From Flash Flood Vulnerability and Risk Assessment to Property Damage Prediction: the Value of Machine Learning

Atieh Alipour (aalipour@crimson.ua.edu), Peyman Abbaszadeh (pabbaszadeh@crimson.ua.edu), Ali Ahmadalipour (aahmada@ua.edu), and Hamid Moradkhani (hmoradkhani@ua.edu)

Center for Complex Hydrosystems Research, Department of Civil, Construction and Environmental Engineering, The University of Alabama

Abstract

- This presentation is a summary of the two papers we published in Journal of Flood Risk Management and Environmental Research Letters. For more details we refer the readers to these articles.
- According to the United States National Weather Service (NWS), a flooding event which is initiated within 6 hours is referred to as flash flood.
- Due to frequent torrential rainfall caused by tropical storms, thunderstorms, and hurricanes, flash flood is a common danger across the Southeast U.S. During the past decades, flash floods imposed 7.5 billion dollars property damages to southeast US (SEUS).
- We present a framework that considers a variety of features explaining different components of risk (i.e., hazard, vulnerability, and exposure), and multiple Machine Learning (ML) models to predict flash flood damages.
- Over 14,000 flash flood events during 1996 to 2017 were assessed to analyze their characteristics including frequency, duration, and intensity.
- The ML model is implemented in two modes: first, as a binary classifier to estimate whether a region of interest is damaged in any particular flood event, and then as a regression model to predict the amount of damage associated with each event. The results indicated effectiveness of the proposed methodology in predicting flash flood damages across the SEUS.

Materials and Methods

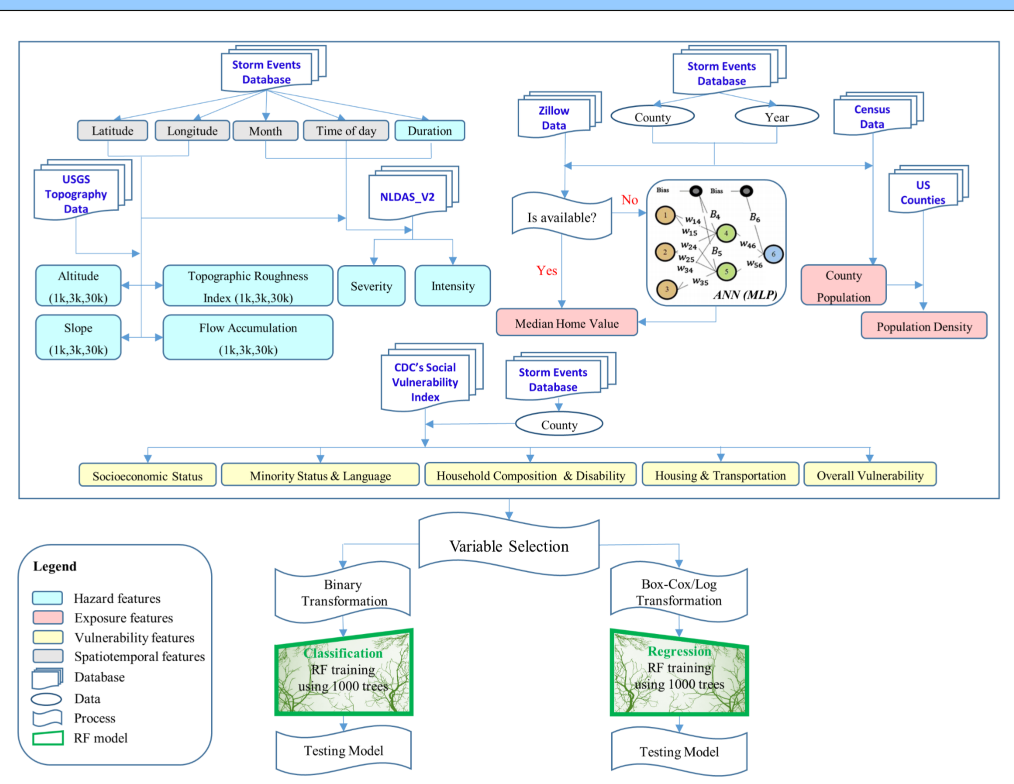


Figure 1. Schematic representation of the proposed framework for flash flood damage prediction. In the figure, ANN (MLP) stands for Artificial Neural Network (Multilayer Perceptron), and RF is Random Forest.

