Civil and Environmental Engineering Theses and Dissertations

Civil Engineering and Environmental Engineering

Fall 12-17-2022

# Leveraging Crowdsourced Navigation Data In Roadway Pluvial Flash Flood Prediction

Arefeh Safaei-Moghadam

asafaeimoghadam@smu.edu

LEVERAGING CROWDSOURCED NAVIGATION DATA IN ROADWAY PLUVIAL

FLASH FLOOD PREDICTION


Approved by:


_____
Prof. Barbara Minsker
Civil and Environmental Engineering


_____
Prof. Khaled Abdelghany
Civil and Environmental Engineering


_____
Prof. Janille Smith-Colin
Civil and Environmental Engineering


_____
Prof. Eric Larson
Computer Science


_____
Prof. David Tarboton
Civil and Environmental Engineering
Utah State University


_____
Prof. Brenda Philips
Electrical and Computer Engineering
University of Massachusetts Amherst

LEVERAGING CROWDSOURCED NAVIGATION DATA IN ROADWAY PLUVIAL

FLASH FLOOD PREDICTION


A Dissertation Presented to the Graduate Faculty of the

Lyle School of Engineering

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Civil and Environmental Engineering

by


Arefeh Safaei Moghadam

B.S., Civil Engineering, Shahid Chamran University, Ahvaz, Iran

M.S., Water Resources Management, Amirkabir University of Technology, Tehran, Iran


December 17, 2022

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my advisor, Dr. Barbara Minsker, for her unconditional support and understanding throughout my education. Without her guidance this journey would not have been feasible.

I would like to thank my committee members, Dr. Khaled Abdelghany, Dr. Janille Smith-Colin, Dr. Eric Larson, Brenda Philips, and Dr. David Tarboton, for their guidance, insight, and support throughout the course of this research. Thanks also to my friends and colleagues and the department faculty and staff for making my time at SMU a great experience.

My heartfelt gratitude also goes to my husband, Dr. Soheil Sohrabi, for his love, support, and being by my side through the ups and downs of my personal and professional growth. While living thousands of miles away from home, I enjoyed a wonderful graduate student life thanks to my husband and wonderful Ashire friends, too many to name here.

Although far from home, I am grateful for having the unconditional love and unfailing encouragements of my parents and brothers.

TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

Chapter 1

INTRODUCTION

This dissertation develops and tests a new data-driven framework for short-term roadway pluvial flash flood (PFF) risk estimation at the scale of road segments using crowdsourced navigation data and a simplified physics-based PFF model. Flash flooding is considered one of the most hazardous natural disasters that affect people worldwide (Kousky, 2018). Analysis of flash floods over the contiguous United States shows that flash flood frequency and property damage have increased in the past two decades (Ahmadalipour and Moradkhani, 2019). Pluvial flash flooding (PFF) is defined as localized floods caused by an overwhelmed natural or engineered drainage system (Carter et al., 2015; Rosenzweig et al., 2018). PFF can reduce the reliability of roadway networks by decreasing capacity, increasing travel time, reducing safe speed, and increasing accident risks and deaths through lane submersion (Agarwal et al., 2005; Smith et al., 2004; Suarez et al., 2005). According to the National Weather Services (2022), 30 cm of standing water can be sufficient to float most cars.

Most urban flood studies have focused on fluvial (river) and coastal flooding rather than PFF. Rosenzweig et al. (2018) identified three reasons for pluvial flooding being less studied: 1−It is assumed that stormwater infrastructure, such as sewers, culverts, and pumps, are sufficient to prevent pluvial flooding, 2−Pluvial flooding is believed to be a nuisance with minimal impacts, and 3−Lack of monitoring data to capture short-duration precipitation over small urban watersheds.

In the past, stormwater minor systems (curbs, gutters, inlets, pipes, and channels) have been designed to minimize nuisance hazards associated with a 10-year or less recurrence interval rainfall (U.S. Department of Transportation FHWA, 1979). More recent roadway facilities are designed and evaluated for 50-year and 100-year events (Mark and Marek, 2011), but in older urban areas, undersized conveyance systems remain (Jack et al. 2021). With climate change, growing urbanization, and increasing imperviousness, the frequencies of extreme rainfall events and nuisance flooding are increasing (Hemmati et al., 2021, 2020; United Nations, 2018), leading to increased risks of pluvial flooding and water logging. Risk of a natural disaster is defined as a combination of hazard, vulnerability, and exposure. According to Gouldby et al. (2005), hazard is a potentially destructive phenomenon; vulnerability refers to the characteristics of a system that describe its potential to be harmed by the hazard; and finally, exposure is the number of receptors that harm may influence. In interpreting the risk of roadway PFF on a road segment, the hazard of the system is precipitation; vulnerability is any local characteristic of a road segment that contributes to PFF, such as stormwater drainage system and hydrologic characteristics, and exposure is the traffic volume that navigates on flooded road segments.

## 1.1 Background

Mobility disruption is a noticeable consequence of storm events and PFF on roadways (Coles and Hirschboeck, 2020; Coles et al., 2017; Douglas et al., 2010; Li et al., 2018; Yin et al., 2016). Some studies have evaluated precipitation impacts on either vehicle speeds or traffic delays. For example, Smith et al. found a correlation between rainfall intensity and free-flow traffic speed (Smith et al. 2004). Hooper et al. analyzed historical traffic data and precipitation radar data on two specific roadway corridors to establish the relationship between rainfall characteristics and traffic speed (Hooper et al., 2014). They showed a single global relationship cannot describe the

relationship between precipitation and traffic speed: Instead, local characteristics of road segments should be taken into account. Some studies have focused on relationships between traffic congestion and precipitation. Chung (2012) used weather and traffic data and to estimate rainfall-related non-recurrent traffic congestion; they proposed a function that relates rainfall conditions and time with average non-recurrent congestion. Zhang quantified the impact of rainfall on congestion and reliability of the transportation system, taking into account the time of the day and traffic condition on a specific corridor and assuming a homogenous speed profile. Praharaj et al.(2021) looked at the historical hydrological and traffic data to quantify the impacts of recurrent flooding on the transportation system. Accordingly, they compared Vehicle-hour traveled (VHT) of rainy days with non-rainy days. They showed that impacts are temporally and spatially heterogeneous, meaning that disruption level may vary by road characteristics and time of the day.

A significant undesirable effect of rainfall on traffic is through lane submersion caused by PFF, a highly localized phenomenon that occurs in low-lying areas. While precipitation is assumed uniform in the study area of the above-mentioned studies, water accumulations and pluvial flooding depth is indeed non-uniform and depends on highly localized characteristics of road segments. Hence, instead of precipitation, some researchers evaluated the correlation between water depth and traffic flow. Pregnolato et al. (2017) used multiple data sources, including experimental, simulation, and observational data, to develop a function that relates the depth of flooding to vehicle speed reduction. They estimated that a driver facing 10 cm of standing water must not drive faster than 40 km/hr to maintain safe driving, stopping, and steering without loss of control.

To warn drivers about rapidly changing flash flood conditions, high-resolution predictive models are needed at navigational scale (road segment and intersection). Simplified terrain models, such as rapid flood spreading model (RFSM) (Lhomme et al., 2008b), height above nearest drainage model (HAND) (Nobre et al., 2011), and hierarchical filling and spilling models (Chu et al., 2013; Samela et al., 2020; Wu et al., 2019a; Zhang and Pan, 2014) can estimate inundation extent in less complex terrains where the dynamics of flow, velocity, and momentum are negligible (Teng et al., 2017). Statistical methods are also able to predict flooding by analyzing historical observations, however, since they learn from the past, updating procedures are required to make them adaptive to accelerated future changes as they are built upon the assumption that similar conditions in the future will cause flooding. A notable advantage of statistical PFF models is their ability to capture impacts of unobserved variables and uncertainties from historical observations, as well as the ability to rapidly update the models as new data become available and system dynamics change. Haghighatafshar et al. (2020) suggested that designing stormwater infrastructure based on storm recurrence intervals is ambiguous while statistical models can provide the basis of a more resilient system by taking uncertainties of vulnerability and hazard of pluvial flooding into account. Many studies have investigated statistical flood modeling to predict flooding by applying statistical and machine learning methods such as classification models, Bayesian frameworks, and Random Forest models (Solomatine and Ostfeld, 2008; Tehrany et al., 2013; Tien Bui and Hoang, 2017; Zahura et al., 2020). Other studies have combined deterministic physics-based models with statistical models for forecasting applications (Li and Willems, 2019; Zhao et al., 2018).

Empirical and data-driven models require flooding observation data with high spatio-temporal resolution. The average duration of flash flooding events in the United States has been

3.5 hours during the last two decades (Ahmadalipour and Moradkhani, 2019), limiting the applicability of aerial imagery to obtain sufficiently frequent flash flooding observations. To fill this data gap, there is increasing interest in the application of newer "crowdsourced" data into flood modeling, monitoring, and impact assessment (Assumpção et al., 2018; Gaitan et al., 2015; Helmrich et al., 2021; Liu et al., 2015; Molinari et al., 2019; Praharaj et al., 2021; Schnebele et al., 2014; Zhu et al., 2022). Previous crowdsourced flood data studies have involved engaging citizens in collecting four types of data: streamflow or rain gauge readings, videos, text messages, and image postings (Assumpção et al., 2018; Cervone et al., 2016; le Coz et al., 2016; Li et al., 2018; Li and Willems, 2019; Moy De Vitry et al., 2019; Zhu et al., 2022). Also, Zhu et al. (2022) and Liu et al. (2015) applied artificial intelligence techniques to extract flooding waterlogging from microblog information shared in crowdsourcing apps. A big challenge in using crowdsourced data is identifying the accurate location and flood extent from posted pictures, videos, and texts. However, even with the challenges mentioned above, researchers have concluded that integrating crowdsourced data into flood models improves the overall performance and timeliness of forecasts, hence increasing flood hazard awareness (Assumpção et al., 2018; Goodrich et al., 2020).

The majority of studies have implemented crowdsourced data into physics-based models as complementary data for model setup, calibration, validation, and data assimilation (Assumpção et al., 2018; Smith et al., 2017; Zahura et al., 2020). Physics-based pluvial flood simulation models can be classified into four categories: (1) 1-D sewer flow models that represent spatially averaged characteristics of overland flow and simulate flow in the stormwater drainage system (Rossman 2010); (2) 1-D or 2-D overland flow simulation models that do not consider runoff interactions with stormwater drainage systems (van der Knijff, Younis, and de

Roo 2010; E. Mignot, Paquier, and Haider 2006); (3) coupled sewer-surface models such as 1-D-surface/1-D-sewer or 2-D-surface/1-D-sewer models that take surface-sewer interactions into account (Noh et al. 2018); and (4) rapid flood spreading models (RFSM) that simplify the physics of flood spreading and use conservation of mass to route floodwater (Bulti and Abebe, 2020; Mignot et al., 2019; Paquier et al., 2015).

Review of the literature indicates that coupled 2D-1D models produce the most accurate flood inundation extent on a hyper-local scale and in an urban environment with complex flow paths. However, their application in real-time flood prediction at road segment scales is limited due to their high data requirements and computational burden (Bulti and Abebe 2020), as well as the highly complex and interconnected variables that contribute to flooding in urban environments (Coles et al., 2017; Rafieeinasab et al., 2015).

Micro topographic features, steep slopes, and varying surface materials can generate different types of flow regimes at small spatial scales. Dual-drainage hydrodynamic models that couple equations for the underground sewer system and surface flow, require detailed layouts of urban drainage systems that can be of varying quality, particularly in older urban areas where PFF is most prevalent (Berndtsson et al., 2019; Haghighatafshar et al., 2020; Sadler et al., 2018; Smith et al., 2017). Finally, catchments that drain into roadways are often very small and ungauged, leading to further uncertainties in estimating road inundation (Versini et al., 2010). Hence accurate high-resolution real-time physics-based hydrodynamic modeling in urban areas is computationally extensive and rarely considered feasible (Mignot et al., 2019).

With recent advances in technology, crowdsourced data is now widely available at a low cost and can supplement high-fidelity datasets. However, by definition, *crowdsourced data are collected by heterogeneous volunteer individuals of varying knowledge, experience, perceptions,*

*and number*. (Estellés-Arolas and González-Ladrón-de-Guevara, 2012). Social Media

Geographic Information (SMGI) is a specific type of Volunteered Geographic Information (VGI)

crowdsourced data that, in addition to geographic coordinates, contains time, user information,

and multimedia content (Campagna, 2016). Therefore, to analyze and assess the quality of SMGI

data, in addition to the spatial analysis, time, user, and multimedia content should also be studied

in an integrated framework. This dissertation focuses on SMGI from alerts posted to the Waze

navigation App, which are geotagged points posted by Waze users, called Wazers, that express

drivers' experience of the road condition. Waze alerts, in addition to multimedia content

describing road surface conditions, contain geographic coordinates, time, and user characteristics

such as reputation and feedback from other users on the user's postings. From the context of

roadway flooding, the focus of this dissertation, Waze alerts are not specifically designed for

flood situational awareness, and Wazers may not be aware that their shared experience will be

interpreted as flooding data. Craglia et al. divided citizen science and VGI into implicitly versus

explicitly volunteered information. Implicitly volunteered information means users did not share

the information with the specific purpose that the data is being used for (Craglia et al., 2012).

Accordingly, Waze alerts are classified as implicitly SMGI crowdsourced data. Hence, to assure

the flood-wise value of a Waze alert, other than intentional and unintentional manipulation in the

crowdsourced data, conceptual misjudgment and misperception should be taken into account.

     To address these issues, this study develops a data curation and computational framework

for data collection, preprocessing, and modeling to estimate the risk of roadway pluvial flash

flooding (PFF) at road-segment scales. A hybrid approach is also developed that couples a

statistical model and a simplified physics-based simulation model in a machine learning (ML)

model to rapidly predict the risk of roadway PFF using Waze alerts in real-time.

In the next chapter, data gaps are addressed by incorporating crowdsourced navigation data from the Waze navigation app as highly localized flood observations into high-resolution data-driven models that can capture the impact of unobserved variables and estimate the vulnerability of road segments to PFF. The third chapter presents an efficient physics-based RFSM model that uses graph-based calculations to estimate roadway PFF potential in the event of stormwater drainage failure. The fourth chapter presents a hybrid ML model that estimates road segment PFF vulnerabilities, computed in the second and third chapters, with hazard and exposure components of risk to predict the risk of roadway PFF based on Waze flood alerts.

# Chapter 2

ESTIMATING THE VULNERABILITY OF ROADWAY PLUVIAL FLOOD BASED ON

CROWDSOURCED TRAFFIC DATA AND DEPRESSION-BASED DEM ANALYSIS

## 2.1 Introduction

In this chapter, we address gaps and limitations of PFF probability estimation on roadways by

incorporating crowdsourced navigation data from the Waze navigation app as highly localized

flood observations into high-resolution data-driven PFF vulnerabilities that can capture the

impact of unobserved variables and estimate the vulnerability of road segments to PFF. The

proposed approach can be updated and implemented rapidly to provide near-real-time

navigational warnings. The initial framework developed in this chapter has three steps. In the

first step, road surface depressions and their upstream catchments are delineated from a high-

resolution digital elevation model using simplified flow-routing and hierarchical fill spill

approaches. In the second step, two statistical and machine learning models─Empirical Bayes

(EB) and random forest (RF)─are developed and tested to predict PFF frequency using roadway,

catchment, depression, and rainfall characteristics. In the third step, roadway flooding

likelihoods (i.e., vulnerability) and flood maps are generated that could be disseminated to

navigation software. To our knowledge, this study is the first to develop real-time PFF

vulnerability maps at road segment scales using data-driven models and crowdsourced traffic

data. With the widespread use of smartphones and crowdsourced applications, this study shows

the benefits of integrating crowdsourced data and statistical modeling approaches into roadway

flood awareness and management systems.

## 2.2 Methodology

The three steps of the framework developed are shown in Figure 2-1. The first step involves data preprocessing to create the dataset needed for modeling. The second step fits statistical and machine learning models to the historical dataset, and the third step performs the roadway flooding likelihood estimation for future storms. These steps are described in more detail in sections below.
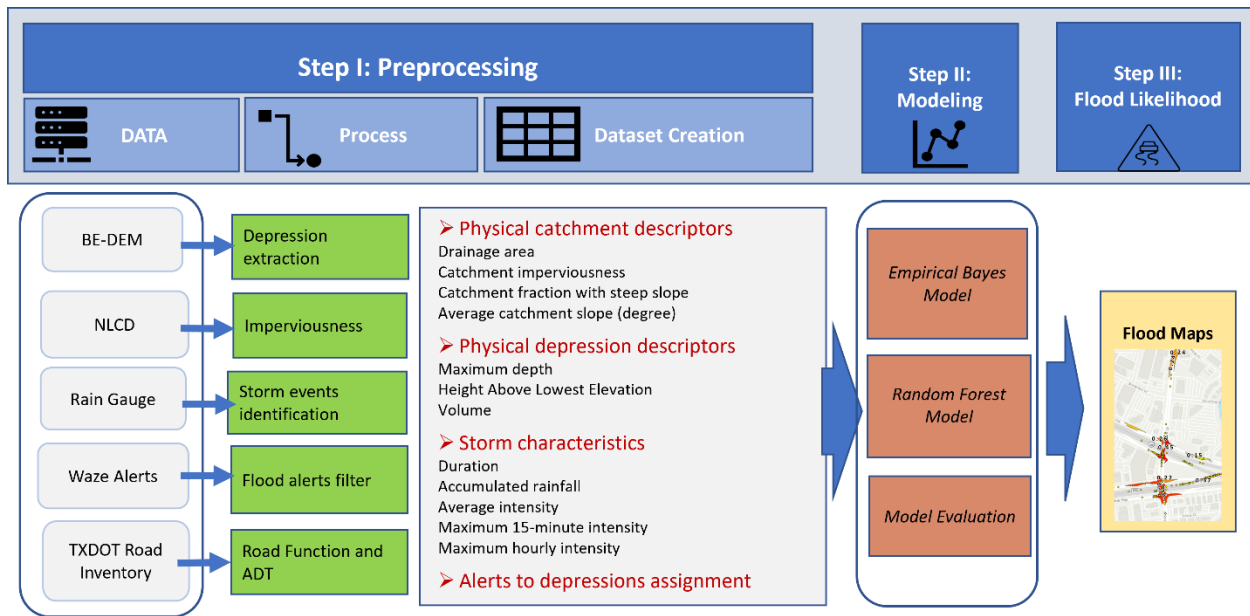


Figure 2-1 Methodology framework (basemap from ESRI-2021)

### 2.2.1 Step I: Preprocessing

The dataset preprocessing in Step I includes three primary components that are described in detail in the sub-sections below and depicted in Figure 2-1. First, road surface depressions and their upstream catchments are delineated. Second, storm events and their characteristics are determined from continuous rain gauge observations; third and last, flood alerts are assigned to corresponding depressions and storm events.

2.2.1.1  Depression Extraction

The first step of data preprocessing is to find road surface depressions that are prone to PFF.

Generally, surface depressions are defined as the difference between the hydrologically-conditioned digital elevation model (DEM) (Lindsay and Dhun, 2015) and the raw DEM. In hydrologically-connected DEM elevations of internally draining sinks are raised to form a flat area that can drain to downstream. Locating surface depressions in a highly urbanized terrain is challenging due to micro-topographic and underground features (such as curbs, stormwater inlets, etc.) that determine the actual flow path. In addition, using a high-resolution DEM (1-meter) introduces hierarchical depressions with different orders of magnitude in spatial scale, from highly localized (minor pits) to surface depressions that cover more than one neighborhood (residual depressions). Therefore, a nested hierarchy of depressions must be considered to extract depressions compatible with urban features.

In this paper, the "sink evaluation" tool of the ArcHydro toolbox (Djokic et al., 2011) is utilized to extract a nested hierarchy of surface depressions. The sink evaluation tool scans the bare earth DEM (BE-DEM) and characterizes low-lying cells. The process of local depression extraction is an iterative process that examines each sink, raises the elevation of low-lying cells to fill the sink, and then reapplies the process on the resulting DEM. This procedure is depicted in Figure 2-2. In the first sink evaluation step, Level-1 depressions are delineated and raised (Figure 2-2-a). In the second step, the DEM resulting from the first level fill (Figure 2-2-e, red areas) is evaluated and Level-2 depressions are delineated. This process can be repeated until the area is fully hydrologically-conditioned and no higher-level depressions remain. The number of steps required in this process is dependent on the resolution of the DEM and the complexity of the depressions in the landscape.

Due to the complexity of urban terrain, the spatial scale of depressions at each hierarchy level is quite variable and depressions at the same level can be as large as a neighborhood or as small as a pothole. Initially, depressions at all hierarchical levels are extracted. Because 6-inches of standing water has minimal impact on most cars (National Weather Service, 2022) depressions with maximum depth smaller than 6-inches are removed from further analysis. Next, those depressions that coincide with urban topographic features that affect flow, such as roadway curbs and gutters, are manually selected as flood-prone depressions. Flood-prone depressions are then selected by examining overlays of the depressions and Waze flood reports, as well as the areas of depressions and road surfaces that the depression covers. Heuristics for this procedure are presented in detail in Section 2.2.1.5. Figure 2-2-e shows 10 depressions (L1-1 to L1-7, L2-1, L2-2, and L3-1) extracted on a road segment with three depression levels. Level-1 depressions and L2-2 appear as small potholes or single-cell pits that could be DEM errors, but regardless are too small on the road surface to cause traffic disruptions. However, L2-1 aligns with road curbs and gutters and could cause traffic disruptions by covering a large area and all lanes of the roadway. Therefore, L2-1 is manually selected as the smallest depression that is prone to PFF and could affect traffic flow on this road segment. (Note that L3-1 includes L2-1, hence it will be filled only after L2-1 has filled and disrupted traffic flow already. Hence, L3-1 does not need to be included in the model for traffic navigation purposes.)

2.2.1.2   Physical Catchment and Depression Descriptors

After delineating road surface depressions, physical descriptors of depressions and their upstream catchments are computed as follows. Two sets of characteristics, summarized in Table 2-1, are defined for every depression that is selected in the previous extraction step: physical depression descriptors (PDD) and physical catchment descriptors (PCD) (Kalantari et al., 2017,

12

2014). PDD features describe the depression topography that is likely to affect water accumulation. These features are area, average depth assuming the depression is filled, and the height of road DEM cell elevations above the lowest elevation of the depression (hereafter called Height Above Lowest Elevation, or HALE). The HALE feature indicates which terrain cells would be inundated first and the accumulated depth required for flood water to reach that grid cell. Figure 2-3 shows a schematic of the HALE and depth features. The PCD features are derived from the upstream catchment that drains into each depression. The extracted features are *average slope*, fractions of the upstream catchment with a steep slope (defined as steeper than 8%), *percentage of imperviousness*, and the *net log-transformed drainage area*, hereafter called net drainage area ($NetDA$) which is computed using Equation 2-1.

Equation 2-1

$$NetDA = Log(CA) \times I$$

where: $CA$ is the catchment area in $m^2$ , and $I$ is the percentage imperviousness of the catchment based on the National Land Cover Dataset (NLCD). $Log(CA)$ represents the nonlinear relationship between catchment area and flood likelihood. This can happen because the larger the drainage area, the higher the impacts of infiltration loss and stormwater drainage that were not considered in this analysis.
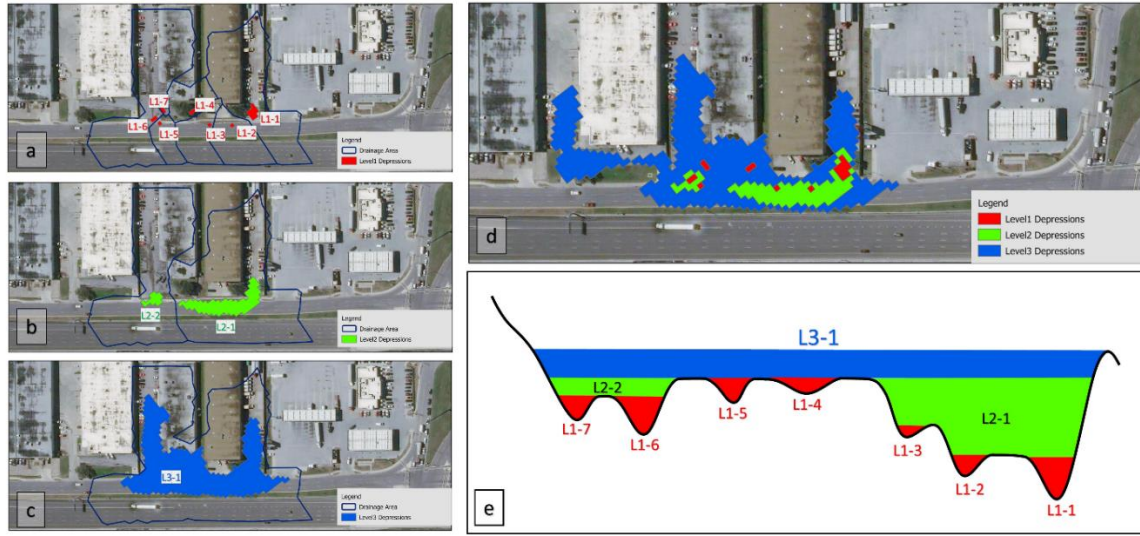
Figure 2-2 Hierarchical filling of surface depressions (basemap from ESRI-2021)
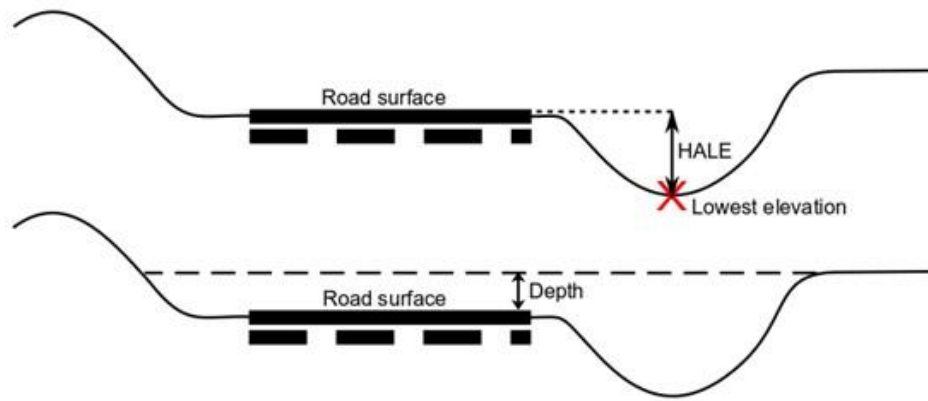


Figure 2-3 Schematic of HALE and depth features

### 2.2.1.3   Traffic Exposure

Crowdsourced data are generated by volunteer contributions, which results in more data availability on roads with higher traffic volumes. Therefore, including a feature in the model that captures roadway traffic exposure to flooded areas is necessary to consider the likelihood of reporting a flooded depression. For this purpose, two additional variables are included in the framework (Table 2-1): (1) the natural logarithm of Annual Daily Traffic (ADT) and (2) the road function as defined by the Texas Department of Transportation (TX-DOT).

14

2.2.1.4  Storm Event Definition and Storm Clustering

Raw precipitation data are obtained from Automated Surface Observing Systems (ASOS) stations in continuous 5-minute interval rain pulse observations. To predict the probability of depression flooding during a storm of particular severity, independent storm events must be derived from the continuous data. In this study, the Minimum Inter-event Time (MIT) method is used to define independent storm events. The MIT approach defines a storm event as rainfall that follows and is followed by a minimum dry (rainless) period called the Minimum Inter-event Time.

