A GIS Tool for Assessing Community Susceptibility to Flash Flooding

Rohan Singh Wilkho[1, +] and Nasir G. Gharaibeh[1, *]
[1] Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, Texas – 77840
[+] Student, [*] Faculty Advisor

Abstract

Flash Floods (FFs) are the leading cause of fatalities due to natural hazards in the US. FFs have small spatial extents and high-water flow velocity, making them a localized phenomenon. Several factors like geological parameters, terrain characteristics, and meteorological conditions may affect the susceptibility of a place to FF. Due to FF's localized nature, causal (direct cause-and-effect) relationships between factors affecting flash flooding may differ from one region to another. Therefore, the localized impacts and nature of flash flooding call for a dynamic (region-based) assessment of the susceptibility of communities to FF. This paper responds to this call by introducing DFFS (Dynamic Flash Flood Susceptibility), a Geographic Information System tool that integrates causal discovery and machine learning to assess FF susceptibility dynamically. DFFS operates in three distinct steps: (1) For a region of interest, it uses the NOAA Storm Events Database to obtain past FF records for census tracts (CTs) (communities) in the region and then performs a hotspot analysis to identify the CTs with significantly high FF frequency (i.e., hot spots) and low FF frequency (i.e., cold spots), (2) DFFS then performs causal discovery to identify conditioning factors that are most significant in the region, and (3) Using the causal factors as independent variables, DFFS trains machine learning algorithms to predict CT susceptibility scores. This paper describes the methodology for assessing FF susceptibility and provides case studies where we test DFFS on three different regions highlighting the importance of region-based susceptibility assessment. Researchers and practitioners can use this methodology to assess the flash flooding risk in their regions at the census tract scale.

1. Introduction

Flash Floods (FFs) occur in small catchment areas (typically smaller than 1000 km2) within a short time (approximately 6 hours) of a trigger event (heavy rain, dam break) [Llasat et al., 2016]. High runoff velocity and little to no lead time make FFs one of the most harmful natural hazards [Hardy et al., 2016]. FFs have been the leading cause of fatalities due to natural hazards in the US [SE data]. Furthermore, increasing urbanization and climate change will likely result in worse and hard-to-forecast FF events [Norbiato et al., 2008]. These factors collectively call for improved FF mitigation strategies to help create targeted policies for the most susceptible regions. This paper responds to these calls by introducing Dynamic Flash Flood Susceptibility (DFFS), a Geographic Information System (GIS) tool to dynamically assess FF susceptibility at the community level.

We measure FF susceptibility based on factors that, in combination, make communities vulnerable to the hazard. There is a consensus in the literature that flash flooding is a localized phenomenon, potentially influenced by several local factors (like precipitation, ground slope, and land use pattern, among others) [Shirzadi et al., 2020]. However, because of the localized nature, the significance of different influencing factors can vary across different regions. Hence, it is essential to assess FF susceptibility dynamically (i.e., on a region-by-region basis). The term "dynamic" in DFSS indicates region-based susceptibility assessment.

DFFS assesses the susceptibility of each census tract (CT) (a surrogate for community) within any given region of interest (group of CTs). DFFS performs causal discovery to identify the causal FF factors for the region and then uses these factors in machine learning (ML) models to predict the FF susceptibility score for each CT within the region. Knowledge of the causal factors and community susceptibility can enable regional authorities to design targeted interventions and prioritize resources to reduce the risk of flash flooding [Tehrany et al., 2015].

The remainder of the paper is organized as follows. Section 2 describes DFFS' methodology. Section 3 presents the results and discussion of a case-study example of the tool that highlights the relevance of dynamic assessment of flash flooding. Finally, section 4 covers the paper's conclusions.

2. Methodology

Figure 1 contains the graphical outline of DFFS' methodology. We divide the methodology into three stages: (1) Input, (2) Back-end, and (3) Output (Figure 1). In the input stage, the user must select a group of neighboring CTs as their region of interest (for example, CTs in the Greater Houston region, Texas).