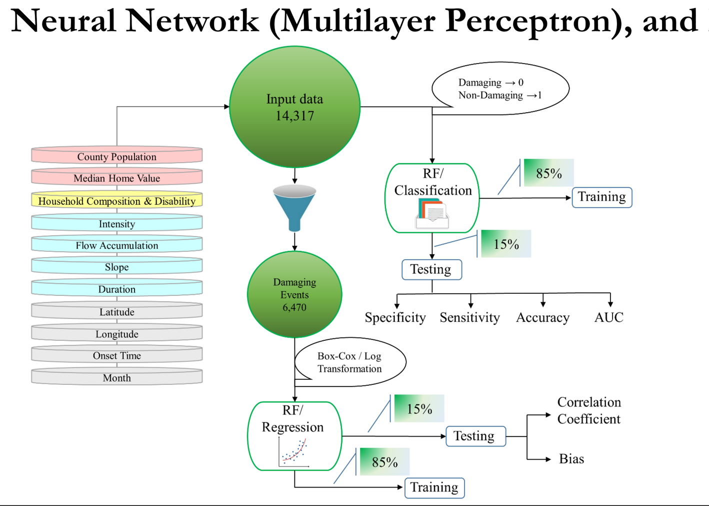
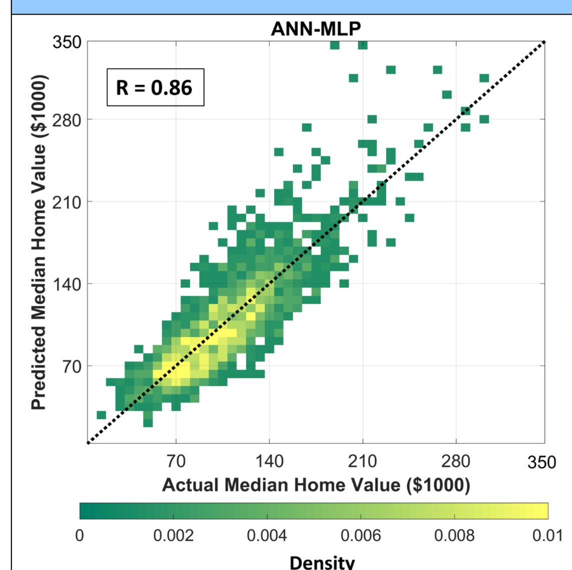


Figure 2. The schematic representation of the flash flood damage prediction framework. In the figure, RF stands for Random Forest, and AUC is the area under relative-operating characteristic curve.

Filling the Gaps in Zillow Dataset



One of the variables used in this study is median home value that explains the flash flood exposure. We utilized Zillow dataset to extract this information for each flash flood event during 1996 to 2017 over the SEUS. Unfortunately, the median home value is not available for all counties and all years in the study period. To overcome this shortcoming, we utilized Artificial Neural Network (ANN) to predict the missing median home values.

Figure 3. Verification result for the ANN-MLP model for the testing period that is used to fill out the missing values of the median home values of the Zillow dataset; R= correlation coefficient.

