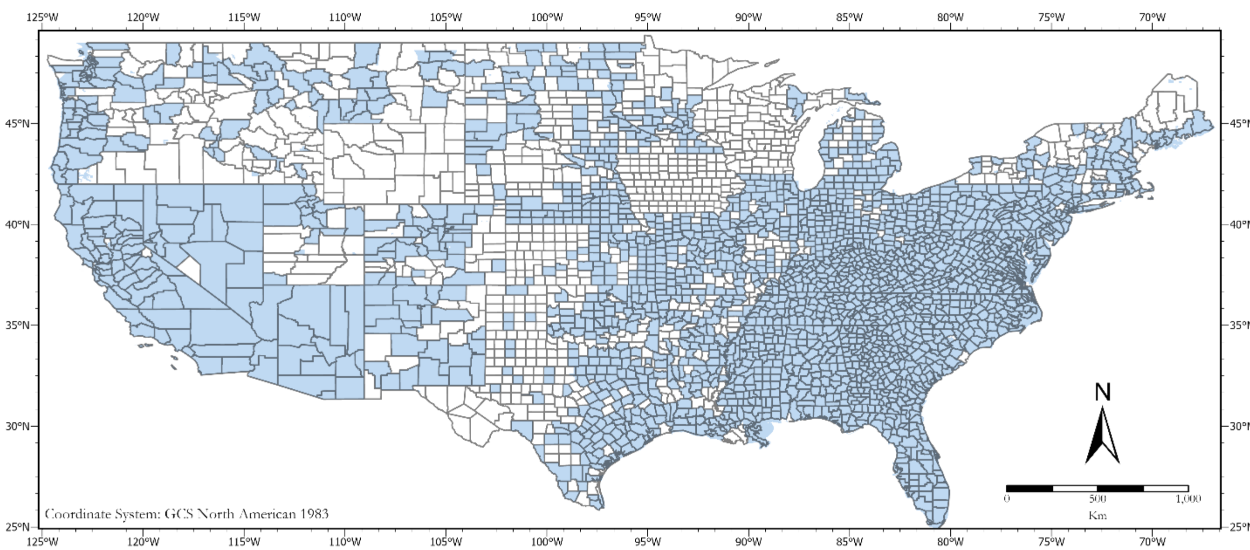

Figure 1. Geographic Distribution of FEMA Flood Insurance Rate Maps in the United States

The study applies Random Forest, a machine learning method, to produce accurate flood maps for areas without them. By analyzing data inputs like DEM and remote sensing variables, precise flood maps are created for Arizona. The freely available data through an interactive ArcGIS map could improve flood risk management, promote further investigation, and inform decision-making to minimize flood impacts.

Methodology

This study analyzes flood risk in Arizona using the National Flood Hazard Layer (NFHL) as the primary data source. The NFHL is a geospatial database developed by FEMA that provides up-to-date information on flood hazards across the United States. The study extracted and processed the data from the most recent version of the Arizona NFHL to create a high-resolution flood map. The map includes classifications of A, AE, AH, AO, D, X, and an "area not included" designation for all counties in Arizona. However, certain counties in New Mexico and Utah have not yet been classified by the NFHL. Figure 2(A) and (B) displays the NFHL classification for these states, highlighting areas with no flood information. The study's findings could assist local communities in making informed decisions and taking necessary steps to prepare for and minimize the impact of floods, ultimately leading to better flood risk management and increased safety for individuals living in flood-prone areas.