Table 2-1 Physical depression/catchment descriptors

| | Depression descriptor | Definition | Unit | Source |
|---|---|---|---|---|
| **PDD** | Depression area | The area of the road surface that the depression covers | Square meters | DEM processing |
| | Average depth | The average depth assuming that the depression is filled | Meters | DEM processing |
| | Maximum depth | The maximum depth assuming that the depression is filled | Meter | DEM processing |
| | Depression volume | The volume that fills the depression | Cubic meters | DEM processing |
| | Minimum depression volume | The volume that generated 6-in depth on the road | Cubic meters | DEM processing |
| | Height above the lowest elevation (HALE) | The average height of the road above the lowest elevation of the depression | Meters | DEM processing |
| **PCD** | Net drainage area | Proxy to the runoff generated from the upstream catchment | N/A | DEM processing |
| | Upstream imperviousness | Average imperviousness fraction of the upstream catchment | Percentage | National Land Cover Dataset |
| | Upstream steep slope | The fraction of the catchment area that has a slope steeper than 8 percent | Percentage | DEM processing |
| | Average upstream slope | The average slope of upstream catchment | Degree | DEM processing |
| **Road** | Log ADT | Natural logarithm of the ADT | Vehicles per day | TX-DOT Inventory* |
| | Road function | The function of the road as 1: interstate, 2: Freeway and Expressway, 3: Principal Arterial, 4: Minor Arterial, 5: Major Collector | N/A | TX-DOT Inventory* |
| **\*- https://www.txdot.gov/data-maps/roadway-inventory.html** | | | | |

The MIT value can be calculated using different approaches. A reasonable estimate of the MIT value is the lag-time at which the serial autocorrelation between rain pulses reaches a pre-set low threshold and remains steady (Asquith et al., 2005). In this study, the MIT value is diagnosed using the correlogram method to visualize the autocorrelation of a rain pulse timeseries to find the lag time that makes a rain pulse independent of its preceding rain pulses. After defining independent storm events, storm characteristics, including accumulated precipitation, duration, average intensity, and maximum 15-minute, 30-minute, and hourly intensities, are calculated.

For storm events with similar characteristics, similar depression PFFs are likely to occur. To capture this phenomenon, storms are clustered into classes with similar severity (light, moderate, severe) using the following storm characteristics: rainfall intensity, rainfall depth, and storm duration. For storm clustering, agglomerative hierarchical clustering is applied using a bottom-up approach that forms a single cluster for each storm event and successively merges clusters based on Ward's linkage method. Ward linkage method minimizes the total increase in within-cluster variance (Edelbrock, 1979) caused by merging clusters. The benefit of using agglomerative clustering is that this algorithm is less sensitive to outliers compared with other commonly used clustering algorithms and avoids creating a large number of small clusters for extreme storm events (Edelbrock, 1979).

## 2.2.1.5   Waze Data Preprocessing

Waze is a GPS-based traffic navigation app that collects crowdsourced information about road conditions. The Waze app aggregates traffic incidents reported by its users as traffic alerts. Traffic alerts are geotagged points with two attributes that specify their lifetime: 'publish date' and 'last seen'. Waze alerts are a type of implicit Social Media Geographic Information (SMGI)

for which, other than intentional and unintentional manipulation, can also be subject to conceptual misjudgment. To distinguish between Waze flood alerts that are reliable as flood observations and the non-flood related Waze alerts, their content, location, and publish time are investigated. The Waze app has no pre-qualification for users to post a report, consequently not all of the flood-labeled alerts are reliable to be used as flood observations. Praharaj et al (2021) showed that 71% of Waze flood alerts are reliable in Norfolk, Virginia. To investigate Waze alerts' authenticity, flood-related alerts are matched to the most recent rainfall event and computed the delay between alerts' publishing and rainfall end-time. A temporal threshold is found by analyzing the cumulative distribution of delays that determines whether a flood report is related to a storm event.

In addition to alert timing, the locations of Waze alerts are compared to publicly available datasets of high-flood-risk locations, including the National Flood Hazard Layer (NFHL), high watermarks and low water crossings data inventories from the North Central Texas Council of Government (NCTCOG), and the road surface depressions computed as described in the methodology section. The NFHL is a spatial dataset that uses river flood hazard information provided by the Federal Emergency Management Agency (FEMA) to generate flood hazard maps showing areas at high risk of flooding. The proximity of Waze alerts to the high flood risk locations were investigated to find the spatial accordance of flood alerts to these locations.
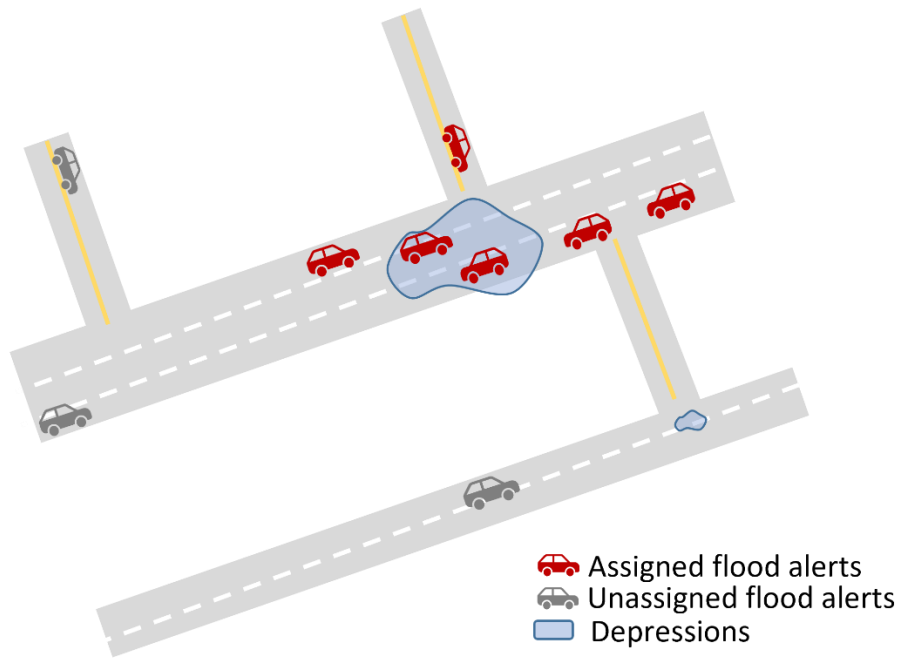
Figure 2-4 Alert assignment

One challenge in adopting Waze flood-related alerts as roadway PFF observations is assigning the alerts to the appropriate flooded location because the coordinates of alert points do not perfectly align with flooded location coordinates. The distance between the flooded location and alerts depends on many unknown factors such as drivers' reaction times, direction, and sight distance. Posting a flood alert requires Waze users to complete three steps (three selections) in the app while driving or riding and users can post a flood alert before or after passing the flooded road segment. Hence assigning flood alerts to the proper depression must be done carefully. Waze data do not provide the direction of travel. Thus, no constraints regarding the travel direction have been used for assigning flood alerts to flooded depressions, since depressions can cross both sides of the road.

In this study, three independent individuals were asked to separately visually assess a map of historical flood alerts laid over surface depressions and assign alerts to depressions using

the following criteria: a cluster of more than two flood alerts should be available near the depression and the depression must be distinct from other nearby surface depressions. Flood alerts posted from bridges and elevated highways are excluded since BE-DEM does not represent bridge surfaces. Figure 2-4 shows a schematic example of alerts that can be assigned to the depicted depression and some that should remain unassigned because they are isolated and too far from a depression.

### 2.2.2   Step II: Modeling

Pluvial flooding on any given surface depression can be modeled as a Bernoulli trial (i.e., a random binary function such as a coin flip) of flood failure (i.e., non-flooded) or success (i.e., flooded). If a depression has one or more Waze flood alerts linked to it, the depression is labeled as flooded (success). Assuming that the probability of being flooded is smaller than the non-flooded situation and that the likelihood of flooding in a particular storm event for each depression only relies on its characteristics and the storm magnitude (i.e., is independent of the probability of flooding on other depressions), a random variable $y_{i,j}$ will define the count of successes (flooding) out of the $N$ trials ($N$ storm events of cluster $j$) on depression $i$. The purpose of this study is to estimate the random variable $y_{i,j}$ using extracted topographic features, road function, and storm severity. Both statistical and machine learning models are implemented to estimate $y_{i,j}$, namely Empirical Bayes and Random Forest. Table 2-2 summarizes the categories of pre-processed independent variables used in the modeling.

Table 2-2 Count dataset of PFF events

| | PDD & PCD | Road function (categorical) | Storm clusters | Count of Flooding[1] (count) |
|---|---|---|---|---|
| *Depression $i^{th}$* | Topographic Features | Interstate Freeway-Expressway Principal Arterial Minor Arterial Major Collector | Light | $y(i, j = light)$ |
| | | | Moderate | $y(i, j = moderate)$ |
| | | | Severe | $y(i, j = severe)$ |

## 2.2.2.1 Empirical Bayes Model

In a highly urbanized area there are numerous uncertain and unobserved site-specific features that affect localized PFF vulnerability, such as nearby storm drain inlets' locations, age, capacity, and condition. For example, consider two road surface depressions (A and B) with similar PDD, PCD, road type, and ADT that experience the same storm. Suppose Depression A is located in a neighborhood with lower infrastructure maintenance services and its drainage system clogs more often. Then, despite similar descriptive features, higher flood frequency should be expected at depression A.

The Empirical Bayes (EB) algorithm, a simplified and faster version of Bayes theory, takes advantage of the historical count of reported flood events from the Waze data to better reflect the impacts of these types of uncertain and unobserved variables. The EB approach has previously been implemented in many fields to address the impacts of unobserved variables in estimating rare events, including hydrology (Fill and Stedinger, 1998; Hauer et al., 2002; Kuczera, 1982; Lord et al., 2005; Smith et al., 2014; Strupczewski et al., 2001). The EB method uses the joint global prior and site-specific counts and produces the posterior probability $y_i$ by employing a weighted average as shown in Equation 2-2 (Fill and Stedinger, 1998; Hauer et al., 2002; Kuczera, 1982; Smith et al., 2014).

Equation 2-2

$$EB(y) = w \times \mu + (1 - w) \times y$$

where $w$ is the EB weight factor, $\mu$ is the expected flood frequency on depressions similar to a given depression, and $y$ is the number of flood events on a given depression. The expected flood frequency for similar depressions ($\mu$) is the global prior probability distribution from a fitted regression model, which in this study is a Negative Binomial regression model. The number of flood events ($y$) is the historical site-specific flood event observation from the Waze data.

2.2.2.2  Negative Binomial Distribution

Based on Waze flood observations, the variance of flood frequencies on depressions with similar PDD, PCD, road type, and ADT is assumed to be greater than the average of flood frequencies (i.e. $E(y) < Var(y)$). This assumption is appropriate given the importance of unobserved variables on the PFF formation on roads such as storm inlet conditions. In other words, among $n$ similar surface depressions, $k$ depressions, where $k \ll n$ experience flooding significantly more than average. This fact leads to an over-dispersed dataset where $E(y) < Var(y)$.

Studies have shown that in the case of over-dispersed data, $y_i$ follows a Poisson distribution with the rate parameter $\lambda_i$, where $\lambda_i$ follows a Gamma distribution with the dispersion parameter $\phi$ and the rate parameter $\phi/\mu_i$. The resulting distribution is Poisson-gamma, also called the Negative Binomial (NB) distribution (Zou et al., 2018). The probability mass function of the NB distribution is given in Equations 2-3 and 2-4. Therefore, in this study, the expected flood frequency on similar depressions in the EB equation (Equation 2-2), is derived from a Negative Binomial (NB) regression model that is fit to the count dataset shown in Table 2-2. NB parameters ($\phi$ and $\beta_i$) are estimated using the Maximum Likelihood Estimation method.

Equation 2-3

$$P(y) = \frac{\Gamma(y + \phi)}{\Gamma(y + 1)\Gamma(\phi)} \left(\frac{\phi}{\phi + \mu}\right)^{\phi} \left(\frac{\mu}{\mu + \phi}\right)^{y}$$

Equation 2-4

$$\mu = exp(\textstyle\sum \beta_k x_k)$$

where $\phi$ is the dispersion parameter of the NB distribution, $y$ is number of flood events on

depression $i$, and $\mu$ is the expected flood frequency on a given depression based on similar

depressions (Equation 2-4). $\beta_k$ is the coefficient of $k^{th}$ regressor variable in fitted regression

model, and $x_k$ is the value of $k^{th}$ regressor on a given depression.

Model selection for the NB regression model is implemented using the Bayesian

Information Criterion (BIC). In model selection, minimizing the BIC to the simplest model with

the least number of exploratory variables is reasonable. Reducing the BIC by adding more

explanatory variables increases the risk of overfitting and loss of generality. Equation 2-5 shows

the calculation of BIC.

Equation 2-5

$$BIC = -2\, log(L) + K.Ln(n)$$

where $L$ is the maximum likelihood of the model representing the overall fit of the model, $K$ is

the number of model parameters, and $n$ is the sample size.

It can be shown that the weight in the EB equation based on the NB regression is

calculated as $\frac{\phi}{\mu+\phi}$, hence we can rewrite Equation 2-2 as Equation 2-6. For more information

regarding the mathematics of  deriving the weight factor of EB, refer to Zou et al. (2018).

Equation 2-6

$$EB(y) = \frac{\phi}{\mu + \phi}\mu + \frac{\mu}{\mu + \phi}\, y$$

where $\phi$ is the dispersion parameter of NB distribution, $y$ is number of flood events on a given

depression, and $\mu$ is the expected flood frequency on a given depression based on similar

depressions (Equation 2-4).

The EB model's predictive power is estimated using the mean absolute error (MAE). The MAE shows the average error of the fitted values across the observations. The lower the MAE, the better the EB estimates fit the observations. The MAE is calculated using Equation 2-7:

Equation 2-7

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

where $n$ is the sample size, $y_i$ is number of flood events on depression $i$, and $\hat{y}_i$ is the EB predicted number of flood events on depression $i$

### 2.2.2.3 Random Forest

Random Forest (RF) is a supervised ensemble machine learning algorithm that uses multiple decision tree learners to increase predictive performance (Breiman, 2001). A decision tree consists of a hierarchy of nodes, each of which represents a conditional decision rule that splits the data into different decision paths. The final prediction of RF is the average prediction of all decision trees; each tree is built from a bootstrap sample (i.e., sampling without replacement from the full dataset) of observations and a subset of features. The RF has been widely used for data-driven modeling in the field of water resources (Sadler et al., 2018). This algorithm can handle large and imbalanced datasets and is well known to be easy to train. An important strength of the RF is that its convergence rate is independent of noise and sparsity in the descriptive variables. RF models are useful for estimating the contribution of features in the target variable (in this case, flood frequency). The node impurity in each node of the RF is an important performance metric for decision trees because it measures the heterogeneity of the target values (predicted count of observed flooding events) at that node, which is the variance of the target values in this regression problem. The count of observed flooding events. The normalized reduction in the node impurity achieved by adding a specific feature to a tree defines

the importance of that feature. In RF, the average of importance of a feature in all trees weighted by the number of samples involved in each split is the overall feature importance.

In this study, RF regression is executed using the Scikit-Learn library in the Python environment (Breiman, 2001; Fabian Pedregosa et al., 2011). The number of decision tree learners in the RF regression is optimized by the algorithm. For hyperparameter tuning and model selection, a randomized cross-validated grid search is applied on a wide range of model parameters and MAE is used to measure parameter performance and select the best-performing parameter set. The resulting parameters are then used to estimate the frequency of PFF at every depression for each storm class using Equation 2-8.

Equation 2-8

$$RF(y) = RF(PDD, PCD, road\ features, storm\ type)$$

where $RF(y)$ is the random forest prediction of number of flood events on a given depression.

## 2.2.2.4   Model Evaluation

To evaluate the performance of the proposed model, the following approaches are used. First, 80% of the historical data, randomly selected, are used in model training. Model testing is then implemented using the remaining 20% of the data held out from the training process. The performance of the models is then assessed using the MAE of the predictions. In order to ensure that the models are stable and their performance does not change with different train-test sets, the models are trained and evaluated for several randomly chosen training sets and the variation in their performance is considered in selecting the best models for the final step of the framework.

Then, to further assess the improvements in PFF event estimation using topographic and historical Waze observations, the EB and RF models are compared with three simple benchmark models. First, the average model (Equation 2-9) assumes that the average PFF counts from historical Waze observations apply to all depressions and all storms without considering storm

type and topographic feature. Second, the storm-based average model uses the average of the PFF count in each storm cluster without considering topographic features (Equation 2-10). Finally, a regression model is used that predicts PFF based on topographic, road type, and storm features but without implementing EB to update the prior probability (Equation 2-4).

Equation 2-9

$$p(i) = \frac{\sum_1^n y_i}{N_t}$$

where $p(i)$ is the likelihood of flooding on depression $i$, $y_i$ is the number of reported floodings on depression $i$, $n$ is total number of depressions, and $N_t$ is number of total storm events.

Equation 2-10

$$p(i,j) = \frac{\sum_1^n y_{i,j}}{N_j}$$

where $p(i,j)$ is the likelihood of flooding on depression $i$ and storm type $j$, $y_{i,j}$ is the number of reported floodings on depression $i$ and storm type $j$, and $N_j$ number of total storms of type $j$.

### 2.2.3   Step III: Flood Probability Estimation

Finally, in Step III, the most accurate model from Step II is used to produce flood probability maps for every storm cluster across the region of interest. The probability of flooding is calculated using Equation 2-11.

Equation 2-11

$$p(i,j) = \frac{\hat{y}_{i,j}}{N_j}$$

where $p_{(i,j)}$ is the likelihood of flooding on depression $i$ in a storm of type $j$, $\hat{y}_{i,j}$ is the predicted number of floodings on depression $i$ and storm type of $j$, and $N_j$ is the number of storms of cluster $j$.
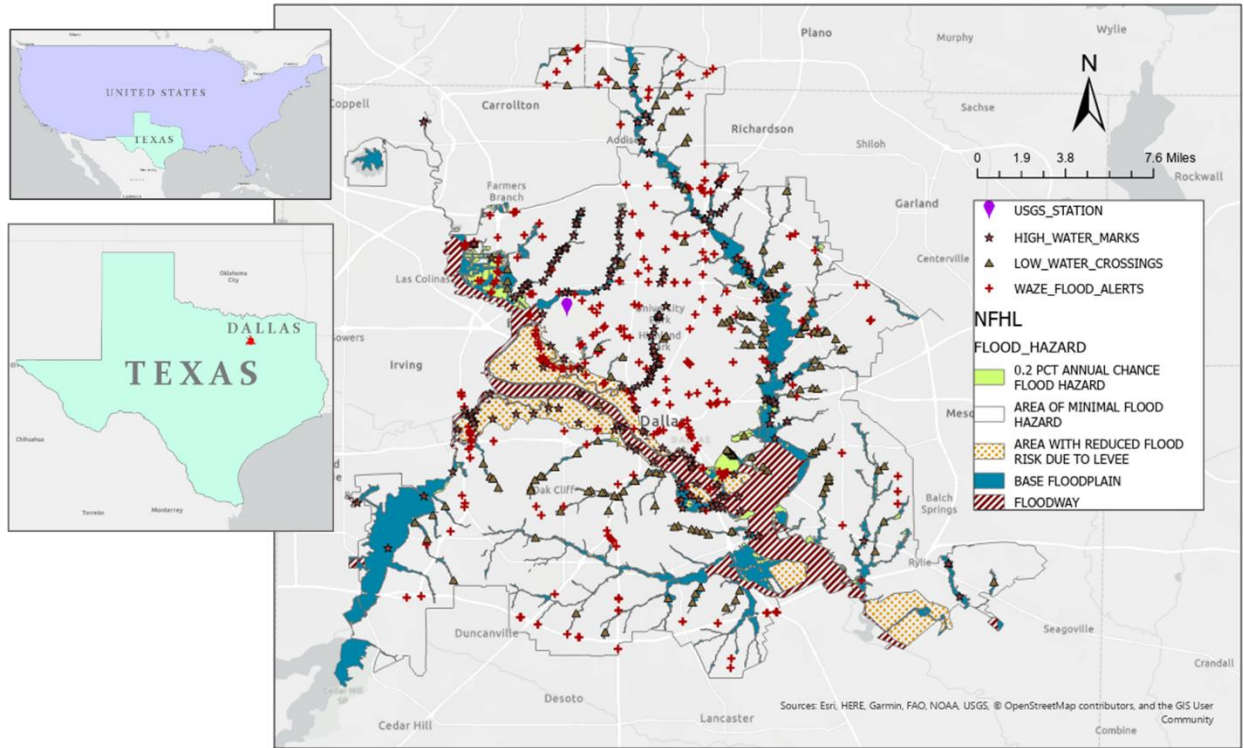
Figure 2-5 Study area and datasets

## 2.3 Case Study Background and Datasets

The described methodology was evaluated in the city of Dallas, Texas, USA (Figure 2-5), which is the third-largest city in Texas with a population of more than 1 million. Dallas elevation ranges from 137 to 168 meters (450 to 550 feet), and it is mostly flat. According to the Texas Department of Transportation (TXDOT), almost 20 percent of crashes, equal to 248 vehicle crashes in the City of Dallas in 2018, happened on either standing water or wet road surface conditions. According to an analysis conducted by the First Street Foundation, flooding can expose 1841 miles of Dallas roadways (out of 6064 miles) to the risk of becoming impassable (Risk Factor, 2021). However, currently available fire-rescue dispatch software, including that used by the Dallas Fire-Rescue Department (DFRD), assumes empty and dry roads for routing rescue vehicles. This has resulted in rescue delays and occasional loss of life on flooded roadways, which provided the motivation for this study.

26
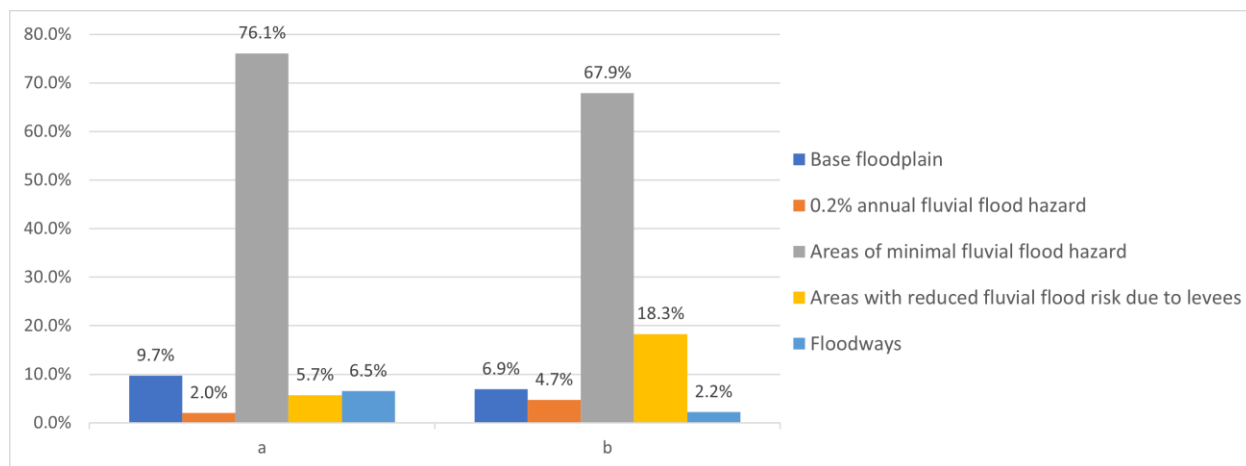
Figure 2-6 a: Flood alerts over NFHL flood zones b: Distribution of NFHL flood zone over the study area

For this case study, several datasets were used. First, a 1-meter resolution Bare Earth Digital Elevation Model (BE-DEM) was obtained from the North Central Texas Council of Government (NCTCOG), which was derived from a Quality Level 2 Lidar survey performed by Digital Aerial Solutions, LLC, in 2018, under contract with the Unites States Geological Survey (USGS)/ National Resources Conservation Services (NRCS). The BE-DEM dataset's name is TX Pecos Dallas 2018 D19, with horizontal accuracy of +/-0.682 meters at a 95% confidence level and non-vegetated vertical accuracy (NVA) of 0.196 meters.

For rainfall, 15-minute precipitation observations were obtained from the USGS ASOS station at Dallas Love Field Airport (DAL) (Figure 2-5). Precipitation observations from January 1, 2017, to March 1, 2020 were used. Next, the US Department of Agriculture's (USDA) National Land Cover Database (2016) (Homer and Fry, 2012) is used to extract catchment imperviousness. The imperviousness raster over Dallas has a 30-meter resolution and ranges from 0 to 100%, with a mean of 33.87% and standard deviation of 32.98%.

Waze alerts were obtained from the NCTCOG, which is a Waze partner in the Waze Connected Citizen Program (CCP). The NCTCOG granted us access to the Waze data for the

27

period of 2018-04-21 (the start of NCTCOG's Waze partnership) to 2020-03-20. Waze alerts are classified into seven main categories: accident, jam, construction, miscellaneous, hazard or weather (hazard-weather), road-closure, and others. The "hazard-weather" data itself is divided into several subcategories. Alerts in the "flood" subcategory and ones which have any form of the word "flood" in their report description, such as "right lane flooded," are potentially flood-related and were included in this study, resulting in *5652* Waze alerts.

The locations of these Waze alerts were shown in Figure 2-5, along with the NFHL river flood zones. Figure 2-6-a shows that the majority (around 70%) of alerts during the study period were posted in areas with minimal river flood hazard, which comprise approximately 76% of the study area (Figure 2-6-b). Another 18% of the alerts were posted in areas of reduced river flood risk due to levees, which were not breached during the study period. This indicates that PFF is likely the cause of most Waze alerts.

To further investigate the potential causes of Waze flood alerts, the high-water marks inventory and low-water crossing dataset were obtained from the Texas Natural Resources Information System (TNRIS). The high-water marks inventory contains historic high water level reports from flooded water bodies or structures at 334 locations across the city of Dallas (Figure 2-5). The low-water crossing dataset includes 175 locations where surface water has crossed roads during high-flow conditions (Figure 2-5). Analyzing Waze alert distances to the nearest high-water mark and low-water crossing shows that the vast majority of alerts are more than 200 meters from both low-water crossings and high-water marks (Figure 2-7).
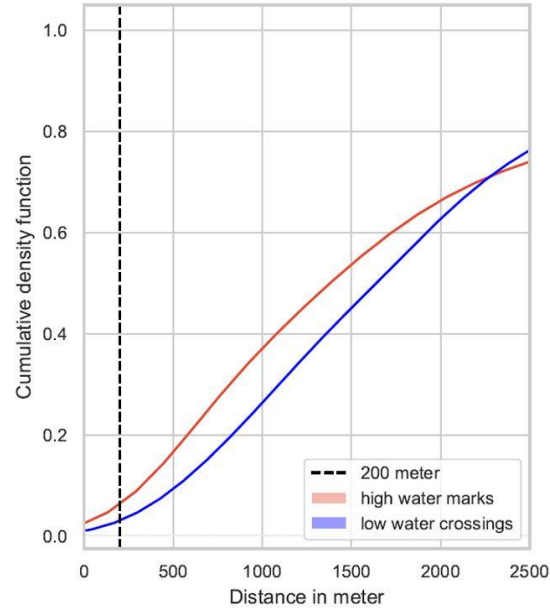
Figure 2-7 Cumulative density of alert distances to closest high-water mark and low-water crossing

These findings show how complementary flood observations such as Waze data are needed to assess roadway conditions more comprehensively than available official datasets. Thus, in order to predict local roadway PFF, it is necessary to consider local surface depressions as low-lying areas where surface runoff can accumulate during storms.