Flash flood Characteristics Across the SEUS

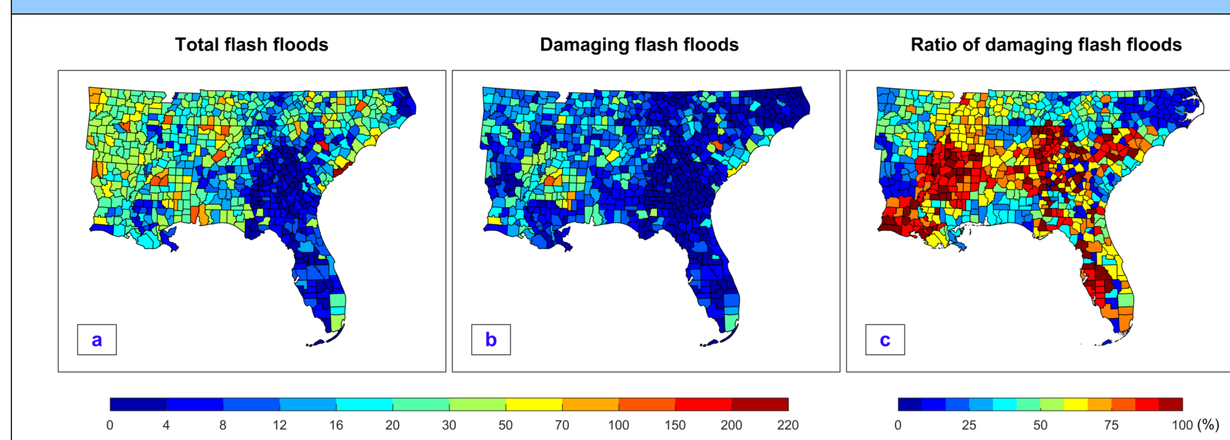


Figure 4. The spatial pattern of flash flood frequency. (a) Frequency of all flash flood events during 1996 to 2017, (b) the number of damaging events for the same period, and (c) the ratio (percentage) of flash floods that caused property damage during 1996-2017.

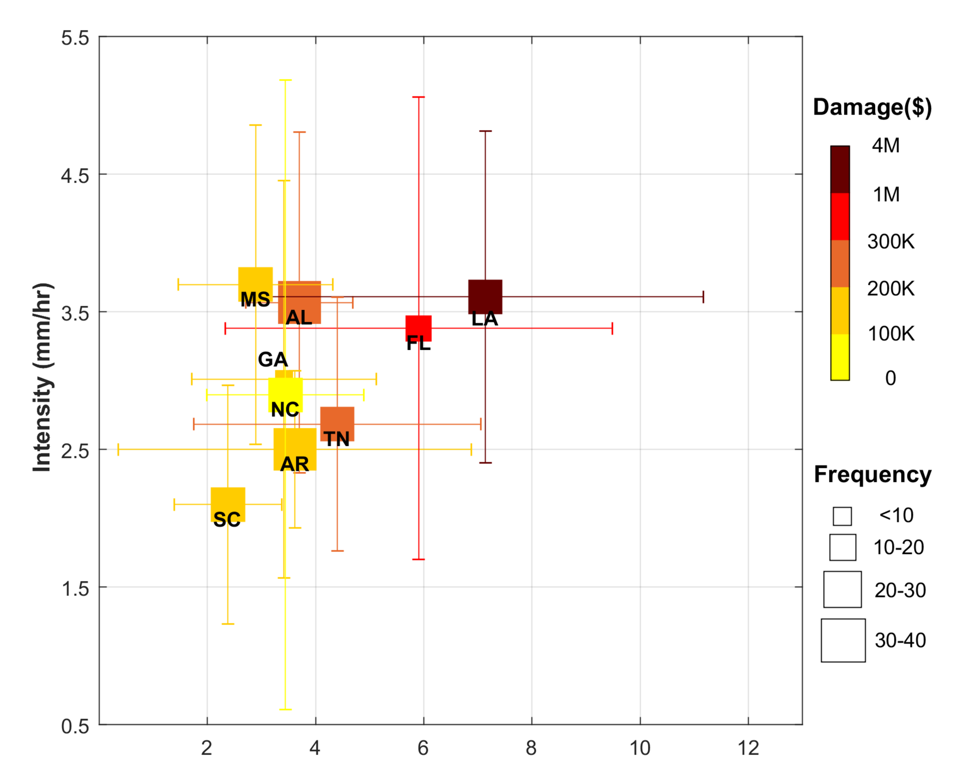


Figure 5. Composite plot of flash flood characteristics, viz., intensity, duration, frequency and property damage across the SEUS states. The marker size, x and y axes, and marker color represent the mean frequency, mean duration, mean intensity, and the mean property damage of flash floods per event, respectively. The error bars indicate the variation (one standard deviation) of duration and intensity (plotted along the corresponding axes) among the counties of each state.