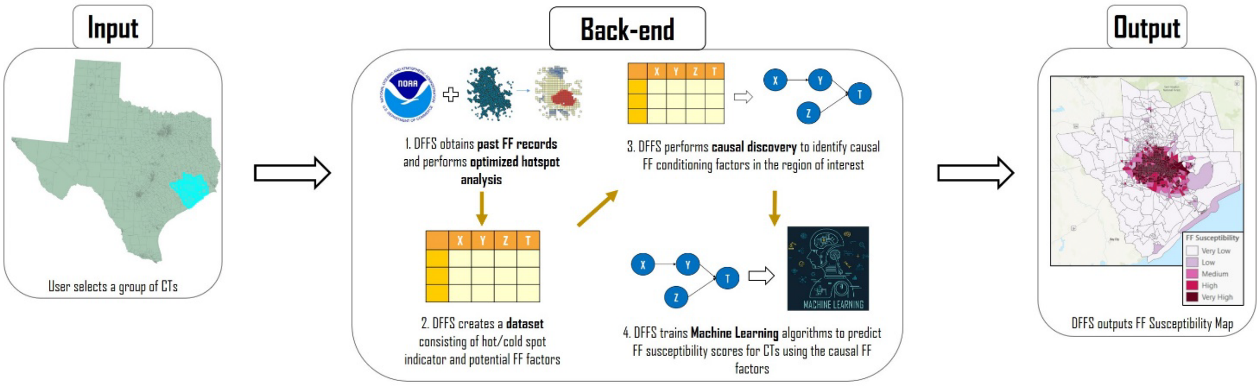


Fig 1: Methodology of the Dynamic Flash Flood Susceptibility (DFSS) tool

After receiving an input region of interest, DFFS performs four back-end operations:

- *Obtaining past FF records and performing optimized hotspot analysis:* DFFS obtains the past FF records (number of FF events in each CT between 2010-19) for the input region from NOAA Storm Events Database [SE Data] and performs optimized hotspot analysis to identify CTs that have significantly higher flash flooding frequency (hot spots) and significantly lower flash flooding frequency (cold spots) as compared to the average in the region.
- *Creating a dataset*: After identifying the hot and cold spots, DFFS creates a dataset that contains the hot/cold class (1/0) of each CT (in the region) and values for 11 possible conditioning factors that can potentially influence flash flooding: rainfall intensity, storm duration, elevation, ground slope, topographic wetness index, surface roughness, land use and land cover, lithology, soil type, population density and median year structure built. Table 1 contains an abridged example of such a dataset.
- *Performing Causal Discovery:* Causal discovery is the process of discovering causal relationships between variables in an observational dataset [Glymour et al., 2019]. DFFS uses the dataset created in the previous step to perform causal discovery. Figure 2 contains an example causal graph that causal

discovery yields. The system represented by Figure 2 has five variables (A, B, C, D, and T) wherein C and D are the causal factors of T, i.e., only a change in either or both of C and D will result in a change in T. DFFS uses the order-independent PC algorithm [Colombo & Maathuis, 2015] to develop causal graphs for the flash flooding phenomenon and identify FF causal factors for a region of interest.

- *Machine learning prediction:* DFFS' final back-end operation is to train and test ML algorithms to predict the FF susceptibility for the CTs in the region. DFFS uses the causal FF factors as predictor variables for this prediction. DFFS trains three models for every region of interest: Random Forest (RF) [Breiman, 2001], Extra Trees (ET) [Geurts et al., 2006], and a combination of the two to predict a CT's susceptibility score. After training, DFFS selects the model that gives the best prediction.