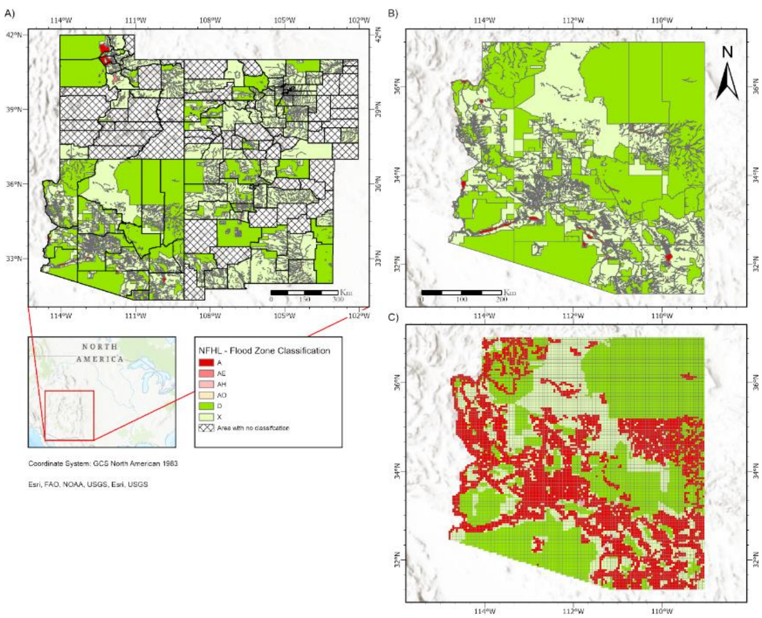

Figure 2. (A) National Flood Hazard Layer (NFHL) Analysis in Arizona, New Mexico, and Utah: Extracting and Processing Data to Create High-Resolution Flood Maps. (B) NFHL Classification for Arizona. (C) NFHL-Based Flood Risk Analysis in Arizona Using Grid Cell Classification for Arizona (worst-case threshold).

This study proposes an approach for predicting flood risk in areas without National Flood Hazard Layer (NFHL) classifications using topographic and weather information. Accurate flood risk assessments are essential for effective disaster risk management and emergency preparedness, especially in regions with limited flood hazard data. We utilize highly detailed USGS 3D Elevation Program (3DEP) Standard Digital Elevation Models (DEMs) to accurately predict flood risk in areas lacking NFHL classifications. The study integrates terrain parameters such as elevation, slope, aspect, hill shade, curvature, flow direction, and flow accumulation into a Random Forest Model to identify potential areas of water accumulation during flood events, thus improving the accuracy of flood risk assessments. Weather variable estimations from the TerraClimate dataset were integrated into the flood risk analysis to enhance understanding of the relationship between weather patterns and flood risk in each area, utilizing the most recent 2020 weather information available to ensure the analysis was based on up-to-date and relevant data. Also, we created a consistent grid with the same resolution as the USGS 3DEP Standard Digital Elevation Models. By spatially matching the climate data with other sources of information such as the NFHL classifications and DEM analysis, we calculated the climate variables for each cell in the grid. The highest NFHL classification within each cell was defined as the representation of the flood risk. The flood risk assessment applied four different thresholds, including the worst-case scenario, 5%, 10%, 25%, 50%, and 75%, based on the area of the classification in each cell and evaluated to select the most appropriate threshold for the study area. The selected threshold was used to improve the accuracy of the flood risk analysis and inform decision-making for disaster risk management. By merging the A and AE categories, the new merged category was designated as "A," which represents areas that exhibit a 1% annual chance of flooding. By using different thresholds, this approach helped to identify the most suitable level of flood hazard within each cell and enhanced the precision of flood risk predictions.

A Random Forest Classification approach was used to estimate the NFHL for each cell in Arizona, utilizing topographic and weather information as input to the model. The model was calibrated using Scikit-Learn's RandomizedSearchCV, and feature importance was evaluated using the Gini index to identify the most influential input variables for predicting flood risk.

Results

This study assessed flood risk in Arizona by dividing the state into standardized cells and assigning flood risk classifications based on various threshold settings. The study developed a Random Forest Classifier using topography and weather data for each cell, demonstrating high performance with an Accuracy Training of 1.0 and an accuracy performance exceeding 0.8 for each threshold setting. The Gini index was used to estimate the importance of variables in different scenarios, revealing the significance of climate-related variables and topographical features in predicting flood risk. The Random Forest Classifier was employed to create flood risk maps for Utah, Colorado, and New Mexico, utilizing a 25% scenario threshold setting (Figure 3). These maps provide crucial information for local authorities to make informed decisions regarding flood risk management and emergency preparedness. The study demonstrates the potential for this approach to be applied to other regions, providing a customizable and scalable solution for flood risk assessment and management.