Variable Selection

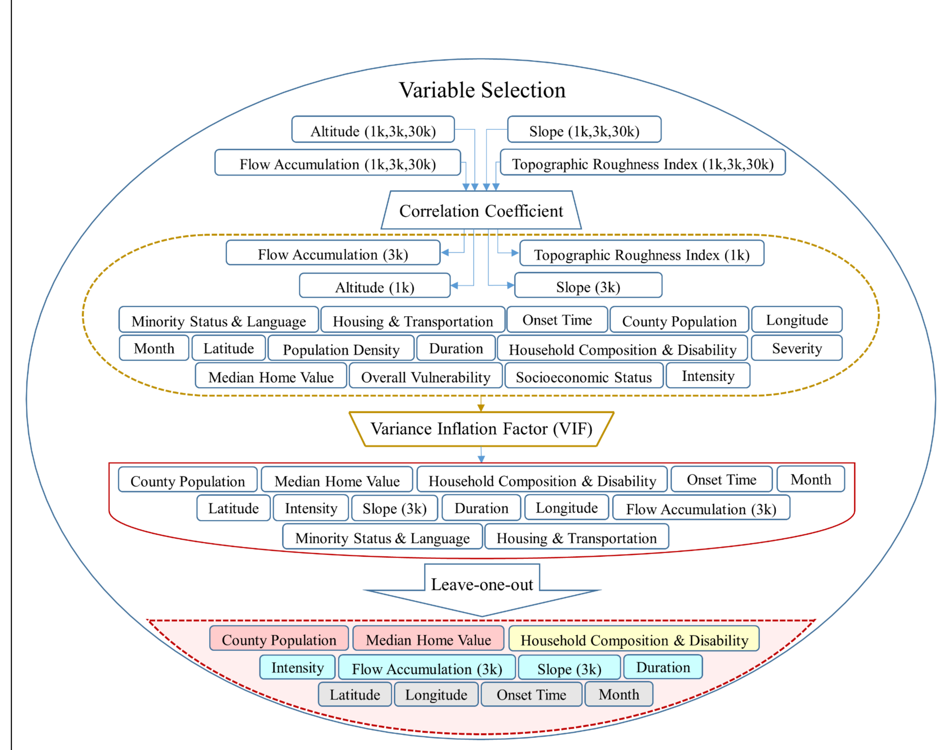


Figure 6. Flowchart of the variable selection method, and the final 11 chosen variables (at the bottom) that are used as input to the Random Forests model. Red, yellow, blue, and gray colors are used for variables representing exposure, vulnerability, hazard, and spatiotemporal features, respectively.