## 2.4    Case Study Data Pre-Processing

### 2.4.1    Depression Extraction

Almost 380,000 surface depressions were extracted over the city of Dallas in three steps of hierarchical filling (described in the methodology).  Only 315 depressions are located on roads and deeper than 6-inches. Among these 315 depressions, 191 depressions were proximal to reliable Waze flood alerts more than twice. To consider only chronically flooding areas, the rest of this analysis is focused only on these 191 surface depressions.

### 2.4.2    Storm Event Definition

As described in the methodology section, the autocorrelation of rain pulses defines the optimal MIT for independent storm events. Figure 2-8 shows that the autocorrelation coefficient

29

first reaches a low value and remains steady at a lag time of 9 hours; accordingly, MIT = 9 hours is chosen to convert the continuous precipitation data into independent storm events.
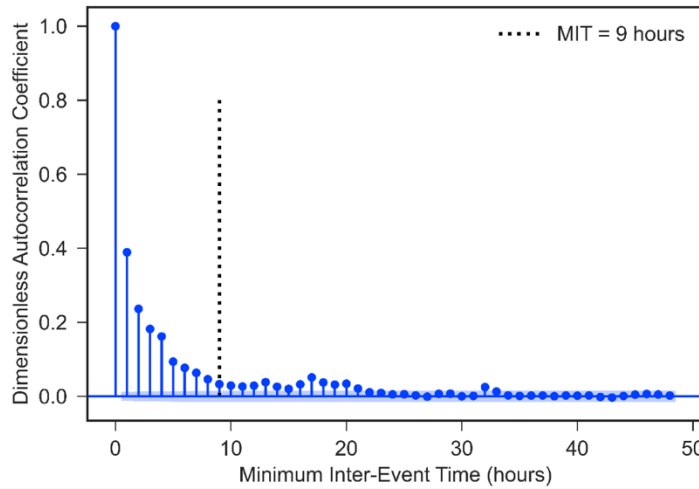


Figure 2-8Autocorrelation of rain pulses

Using MIT=9 hrs, 236 independent storm events are extracted from January 1, 2017 to March 1$^{st}$ 2020. For each storm event, duration; total accumulated precipitation; average hourly intensity; and maximum 15-minute, 30-minute, and one-hour interval intensities were computed. Once storms were defined, they were clustered using the agglomerative clustering algorithm discussed in Section 2.2.1.4. The maximum 15-minute interval intensity and the total accumulated precipitation were found to have the most significance in separating storms into the clusters. To define the optimum number of clusters, Ward linkage method was used to minimize the total within-cluster variance (Edelbrock, 1979). Figure 2-9 shows a dendrogram that illustrates how clustering the storms into three groups captures acceptable dissimilarity between storms, which are defined as light, moderate, and severe storms. The vertical axis of the dendrogram depicts the dissimilarity (Ward distance) between storms, and the horizontal axis represents storms. The position of each split on the vertical axis shows the dissimilarity of the two clusters on each side of the split. Table 2-3 shows summary statistics for the three storm clusters.

30
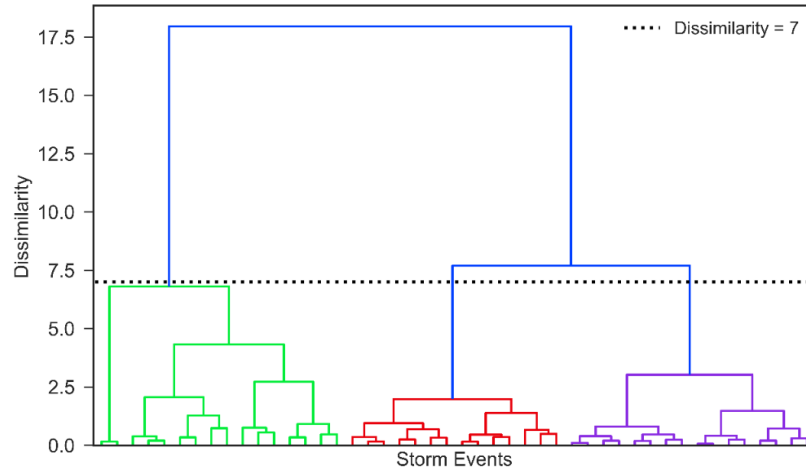
Figure 2-9 Tree-based dendrogram of agglomerative clustering

Table 2-3 Summary statistics of storm clusters

| Storm cluster | Number of storms | Mean of Maximum 15-min intensity (in/15 min) | Mean of total precipitation (in) | Mean of duration (hours) |
|---|---|---|---|---|
| Light | 142 | 0.05 | 0.12 | 4.08 |
| Moderate | 70 | 0.29 | 0.85 | 8.89 |
| Severe | 24 | 0.72 | 2.99 | 18.59 |

### 2.4.3   Waze Data Preprocessing

Potential flood-related alerts posted in the timespan of 2018-04-21 to 2020-03-20 were matched to their most recent preceding storm, and the delay between the time of each alert's posting and the end of rainfall was calculated.  Figure 2-10 gives the distribution of delays between alert's published time and storm end. Figure 2-10 shows that more than 90% of Waze flood alerts are posted within 5 hours of storms. Therefore, potential flood-related alerts posted later than 5 hours after storms were considered outliers (noise) and removed from the analysis. This processing left 4,996 flood-related alerts out of the initial 5,652 alerts. The number of flood alerts posted per storm event ranged from 0 to 375, with the distribution depicted in the Figure 2-

11. During the study period, 150 storms occurred, but only 98 storms were associated with Waze flood alerts. On average, each storm event had ten flood alerts.

The process of flood alert assignment, explained in Section 2.2.1.5 of the methodology section, was performed for the 4,996 flood alerts in the Dallas case study by three independent annotators. When more than four alerts were clustered around a depression, 100% agreement between the annotators was observed in the assignment of alerts to depressions. Disagreement between annotators was observed in locations where fewer than four alerts are clustered around a depression. The author reviewed alerts that indicated disagreement, and if the specified criteria for making the assignment were not met, alerts were removed from the analysis. Among the 4,996 flood alerts that were filtered, 2,665 alerts were assigned to 191 independent surface depressions using the approach described in the methodology section.
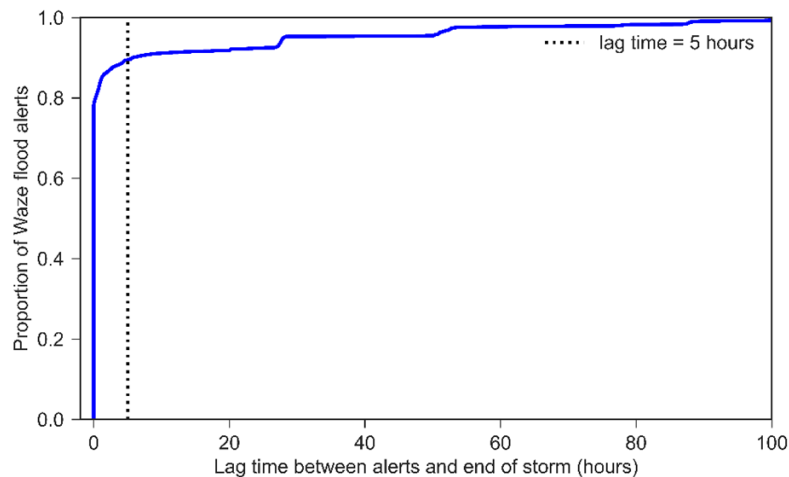


Figure 2-10 Distribution of delay in alert posting from storm end

## 2.5 Results

The performance of the proposed framework in estimating flood frequency is evaluated using both the Empirical Bayes (EB) and Random Forest (RF) models and compared to the baseline

models. Results from the best-performing model, EB, are then examined in more detail in the following sections.
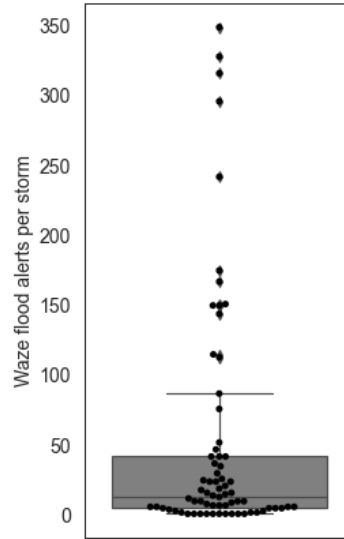


Figure 2-11 Total number of flood alerts per storm

### 2.5.1 Model Parameters and Performance

A random 80%-20% train/test split is implemented to evaluate models. Models are fit using a randomly-selected training dataset that represents 80% of the processed flood alert dataset, with the remaining 20% of the dataset (the testing dataset) used for assessing the predictive power of the models. Parameters for the fitted NB model (Equation 2-4) are presented in Table 2-4.

Table 2-4 NB model estimation results

| Variable | Coefficient | Standard Error | Z value | P-value |
|---|---|---|---|---|
| Constant | -9.3e+1 | 3.17e-1 | -295.3 | 0.000*** |
| Moderate storm | 5.1e+1 | 2.1e-2 | 3.1 | 0.000*** |
| Severe storm | 7.6e+1 | 2.4e-4 | 3.3 | 0.000*** |
| Net DA | 8.1e-3 | 1.2e-1 | 445.9 | 0.001** |
| Average slope | 6.3e-2 | 1.1e-1 | 686.2 | 0.003** |
| Log ADT | 8.2e-2 | 2.7e-2 | 3.0 | 0.003** |
| | Goodness of fit | | | |
| BIC | 1836.31 | | | |
| MAE | 1.74 | | | |
| *** significant with more than 99% confidence <br> ** significant with more than  95% confidence | | | | |

The dispersion parameter of the fitted NB regression model ($\phi$ of Equation 2-3) is 2.943. A value of $\phi > 1$ demonstrates that the over dispersion assumption is valid, whereas $\phi < 1$ shows an under-dispersed dataset. The MAE value achieved from fitting the NB distribution is 1.74, which shows that the flood frequencies fit to the prior probability distribution have an average error equal to 1.74 flood events out of 150 storms. The EB estimate of the fitted NB regression model, computed based on Equation 2-6, reduces the MAE on the training set to 0.88 flood events.

For the RF model, hyperparameter tuning is implemented using a 3-fold cross-validated randomized search in the Scikit-Learn library in Python programming environment. The best-performing model is found to have ten trees. The features with the highest importance (based on impurity-based feature importance calculated by the Scikit-Learn library) in the RF model are severe storms, maximum depth, average upstream slope, *log ADT*, and the *net drainage area*. The MAE of RF estimates on the training set is 0.73.

The predictive power of both models is evaluated on the held-out test dataset. The EB approach predicts the number of flood events for unseen situations with MAE=0.92, while the RF models' evaluation MAE is considerably higher, with MAE=2.1. To minimize the impact of particular train-test datasets on the models performance, the dataset is randomly split 50 times and the model performance statistics are re-evaluated for each split. Across these splits, the EB model has an average MAE of 0.89, as opposed to the average MAE of 1.92 attained by the RF model. EBs predictive capability is also more stable across the 50 runs than the RF model, with MAE standard deviations of 0.11 and 0.18, respectively. Figure 2-12 shows the prediction power of the models on the train and test datasets.

It can be seen that the RF model is a better fit on the training dataset but its lower performance on the test set shows that it is overfitting on the training set while the EB approach has more consistent performance on both datasets. The superiority of the EB model shows that the unobserved features play a significant role in PFF formation on road segments and a Bayesian approach is more successful in capturing the effects of these features by incorporating site-specific historical observations.

Next, the EB model that is found superior to the RF model is compared with the simple benchmark models given in the methodology section (2.2.2.4). Figure 2-13 demonstrates how the flood counts will be predicted on the test dataset using each benchmark model, NB regression, and EB model. Table 2-6 summarizes the performance of the EB approach, NB regression, and benchmark models. It can be seen that the MAE for both training and testing sets improves by adding storm clusters to the average model. This increase is more noticeable in light storms (almost 50% improvement for both training and testing datasets).

However, adding topographic and observed flooding variables, as in the EB model, increases the accuracy of PFF count estimation for severe storms more than moderate and light storms. This shows that topographic features are more important in the formation of PFF when storms are more severe. Also, if PFF is observed at a particular location, then it is more likely to be observed at that depression again.
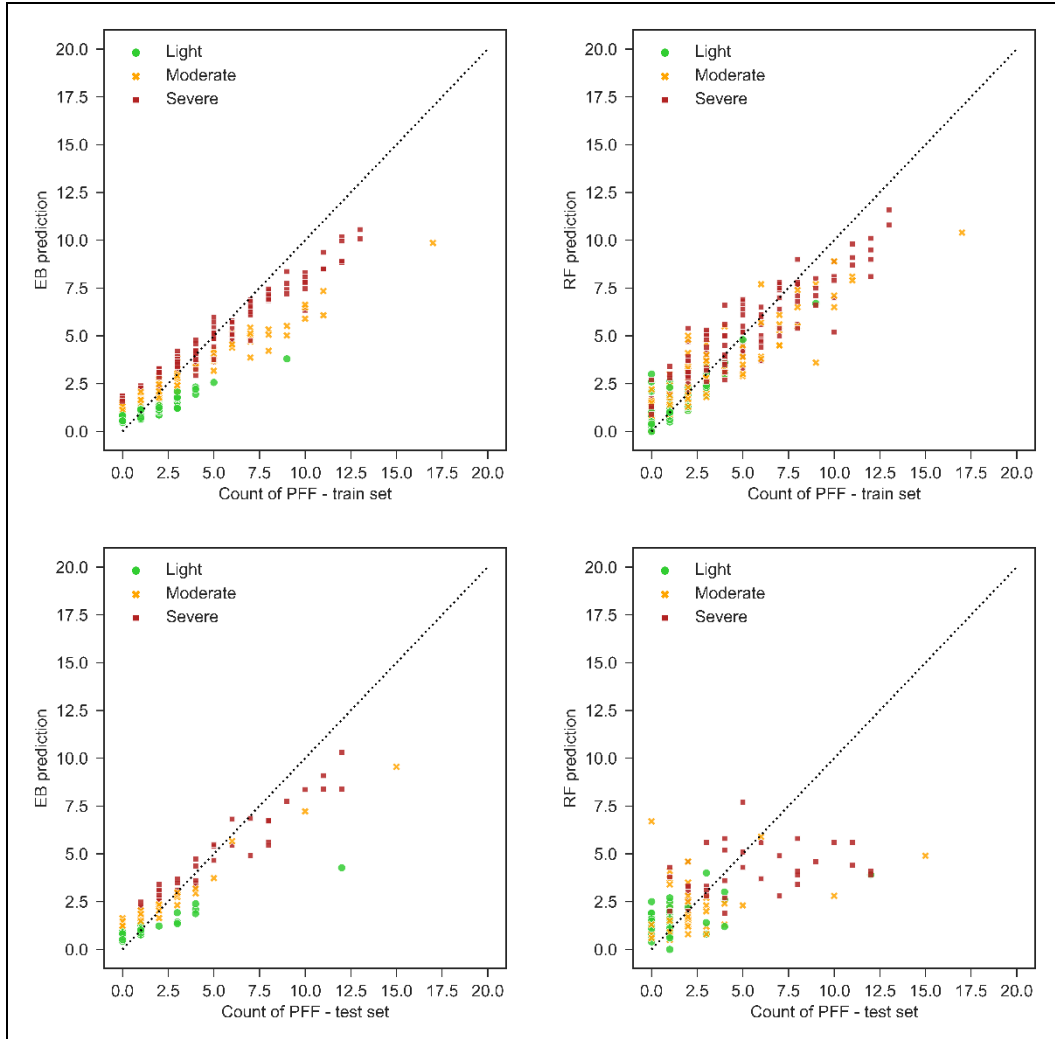
Figure 2-12 Prediction of number of roadway PFF events

Table 2-5 Summary performance of the EB approach, NB regression, benchmark and RF models

|  | MAE of train set | | | | MAE of test set | | | |
|---|---|---|---|---|---|---|---|---|
|  | Light | Moderate | Severe | Total | Light | Moderate | Severe | Total |
| **Total average** | 1.88 | 2.01 | 2.53 | 2.14 | 2.19 | 1.93 | 3.04 | 2.37 |
| **Storm cluster based average** | 0.95 | 1.97 | 2.52 | 1.82 | 1.16 | 1.86 | 2.72 | 1.89 |
| **NB regression** | 0.94 | 1.91 | 2.37 | 1.74 | 1.16 | 1.65 | 2.75 | 1.82 |
| **Empirical Bayes** | 0.69 | 1.01 | 0.93 | 0.88 | 0.84 | 0.85 | 1.09 | 0.92 |
| **Random Forest** | 0.68 | 0.98 | 0.91 | 0.86 | 1.34 | 1.66 | 2.76 | 1.92 |
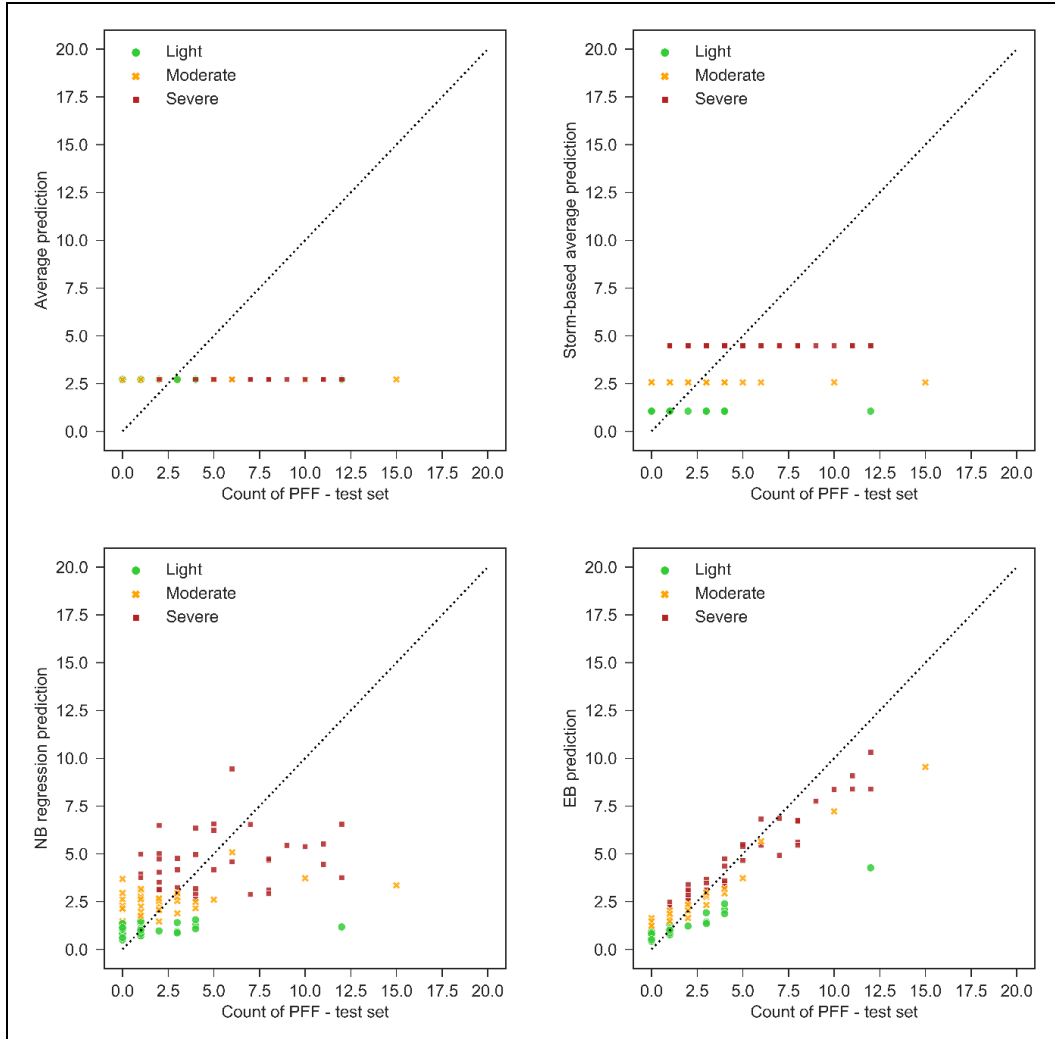
Figure 2-13 Prediction of number of PFF using benchmark models and EB

### 2.5.2   Flood Vulnerability Estimation

The EB approach is superior in predicting the total number of flood events; hence, this approach

is used to estimate flood likelihoods (i.e., vulnerabilities) from the frequency of PFF events

(Equation 2-11). Figure 2-14 shows a higher PFF likelihood during severe storms compared to

light and moderate storms. Generally, we can see that flood likelihoods are higher when flooding

has been posted. However, as discussed in the methodology section, true negative situations

cannot be identified with voluntary crowdsourced data (i.e., there could be flooding that no Waze

user has reported). Figure 2-15-a shows an example of a flood probability map for severe storms,

along with historical flood-related alerts and traffic jams reported by Waze during one severe

37

storm that occurred on September 22nd, 2018. Figures 2-15-b and 2-15-c show the same information during the same time and day of the week for the following and preceding weeks.

Waze traffic jam reports include severity and congestion levels ranging from 1 (lowest) to 5 (highest), which denote the level of traffic slow down or complete shutdown. Negligible, low, moderate, and high flood probabilities are defined as less than %10, less than %25, less than %50, and higher than %50, respectively. In Figure 2-15, high traffic levels (Waze jam levels of 5) can be seen near a depression with high PFF probability (more than %50). Figure 2-15 indicates that traffic jams during this severe storm are noticeably higher than at similar time intervals before and after the storm. These maps suggest that the traffic jam on the storm date, which agrees with the flood likelihood, is likely to be an anomaly due to flooding. This finding is consistent with the flood alerts and predictions of severe flooding at this location during the storm.
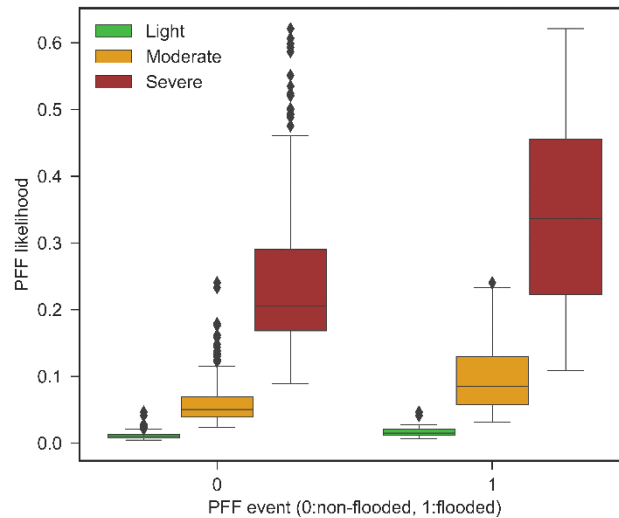


Figure 2-14 Distribution of flood likelihoods in reported versus non-reported floods

## 2.6    Discussion

The EB model is superior compared to the RF and benchmark models in predicting the number of flood events; hence this model is used to estimate flood probabilities for storm clusters. The

distribution of estimated flood probabilities (Figure 2-14, and Table 2-5) are plausible given the magnitude of the storms. For example, light storms have an average duration of 4 hours, and average total precipitation of 0.1 inches, which is quite low and flooding would not be expected during these storms. Flood-related alerts that are posted during these rainfall events can be assumed to be noise and disregarded for future studies. Based on the NB regression line that is fitted to the count of observed flood events, we expect to see 7.6 and 5.2 times more flood events in moderate and severe storms, respectively, compared to light storms. The NB model also shows that increases in the upstream *net drainage area* and *average slope* increase the probability of flooding, as would be expected. Furthermore, *log ADT* has a direct relationship with the probability of observing a PFF event because frequently-traveled roads are more likely to have Waze postings. This finding shows the limitations of estimating flood events from crowdsourced Waze datasets that tend to neglect flood events on less-traveled roads. The superior performance of the EB approach shows the significant impact of unobserved site-specific features such as stormwater inlet conditions in predicting the likelihood of PFFs on roadways. By using historical observations, the EB approach better identified frequently-flooded locations (road surface depressions), perhaps due to site-specific features such as under-sized stormwater inlets. Data were not available on these features for this study. In highly urbanized areas, these types of uncertainties in engineered structures, particularly in older areas of the city where record keeping can be poor, add to temporal uncertainties such as changing climate and land use that can affect flood formation. Despite these limitations, this study showed that localized traffic-related flood alerts are helpful in estimating PFF probabilities over a three-year period. For longer periods, periodically retraining the model to account for changes in infrastructure and climate is recommended.

To make effective use of crowdsourced traffic data, extensive preprocessing is needed to evaluate the reliability of the data and map flood alerts, which are not necessarily posted at the exact location of the flooding, to plausible nearby depressions. This process, which was done manually in this study, can introduce errors and bias to the analysis. With more data and integration of other data sources (e.g., flood sensors and stormwater inlets), an automated mapping process could be developed that could potentially reduce these errors.