Table 1: Abridged example of the temporary dataset created by DFFS

Hot/Cold Class*	Rainfall Intensity	Population Density	Mean TWI	% area in HSG class A
1	0.75	3658.45	3.57	2.69
1	0.58	5114.87	5.68	5.87
0	0.25	168.97	1.25	1.36
1	0.38	357.84	19.57	11.42
0	0.66	511.78	8.48	0.19

*Hot/Cold Class: CT in hot spot = 1; CT in cold spot = 0; TWI: Topographic Wetness Index; HSG – Hydrologic Soil Group

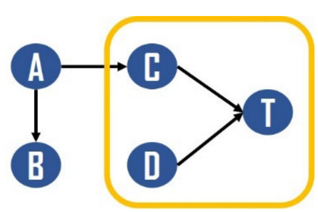


Fig 2: Causal graph example

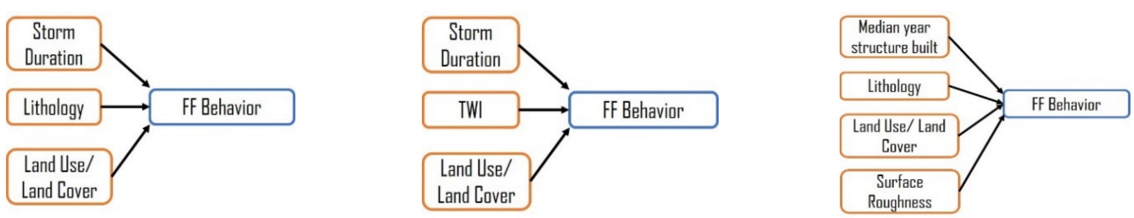
In the output stage, DFFS produces a color-coded map of the region where the CTs are coded with varying levels of FF susceptibility scores (output in Figure 1).

3. Case-study results and discussions

To demonstrate the practical utility of DFFS, we performed a case study on three urban regions in Texas: Dallas-Fort Worth (DFW - 1311 CTs), Greater Austin (GA - 769 CTs), and Greater Houston (GH - 1072 CTs). The results of these case studies are discussed next.

3.1 Causal factors in the study regions

Figure 3 contains the causal graphs for the flash flooding behavior in the three study regions. For brevity, we show only the causal factors and not the entire graph. These figures show that while certain factors (e.g., land use land cover - LULC) are common in the three regions, the set of causal factors varies across regions. Additionally, certain factors are unique to particular regions. For example, the Topographic Wetness Index (TWI) is significant in GA only and surface roughness is significant in GH only. This leads us to the inference that causal FF factors vary across regions, affirming the fact that flash flooding is a localized phenomenon.



(a) DFW Region (b) GA Region (c) GH Region

Fig 3: Causal factors

3.2 ML Results and FF Susceptibility Maps

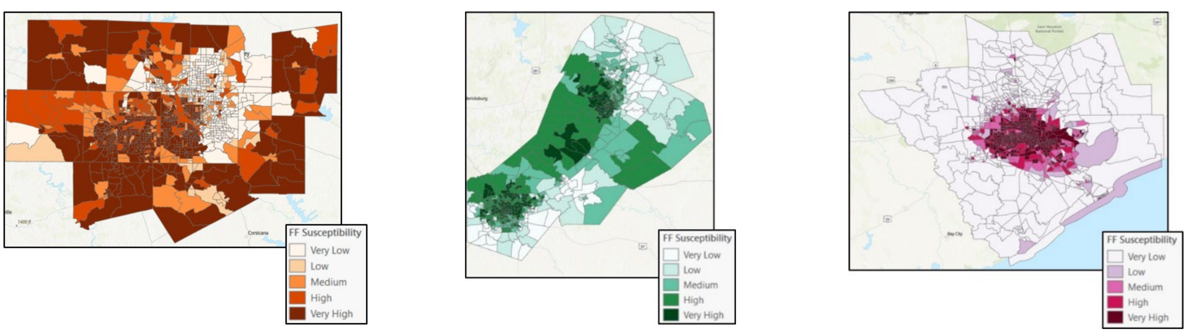
Table 2 contains the accuracy of the three ML models DFFS trained for the three study regions. DFFS measures accuracy as the ratio of the number of correct predictions to the total number of predictions that it makes for an unseen testing set (not used for training). Different ML techniques yielding superior performances in the study regions further illustrate the localized nature of flash flooding.

Table 2: ML model performances

Model	DFW	GA	GH
	Accuracy*	Accuracy	Accuracy
RF	**0.875**	0.804	0.872
ET	0.830	**0.822**	0.856
RF + ET	0.860	0.813	**0.883**

*Accuracy: Ranges from 0 to 1. A higher value denotes better performance.