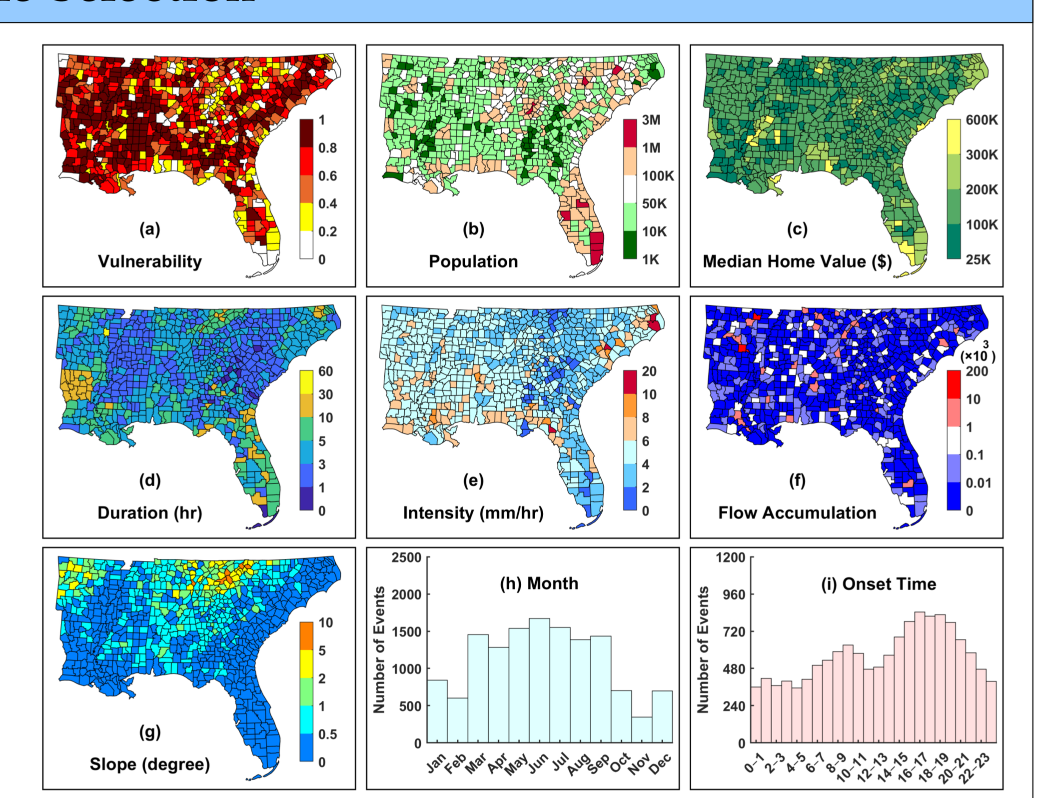
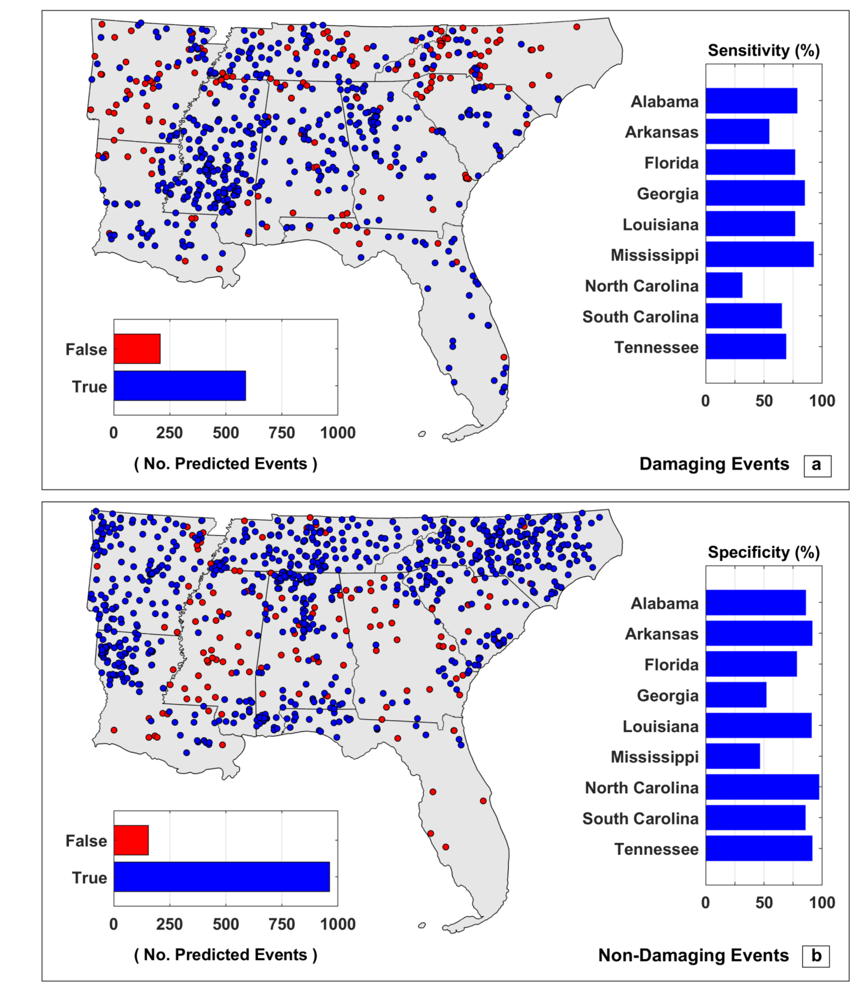


Figure 7. The spatial variation of input features that are used for predicting flash flood damages. a) The 2016 relative vulnerability index (household composition & disability); b) Population of each county in 2017; c) Median home value in 2017; d) Mean duration of flash floods during 1996 to 2017; e) Long-term average intensity of flash flood events during 1996 to 2017; f) flow accumulation; g) slope for each county; and the monthly (h) and diurnal (i) distribution of flash flood events during 1996 to 2017.