Furthermore, the approach taken in this study only considers flood-prone locations reported by Waze users. Numerous parameters affect human exposure to flooded locations, and the number of Waze users that pass a road segment, such as road type, road function, day of week, and time of day. Hence, a similar flood extent on the road can cause significantly different magnitudes of traffic disruption at different times and locations, and, therefore, different flood reports. Data-driven models also have limitations due to the previously discussed dataset constraints. The EB model accounts for heterogeneity by utilizing historical frequencies. However, because of the bias and uncertainty in the Waze data, as discussed in Section 2.2.1.5, the EB model estimates will be less accurate for depressions situated on local and less-traveled routes. While major routes are more important than minor routes for minimizing exposure and risk, these limitations must be acknowledged. It is possible that, with more data, an approach to extrapolating findings on major roads to minor roads could be developed. To develop a more unbiased flood prediction model, we suggest that crowdsourced data be used as complementary data in conjunction with other data sources and models to account for less frequently traveled areas and times (e.g., during the Covid-19 pandemic, which was not included in this study when traffic was significantly reduced).
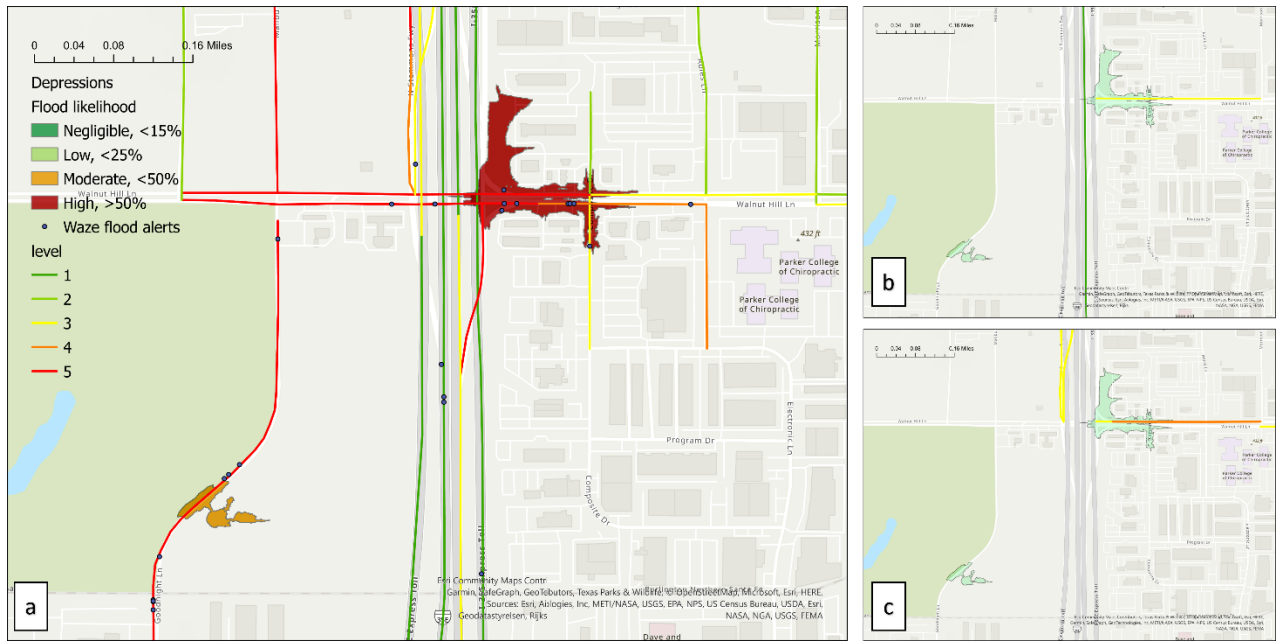
Figure 2-15 PFF probability map versus flood alerts and traffic jams from 1:00 PM to 5:00 PM on September 22, 2018, during the severe storm, b. Friday, September 29, 2018, under dry conditions, and c.. Friday, September 15ᵗʰ, 2018, also dry

## 2.7    Conclusion

This analysis is a first step in exploring approaches to implement crowdsourced data from the Waze app into flash-flood prediction. For this case study, Waze flood alerts were primarily posted in areas outside of mapped river flood hazards and low water crossings, suggesting the need for and importance of modeling rainfall-induced or pluvial flash flooding (PFF). The statistical and ML models implemented in this study demonstrated the feasibility of modeling PFF in terrain depressions based on storm, catchment, and road properties. The EB approach is found to be superior in terms of predictive power compared to RF, which shows the importance of unobserved site-specific features on roadway PFF, which the EB approach captures by incorporating historical site-specific PFF observations into the posterior probability. Both statistical and machine learning models achieve smaller MAEs for severe storms compared with moderate and light storms. This shows that the modeled depression and catchment descriptors are more explanatory in severe storms when infiltration is reduced and drainage systems are

more likely to be overwhelmed. The low MAE of the proposed methodology in the Dallas case study shows that crowdsourced traffic data has value for high spatio-temporal resolution flash flood prediction. Stakeholders and decisionmakers could benefit from the developed model for identifying locations that require stormwater utility maintenance or capital investment. Further research is needed to further explore the benefits of crowdsourced data as a complementary data source with more authoritative data and physics-based models

Acknowledgements

Chapter 3

GRAPH-BASED RAPID FLOOD SPREADING MODEL FOR REAL-TIME ESTIMATION

OF HYPER-LOCAL ROADWAY FLOODING VULNERABILITY

This study develops and tests a simplified pluvial flash flooding (PFF) simulation model that can

estimate inundation on roadways caused by direct rainfall. Localized floods generated by water

accumulation in low-lying areas and overwhelmed natural or manmade drainage systems can

degrade roadway network resilience by reducing road capacity, increasing travel time, limiting

safe speeds, and raising road safety concerns and casualties through lane submersion

(Rosenzweig et al., 2018; Smith et al., 2004; Suarez et al., 2005).

## 3.1    Introduction

To alert drivers about rapidly changing flash flood conditions on roadways, an efficient

simulation model is needed that can be used in real-time and at hyper-local scales where data

scarcity and sparsity can be a problem. To simulate interaction between surface water and

underground stormwater drainage system in an urbanized area, coupled sewer-surface models

such as 2-dimentional surface runoff models coupled with 1-dimentional sewer model (aka 2D-

1D models) are found to have the highest accuracy (Noh et al., 2018), however their application

is limited due to computational tie and high data requirements (Bulti and Abebe, 2020).

Although researchers have progressed in the mitigation of computational time and cost by

incorporating techniques such as current Graphical Processing Units (GPU)/ Central Processing

Units (CPU), Cloud computing, and dynamic grids and time-steps, data scarcity continues to be a

challenge for hyper-resolution navigational scale urban flood modeling(Noh et al., 2018; Smith

et al., 2015).

Given the challenges associated with 1D/2D flood inundation models in an urban environment and the growing access to high-resolution Light Detection and Ranging (LiDAR) derived DEMs, researchers are actively investigating alternatives that can simulate flood inundation without a large amount of input data or solving computationally extensive equations (Manfreda and Samela, 2019; Zheng et al., 2018). Rapid Flood Spreading Models (RFSM) are common approaches in this category that route the water over the terrain to generate a fast flood depth estimation in the event of storm drainage failure (Lhomme et al., 2008a).

Instead of grid cells, RFSM, also known as Hierarchical Filling and Spilling Models (HFSM), treat DEM depressions and their catchments as hydrologic units. Surface depressions in DEMs cause inward flow and hydrologically isolate flow pathways; hence they are frequently removed during DEM preprocessing (pit-filling or breaching). However, in urban PFF modeling, surface depressions have a substantial influence on the response of the urban catchment to rainfall due to runoff retention. Hence, the number of studies that employ depressions as hydrologic units rather than grid cells are increasing (Guidolin et al., 2016; Lhomme et al., 2008b; Manfreda and Samela, 2019; Samela et al., 2020; Shen et al., 2016; Yong-He et al., 2015). Surface depressions, as hydrologic units, spill and merge into other surface depressions in a sequence that depends on the terrain. The sequence of depressions spilling into each other can be described as the horizontal hierarchy, while the merging process and forming higher-level depressions is the vertical hierarchy (Balstrøm and Crawford, 2018; Chu et al., 2013; Le and Kumar, 2014; Manfreda and Samela, 2019; Samela et al., 2020; Wu et al., 2019b).

In RFSM models, surface depressions and their nested hierarchy are first delineated, then the flood volume is distributed across them. Different methods for extracting surface depressions and identifying their nested hierarchy have been proposed in the literature. Le and Kumar (2014)

used the D8 algorithm to find local minima, then applied a search algorithm to find overflowing elevations where water spills out of a depression. Wu and Lane (2017) modeled the hierarchy of hydrologic connections between depressions by establishing one horizontal and one vertical tree based on graph theory. Wu et al. (2019b) developed the level-set approach as an effective raster-based method for identifying surface depressions and their hierarchy. In the level-set approach, a hypothetical sliding plane descends from the pour-point to the lowest cell of depression and scans the depressions' nested complexity as a graph with vertical and horizontal hierarchy. While this procedure is effective and fast, it does not identify the contributing catchment of depressions.

Following the delineation of depressions, frameworks are developed to disperse flood volume amongst them based on mass conservation. Lhomme et al. (2008a) created an RFSM to route excess water from a failed defense (e.g., overwhelmed drainage or dam failure) amongst depressions by filling and spilling process. Zhang and Pan (2014) developed an algorithm for rapid urban flood inundation called the urban storm inundation simulation method (USISM), which considers filling and spilling regardless of merged depressions and vertical hierarchy. Balstrøm and Crawford (2018) created a GIS-based toolbox that depicts the overland surface as a one-dimensional geometric network. In which nodes indicate depressions, while edges represent streamlines through which spillovers flow to the downstream depression. However, they did not address merging depressions or the vertical hierarchy of depressions in their methodology. Samela et al. (2020) developed an HFSM to identify pluvial flood hazards across large urban areas called Safer-RAIN. Safer-RAIN considers the vertical hierarchy of depressions that the other approaches ignore.

Despite the mentioned advantages of simplified RFSM, one of the significant limitations of approaches discussed thus far is that they only estimate the maximum inundation extent from

total net rainfall at the end of the storm. However, for real-time flood vulnerability estimation, the temporal flood accumulation process must be modeled. To address this limitation, Chu et al. (2013) proposed an approach called P2P filling-spilling-merging-splitting model that simulates the temporal inundation evolution by updating water level in depressions at each timestep. However, their approach is employed on a laboratory basin with constant rainfall and loss rate. Thrysøe et al. (2021) developed an approach denoted FloodStroem where they coupled a one-dimensional (1D) drainage network model (MIKE URBAN) with the filling and spilling process. They used the time series of drainage system surcharges simulated by the 1D model to generate flood maps by routing surcharges across the downstream terrain using a filling-spilling process.

Given that precise information about the layout and features of stormwater drainage systems is often lacking, as well as the complex and interconnected flow paths of real-time hyper-local scale applications(Chegini and Li, 2022), RFSM models can be useful for estimating roadway flooding potential in the event of stormwater drainage failure. This study addresses the limitations of the previously introduced RFSM approaches and proposes a new model for local-scale flooding prediction that accounts for merging depressions as well as taking catchment size and land cover into account. Neglecting the impacts of the momentum of flow in roadway flooding, a simplified graph-based RFSM (GB-RFSM) is developed to simulate the vulnerability of road surface depressions to water accumulation based on real-time rainfall and, topography. The developed approach divides terrain into small hydrologic units and has the capability to account for averaged infiltration and drainage from each unit. By converting the nested hierarchy of depressions and their catchments as a directed graph that represents spilling and merging hierarchy, GB-RFSM benefits from simple graph-based calculations to account for the temporal evolution of inundation.

The performance of GB-RFSM is evaluated for a severe storm event in the City of Dallas when a noticeable traffic disruption has been reported. The estimated vulnerabilities are compared with overflows achieved from a commonly used model (SWMM) as the benchmark. The proposed tool can easily be used by decision-makers to generate flood warnings and seek proactive actions to mitigate their consequences.

## 3.2    Methodology

The proposed methodology consists of two distinct steps: (1) preprocessing a DEM to establish the nested hierarchy of depressions and the sequence of filling-spilling-merging and (2) generating the DEM-graph and executing the GB-RFSM. In this study, an ArcGIS Python toolkit is developed to perform the first step, and a Python program is presented for the GB-RFSM implementation. The suggested framework and the connection between the two processes are depicted in Figure 3-1. A significant advantage of the GB-RFSM is that the DEM preprocessing is a one-time process for every study area, at least until the DEM is updated. After preprocessing, the outputs of the first stage (GIS Python toolbox) are utilized as input for flood inundation simulation in the second step. Each step is explained in the subsequent sections.

### 3.2.1    Step I: DEM To Depressions Hierarchy Preprocessing

In a DEM, depressions can be of different levels, with smaller pits nested within bigger depressions that construct a hierarchy. Figure 3-2 depicts DEM components that determine depressions and their hierarchy, including sinks, pour-points, depressions, and their drainage area. Sinks, also known as blue-spots in the literature, are low-lying cells on the DEM with no outward flow direction in the D8 flow direction method (Balstrøm and Crawford, 2018; Samela et al., 2020). The catchment is the drainage area that drains into a sink, and the pour-point is the lowest elevation on the catchment's perimeter. Depressions are the areas of a catchment where

48

water collects owing to an elevation below the pour-point. A filled depression drains into a downstream catchment through its pour-point. As depicted in Figures 3-2-b,c, and d, depressions on terrain can merge into each other to form larger-scale depressions. The sequence of depressions spilling into each other can be described as the horizontal hierarchy, while the merging process and forming higher-level depressions is the vertical hierarchy (Balstrøm and Crawford, 2018; Chu et al., 2013; Le and Kumar, 2014; Manfreda and Samela, 2019; Samela et al., 2020; Wu et al., 2019b).
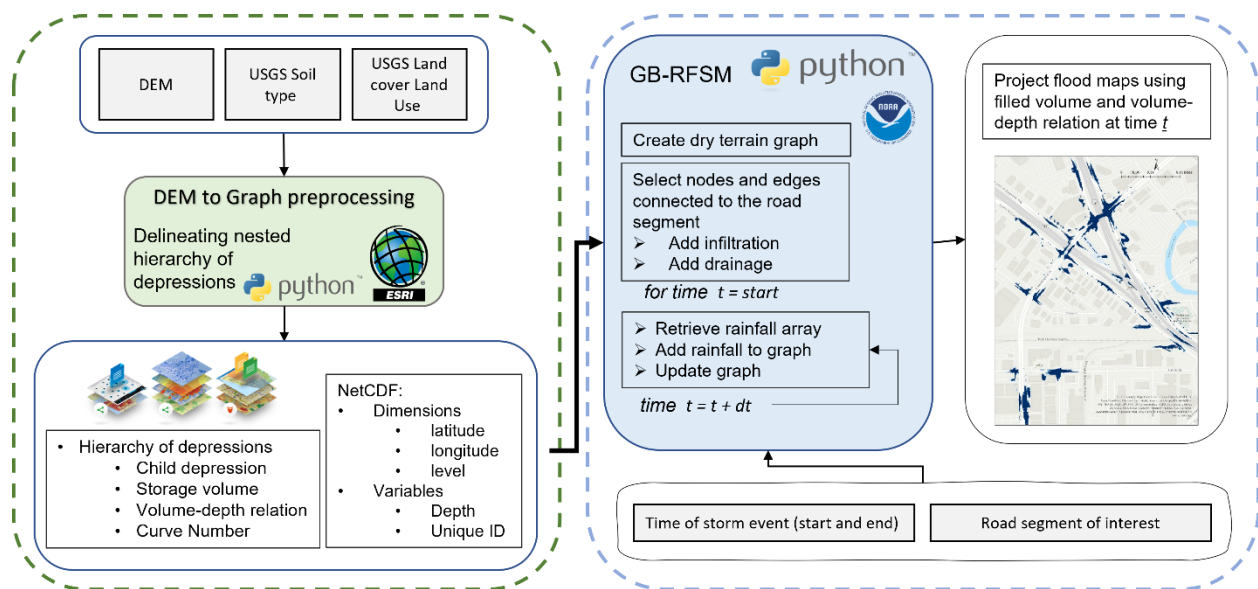


GB-RFSM scripts can be found at https://github.com/EISALab/GB-RFSM

Figure 3-1 Framework of the GB-RFSM

The GB-RFSM requires a preprocessing phase to identify the nested hierarchy of depressions, their upstream catchment, and the filling-spilling-merging process sequence. For this purpose, an ArcGIS Python toolbox is developed in this study. The D8 flow routing technique introduced by Greenlee (1987) is used to extract the depressions, sinks, and catchments (Greenlee, 1987). In the DEM Preprocessing toolbox developed in ArcGIS, extraction of the depression hierarchy is achieved in an iterative process consisting of identifying

sinks, delineating catchments, and raising elevations of depressions to their catchment pour-point (Table 3-1). (Balstrøm & Crawford, 2018; Samela et al., 2020)

The first step in the toolbox involves determining the sink points on the raw DEM, catchments, and pour-points that will produce the first level of depressions. Next, elevations of depression cells are raised to their pour-point level, and a new DEM is achieved that delineates the next level of depression (Figure 3-2-d). This cycle goes on until no depression remains in the study area (Figure 3-2-b and Figure 3-2-c). In each step of the cycle, the sequence of filling and spilling is extracted by locating the downstream cell of each pour-point (Figure 3-2-d). When the lowest depression nested inside a higher level of depression is filled, the depressions merge. In the example shown in Figure 3-2, the depression L1-1 is the lowest depression nested within L2-1; hence once it is filled depressions L1-1, L1-2 and L1-3 merge into L2-1.

### 3.2.1.1   Preprocessing Outputs, GB-RFSM Inputs

In addition to the filling-spilling-merging sequence, depression characteristics that impact the excess runoff volume and inundation formation are required for the GB-RFSM. Physical Catchment Descriptors (PCDs) include area, Curve Number (CN), i.e., a combination of hydrologic soil group and imperviousness, and stormwater drainage rate to culverts. To project the inundation extent of the collected runoff volume in each depression, the depth-volume relations for all the depressions must also be obtained. This procedure is carried out using the level-set approach presented by Wu et al. (Wu et al., 2019c). This approach generates a hypothetical horizontal stepwise plane with elevations equal to the sinks in each catchment. The volume of collected water for each depression is derived by sliding this plane upward from the bottom to the pour-point of catchments in pre-defined increments (15 centimeters in this study).

The ArcGIS Python toolkit that executes this DEM preprocessing outputs a Network Common Data Form (NetCDF) and depression descriptor tables that contain the resulting information needed in the GB-RFSM. The NetCDF has three dimensions: latitude, longitude, and hierarchy level, and it holds a raster of depression depths and catchment identifiers. Table outputs include the hierarchy of depressions as well as the properties of each depression, such as the CN, storage volume, and depth-volume relationships. The application of the generated information in Step 2 of the proposed framework is presented in the following section.

Table 3-1 Pseudocode of the iterative depression delineation process

| | |
|---|---|
| 1 | $Level = 1$ |
| 2 | $D8\ flow\ direction\ (DEM)$ |
| 3 | $\rightarrow sinks\ =\ cells\ with\ no\ outward\ flow\ direction$ |
| 4 | $\rightarrow catchments =\ delineate\ catchments\ of\ sinks$ |
| 5 | $\rightarrow surface\ elevations\ =\ lowest\ elevations\ on\ catchment\ boundary$ |
| 6 | $\rightarrow raised\ DEM\ =\ Max(DEM, surface\ elevations)$ |
| 7 | $\rightarrow depth\ raster\ =\ raised\ DEM\ -\ DEM$ |
| 8 | $\rightarrow If\ \ D8\ flow\ direction\ (raised\ DEM)\ contains\ sinks:$ |
| 9 | $\rightarrow \rightarrow DEM\ =\ raised\ DEM$ |
| 10 | $\rightarrow \rightarrow locate\ downstream\ catchments\ from\ flow\ accumulation\ (DEM)$ |
| 11 | $\rightarrow \rightarrow Level\ =\ Level\ +\ 1$ |
| 12 | $\rightarrow \rightarrow\ back\ to\ line\ 3$ |

### 3.2.2  Step II: GB-RFSM

The GB-RFSM step converts a DEM to a directed graph dataset (DEM-graph). The graph is a data structure in which nodes represent data objects (in this case rain, subbasins, and depressions) and edges reflect relationships between nodes. A schematic of graph components is depicted in Figure 3-2. Nodes in a graph dataset can have attributes that represent each nodes characteristics and edges connecting nodes to each other can control the relation between nodes attributes. In this study, depending on the node type, attributes and edge weights can vary according to Tables Table 3-2 and 3-3. The sequence of filling-spilling-merging is shown by

51

edges that are weighted according to their water volume retention capabilities in the DEM-graph. The GB-RFSM applies runoff volume to the DEM-graph and calculates the filled volume of depressions by routing runoff through the nodes and calculating the water mass balance until all rainfall is captured.
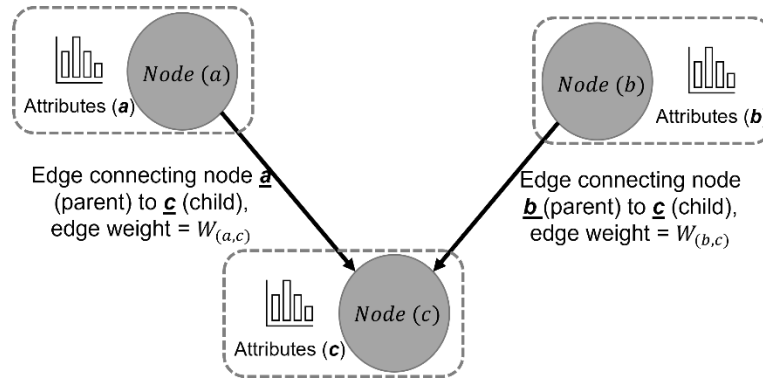


Figure 3-2 Primary components of the graph data structure used to represent the DEM

The graph representation of a DEM in this study has four types of nodes (Table 3-2), including Rain, Subbasins, Level-1 depressions, and Merged depressions (depressions of Level-2 or higher). Uni-directional edges connect nodes to their downstream depression, hereafter called child depression which has a weight equal to the capacity of the parent node to capture runoff volume. A detailed description of DEM-graph components is presented in the next section. In the nested hierarchy of depressions, Level-1 depressions are always the direct receivers of surface runoff (Figure 3-3-a). That is, a higher level (level>1) never gets inundated unless all of the lower-level depressions nested within it spill over and merge to form a higher-level depression. To model this process, only Level-1 depressions and their upstream catchment interact with runoff contributors. Runoff contributors accounted for in this model are precipitation and direct abstractions, including infiltration and stormwater drainage. Merged depressions, however, do not interact directly with runoff contributors and only receive spillover from Level-1 depressions. Therefore, the drainage area of Level-1 depressions is used as

52

subbasins that divide DEM into smaller units receiving rainfall. For example, Level-1 catchments boundaries (blue polygons) depicted in Figure 3-3-a are employed as subbasins. With this approach, each subbasin only contains one Level-1 depression that can spill or merge with other depressions through its pour-point (Figure 3-3-d).
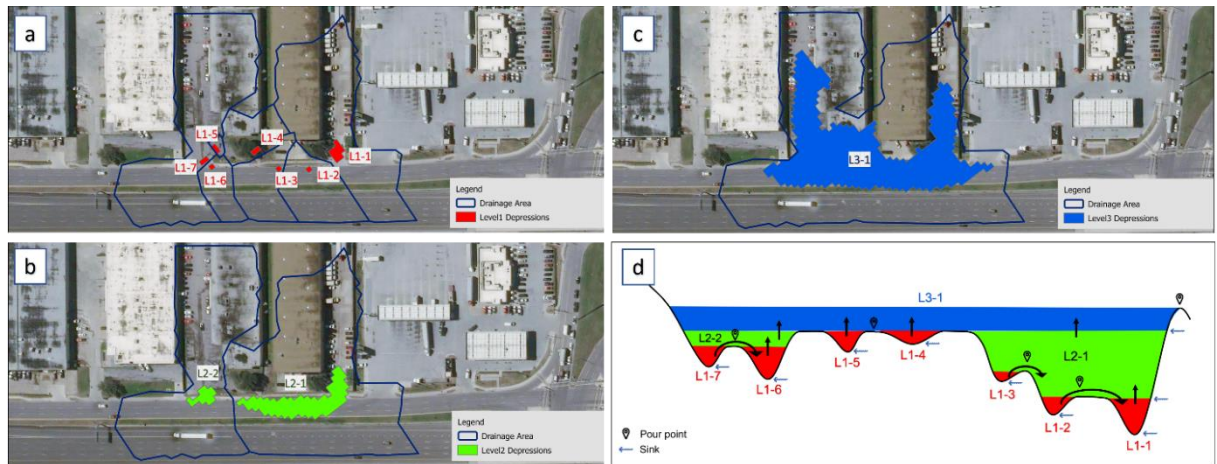


Figure 3-3 Depression components of a DEM

### 3.2.3 DEM-Graph Components and Attributes

The DEM-graph developed as noted above has components that carry information about the DEM's topographic properties in the format of attributes and weights in nodes and edges, respectively. Edge weights of a DEM-graph represent the capacity of their parent node to hold runoff water volume. In the proposed approach, a DEM-graph has different node types to model the physical process of inundation formation. Based on the node type, edges have different attributes. Tables 3-2 and 3-3 summarize the information that is stored in the nodes and edges of a DEM-graph.

Table 3-2 Types of nodes in a DEM-graph

| Nodes | | Attributes |
|---|---|---|
| Rain | | N/A |
| Subbasins | | Subbasin Area, CN, and stormwater drainage rate (Volume/time) |
| Level-1 depressions | | Volume-depth relation, pour-point elevation |
| Merged depressions | | Volume-depth relation, pour-point elevation |

Except for the rainfall volume, other edge weights and node attributes presented in Tables 3-2 and 3-3 are outputs of preprocessing ArcGIS Python toolkit. Rainfall data for every time interval is gathered from the Next Generation Radar (NEXRAD) data made available by the National Center for Environmental Information (NCEI). NEXRAD system is a network of 160 high-resolution weather radars managed by the National Weather Service (NWS), the Federal Aviation Administration (FAA), and the United States Air Force ('Nelson, 2017). Net rainfall from NEXRAD data is assigned to subbasin nodes at the start of each timestep and directed downstream the graph by updating edge weights, including rainfall volume, subbasin capabilities to absorb runoff water, and depression storage capacity, as part of the GB-RFSM framework. Detailed calculations of implementing rainfall and updating edges is presented through a synthetic DEM example, as presented in Figure 3-4.

Table 3-3 Information stored in an edge

| Edges | | |
|---|---|---|
| **Parent node type** | **Child node type** | **Edge weight** |
| Rain | Subbasins | Rainfall volume (0 in the rainless condition) ($m^3$) |
| Subbasin | Depression Level-1 | Subbasin capacity to abstract water ($m^3$) |
| Depressions | Depressions | Storage capacity of parent depression ($m^3$) |

The landscape of the synthetic DEM presented in Figure 3-4-b is made of four Subbasins, four Level-1 depressions, two Level-2 depressions, and one Level-3 depression. The conversion of the DEM (Figure 3-4-b) into a rainless DEM-graph is demonstrated in Figure 3-4-a. To implement rainfall to the DEM-graph and update edges, at each timestep a node is added to the DEM-graph reflecting the rainfall source as shown in Figure 3-4-c. This node is then connected to subbasin nodes through new edges.

At the end of each timestep, a DEM-graph with updated weights is derived that shows the remaining capacity of each depression (Figures 3-4-d and 3-4-e). The framework and calculations of the filling, spilling and merging is presented in Figure 3-5. At every timestep,

54

rainfall at the centroid of each catchment is retrieved from NEXRAD products, and weights of

the DEM-graph ($W_{(parent,child),t}$) are calculated as presented in Equations 3-1 to 3-6 and Figure

3-5.

Equation 3-1

$$W_{(rain,sub_n),t} = -(P_{(lat_n,lon_n,t)} * A_n)$$

where $W_{(rain,sub_n),t}$ is the weight of the edge connecting $n^{th}$ subbasin to the $Rain$ node at

timestep $t$, $P_{(lat_n,lon_n,t)}$ is the accumulated rainfall during timestep $t$ at the centroid latitude and

longitude of the $n^{th}$ subbasin, and $A_n$ is the area of the $n^{th}$ subbasin. The depth of rainfall that is

drained by the stormwater drainage system and infiltration ($H_{n,t}$) can be calculated as shown in

Equation 3-2.

Equation 3-2

$$H_{n,t} = \left(1 - R_{(n,t)}\right)P_{(lat_n,lon_n,t)} + D_n$$

where $P_{(lat_n,lon_n,t)}$ is the rainfall to subbasin $n$ at timestep $t$ in $\frac{m}{\Delta t}$, $D_n$ is the rate of stormwater

drainage at subbasin $n$ in $\frac{m}{\Delta t}$, and, $R_{(n,t)}$ is the ratio of the accumulated runoff to total rainfall

since the beginning of storm according to Equation 3-3. The rate of stormwater drainage can be

calculated using the hydraulics of the drainage system. However, for the case study considered in

this work, the drainage system is neglected due to lack of data.

Equation 3-3

$$R_{(n,t)} = \frac{Q}{P}$$

where $P$ is the accumulated precipitation from the beginning of the storm event $(m)$, $Q$ is the

surface runoff corresponding to $P$ based on the CN method $(m)$. $Q$ may be determined using

Equation 3-4 according to the SCS CN approach.

Equation 3-4

$$Q = \begin{cases} 0 & for\ P < 0.2S \\ \dfrac{(P - 0.2S)^2}{P + 0.8S} & for\ P > 0.2S \end{cases}$$

where, $S$ is the potential maximum retention after runoff begins, estimated using the hydrologic

soil group and land cover/land use of the basin by using the CN according to Equation 3-5

(Homer et al., 2012).