Figure 4 contains the FF susceptibility maps that DFFS creates after performing causal discovery and ML training. These maps highlight the communities that are more susceptible to flash flooding in the region, which can help practitioners and policymakers concoct targeted mitigation and preparedness policies to improve these communities' resilience. These maps also exhibit possible outlier behavior where neighboring communities have different FF susceptibility levels. Furthermore, a visual comparison of the maps reveals that the same community type can be at different susceptibility levels in two regions, reiterating that FF behavior is region-specific. For example, in GH, the larger CTs (less population density) are not at high susceptibility levels, whereas they are at higher levels in DFW and GA.



(a) DFW Region (b) GA Region (c) GH Region

Fig 4: Flash Flood Susceptibility Maps

4. Conclusions and Future Work

This study suggests that the dynamic and localized nature of FFs warrants a region-specific susceptibility analysis. We introduce DFFS, a GIS tool to dynamically assess FF susceptibility at the community level.

DFFS' novelty lies in performing causal discovery to identify FF causal factors. As a result, every time a user selects a region (group of census tracts), DFFS first identifies the causal FF factors (from a list of potential factors) for the region of interest and then uses them and ML to assess susceptibility. The term "dynamic" implies the region-specific assessment of causality and susceptibility.

This study further showcases a use-case example of DFFS for three urban regions in Texas: Dallas-Fort Worth, Greater Austin, and Greater Houston. DFFS yields different causal factors and susceptibility maps for these regions. These results affirm the localized nature of flash flooding. The susceptibility maps also help identify communities that are most at risk of flash flooding within their region, which can uncover further insights about such communities, e.g., demographics and socioeconomic characteristics, among others.

Currently, DFFS is restricted to Texas. Our ongoing work includes extending it to the entire United States and making it available for public use. In addition, future works may include undertaking outlier analysis to understand why specific communities react differently to FFs than their neighbors.

References:

Llasat, M. C., Marcos, R., Turco, M., Gilabert, J., & Llasat-Botija, M. (2016). Trends in flash flood events versus convective precipitation in the Mediterranean region: The case of Catalonia. *Journal of Hydrology, 541(October), 24–37.* https://doi.org/10.1016/j.jhydrol.2016.05.040

Hardy, J., Gourley, J. J., Kirstetter, P. E., Hong, Y., Kong, F., & Flamig, Z. L. (2016). A method for probabilistic flash flood forecasting. *Journal of Hydrology, 541, 480–494.* https://doi.org/10.1016/j.jhydrol.2016.04.007

SE Data. *National Climatic Data Center, 2021*. Retrieved from http://www.ncdc.noaa.gov/stormevents/. Accessed Apr 06, 2023

Norbiato, D., Borga, M., Degli Esposti, S., Gaume, E., & Anquetin, S. (2008). Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins. Journal of Hydrology, 362(3–4), 274–290. https://doi.org/10.1016/j.jhydrol.2008.08.023

Shirzadi, A., Asadi, S., Shahabi, H., Ronoud, S., Clague, J. J., Khosravi, K., Pham, B. T., Ahmad, B. Bin, & Bui, D. T. (2020). A novel ensemble learning based on Bayesian Belief Network coupled with an extreme learning machine for flash flood susceptibility mapping. Engineering Applications of Artificial Intelligence, 96(August 2019), 103971. https://doi.org/10.1016/j.engappai.2020.103971

Tehrany, M. S., & Kumar, L. (2018). The application of a Dempster–Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods. Environmental Earth Sciences, 77(13), 1–24. https://doi.org/10.1007/s12665-018-7667-0.

Hot Spot Analysis. https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-statistics/optimized-hot-spot-analysis.htm. Accessed April 06, 2023.

Colombo, D., & Maathuis, M. H. (2015). Order-independent constraint-based causal structure learning. *Journal of Machine Learning Research*, *15*(2010), 3741–3782.

Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, *10*(JUN), 1–15. https://doi.org/10.3389/fgene.2019.0052

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32. https://doi.org/10.1023/A:1010933404324

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1