Damaging vs. Non-damaging Classification



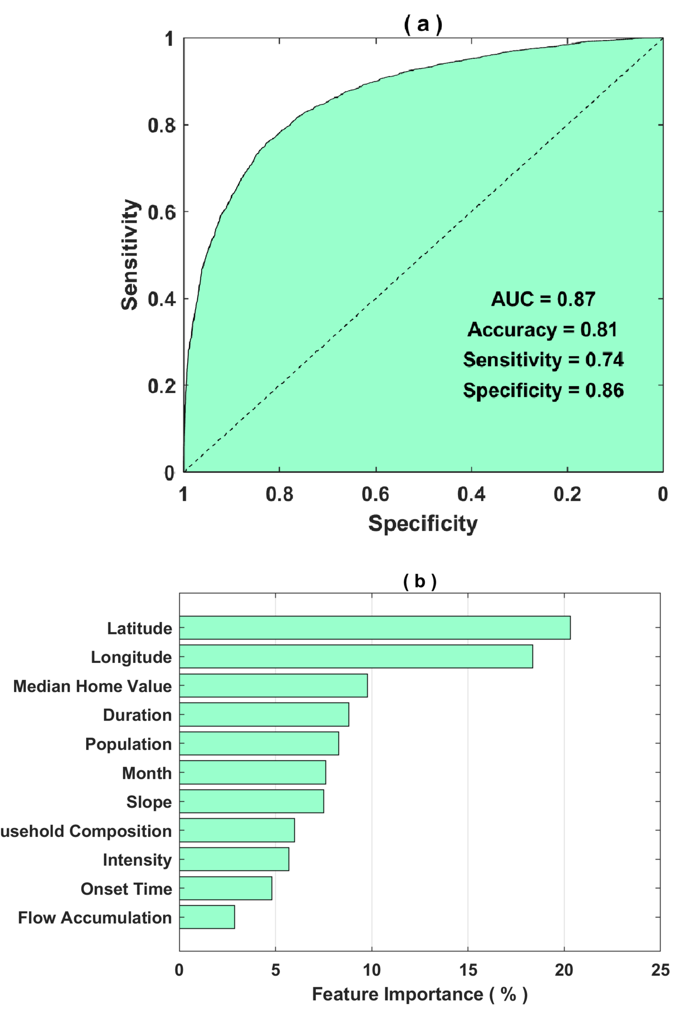


Figure 8. The performance of the proposed binary damage classification approach for (a) damaging and (b) non-damaging flash flood events. The blue and red colors illustrate the true and false predictions, respectively. The blue and red colors indicate the true and false predictions, respectively. The points on the map show the location of flash flood events. The total number of correct and incorrect predicted events are shown for both cases. On the right side of the panels, the sensitivity and specificity of the model are shown for each state.

Figure 9. a) The Relative-Operating Characteristic (ROC) curve of the proposed Random Forest classifier; AUC = Area Under Curve. b) The relative importance of features for the random forest classifier model.