Equation 3-5

$$S = \frac{25400}{CN} - 254\ (mm)$$
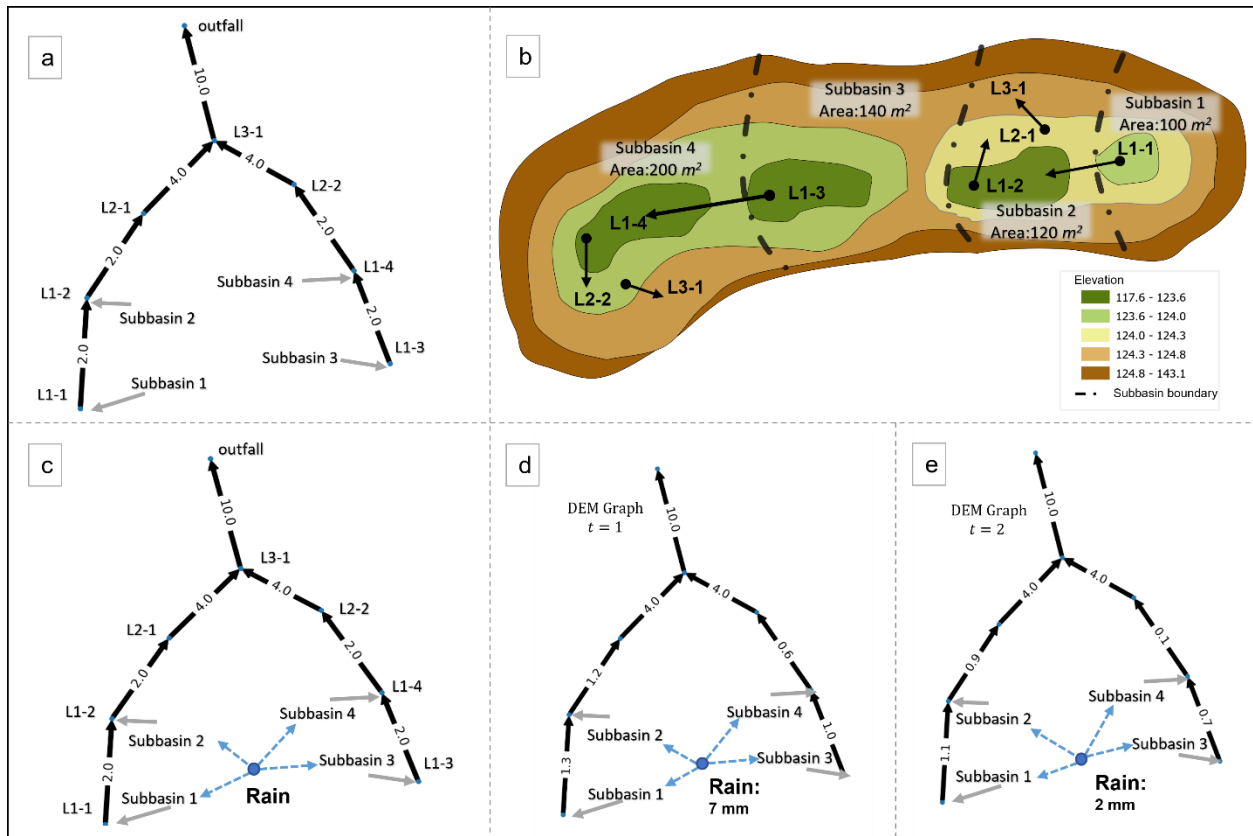


Figure 3-4 Synthetic terrain and DEM-graph analysis

The CNs of subbasins are calculated using the USGS Land Cover Land Use and Global

Hydrologic Soil Groups (HYSOGs250m) datasets (Ross et al., 2018) using Table 2-2a presented

in Urban Hydrology for Small Watersheds, Technical Release 55 (TR 55) (United States Department of Agriculture, 1986).

For example, in the synthetic example presented in Figure 3-4-d, $W_{(Rain,sub_1),t=1} = -0.00695 * 100 = 0.7 \, m^3$ is the generated runoff in Subbasin-1. With no abstraction from Subbasin-1 (assuming CN=100 and no stormwater drainage), $0.7 \, m^3$ water is captured by depression L1-1, and the weight connecting L1-1 to L1-2 is reduced from 2 (Figure 3-4-c) to $2 - 0.7 = 1.3$ (Figure 3-4-d). The same calculation is performed for all the edges at every timestep according to the framework presented in Figure 3-5.

### 3.2.4 Inundation Map

Finally, the depth of water in each depression is retrieved from the depth-volume relations extracted in the preprocessing stage. This is represented in the DEM-graph as inundation extent maps at each timestep. Inundated grid cells are those with elevations lower than the water surface elevation in each catchment, based on the level of standing water.
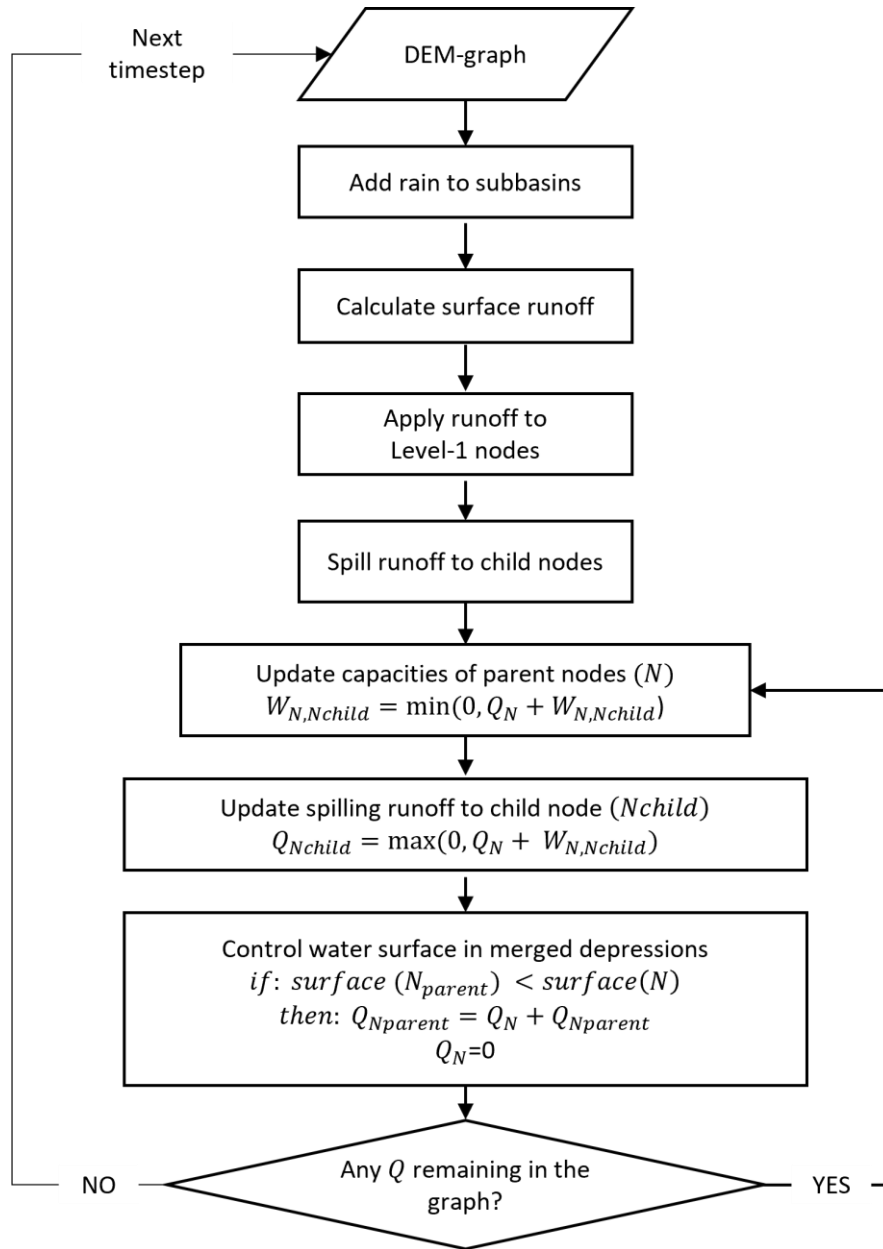
Figure 3-5 Flowchart of updating the DEM-graph at each timestep

### 3.2.5 Model Comparison

To compare the GB-RFSM approach to previously accepted models, inundation vulnerabilities

during a historic rainstorm event, assuming the overflow of the drainage system, are estimated.

The findings are then compared with simulation results of a SWMM model, which is a widely

accepted model for urban stormwater management (Rossman, 2010). To do so, overflow

volumes of sewer inlets computed by the SWMM model are imposed to the subbasins where inlets are located in. This is done by connecting an overflow source node to the DEM-graph and connecting it to subbasins by unidirectional edges. The edge weight connecting the source to overflowing subbasins at every time interval is the overflow simulated by SWMM model. The overflow volume at each timestep is routed through the DEM-graph according to the framework presented in Section 3.2.1 (Figure 3-5). Even though overflow spills to the adjacent subbasins, it is assumed that no other inlet exists that can abstract spilled overflow from adjacent depressions; hence overflow fills, spills, and merges depressions that are hydrologically connected to the inlet's subbasin. Overflows are routed on the DEM using the GB-RFSM approach, and the resulting inundation maps are compared with the inundation extents computed by the GB-RFSM.

The Jaccard similarity score is used to compare the inundation maps resulting from both models (Deng et al., 2014). The Jaccard similarity score is defined as the intersection between two rasters divided by the union of the two rasters as presented in Equation 3-6 where *A* and *B* are inundation extents computed from GB-RFSM and SWMM, respectively. In this study, the intersection is the number of grid cells that GB-RFSM and SWMM inundations have in common (i.e., the number of overlapping flooded cells). The union is the total number of flooded cells (belonging to either the GB-RFSM flooded cells or the SWMM flooded cells). The Jaccard similarity score can range between 0 when there is no intersection between the two inundations and 1 when the inundation extents are accurately similar.

Equation 3-6

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

To compare inundation extents, inundation levels are categorized into very shallow (i.e., less than 15 cm or roughly 6 in), less than 1 foot (15 to 30 cm), less than 2 feet (30 to 60 cm), and deeper than 2 feet. Then, inundation categories of the two inundation maps are compared

respectively. Inundation of less than 15-cm (half a foot) has little hazard for vehicles (Ford et al., 2017); therefore, it is not included in the evaluation. This inundation classification makes inundation extents more interpretable in terms of roadway navigation.

### 3.3 Case Study and Datasets

In this study, an underpass in downtown Dallas, Texas, U.S., which has previously been identified as highly likely to flood in moderate and severe storms (Safaei-Moghadam et al., 2022), was chosen to analyze the GB-RFSM roadway flooding vulnerability during an historical storm event (Figure 3-6). Dallas is relatively flat and has an elevation range of 137 to 168 meters (450 to 550 feet). According to research conducted by the First Street Foundation, floods may cause 1841 miles (out of 6064 miles) of Dallas roadways to become impassable. (Risk factor, 2021).

For this case study, the North Central Texas Council of Government (NCTCOG) provided a 1-meter resolution Bare Earth Digital Elevation Model (BE-DEM) derived from a Quality Level 2 Lidar survey conducted by Digital Aerial Solutions, LLC, in 2018 under contract with the United States Geological Survey (USGS) National Resources Conservation Services (NRCS). TX Pecos Dallas 2018 D19 is the name of the BE-DEM dataset, which has a horizontal accuracy of +/-0.682 meters at a 95% confidence level and a Non-vegetated Vertical Accuracy (NVA) of 0.196 meters.

The high-resolution national hydrography dataset (NHD) provided by the U.S Geological Survey (USGS) (Buto et al., 2020) is used to find the contributing area to the intersection of interest. Two catchments are chosen as the maximum boundary that will drain into the intersection, as depicted in Figure 3-6.The national land cover dataset (NLCD) and the Global Hydrologic Soil Groups (HYSOGs250m) datasets were utilized in the curve number grid

computation (Homer et al., 2012; Ross et al., 2018). Landcover of this study area is mainly

developed with medium to high intensity, and the hydrologic soil group belongs to moderately

high runoff potential (i.e., 50 percent sand and 20-40 percent clay). Hence, the CN of more than

95 percent of study area is higher than 93. The NEXRAD product used in this case study is the

digital precipitation rate (DPR) that displays the 2-minute interval precipitation rate using the

dual-polarization QPE algorithm provided by NCEI (Branch, 2012). Since the drainage system's

layout and characteristics are not publicly available in the City of Dallas and could not be

obtained for this study, and roadway pluvial flash floodings are not measured and recorded, the

runoff abstraction by sewer drainage system ($D_{n,t}$ in Equation 3-2) is neglected in this analysis.

To compare the GB-RFSM results to SWMM, a previously-developed SWMM model is

used. Texas A&M AgriLife developed the model for the Nature Conservancy (TNC) in

collaboration with the city of Dallas to identify overflowing drainage inlets and potential

candidate areas for Green Stormwater Infrastructure (GSI) implementation to mitigate the

overflows. The stormwater drainage network features were modeled based on the maps and

engineering plots provided by the city of Dallas (Jack et al., 2021). The simulation outputs from

the model include the location and time-series of overflow volume at stormwater inlets showing

overflow during the rainfall of interest (Figure 3-6). (Note that the locations of the stormwater

inlets used in SWMM were from a previous project and could not be shared with this study.) The

GB-RFSM and SWMM models are used to simulate flood inundation during a record storm

event on September 21, 2018. The storm duration was 24 hours and resulted in the third largest

24-hour rainfall event in Dallas history. According to NCEI data, the total precipitation for this

storm event over Dallas was 155 millimeters, which is consistent with a 50-year storm prediction

(Perica et al., 2018).
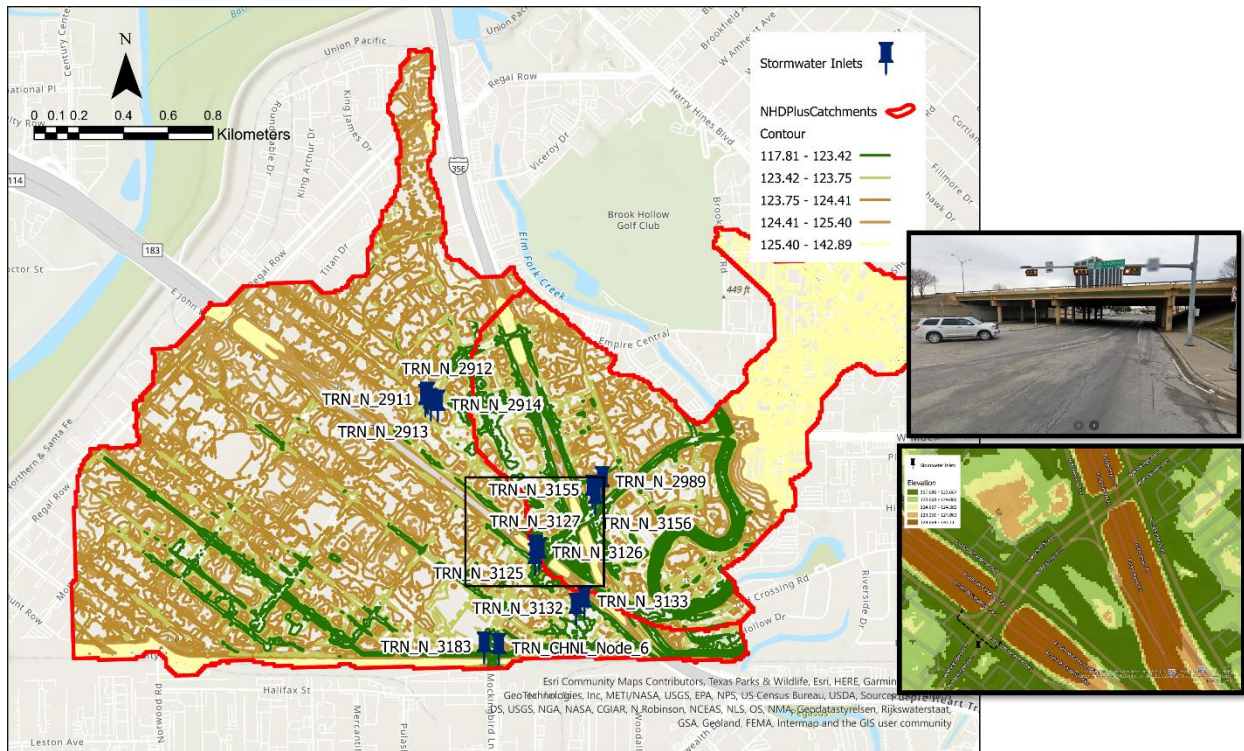
Figure 3-6 Case study area

## 3.4 Results

The GB-RFSM and the SWMM model are used to simulate the inundation with 12 minutes

timesteps and stormwater overflows in the case study's historical storm event, respectively, and

their results are compared. The nested hierarchy of depressions and depth-volume relationships

by implementing the developed preprocessing ArcGIS python toolbox (Step1, preprocessing).

The study area gets hydrologically connected (hydro-conditioned) after five levels of filling

depressions and raising elevations as described in the Section 3.2.1 (Figure 3-7). Subbasins and

catchments of merged depressions of Level-2 to Level-5 are illustrated in Figure 3-7. The

preprocessing step resulted in 1659 subbasins and Level-1 depressions, as well as 221, 48, 10,

and one merged depressions in Level-2 to Level-5, respectively. The distribution of the

catchment area and depression volume depths for each level are presented in Figures 3-9 and 3-

10, respectively. After conducting the preprocessing step, the DEM-graph is created according to

the framework presented in the methodology section using the Python toolbox developed in this

study (Step2, GB-RFSM).
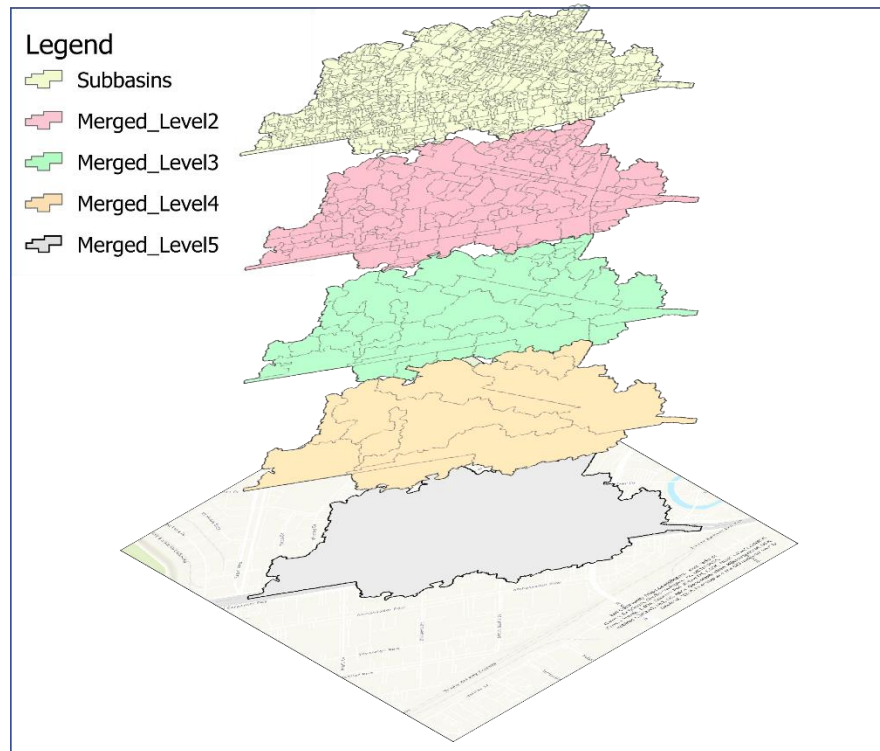


Figure 3-7 Subbasins and merged depression catchments upstream of the underpass
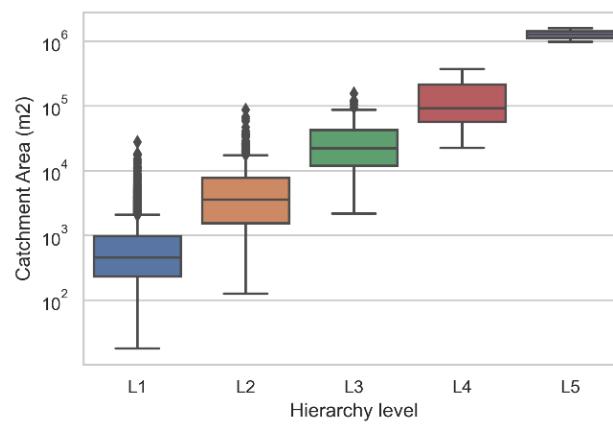


Figure 3-8 Distribution of catchment areas in different hierarchy levels of depressions
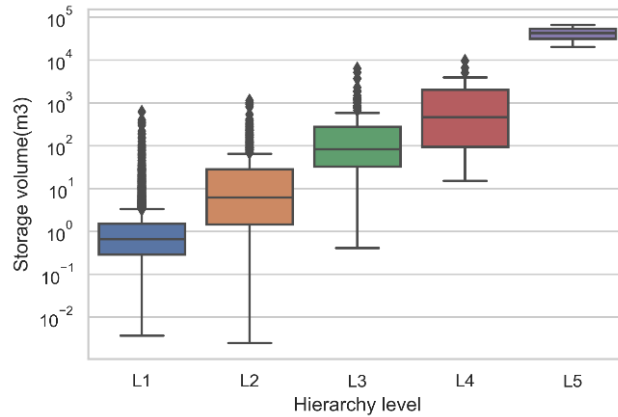
Figure 3-9 Distribution of volume capacity of depressions in different hierarchy levels

As described in Section 3.3, the QPE precipitation data from the NEXRAD radars are used to run the SWMM and GB-RFSM models for the time duration of the case study rainstorm event in 5-minute interval timesteps. The SWMM model simulation found three stormwater inlet failures due to insufficient capacity in the study period. The overflowing stormwater inlets are TRN-N-2911, TRN-N-2914, and TRN-N-3126 (Figure 3-6). The cumulative hyetograph of precipitation over the study area and the cumulative overflow for the three failed inlets are presented in Figure 3-10. Figure 3-10 shows that the three stormwater inlets all start overflowing approximately when the rainfall accumulation reaches 50 mm. Also, Figure 3-10 depicts that TRN-N-2914 and TRN-N-2911, which are almost 60 meters apart (Figure 3-5), have very similar overflow patterns with two sharp and stepwise increases in the overflows at 3:30 AM and 6:00 AM, 2018, September 22. According to the DEM-graph, overflow of TRN-N-2914 spills into the subbasin of TRN-N-2911.
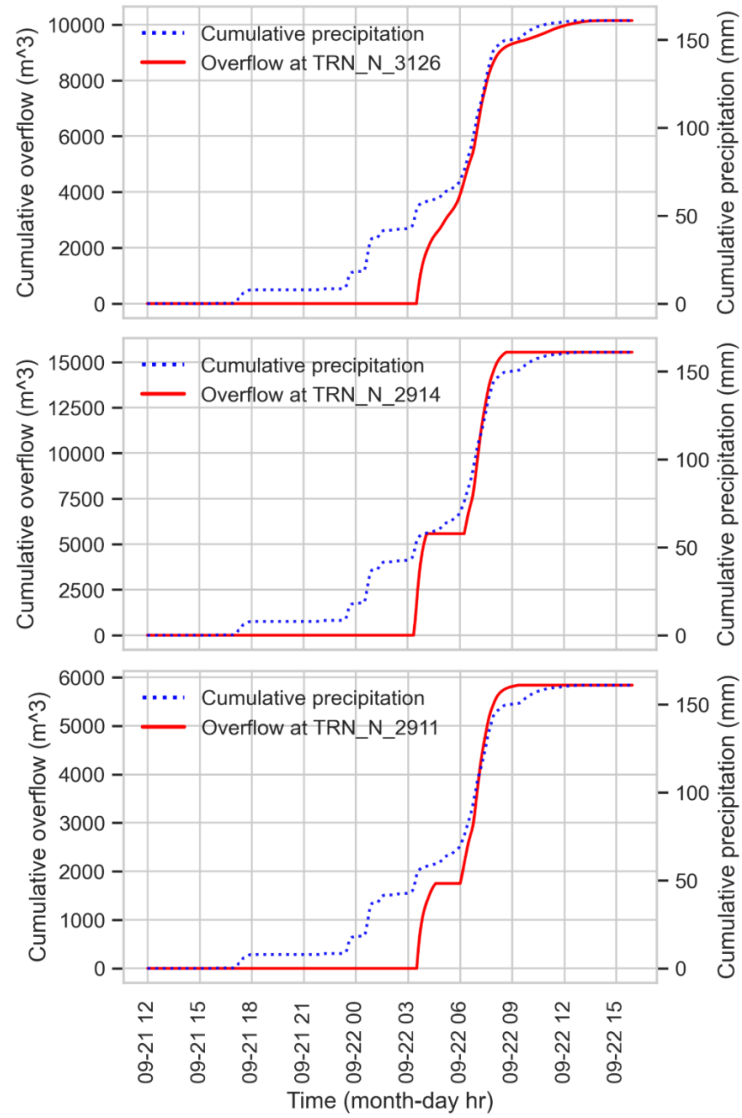
Figure 3-10 Cumulative hyetograph versus cumulative overflow at identified failed inlets calculated by the SWMM model
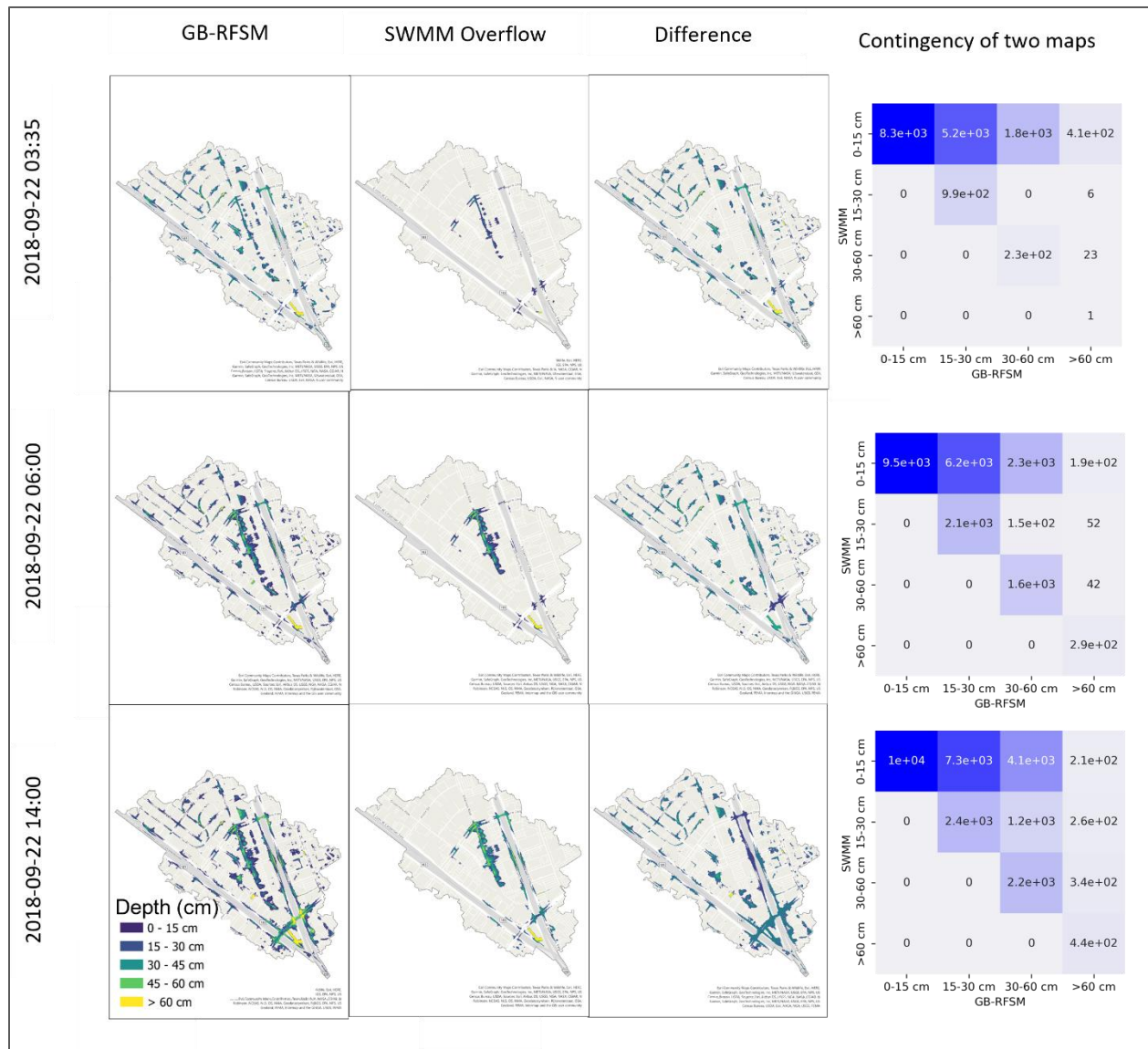
Figure 3-11 Inundation maps derived from GB-RFSM and SWMM results

Water accumulation volumes obtained by GB-RFSM and SWMM are transformed into inundation maps using depth-volume relations of the depressions determined during the preprocessing stage. Inundation maps for three timesteps are depicted in Figure 3-11, including 2018-9-22 at 3:35 when inlets just start to overflow, 2018-9-22 at 6:00, just before the second spike of overflow, and 2018-9-22 at 14:00 when overflows reach their maximum stage.