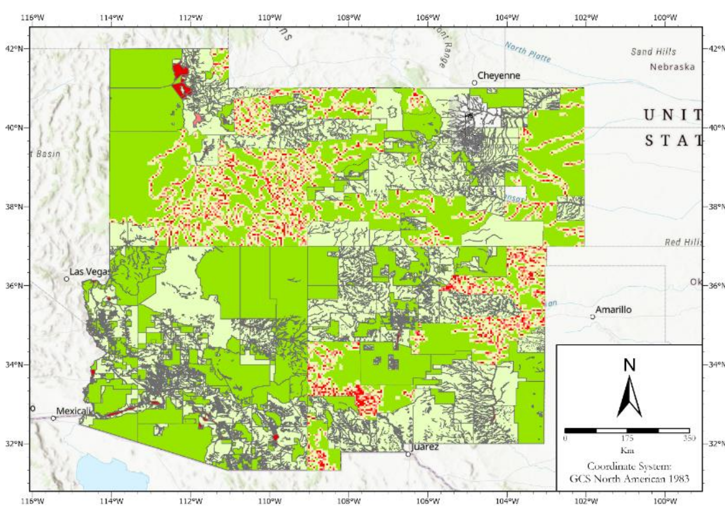

Figure 3. New Flood Risk Maps for Utah, Colorado, and New Mexico Based on Random Forest Classification Using Arizona Flood Risk Map Calibration

Our study created an interactive map accessible to the public, estimating flood hazard levels for each cell in Arizona, New Mexico, and Utah using topographic and weather data with a random forest model. The map can be used by local authorities, emergency management agencies, and residents to plan and prepare for potential flooding. Access the map at https://shorturl.at/jyKPY.

Discussion and Conclusion

This study developed a Random Forest-based model to estimate flood risk in Arizona, New Mexico, and Utah using topographic and weather data. The study demonstrated that this approach can estimate flood risk in areas lacking National Flood Hazard Layer (NFHL) classifications, and can provide a comprehensive picture of the overall impact of weather on flood risk in each area. The study has important implications for decision-makers in flood-prone areas, providing a valuable tool for planning and preparedness efforts. The methodology presented in this study could be extended to other regions and refined using other machine learning algorithms to improve accuracy and robustness. However, limitations include the accuracy of the input data and the potential inability of the model to capture all complex interactions between input variables and flood risk. The availability of the results through an ArcGIS interactive map makes the data accessible to the public, allowing local authorities, emergency management agencies, and residents to make informed decisions about flood risk. Overall, machine learning is a powerful tool that enhances the stakeholder process for flood risk management, enabling us to estimate flood risk in areas lacking accurate data and improve our understanding of the complex interactions between topography, weather patterns, and flood risk.

Bibliography

Burn, D. H., & Whitfield, P. H. (2023). Climate related changes to flood regimes show an increasing rainfall influence. *Journal of Hydrology*, *617*, 129075. https://doi.org/https://doi.org/10.1016/j.jhydrol.2023.129075

Chaminé, H. I., Pereira, A. J. S. C., Teodoro, A. C., & Teixeira, J. (2021). Remote sensing and GIS applications in earth and environmental systems sciences. *SN Applied Sciences*, *3*(12), 870. https://doi.org/10.1007/s42452-021-04855-3

Chen, Y. (2022). Flood hazard zone mapping incorporating geographic information system (GIS) and multi-criteria analysis (MCA) techniques. *Journal of Hydrology*, *612*, 128268. https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.128268

Kam, P. M., Aznar-Siguan, G., Schewe, J., Milano, L., Ginnetti, J., Willner, S., McCaughey, J. W., & Bresch, D. N. (2021). Global warming and population change both heighten future risk of human displacement due to river floods. *Environmental Research Letters*, *16*(4), 044026. https://doi.org/10.1088/1748-9326/abd26c

Managers, A. o. S. F. (2020). Flood Mapping for the Nation: A Cost Analysis for Completing and Maintaining the Nation's NFIP Flood Map Inventory. In: Association of State Floodplain Managers Madison, WI, USA.

Mudashiru, R. B., Sabtu, N., Abustan, I., & Balogun, W. (2021). Flood hazard mapping methods: A review. *Journal of Hydrology*, *603*, 126846. https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126846

National Research, C., Water, S., Technology, B., Board on Earth, S., Resources/Mapping Science, C., & Committee on, F. F. M. (2009). *Mapping the Zone : Improving Flood Map Accuracy*. National Academies Press. http://ebookcentral.proquest.com/lib/uaz/detail.action?docID=3378492

Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A., & Slayback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, *596*(7870), 80-86. https://doi.org/10.1038/s41586-021-03695-w