Damage Prediction Model

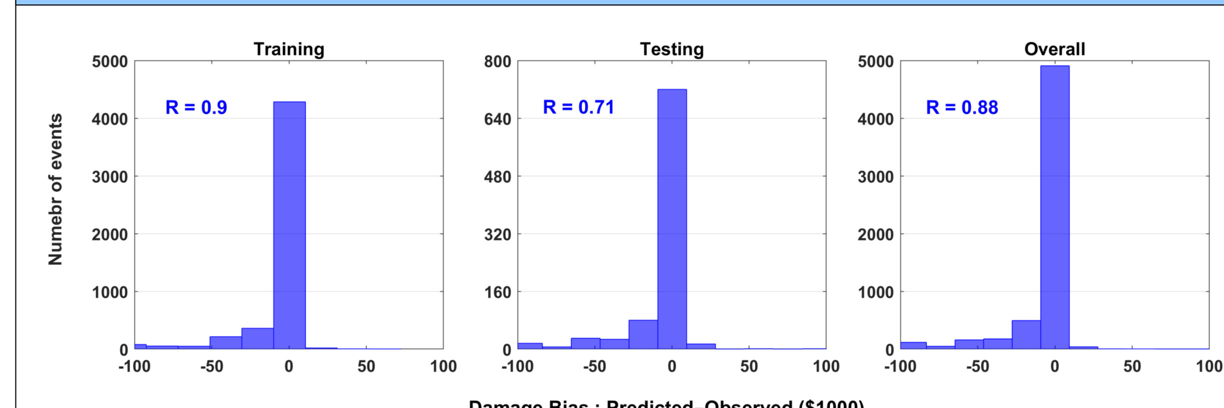


Figure 10. The performance of the Random Forest model in prediction of flash flood damage over the SEUS. The subplots indicate the histogram of bias during the training, testing, and entire dataset (totaling 5500, 970, and 6470 events, respectively).

Conclusions

- Analysis of flash flood characteristics shows that the frequency of flash flooding is increasing across the SEUS. It is higher during summer and spring. The intensity of flash flooding is increasing and overall, flash flood hazard in Louisiana is higher than other states in the SEUS.
- Here, we proposed a risk-based and physically informed model for predicting flash flood property damage across the SEUS using several influential factors of geographic, socioeconomic, and climatic features and machine learning techniques.
- The findings of this study suggest the applicability and usefulness of ML models for prediction of property damages associated with flash flood events over a large domain.

References

Alipour, A., Ahmadalipour, A., Moradkhani, H., 2020. Assessing flash flood hazard and damages in the southeast United States. Journal of Flood Risk Management. https://doi.org/10.1111/jfr3.12605

Alipour, A., Ahmadalipour, A., Abbaszadeh, P., Moradkhani, H., 2020. Leveraging machine learning for predicting flash flood damage in the Southeast US. Environ. Res. Lett. 15, 024011. https://doi.org/10.1088/1748-9326/ab6edd

Received: 22 February 2019 | Revised: 27 November 2019 | Accepted: 5 February 2020

DOI: 10.1111/jfr3.12605

ORIGINAL ARTICLE

CIWEM Chartered Institution of Water and Environmental Management | Journal of Flood Risk Management | WILEY

Assessing flash flood hazard and damages in the southeast United States

Atieh Alipour | Ali Ahmadalipour | Hamid Moradkhani

Center for Complex Hydrosystems Research, Department of Civil, Construction and Environmental Engineering, University of Alabama, Tuscaloosa, Alabama

Correspondence
Atieh Alipour, Center for Complex Hydrosystems Research, Department of Civil, Construction and Environmental Engineering, University of Alabama, Tuscaloosa, AL.
Email: aalipour@crimson.ua.edu

Abstract

Flash floods are common natural hazards in the southeast United States (SEUS) as a consequence of frequent torrential rainfall caused by tropical storms, thunderstorms, and hurricanes. Understanding flash flood characteristics is essential for mitigating the associated risks and implementing proactive risk management strategies. In this study, flash flood characteristics including frequency, duration, and intensity are assessed in addition to their associated property damages. The National Oceanic and Atmospheric Administration (NOAA) Storm Events database as well as hourly precipitation data of the North American Land Data Assimilation System project phase-2 (NLDAS-2) are utilised, and more than 14,000 flash flood events during 1996–2017 are analysed. Flash flood hazard is investigated at county, state, and regional levels across the SEUS. Results indicate increasing pattern for the frequency and intensity of flash flooding over the SEUS. The frequency of flash flooding is found to be higher in spring and summer, whereas the duration and intensity of events are higher during winter and fall, respectively. The western parts of the SEUS are prone to more frequent and intense flash flooding compared to the eastern parts. Overall, our analyses suggest that flash flood hazard in Louisiana is higher than other states in the SEUS.