With GB-RFSM, however, all stormwater inlets are assumed to fail from the beginning of the storm event; hence from the early stages of the storm, GB-RFSM assigns flood

vulnerability to depressions while the SWMM model simulates no overflow and therefore no inundation. However, at 3:35 AM, after overflow starts, the SWMM model begins estimating inundation. The contingency matrices in the last column of Figure 3-11 show the basic interrelations and interactions between the two estimated maps. As the matrices are always upper-triangular, it shows that the GB-RFSM always estimates higher inundation compared to the SWMM model. The upper non-diagonal elements of the contingency matrix show the frequency of higher vulnerability estimates from the GB-RFSM.

Given the stormwater inlet assumption in the GB-RFSM approach, it is expected that as catchment areas grow, the GB-RFSM will have greater differences from SWMM due to the higher number of engineered and natural drainage inlets being modeled by SWMM. As predicted, Figure 3-12 shows that the catchment areas of depressions with different inundations between SWMM and GB-RFSM are larger. Conversely, the agreement between the GB-RFSM and SWMM predictions are higher in smaller catchments. Comparing the inundation volumes predicted by the SWMM model and the GB-RFSM, the SWMM model, on average, had 0.71 $\frac{mm}{m^2}$ higher runoff abstraction in every 12-minute time interval.

Figure 3-13 illustrates how the three categories of inundation explained in the methodology (15 to 30 cm, 30 to 60 cm, and deeper than 60 cm), are similar in the two models based on the Jaccard similarity score. Figure 3-13 shows that at depths less than 60 cm, the Jaccard score increases rapidly, shortly after overflow begins; however, Jaccard similarity score for grid cells deeper than 60 cm increase more slowly. After the sharp increase in the stormwater inlet overflows at 09-22 3:35 (Figure 3-10), the Jaccard scores starts raising from no similarity. Figure 3-13 shows that when the cumulative overflow becomes steady in TRN-N-2911 and TRN-N-2914, between 9-22 3:35 to 9:22 6:00 in Figure 3-10, which resembles that stormwater

inlets drain additional rainfall, Jaccard score decreases, but has another jump after the next spike

in overflow at 09-22 6:00.
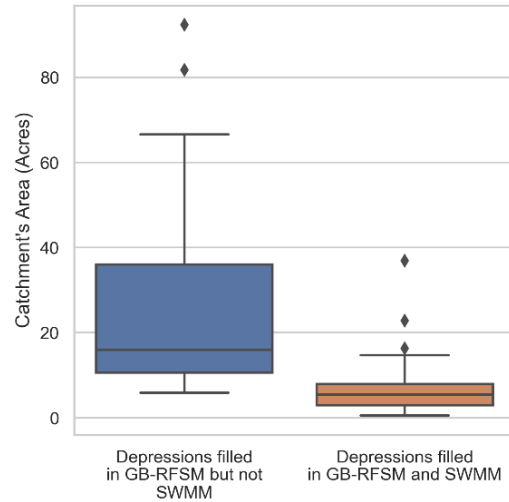


Figure 3-12 Distribution of catchment areas that are consistently and inconsistently filled in SWMM and GB-RFSM
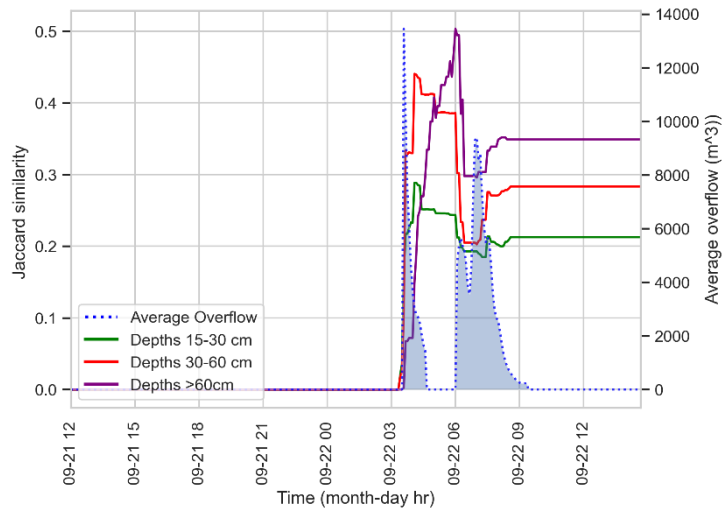


Figure 3-13 Jaccard score versus overflow over time

### 3.4.1 Ground Truth Assessment

Given that the SWMM model is a simplified model itself, without having the actual ground truth

observations, it appears to be inaccurate to assume the SWMM inundations are closer to the

actual roadway flooding situations. To further assess the differences between the models in terms

of actual flooding at the Dallas underpass during the case storm event, I obtained flooding

information from several data sources: Twitter postings, a video from a local news outlet, flood alerts posted to the Waze crowdsourced navigation app, and flood-related city service requests (i.e., City of Dallas 311 service requests available at https://www.dallasopendata.com/browse). Keywords used to search for flood observations include 'flood', 'water logging', 'high water', 'ponding', 'deep water', 'pooling', and 'turn around don't drown'. No Twitter posting or city service requests through the 311-reporting system were found for this area but Waze flood alerts and media outlets proved flooding in the case study situation. Figure 3-14 shows the gathered data at the case study intersection. Locations tagged as 1, 2, and 3 were reported in the news, and the flood extent captured from the videos in Figure 3-14-b appears to be comparable with the GB-RFSM outputs shown in Figure 3-14-a.
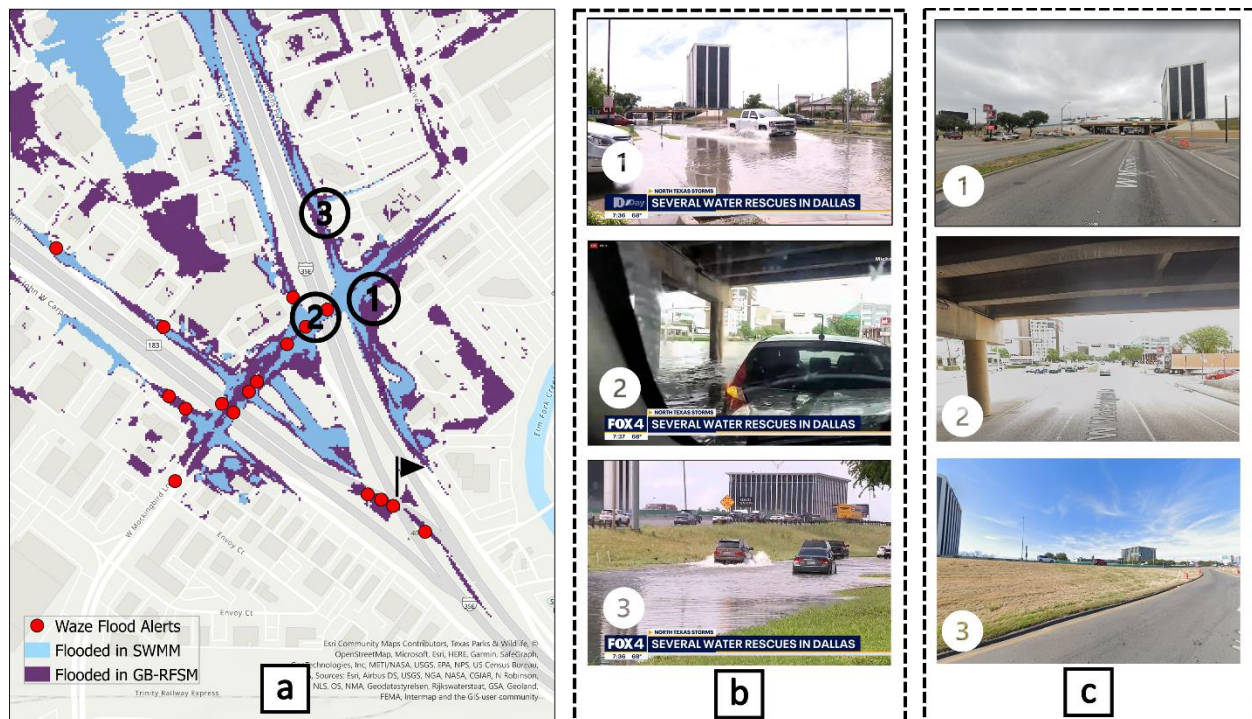


Figure 3-14 Map of inundations, Waze flood alerts, and locations covered in media, (a) video clips at locations 1, 2, and 3 during the flood event (b) and after (c).

The Waze flood alerts posted during the storm (shown as red points in Figure 3-14-a) also show acceptable agreement with the flooded locations. However, the alerts on the highway

ramp (marked with a black flag in Figure 3-14-a) is modeled as flooded by the GB-RFSM but not the SWMM model. The superiority of the GB-RFSM in capturing flooding at the flagged point could have various causes. For example, in SWMM model development, to maintain hydrologic connectivity of every DEM cell to stormwater inlets, depressions upstream of inlets must be filled; this simplification, as well as rare events of clogged stormwater inlets are possible reasons behind GB-RFSM's better performance at the flagged point.

## 3.5    Discussion and Conclusions

This study develops a new approach to addressing the computational complexities and data requirements of common hydraulic models for local scale flooding prediction. An efficient, simplified physics-based model is developed that can be used to estimate the pluvial flood vulnerability on a highly localized scale. The resulting GB-RFSM estimates the extent of potential roadway flooding in the event of stormwater system failure without relying on the stormwater drainage system characteristics and layout. The graph-based structure of the GB-RFSM makes it suitable for capturing the temporal formation of flooding by updating the graph at every timestep.

Although the GB-RFSM considered in this study accounts for infiltration based on the CN and neglects the stormwater drainage rate due to lack of data availability, the methodology could be modified to include averaged runoff drainage over subbasins. The flood estimated by GB-RFSM in an historic 50-year storm event with 24-hour duration showed general agreement with the inundations produced from SWMM when overflowing occurred and roadway flooding is likely to be most acute. One significant difference between the two inundation extents appears to be a consequence of ignoring drainage to the stormwater system at the beginning of rainfall, prior to the start of SWMM's predicted overflow. As depicted in Figure 3-11, failed inlets all

start to overflow just after accumulated precipitation reaches 50 millimeters; this suggests the possibility of making meaningful conjecture to account for a defeat point that can mirror when drainage system reaches its capacity and overflow starts to occur, when implementing GB-RFSM. Moreover, it is possible to find the optimum runoff abstraction by the sewer drainage system at every subbasin and time interval ($D_{n,t}$ in Equation 3-2) by calibrating the GB-RFSM parameters according to ground truth inundations; however, the records of flood observations on a road segment scale are scarce.

While these findings are promising, this study has certain limitations. In addition to neglecting the dynamics of flow, runoff abstraction by drainage systems, and other possible losses such as retention and evaporation, the natural drainages and building footprints are not modified on the DEM. DEM modifications to account for building footprints, natural rivers and streams, walls, and sinks at major stormwater defense structures such as tunnels can help to define more precise subbasins and would likely improve model performance. In addition, the preprocessing steps require periodic updates to account for changes in land cover and topography due to construction.

Overall, this study showed that GB-RFSM could detect areas with high flooding inundation when drainage systems have failed, particularly as shown by the media and crowd-sourced flood observations A hybrid model that combines GB-RFSM with a data-driven model using the types of crowdsourced data examined in this paper could improve the model's performance by considering unmodeled factors such as those outlined above. Such a rapid approach that can forecast real-time hyper-local flooding on roadways would be a valuable tool for drivers, first responders, and decision-makers to monitor flood conditions more effectively and take proactive actions to minimize risks and losses.

# Chapter 4

## PREDICTING ROADWAY PLUVIAL FLOOD RISK; A HYBRID MACHINE LEARNING MODEL COUPLING GB-RFSM AND HISTORICAL VULNERABILITIES BASED ON WAZE DATA

Using Waze flood alerts and the Digital Elevation Model (DEM), in previous chapters data-driven and simplified physics-based model were developed to represent the vulnerability of a road segment to PFF. In this chapter, however, hazard, vulnerability, and exposure are merged to forecast the overall risk of roadway PFF. Predicting PFF risk is essential to facilitate road safety management, increase travelers' awareness, and empower decision-makers to take proactive actions.

## 4.1    Introduction

The framework developed in this chapter adopts the risk definition proposed by Gouldby et al. (2005): Risk is as a combination of hazard, vulnerability, and exposure. According to this definition, the hazard is a phenomenon that has the potential to be destructive; vulnerability specifies how likely the system is to be damaged by the hazard; and exposure is the total number of receptors that the destruction may affect.

In interpreting the risk of roadway flooding, the hazard of the system is the precipitation; vulnerability is the potential of a road segment to accumulate water and form PFF in the worst-case event of stormwater drainage system failure, which depends on the topographic and hydrologic characteristics of the roadway and catchments; and exposure is the traffic volume that confronts flooded road segments. In a highly urbanized area, flood formation is a highly complex and uncertain process due to the lack of information about underground and overland flow interaction. The impacts of flooding on roadway mobility also depend on various temporal and

spatial variables. Moreover, the rapid onset and short lifetime of roadway PFF lead to a lack of observational data to quantify the historical PFF risk at a local scale. Hence, a comprehensive approach is needed that reflects the topographic and hydrologic specifications of road segments as well as temporal traffic levels to capture the heterogeneity of risk.

On this basis, this research proposes to address the shortcomings of previous PFF models (reviewed in Section 3.1) by incorporating multiple data sources into an end-to-end real-time prediction of the risk of roadway PFF on low-lying road segments. To achieve this, computationally efficient hybrid machine learning (ML) models are developed that combine a statistical model based on historical PFF observations and hydrologic characteristics (developed in Chapter 2) with GB-RFSM (developed in Chapter 3). Crowdsourced flood alerts posted to Waze are used as a risk proxy to train the ML models.

## 4.2 Methodology

For Waze flood alerts to be a reasonable proxy to the risk of roadway PFF, the alerts need to be associated with the hazard, vulnerability, and exposure. In Chapter 2 (Safaei-Moghadam et al., 2022, in review), we demonstrated that the total number of flood alerts correlates with the rainfall's duration and maximum intensity. Compared to moderate and light storm clusters, rainfall events that were part of the severe storm cluster had a higher incidence of alerts related to flooding, as expected. Regarding the vulnerability, Chapters 2 and 3 showed that Waze flood alerts are clustered around low-lying areas and depressions that can accumulate a large volume of excess runoff and are more prone to flooding in the future. Concerning exposure, our preliminary analysis in Chapter 2 showed that Waze flood alerts are primarily posted from frequently traveled roads. The findings showed a clear correlation between Waze flood alerts and exposure to road PFF, with increased traffic volume increasing the likelihood of observing a

73

Waze roadway flooding alert under similar circumstances. Hence, it is reasonable to assume that the existence of Waze flood alerts is a function of hazard, vulnerability, and exposure and can be used as a proxy to the risk of roadway PFF. Figure 4-1 shows the risk components and their relationship with Waze flood alerts and roadway PFF risk. More details on the approach taken for each of the three risk components, as well as the hybrid model and model evaluation criteria, are given in Sections 4.2.1 to 4.2.5.
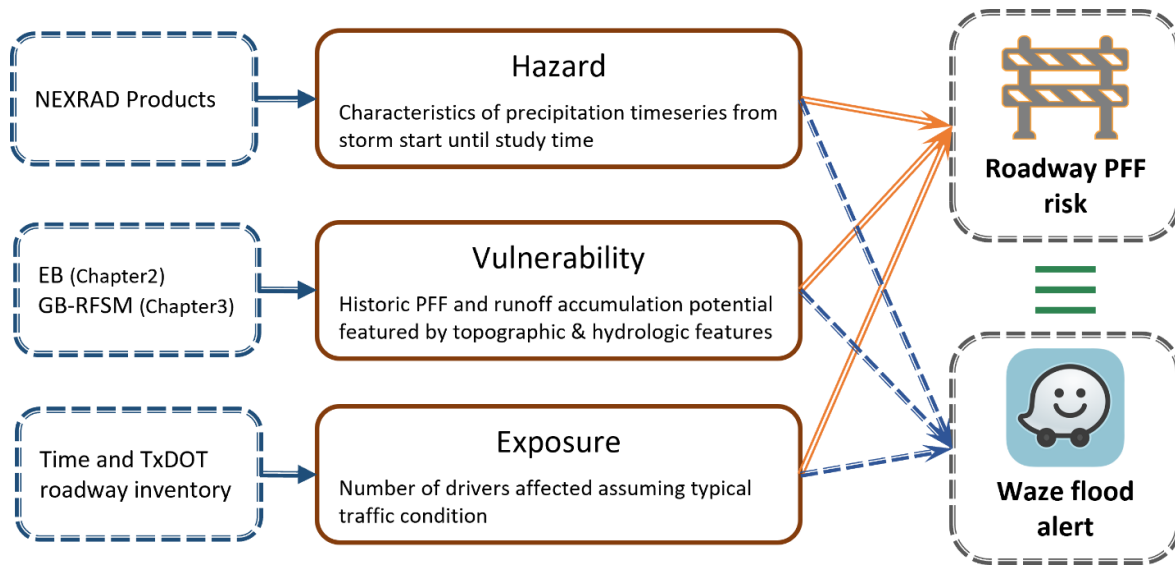


Figure 4-1 Methodology flowchart

### 4.2.1 Hazard

High spatiotemporal precipitation data are required due to the roadway PFF phenomenon's rapid onset, brief lifespan duration, and high spatial resolution. The precipitation time series from the start of the storm event until the study time interval is utilized to extract descriptive features that can characterize the storm event in order to estimate the probability of PFF at each time interval. In addition to total rainfall, temporal patterns of storms play a significant role in PFF formation. When modeling PFF, it is important to take into account the consistency, peaks, minimums, variations, zeros, and immediately preceding rain pulses of precipitation timeseries. As a result, features that may capture these qualities are calculated. Table 4-1 displays these features, where

$p_t$ is the precipitation timeseries from the storm start until time $t$, and $\mu$ is the average precipitation intensity since the storm started.

Table 4-1 Features representing the hazard

| Feature | Equation | Definition |
|---|---|---|
| Precipitation | $P_T$ | Precipitation at time T (mm) |
| Maximum preceding rain pulse | $Max(p_t)\ t\ in\ [0,T]$ | Maximum precipitation per timestep since the beginning of storm |
| Total accumulation | $\sum_{0}^{T} p_t$ | Total accumulated precipitation since the beginning of storm |
| n-hours preceding accumulation | $\sum_{t-n}^{t} p_t$ | Total accumulated precipitation in last $n$ timesteps |
| Count of rain pulses above h mm | $Count(p_t \geq h)$ | Count of timesteps with precipitation higher than $h$ millimeter |
| Count of rain pulses above mean | $Count(p_t \geq \mu)$ | Count of timesteps with precipitation higher than average precipitation ($\mu$) since the beginning of storm event |
| Time to the last maximum rain pulse | $t - t_{p=\max(p)\_}$ | Time interval between timestep of study and the last timestep with maximum precipitation |
| Time to the last minimum rain pulse | $t - t_{p=0}$ | Time interval between timestep of study and the last timestep with 0 precipitation |
| Count of rainless intervals | $Count(p_t = 0)$ | Count of rainless timesteps since the beginning of storm |

## 4.2.2 Vulnerability

The vulnerability of a road segment to roadway flooding is a function of its topographic and hydrologic characteristics that contribute to the excess runoff accumulation as well as the performance of the stormwater drainage system. While high-resolution hydrodynamic models of surface runoff and drainage systems may be used for this, their application for real-time prediction at this scale can be computationally prohibitive and required data may not be available. Therefore, in this work, we employ the PFF likelihoods produced from the Empirical Bayes (EB) model developed in Chapter 2 with GB-RFSM, the simplified DEM-based Hierarchical Filling and Spilling Algorithm (HFSA) described in Chapter 3, to characterize the vulnerability of a road segment to PFF. The GB-RFSM estimates the potential inundation depth and inundation area in the worst-case event when the stormwater drainage system fails, and the

EB-derived PFF likelihoods represent the contribution of site-specific unobserved variables, such as chronic debris generation in a location, in the roadway PFF vulnerability.

The GB-RFSM requires a DEM-preprocessing phase to identify the nested hierarchy of depressions in the DEM of the watershed upstream of a road segment. The ArcGIS Python toolbox developed in Section 3.2.1 is used for this purpose. The toolbox outputs a Network Common Data Form (NetCDF) and depression descriptor tables that contain the resulting information needed in the GB-RFSM step. The GB-RFSM step then converts the DEM to a directed graph dataset (DEM-graph) and applies rainfall to the DEM-graph by routing runoff through nodes and calculating the water mass balance until all rainfall is captured. Once the runoff is routed, the model computes the filled volume of the depressions. Finally, water depth in each depression is retrieved from the depth-volume relationships extracted in the preprocessing stage and inundation maps are generated. In addition to the preprocessing outputs, a binary zone raster indicating road surface grid cells is used to extract statistics of the flooded road segment and features that represent the vulnerability. The list of attributes that reflect vulnerability is shown in Table 4-2, where $d_t$ is the flood depth raster on road surface determined with the GB-RFSM, $x$ is the resolution of the road surface grid cells, $p_{(i,j)}$ is the likelihood of flooding on depression $i$ during storm type $j$, $\hat{y}_{i,j}$ is the predicted number of floodings on depression $i$ and storm type of $j$ using the EB model (Section 2.2.2.1) and $N_j$ is the number of storms in cluster $j$.

Table 4-2 Features representing vulnerability

| Feature | Equation | Definition |
|---|---|---|
| **Maximum depth** | $Max(d_t)$ | Maximum estimated depth on road surface |
| **Area** | $Count(d_t > 0) \times x^2$ | Inundation area on road surface |
| **Percentage of road segment inundated** | $\dfrac{Count(d_t > 0)}{Count(d_t)}$ | Volume of standing water on road surface |
| **PFF likelihood-moderate storms** | $p(i,j) = \dfrac{\hat{y}_{i,j}}{N_j}, j = moderate$ | PFF likelihoods achieved from Empirical Bayes regression for moderate storms |
| **PFF likelihood- severe storms** | $p(i,j) = \dfrac{\hat{y}_{i,j}}{N_j}, j = severe$ | PFF likelihoods achieved from Empirical Bayes regression for severe storms |

### 4.2.3 Exposure

Next, the exposure attributes are established, which define the number of drivers affected by a flooded road segment assuming typical prevailing traffic conditions and count of vehicles passing the road segment. Since access to historical traffic data is not publicly available at low cost, public data sources that are a proxy to the traffic volume on a road segment are implemented in this study. *Road classification*, average annual daily traffic (*AADT*), time of day (*TOD*), and *weekdays* are characteristics used in this work (Table 4-3).

Table 4-3- Features representing exposure

| Feature | Definition | Values |
|---|---|---|
| **Average annual daily traffic (AADT)\*** | Commonly used measure showing the traffic load calculated by dividing the total annual volume of vehicle traffic by 365 | Integer between 0 to 999999 |
| **Functional system classification (F_system)\*** | Categories of roadways and highways based on the service they provide, such as volume of traffic and trip types | 1:Interstate<br>2: Other freeway and expressway<br>3: Other principal arterial<br>4: Minor arterial<br>5: Major collector<br>6: Minor collector<br>7: Local road |
| **Weekday** | Whether the timestep is during weekends or weekdays | 1: Weekends<br>2: Weekdays |
| **Time of day (TOD)** | The time of timestep | 1: After midnight (00:00 to 4:00 AM)<br>2: Early morning (4:00 to 7:00 AM)<br>3: Morning (7:00 AM to 12:00 noon)<br>4: Afternoon (12:00 to 4:00 PM)<br>5: Evening (4:00 to 8:00 PM)<br>6: Night (8:00 PM to midnight) |

\*. Texas Department of Transportation (TxDOT) Roadway Inventory, 2020-Current

### 4.2.4 Modeling

Three Machine Learning (ML) classification algorithms are employed in the PFF risk prediction (i.e., the existence of flood-related alerts): Random Forest Classifier (RFC), Extreme Gradient Boosting Decision Tree (XGBoost), and Support Vector Machine Classifier (SVM).

RFC and XGBoost are decision tree-based algorithms consisting of decision nodes that employ if-else rules to split data into different decision paths. As noted in Chapter 2, one

advantage of tree-based algorithms is their ability to determine feature importance in model prediction power based on node impurities. Node impurity is a measure of heterogeneity in the target values, PFF observations in this application, at each splitting node. The normalized decrease in node impurity estimates the significance of a given feature when it is added to a tree (Breiman, 2001).

Random Forest (RF) is a supervised ensemble machine learning algorithm that uses multiple decision tree learners to increase predictive performance (Pedregosa et al. 2011). The final prediction of RF is the average prediction of all decision trees; each tree is built from a bootstrap sample of observations and a subset of features. Bootstrap sampling involves random sampling with replacement, meaning that a particular observation may not be picked for sampling while it is allowed to appear once or more than once in training samples. RF has been widely used for data-driven modeling in the field of water resources (Sadler et al. 2018; Tyralis, Papacharalampous, and Langousis 2019). This algorithm can handle large and imbalanced datasets by combining bootstrap sampling and ensemble learning to train each tree on a more balanced subsample as well as its capability of being cost-sensitive by assigning class weights in node impurity calculations. The aggregation of several trees and the binning process in decision trees makes RF resilient to bias and overfitting.  In RFC, the overall feature importance is calculated as the average of the importance of a feature over all trees, weighted by the number of samples used in each split across all trees. Tuned RFC hyperparameters in this study are number of trees in forest, maximum depth of trees (maximum number of allowed splits), maximum number of features allowed to be used in each tree, and class weights used in favor of the minority class when calculating the impurity score of a split.