KEYWORDS

duration, flash flood, frequency, hazard, intensity

1 | INTRODUCTION

Flash floods are among the most devastating natural hazards, which threaten human lives and properties in various regions of the world (Ahmadalipour & Moradkhani, 2019; Bezak, Šraj, & Mikoš, 2016; Miao, Yang, Yang, & Li, 2016). According to the National Weather Services (NWS), flash floods generally initiate within a few minutes up to less than 6 hr of an intense rainfall (Jalayer, Aronica, Recupero, Carozza, & Manfredi, 2018). The rapid onset of flash floods limits effective and timely decision making, and causes the highest number of casualties (on average) compared to the other types of flooding (e.g., coastal floods [storm surge] and river floods) (Jonkman, 2005). Flash floods caused the highest number of casualties among various flood events in the United States (Ashley & Ashley, 2008; Terti, Ruin, Anquetin, & Gourley, 2017). The frequency of heavy precipitation has been shown to increase under climate change (Halmstad, Reza, & Moradkhani, 2013; Ma et al.,

Environmental Research Letters

LETTER

Leveraging machine learning for predicting flash flood damage in the Southeast US

OPEN ACCESS

RECEIVED
29 September 2019

REVISED
21 January 2020

ACCEPTED FOR PUBLICATION
22 January 2020

PUBLISHED
7 February 2020

Atieh Alipour, Ali Ahmadalipour, Peyman Abbaszadeh and Hamid Moradkhani

Center for Complex Hydrosystems Research, Department of Civil, Construction and Environmental Engineering, University of Alabama, Tuscaloosa, AL, United States of America

E-mail: aalipour@crimson.ua.edu

Keywords: flash flood, risk, flood damage, machine learning

Supplementary material for this article is available online

Abstract

Flash flood is a recurrent natural hazard with substantial impacts in the Southeast US (SEUS) due to the frequent torrential rainfalls that occur in the region, which are triggered by tropical storms, thunderstorms, and hurricanes. Flash floods are costly natural hazards, primarily due to their rapid onset. Therefore, predicting property damage of flash floods is imperative for proactive disaster management. Here, we present a systematic framework that considers a variety of features explaining different components of risk (i.e. hazard, vulnerability, and exposure), and examine multiple machine learning methods to predict flash flood damage. A large database of flash flood events consisting of more than 14 000 events are assessed for training and testing the methodology, while a multitude of data sources are utilized to acquire reliable information related to each event. A variable selection approach was employed to alleviate the complexity of the dataset and facilitate the model development process. The random forest (RF) method was then used to map the identified input covariates to a target variable (i.e. property damage). The RF model was implemented in two modes: first, as a binary classifier to estimate if a region of interest was damaged in any particular flood event, and then as a regression model to predict the amount of property damage associated with each event. The results indicate that the proposed approach is successful not only for classifying damaging events (with an accuracy of 81%), but also for predicting flash flood damage with a good agreement with the observed property damage. This study is among the few efforts for predicting flash flood damage across a large domain using mesoscale input variables, and the findings demonstrate the effectiveness of the proposed methodology.

1. Introduction

The Southeast US (SEUS) is known to be susceptible to flash flooding due to the frequent high intensity rainfall triggered by tropical storms, thunderstorms and hurricanes (Orville and Huffines 2001, Czajkowski *et al* 2011, Smith and Smith 2015). During the last two decades, widespread flash flood events have caused significant economic damage in this region. Recent studies have shown that the frequency of flash flooding is increasing in the SEUS (Alipour *et al* 2020). Therefore, predicting property damage of flash floods is crucial for attaining proactive disaster management in this region.

Generally, risk refers to the potential losses of a particular hazard (Cardona *et al* 2012, Armenakis *et al* 2017, Ahmadalipour *et al* 2019), which is characterized as a function of three major components: hazard, vulnerability, and exposure (Adger 2006, Dang *et al* 2010, Winsemius *et al* 2013, Budiyono *et al* 2015, Koks *et al* 2015). Assessing flash flood risk components has been the subject of several studies. Recently, Ahmadalipour and Moradkhani (2019) investigated the spatio-temporal characteristics of flash flooding hazard over the Contiguous United States (CONUS). Also, Khajehei *et al* (2020) assessed the socioeconomic vulnerability of flash flooding at the county scale across the entire CONUS while accounting for flash