XGBoost is one of the most preferred ensemble tree boosting ML models because it has been shown to give state-of-the-art results in different fields (Chen and Guestrin 2016), including water resources (Huang et al., 2021; Janizadeh et al., 2021; Sanders et al., 2022). In the Gradient Boosting approach, the model's loss function is minimized by adding weak learners trained on the remaining residuals of existing learners. The tuned hyperparameters include maximum depth of decision trees, number of trees, and class weights. In XGBoost, the contribution of each decision tree to the final prediction is calculated by minimizing the prediction error of the training dataset (called the training loss).

The SVM classifier is an optimization-based learning technique that divides classes by locating an optimum hyperplane with the greatest marginal distance between the two classes (Vapnik, 1999). While SVM classifier was originally limited to linearly separable datasets, application of a kernel function transforms the data from a nonlinear input space into a linear representation to make the data separable. Standard kernel functions commonly used in SVM include linear, polynomial, radial basis function (rbf), and sigmoid function. The choice of kernel function is determined during the random search process of model selection. Details of SVM and kernel functions are provided by Kecman (2001). Application of SVM in water resources systems, flood prediction, and flood susceptibility mapping has been extensive (Han et al., 2007; Ke et al., 2020; Liu et al., 2022; Saha et al., 2021; Xiong et al., 2019)

In this study, the RF and SVM classifiers are executed using the Scikit-Learn library (Pedregosa et al., 2011) and XGBoost is implemented using the XGBoost package (Chen and Guestrin, 2016) in the Python environment. Recursive feature elimination is used to pick the most important attributes for modeling in order to keep the model simple. This is a repetitive process in which the model is trained, and the least significant feature is dropped in every trial

until affect performance significantly. A randomized grid search on a variety of model

parameters is utilized for hyperparameter tuning and 5-fold cross validation is performed to

choose the optimal model and parameter combination.

### 4.2.5   Model Evaluation

A total of 15 historical storm events are used to train the models and 5 storm events are used to

test the models. The trained models predict PFF risks during test storms; the predicted risks are

then compared with flood-related Waze alerts to evaluate the models' performance.

Given that flood risks and posting Waze flood alerts are rare binary events according to

the Waze data, the classification model to predict the risk of roadway PFF deals with a highly

imbalanced dataset (i.e., numerous negative examples of no flood alerts). To make models more

sensitive to class imbalance and give higher importance to predicting instances when PFF risk

has been observed (positive class), two steps are taken. First, the model is trained in a cost-

sensitive approach by weighting loss values for minority (PFF risk observed)  and majority (no

PFF risk) classes. This method applies different weights to the loss computed for samples that

are in the minority class when calculating the loss function. In this study, a higher weight is

given to the misclassification of timesteps where PFF risk is observed than it is to the

misclassification of timesteps where no flood alert was posted. In the model evaluation phase,

frequently used performance metrics like accuracy might produce misleading assessments since

they are insensitive to skewed data and give minority and majority classes the same priority.

Therefore, metrics that are sensitive to class imbalance are used instead: $F\beta$ score, area under the

precision-recall curve (PRC-AUC), and area under the Receiver Operating Characteristic curve

(ROC-AUC) (Buckland and Gey, 1994; Saito and Rehmsmeier, 2015a).

Fβscore is an abstraction of F1score in which recall's importance can be adjusted by a coefficient named β (Buckland and Gey, 1994). F1score, the harmonic mean of recall (completion of retrievals, i.e., the probability of predicting PFF given its existence) and precision (purity of retrievals, i.e., the probability that PFF occurs given a predicted PFF), is a commonly used metric for imbalanced classification. However, utilizing the F1score assumes that recall and precision are equally significant, suggesting that stakeholders consider both false positive and false negative predictions similarly undesirable. Due to the ambiguity surrounding the crowdsourced proxy variable (Waze flood alerts), we cannot properly define the negative class. That is, the lack of a flood alert could indicate either no flooding or no driver posting an alert. As a result, prioritizing the positive class and reducing false negatives with the cost of increased false positive predictions can give us a higher confidence level overall. Consequently, higher importance should be assigned to recall that focuses on the completion of minority class prediction.

To address this concern, precision, recall, and Fβscore are computed as shown in Equations 4-1 through 4-3.

Equation 4-1

$$precision = \frac{tp}{tp + fp}$$

Equation 4-2

$$recall = \frac{tp}{tp + fn}$$

where $tp$ represent the number of observations in positive instances retrieved correctly, $fp$ represents the number of negative observations retrieved incorrectly, and $fn$ is the number of positive instances retrieved incorrectly. Fβscore is a metric that balances the recall and precision between 0 for the worst score and 1 for the perfect score (Equation 4-3).

Equation 4-3

$$F\beta score = (1 + \beta^2)\frac{precision \times recall}{\beta^2 precision + recall}$$

where $\beta$ is a coefficient that controls the balance between precision and recall, with $\beta < 1$ when minimizing false positives and $\beta > 1$ when minimizing false negatives is the priority. In this study, $\beta = 2$ is to emphasize predicting observed risk incidents by making recall the priority. Given its benefits, F$\beta$score, with $\beta = 2$ (hereafter called F2score)is utilized in cross-validated random search to rank models and find optimum hyper-parameters.

Precision, recall, and F2score are single threshold measures and dependent on a cutoff threshold for the classifier to separate positive and negative classes, while the appropriate cutoff may vary depending on the application and dataset (Saito and Rehmsmeier, 2015b). Model-wide threshold-free measurements are therefore needed to comprehensively assess the model's performance on an unknown test dataset. The Precision-Recall Curve (PRC) and Receiver Operating Characteristic (ROC) plots are model-wide measures that compare a model with a baseline no-skill performance benchmark, The ROC curve depicts the false positive rate versus the true positive rate over a range of thresholds. Although ROC is not biased by either the minority or the majority class, it might be deceptively optimistic when there are few positive samples focusing only on the positive class. the PRC curve is a more accurate measurement for unbalanced binary classification. The baseline performance of a model in the PRC curve can be calculated as Equation 4-4, where P and N represent the number of positive and negative samples, respectively. Since the baseline no-skill performance in the PRC curve is proportional to the number of the minority class, it is a better metric for imbalanced binary classification (Buckland and Gey, 1994; Saito and Rehmsmeier, 2015a).

Equation 4-4

$$y = \frac{P}{P + N}$$

## 4.3    Case Study and Dataset

The methods described above are evaluated at intersections in the City of Dallas, Texas. The

intersections are randomly selected from road segments prone to PFF, as identified in Chapter 2.

Figure 4-2 shows the locations of the selected case study road segments. The Texas Department

of Transportation's (TX-DOT) highway inventory and asset dataset contains features that

describe traffic volume and road segment characteristics. The 15 road segments include arterials,

major collectors, freeways, and interstates. (Figure 4-2). The *AADT* of road segments ranges

between 3,723 and 240,182 vehicles per day.

Since merged depressions and delineated catchments extracted by ArcGIS Python

toolbox (Section 3.2.1) from 1-, 2-, and 3-meter resolution DEMs were identical, to simplify the

simulation process 2.3 a 3-meter resolution DEM is employed in the GB-RFSM model. Other

data required in the GB-RFSM model, including landcover, land use, road network, and high-

resolution catchments, are collected as described in Section 3.3. The roadway dataset provided

by Open Street Map (Mooney and Minghini, 2017) is used to generate road surface grid cells

required for calculating water depth statistics. The buffer width around the road centerline is

calculated according to the number of lanes, assuming that each lane has a width of 12 feet on

average (RD Manual, 2004). The NEXRAD precipitation data are obtained from NCEI (Section

3.3) for every 12-minute time interval and case study road segment. Twenty storm occurrences in

total are selected from August 2018 to April 2022, although the Covid-19 shutdown period

(March 1, 2020, to February 1, 2021) is not included in the research because of the anticipated

anomaly in traffic levels impacting the crowdsourced data at that time (Bureau of Transportation Statistics, 2021).
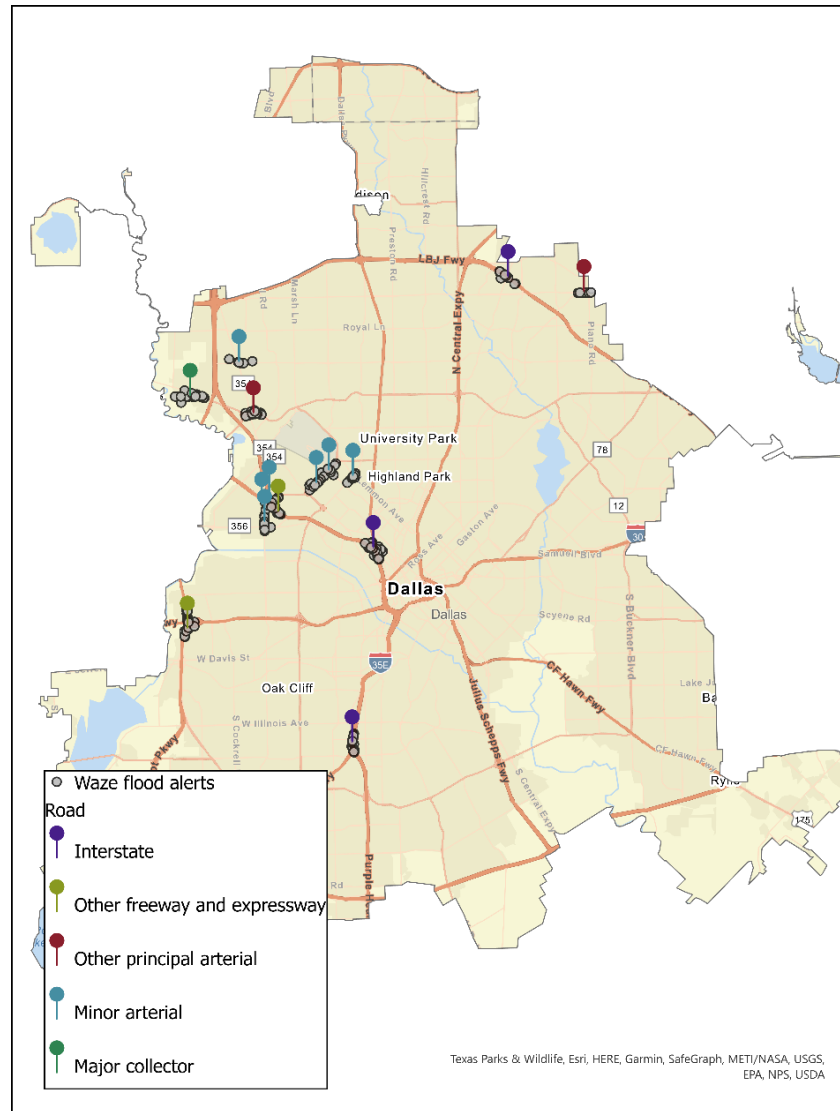


Figure 4-2 Case study intersections and Waze flood alerts

## 4.4 Results

In accordance with the methodology (Section4.2), the RFC, SVM classifier, and XGBoost models are trained on the training dataset and their performance is evaluated on the unseen test dataset generated using storm events different from the training storms.

### 4.4.1  Data Preparation

Fifteen of the storm events are used to train the models and five storm events are held out for testing the models. After retrieving precipitation data from NEXRAD, the optimum value of parameters $n$ and $h$ from Table 4-1 should be determined to generate hazard attributes. The correlation between the risk observation (target feature) and the accumulated rainfall during the previous $n$ hours is used to determine the optimal n, testing several $n$ values of 5, 4, 3, 2, and 1Table 4-1. the correlation is found to be greatest at $n = 2$ hours. In a similar process, the optimum $h$ in Table 4-1 is found to be $h = 4$ mm with a correlation of 0.085. Hence the dataset is created using these thresholds.

Raw precipitation timeseries are used in the GB-RFSM model to estimate vulnerability features, including *inundation area* and the *maximum depth* on the accumulated road surface for 12-minute intervals (Table 4-2). Figure 4-3 shows the GB-RFSM inundation maps on four sample road segments at the end of a severe storm that happened on September 21, 2021.

### 4.4.2  Modeling

Distributions of observed roadway PFF risk in the training and testing datasets are depicted in Figure 4-4. As can be observed, both datasets have an uneven distribution of positive and negative classes, with negative classes being almost ten times more prevalent than positive classes, showing the imbalance in the datasets. Random grid search found the best model configuration for each algorithm. The most important features are identified from models following the procedure described in Section4.2.4. Table 4-4 shows the best model configuration and the six most important attributes for each algorithm. XGBoost and SVM classifier have at least one feature from hazard, vulnerability, and exposure, while no exposure attribute is significantly important for RFC. The important features also show that the most significant

85

exposure attribute is whether the flood event occurred in the *afternoon TOD* or not. It appears

that Waze users are more likely to post flood alerts between 12:00 and 4:00 PM.
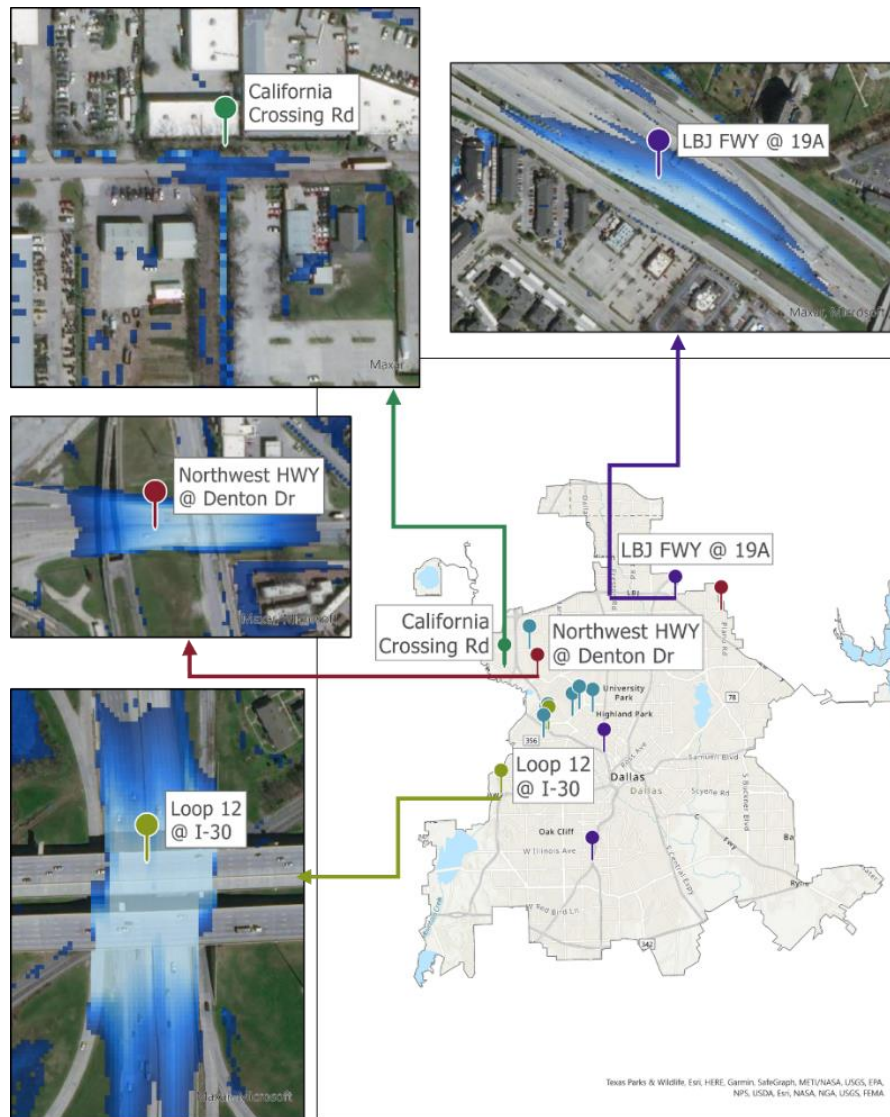


Figure 4-3 Inundation extent on four road segments at the end of a severe storm on September 9, 2021

Among the historical vulnerability attributes, *PFF likelihood in a moderate storm* and

*PFF likelihood in a severe storm*, all three models found EB estimations in a moderate storm

class more significant than EB estimations in severe storms in predicting the risk. This may be

due to the larger number of moderate storms used in calculating PFF likelihoods compared to

severe storms. The larger sample size may cause the estimated likelihoods in the moderate

cluster to be closer to reality.
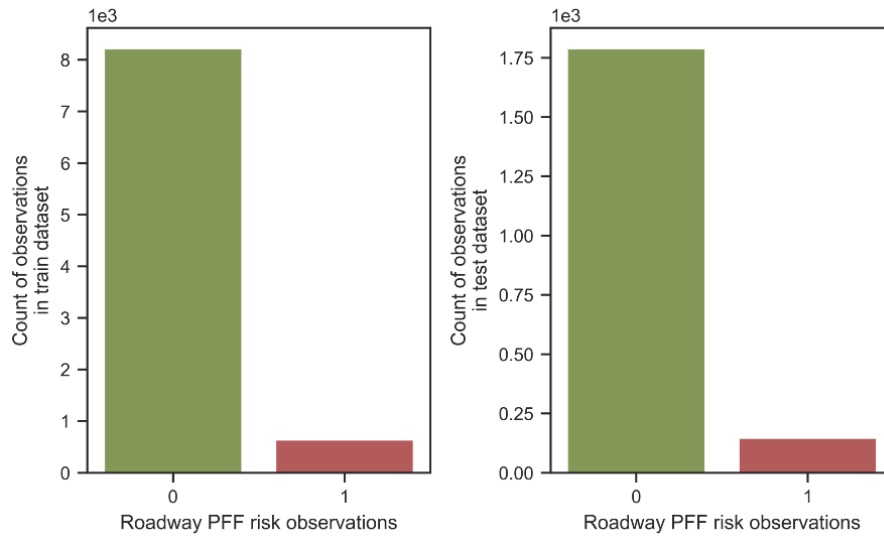


Figure 4-4 Distribution of risk observations in training and testing datasets

Table 4-4 Configuration of tuned ML models

| Model | PRC-AUC | Six most important features | Model configuration |
|---|---|---|---|
| XGBoost | 0.68 | 2-hours preceding accumulation[1]<br>Maximum inundation depth[2]<br>PFF likelihood in moderate storms[2]<br>Afternoon[3]<br>Time to the last maximum rain pulse[1]<br>Maximum preceding rain pulse[1] | Maximum depth: 3<br>Number of trees: 1711<br>Positive class weight = 0.9 |
| RFC | 0.60 | 2-hours preceding accumulation[1]<br>Time to the last maximum rain pulse[1]<br>Maximum preceding rain pulse [1]<br>Maximum inundation depth[2]<br>PFF likelihood in moderate storms[2]<br>Count of rainless intervals[1] | Maximum depth: 4<br>Number of features allowed: 6<br>Number of trees: 741<br>Positive class weight: 0.93 |
| SVM classifier | 0.43 | Maximum inundation depth[2]<br>Afternoon[3]<br>PFF likelihood in moderate storms[2]<br>Time to the last maximum rain pulse[1]<br>2-hours preceding accumulation[1]<br>Maximum preceding rain pulse [1] | Kernel=rbf<br>gamma = 0.1<br>class weight=0.90 |

1 – Hazard features

2 – Vulnerability features

3 – Exposure features

Figure 4-5 and Figure 4-6 show PRC and ROC curves for three tuned ML models compared with a no-skill model. According to both the ROC and PRC curves, XGBoost performs better than RFC and SVM classifier on testing dataset, as shown by the larger area under the curve (AUC) (Table 4-4 and Table 4-5). As mentioned in Section 4.2.5, the desired classification threshold (used to map predicted risk probabilities into risk classes) may vary depending on the data and application. Given the substantial uncertainty in the negative class due to the characteristics of crowdsourced data described in Section 1.1 (smaller confidence in the negative class), forecasting the positive class is most important, since it has more confidence. Therefore, to evaluate misclassification of each model when 90% of risk instances are correctly predicted, the classification threshold is adjusted to map the projected probabilities to PFF risk and no risk classes by setting the minimum required recall score to be 0.9. Using the adjusted threshold, the RFC, XGBoost and SVM classifier predict 73%, 69% and 63% of risk observations in the test storm events respectively, suggesting the superiority of the RFC model.

Table 4-5 shows the model performances and the adjusted threshold. XGBoost performs better on the testing and training dataset for the F2score. However, the recall score on the testing dataset is the highest for the RFC model. This conforms with the PRCs depicted in Figure 4-5, which shows that the superiority of XGBoost to RFC in terms of PRC-AUC is due to higher performance on smaller recalls. However, in recalls higher than 0.2 their performance is almost identical on the test dataset.

The variability of important attributes representing hazard, vulnerability, and exposure (namely, *last 2-hours precipitation*, *maximum inundation depth*, and *TOD*) in true and false positive and negative predictions of the test dataset ($tp, fp, tn, fn$ in Equation 4-2 and 4-3) are shown in Figures 4-7 to 4-9 for XGBoost, RFC, and SVM classifier. The marginal histograms

shown on these figures show the distributions of *maximum depth* estimated by GB-RFSM and

*last 2-hours precipitation* in each set of predictions, respectively.
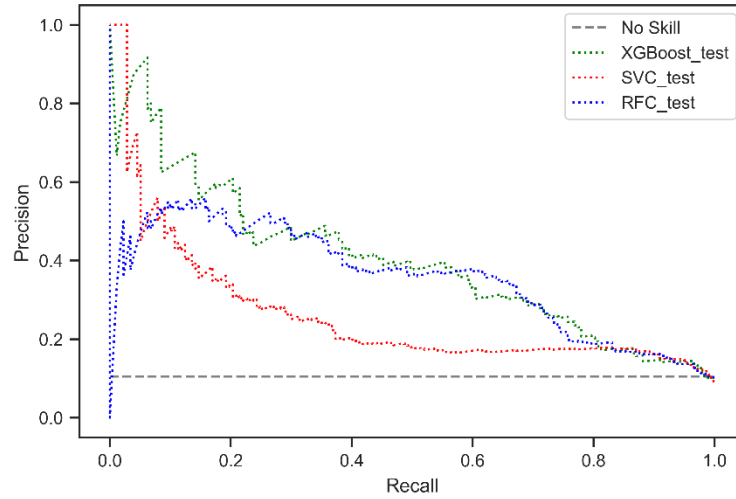


Figure 4-5 Precision-recall curves of ML models
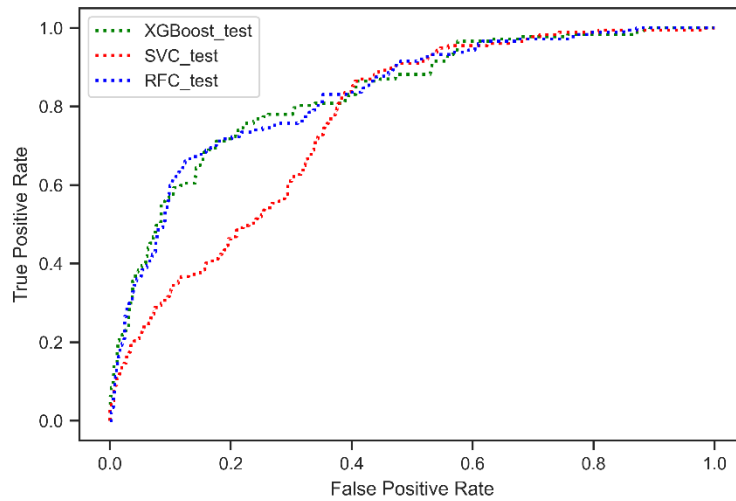


Figure 4-6 ROC curve of ML models

Table 4-5 ML models performance after tuning thresholds

| Model | PRC-AUC | ROC-AUC | Threshold | F2score | Precision | Recall |
|---|---|---|---|---|---|---|
| **XGBoost** | Train = 0.68<br>Test = 0.40 | Train = 0.94<br>Test 0.83 | 0.48 | Train: 0.70<br>Test: 0.54 | Train: 0.37<br>Test: 0.29 | Train:0.89<br>Test: 0.69 |
| **RFC** | Train = 0.60<br>Test = 0.35 | Train=0.91<br>Test=0.83 | 0.62 | Train: 0.65<br>Test: 0.52 | Train: 0.32<br>Test: 0.25 | Train: 0.89<br>Test: 0.73 |
| **SVM classifier** | Train = 0.38<br>Test = 0.18 | Train = 0.85<br>Test = 0.74 | 0.08 | Train: 0.47<br>Test: 0.29 | Train: 0.17<br>Test: 0.16 | Train: 0.89<br>Test: 0.63 |

The distribution of features in XGBoost and RFC are fairly similar; for example, both predict minimal risks in instances where the accumulated precipitation in the last 2 hours is less than 10 mm (Figures 4-7-b, 4-7-d, 4-8-b, 4-8-d). Also, they cover a wide range of maximum inundation depths in the $tn$ category (Figures 4-7-a, 4-8-a),which shows they successfully captured attributes other than *maximum depth* that impact PFF risk. These figures also suggest that XGBoost is not superior to RFC in predicting the positive class, given that it has fewer $tp$ and more $fn$. Rather, its dominance over RFC comes from its ability to distinguish negative classes better, which are uncertain anyway (comparing Figures 4-7-a, 4-8-a).

Unlike XGBoost and RFC, SVM classifier does not show a wide range of *maximum depths* in the $tn$ group (Figure 4-9-a) suggesting that it has a wider range of *maximum depth* in the $fp$ group (Figure 4-9. Figure 4-9-b and 4-9-d show that although SVM classifier predicts low PFF risk in instances with the last 2-hours rainfall less than 10 mm, correctly retrieving 90% of risk instances causes misclassification of low probabilities.

Looking closely at Figures 4-7-a and 4-8-a, it can be seen that instances with large maximum depths in the $tn$ groups mostly occur after midnight and in the morning, when fewer vehicles are on the roads and preceding rainfall is not as high. This implies that these no risk observations have been due to low exposure and low hazard, while vulnerability is still high at locations that could cause roadway PFF.

In addition, the $fn$ group of model predictions (Figures 4-7-c, 4-8-c, 4-9-c) all have low *maximum depth* and *recent 2 hours of precipitation*, which suggests low vulnerability and hazard, respectively. The range of recent precipitation between 30 and 40 mm is identical for the three models in the $fp$ group (Figures 4-7-b, 4-8-b, and 4-9-b). Even though all models indicate a high risk of PFF for these datapoints, there are no risk (Waze) observations available. All of

these timesteps occur at night (8 pm to midnight) and after midnight, when there is less traffic and presumably fewer Waze users reporting risk observations.
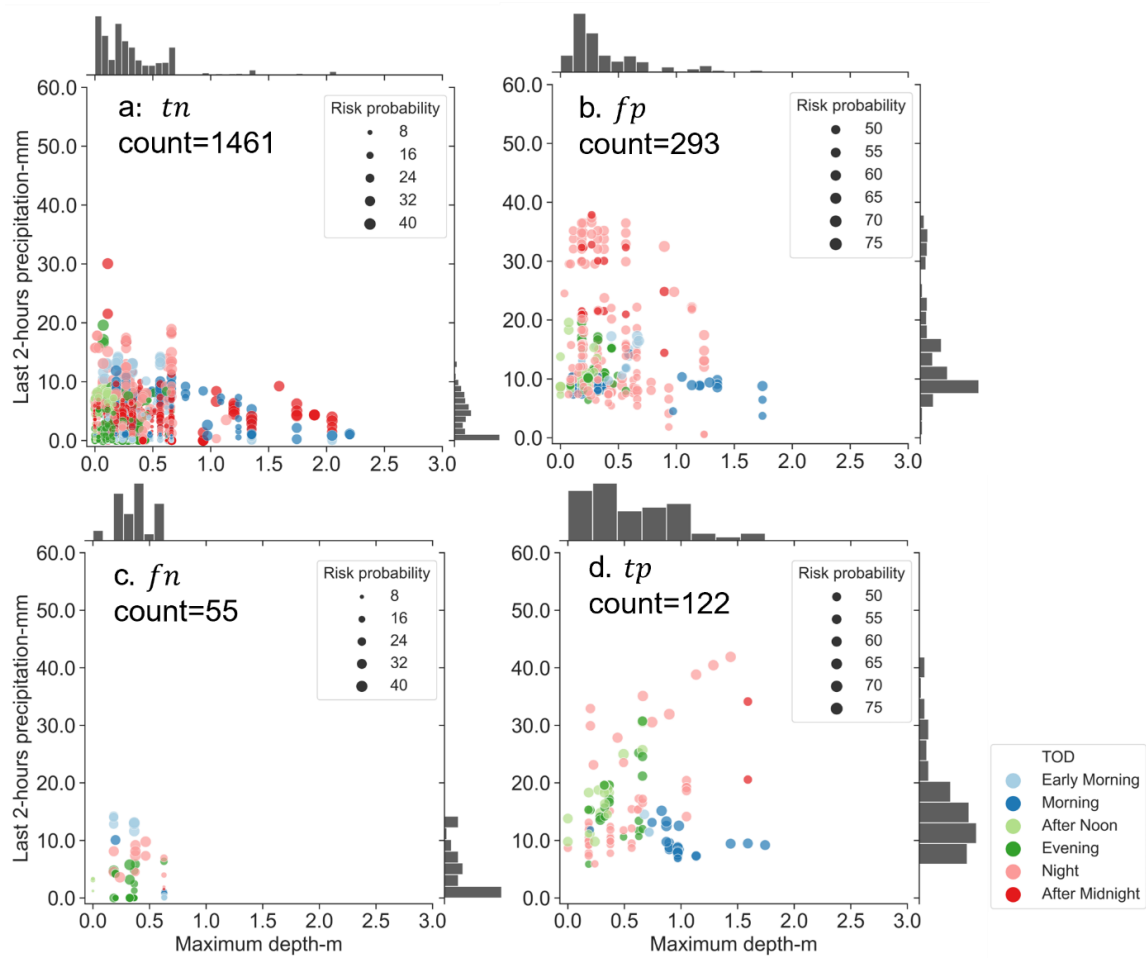


Figure 4-7 Features distributions in confusion matrix of test predictions, a: True negative (tn) predictions, b: False positive (fp) predictions, c: False negative (fn) predictions, and d: True positive (tp) predictions – XGBoost
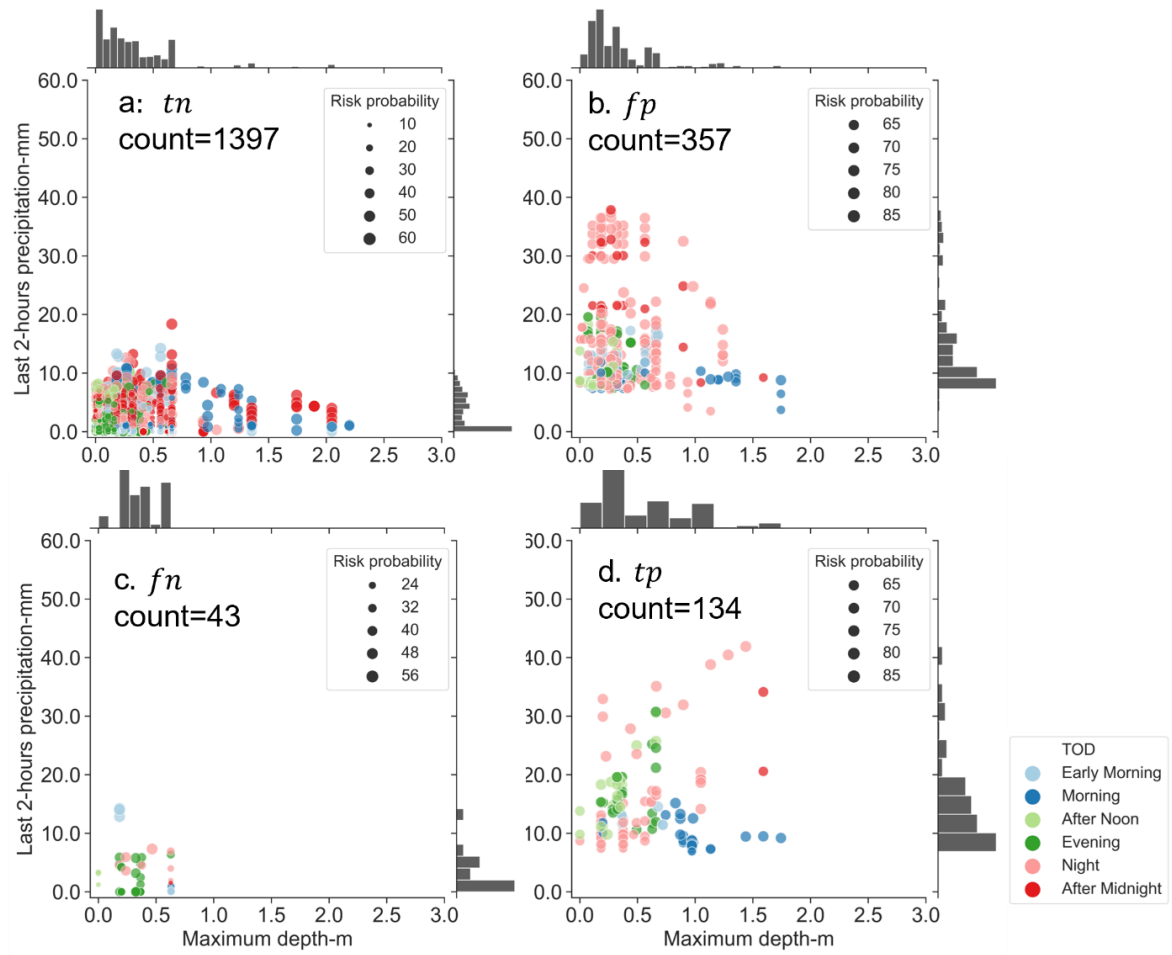
Figure 4-8 Features distributions in confusion matrix of test predictions, a: True negative (tn) predictions, b: False positive (fp) predictions, c: False negative predictions (fn), and d: True positive (tp) predictions -RFC
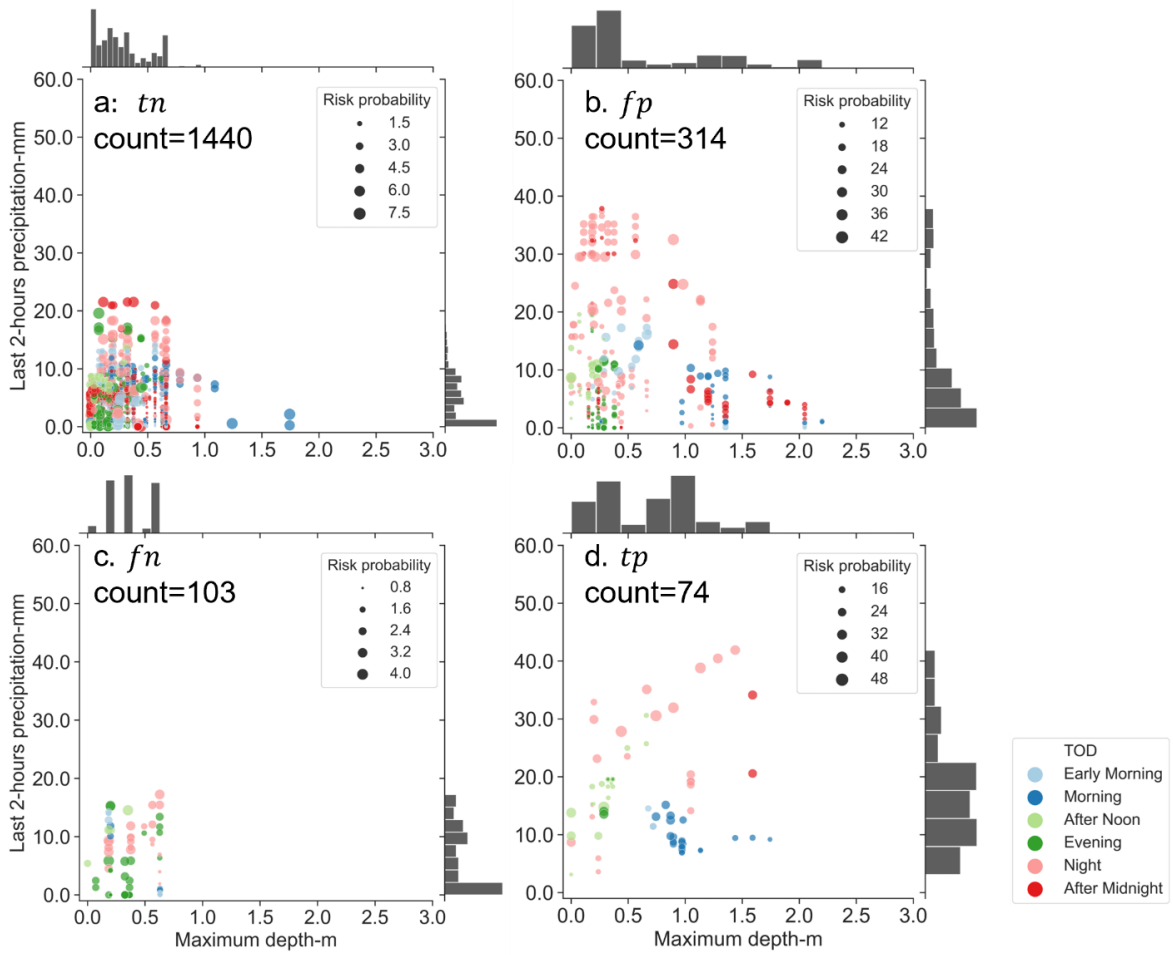
Figure 4-9 Features distributions in confusion matrix of test predictions, a: True negative (tn) predictions, b: False positive (fp) predictions, c: False negative (fn) predictions, and d: True positive (tp) predictions – SVM classifier

## 4.5    Conclusions

This chapter shows that crowdsourced traffic flood alerts, specifically Waze flood alerts, may be leveraged as a proxy to roadway PFF risk, which is a combination of hazard, vulnerability, and exposure. Findings of previous chapters, including likelihoods of PFF in storm clusters (light, moderate, and severe) calculated by the EB regression model developed in Chapter 2 and maximum inundations calculated by GB-RFSM (Chapter 3), represent site-specific vulnerabilities of road segments to PFF. Hazard of a road segment to PFF is estimated based on NEXRAD precipitation and time, *AADT*, and *road classifications* are used to estimate exposure.

Three ML models (XGBoost, RFC, and SVM classifier) are trained using the curated dataset to predict the risk of roadway PFF. The three models consider the features *maximum inundation depth*, *2 hours of preceding precipitation*, *PFF likelihoods* during moderate storms, and *time to the most recent maximum rain pulse* to be representative of significant hazards and vulnerabilities. While *AADT* and *road classification* were not significant in either model, *time of day* is the most important factor among those that indicate exposure. Several strategies are employed to train the models in a way appropriate for imbalanced data and uncertainty in the negative class (no flood risk); First, cost-sensitive training is carried out, which assigns a higher weight to the minority (positive) class' mispredictions. Second, in order to emphasize recall over precision when adjusting model hyperparameters, F2score is utilized in the model selection and cross-validation process. Finally, the classification threshold is modified to extract a training recall score of 0.9 in order to compare models in the high recall area of the PRC and select the model that is more effective at predicting positive risk cases, which represent more risk to motorists. F2score is then used to compare the models after mapping probabilities using the updated threshold.

The modeling results showed that the presented approach could predict reported flood events, but with the cost of a large number of false positive predictions. Further analysis is required to estimate the cost of misclassifications and compare it with costs associated with driving incidents related to roadway flooding. XGBoost and RFC both performed superior to the SVM classifier model by having higher PRC-AUC and ROC-AUC. Even though the XGBoost and RFC models both had greater AUCs, the threshold change showed that by setting the required training recall to 0.9, the XGBoost model outperforms the RFC model in terms of test precision while the RFC model outperforms the XGBoost model in terms of test recall.

Therefore, deciding between the two models depends on the level of uncertainty in the negative class and the cost of false positive predictions. In vehicle routing applications, depending on changes in travel costs and times when rerouting vehicles, higher recall or higher precision may be preferred. Generally, if rerouting does not significantly increase travel costs and times, predicting false positives is justified to minimize the probability of exposure to a flooded road, hence a higher recall is preferred. However, if false positives significantly raise travel costs or turnaround times, it will be more reasonable to avoid predicting risk when there is no risk ($fp$), even if doing so results in missing certain risk instances with low probabilities. As a result, the model would favor higher precision. Overall, the RFC model predicts 73% of risk observations in the test storm events. The presented modeling approach is informative to roadway PFF awareness that could increase travel safety. Mispredictions with high hazard and vulnerability mostly occur at night when exposure is lower and fewer Waze alerts are available.

In terms of limitations, the vulnerability features, maximum inundation depth, and EB-derived PFF likelihoods are simplified attributes. For instance, the GP-RFSM assumes that there is no subsurface drainage system for excess runoff (due to lack of data) and therefore accumulates all excess runoff into low-lying regions. Additionally, developing the site-specific PFF likelihoods involves manual preprocessing and judgment, adding additional uncertainty to the data. Using reliable, high-fidelity records of historical flood depth measurements and storm drainage system configurations and performance in the models could aid in overcoming these limitations. Finally, employing more complex calibrated hydraulic and hydrologic models that account for more PFF mechanisms may improve predictive performance, at the cost of significantly more detailed data requirements and higher computational effort.

# Chapter 5

## SUMMARY, LIMITATIONS, AND FUTURE WORK

### 5.1    Summary and Findings

Water ponding and pluvial flash flooding (PFF) on roadways can pose a significant risk to

drivers. PFF can reduce roadway networks' reliability by decreasing capacity, increasing travel

time, reducing safe speed, and increasing accident risks and deaths through lane submersion

(Agarwal et al., 2005; Smith et al., 2004; Suarez et al., 2005). According to the National Weather

Service, 30 cm of standing water is sufficient to float most cars (National Weather Service,

2022).

PFF occurrences have become more frequent as a result of climate change, expanding

urbanization, growing imperviousness, and inadequate stormwater infrastructure. Physics-based

hydraulic and hydrologic models need precise data, which is frequently unavailable at fine

scales, to estimate PFF at the scale of a road segment suitable for navigation purposes.

Simplifications are frequently employed in hydraulics and hydrologic modeling to address data

shortages and computational limitations. This introduces uncertainty into the results, particularly

in densely urbanized areas where micro-topographic characteristics generally govern flow

dynamics.

Statistical and data-driven methods are alternative techniques to predict PFF based on

historical observations. A notable advantage of statistical models is their ability to account for

possible effects of uncertainties and unobserved variables and rapidly update as new data

become available and system dynamics change. However, they rely on historical PFF

observations with high spatio-temporal resolution, which often is not readily available. In

general, deploying widespread sensors to monitor and record PFF on road segments is

uncommon. Research indicates that stakeholders are more concerned about fluvial and ocean flooding than pluvial flooding (Rosenzweig et al., 2018), which has led to pluvial flooding being a data scarce phenomenon.

With recent advances in technology and the widespread availability of smartphones, social media geographic information and crowdsourced data are now widely available at low cost and can supplement high-fidelity datasets. However, by definition, "crowdsourced data are collected by heterogeneous volunteer individuals of varying knowledge, experience, and perceptions." (Estellés-Arolas and González-Ladrón-de-Guevara, 2012). Therefore, employing these data to supplement PFF observations is challenging since it is dispersed over the research region and time periods and is provided by non-experts (including Waze flood alerts).

To identify the inherent value and applicability of Waze flood alerts as PFF data, a preliminary analysis was performed that found a significant spatial correlation between alerts and road surface depressions. Findings showed that almost 90% of flood alerts are located less than 200 meters away from road surface depressions deeper than 15 cm. Furthermore, no significant correlation exists between Waze flood alerts and National Flood Hazard Layer (NFHL) regions, which implies that Waze flood alerts are consequences of PFF in low-lying areas rather than fluvial flooding (Section 2.4).

Accordingly, Waze flood alerts were utilized to estimate vulnerabilities of low-lying areas to PFF in Chapter 2. Surface depressions and their catchments were delineated from DEMs, along with their physical, hydrologic, and topographic descriptors. These data were used to construct a machine learning model called Random Forest (RF) and a statistical regression model based on Bayes theory called Empirical Bayes (EB) to estimate PFF likelihoods for various storm types. The superiority of EB over RF demonstrated the importance of site-specific

97

unobserved variables in determining the likelihood of PFF. EB successfully captured the

potential effects of such unobserved variables indirectly by updating prior probabilities to

posterior probabilities in accordance with historically observed frequency of PFF.

However, this approach is limited by assuming that the process of filling depressions with

excess runoff is independent of rainfall, which ignores the hydrologic processes involved. As

depressions fill, the flow directions change and merged depressions form with a larger

catchment. To address this issue, a new approach was adopted in Chapter 3 to tackle the hurdles

and data needs of common hydraulic and hydrologic models for local scale flood simulation

while also addressing the limitations of Chapter 2.

An efficient, simplified physics-based rapid flood spreading model (RFSM) is developed

based on the hierarchical filling and spilling process in PFF formation called graph-based RFSM

(GB-RFSM). GB-RFSM simulates water accumulation on road surface depressions in the case of

drainage system failure without relying on the stormwater drainage system's characteristics and

layout, disregarding the effects of flow momentum. GB-RFSM-derived inundation represents the

vulnerability of a road segment to PFF at a given time, considering rainfall and merging

depressions as accumulations increase. By updating the DEM graph at each timestep, the graph-

based structure allows GB-RFSM to capture the temporal evolution of floods. To test the model,

a comparison was made with the Storm Water Management Model (SWMM) simulated

overflows and the projected inundation extent of the GB-RFSM model exhibited satisfactory

agreement on a storm event case study in the City of Dallas.

The vulnerability of road segments to PFF was the emphasis of Chapters 2 and 3, but the

entire risk that PFF poses to drivers goes beyond vulnerability and depends on hazard and

exposure as well. Chapter 4 showed how Waze flood alerts can be utilized to train a model that

considers hazard, vulnerability, and exposure features to estimate the risks of PFF on a roadway in real time. Assuming that the developed GB-RFSM model and EB-derived PFF likelihoods are reasonable indicators of vulnerability, the overall risk is estimated by integrating them with hazard and exposure. Accordingly, a pipeline is developed with data process and a hybrid machine learning model. The hybrid model uses ML classifiers that are trained on Waze flood alerts to combine GB-RFSM and EB results with significant hazard and exposure variables (*maximum inundation depth*, *PFF likelihood in moderate storms*, *last 2 hours rainfall*, *maximum preceding rain pulse*, and *afternoon TOD*).

The results showed that the RFC and XGBoost models performed better than SVM classifier and predicted approximately 70% of the observed PFF risk instances (Waze alerts) during 5 unseen storm events. Compared to what Waze users have reported, however, the overall number of timesteps with PFF risk is roughly three times higher (precision around 0.3).

Overall, this dissertation contributes to advances in PFF understanding and modeling by: 1. introducing and evaluating a novel data source to this data-poor phenomenon and 2. developing statistical and simplified physics-based models to employ this novel data source in PFF situational awareness. The conclusions of this dissertation demonstrate that, despite the substantial uncertainty and complexity of localized PFF formation, historical data can advance our understanding and contribute to mitigating the losses caused by this recurrent hazardous phenomenon. This may incentivize cities and transportation administrations to invest in generating more high-fidelity datasets using crowdsourcing approaches or monitoring sensors.

## 5.2   Limitations

This dissertation has two primary limitations: the first is limits related to Waze and stormwater drainage data, and the second is methods and modeling limitations.

### 5.2.1 Data Limitations

It takes significant preprocessing to assess the Waze data's credibility and map flood alerts—which are not always posted at the exact location of flooding—to plausible neighboring flood-prone surface depressions and use crowdsourced traffic data efficiently. This manual preprocessing is subjective and could introduce errors and bias to the analysis. Additionally, the method used in this study only considers flood-prone areas that Waze users have reported. Wazers' exposure to flooded areas is influenced by a variety of factors, including road type, function, and traffic conditions. As a result, a flood with a comparable extent on the road might result in traffic disruptions of dramatically varying sizes and Waze flood reports at various times and places. The approach taken in this study is less credible for less traveled local roads, which generally have far fewer Waze reports.

The lack of data to model the stormwater drainage system is another important limitation of this study, as well as many other studies because cities often do not have complete knowledge of their drainage systems. The approach taken by Chegini and Li (2022) to estimate urban stormwater topology from surface features could be promising for addressing this limitation in future work.

### 5.2.2 Modeling Limitations

Despite incorporating data-driven site-specific variables estimated by EB to capture the impact of unobserved factors, which may include the functionality of the storm drainage system, the drainage system's temporal variation is still missing. In other words, when calculating the vulnerability of road segments, it is assumed that all drainage inlets are overwhelmed from the start of a storm, although their capacity actually varies spatially and over the storm duration.

Another modeling consideration is the following: Although DEM processing is a one-time step in the GB-RFSM graph extraction procedure, it should be repeated on a regular basis to accommodate for updated DEMs and landscape changes. The vulnerabilities derived from the EB model also require frequent retraining in order to account for changes in infrastructure condition (e.g., aging, damage, or upgrades of stormwater drainage inlets), climate change, and emerging unobserved variables, such as ongoing nearby construction that may increase the possibility of debris and blockages. In addition, the hybrid ML risk prediction model developed in this dissertation is limited by inconsistencies in the negative class (no flooding) of volunteered crowdsourced data. Furthermore, Wazers and other drivers may not have reported roadway flooding, despite being impacted by it., although significant traffic slowdowns could be captured by current navigation apps and result in rerouting around flooded areas. The case study showed that anomalous Waze traffic jams could be helpful for identifying potentially flooded areas that were not reported. Although allowing the trained model to make false-positive risk predictions reduces this bias, it requires assessing the confidence in false positive predictions by comparing them with ground truth data.

## 5.3 Implications

Crowdsourcing is becoming more popular due to improvements in smartphone accessibility and the extensive availability of the wireless internet. As benefits of citizen observation become widely acknowledged, more citizens are eager to participate, which brings us closer to securing traffic safety during storm events. The outcomes of this dissertation could be used in safe route finding to enhance navigation systems and identify the safest route for vehicles by considering PFF risk in routing algorithms (Sohrabi and Lord, 2022). However, the hybrid models developed in Chapter 4 are conservatively skewed towards reducing false negatives (i.e., predicting no risk

when flooding has actually occurred), which could cause unnecessary rerouting of drivers. The models would need to be adjusted to balance these risks in practice. Asking drivers to provide feedback when predictive errors occur could help with this process.

In addition to routing, communities and stakeholders may utilize the developed risk prediction framework to take preventative measures in the event of a high risk of PFF, such as issuing notices and advisories, deploying first responders, and using infrastructure to drain flood waters. Furthermore, detecting road segments that have high vulnerability to PFF allows cities to monitor infrastructure conditions efficiently and make better infrastructure investment plans.

## 5.4    Future Work

As explained earlier, roadway PFF formation is a highly localized and data-poor phenomenon with numerous unpredictable and unobserved contributors. Therefore, utilizing more data would likely significantly increase the accuracy of this approach. On the other hand, scarcity of data raises the need to continue identifying and integrating novel data sources and retrieving implicit information from readily available data (such as extracting flood levels from images and videos). Better assessment of cases of false positive PFF risk predictions could also be achieved using additional ground truth data sources such as flood surveys, news, and social media postings (e.g., Twitter). This would provide the opportunity to retrain the models and quantify the level of confidence in the predictions.

In addition, even though the GB-RFSM algorithm can account for stormwater drainage through runoff abstraction, the lack of stormwater drainage system data has been challenging and imposed notable limitations on this dissertation. In a case study where the necessary data are available, considering the layout of the drainage system and the characteristics of the stormwater inlet can allow removal of the assumption that the drainage system was overloaded before a

storm event started. Barring the availability of such data, as in many cities, an alternative approach would be to infer drainage system locations and characteristics from other data sources, such as aerial imagery, Google Streetview images, and mapped street locations where stormwater pipes are routed (Chegini and Li, 2022).

Furthermore, increasing the resolution of precipitation data used in this study could also improve accuracy. Precipitation data were retrieved from NEXRAD products and the Dallas Love Field ASOS rain gauge station. Using high-resolution network radars such as Collaborative Adaptive Sensing of the Atmosphere (CASA) data, (Chen and Chandrasekar, 2015) can improve results through more accurate precipitation data at the intersection scale.

Another future research avenue would be to incorporate higher resolution land cover and zoning data in vulnerability assessments. The Curve Number used in this study is designed to capture the impact of land use. However, the CN datasets have large resolution compared to road segment scales and higher resolution data could improve vulnerability estimations. For instance, it can be hypothesized that a crowded business area can generate more debris compared to a region with similar land cover and soil type but less busy, which leads to higher site-specific vulnerability for blocked stormwater inlets.

# Chapter 6

## REFERENCES

Agarwal, M., Maze, T.H., Souleyrette, R.R., 2005. Impacts of Weather on Urban Freeway Traffic Flow Characteristics and Facility Capacity. Proceedings of the 2005 Mid-Continent Transportation Research Symposium 14p.

Ahmadalipour, A., Moradkhani, H., 2019. A data-driven analysis of flash flood hazard, fatalities, and damages over the CONUS during 1996–2017. J Hydrol (Amst) 578, 124106.

Asquith, W.H., Roussel, M.C., Thompson, D.B., Cleveland, T.G., Fang, X., 2005. Summary of dimensionless Texas hyetographs and distribution of storm depth developed for Texas Department of Transportation Research Project, Texas Department of Transportation.

Assumpção, T.H., Popescu, I., Jonoski, A., Solomatine, D.P., 2018. Citizen observations contributing to flood modelling: Opportunities and challenges. Hydrol Earth Syst Sci 22, 1473–1489.

Balstrøm, T., Crawford, D., 2018. Arc-Malstrøm: A 1D hydrologic screening method for stormwater assessments based on geometric networks. Comput Geosci 116, 64–73.

Berndtsson, R., Becker, P., Persson, A., Aspegren, H., Haghighatafshar, S., Jönsson, K., Larsson, R., Mobini, S., Mottaghi, M., Nilsson, J., Nordström, J., Pilesjö, P., Scholz, M., Sternudd, C., Sörensen, J., Tussupova, K., 2019. Drivers of changing urban flood risk: A framework for action. J Environ Manage 240, 47–56.

Branch, W.D.T., 2012. Dual-polarization radar principles and system operations. NOAA.[Available online at http://www. wdtb. noaa. gov/courses/dualpol/doc uments/DualPolRadarPrinciples. pdf].

Breiman, L., 2001. Random Forests. Mach Learn 45, 5–32.

Buckland, M., Gey, F., 1994. The Relationship between Recall and Precision. Journal of the American Society for Information Science.

Bulti, D.T., Abebe, B.G., 2020. A review of flood modeling methods for urban pluvial flood application. Model Earth Syst Environ 6, 1293–1302.

Bureau of Transportation Statistics, 2021. Daily Vehicle Travel During the COVID-19 Public Health Emergency [WWW Document]. URL https://www.bts.gov/covid-19/daily-vehicle-travel (accessed 7.31.22).

Buto, S.G., Anderson, R.D., Survey, U.S.G., 2020. NHDPlus High Resolution (NHDPlus HR)---A hydrography framework for the Nation, Fact Sheet. Reston, VA.

Campagna, M., 2016. Social Media Geographic Information: Why social is special when it goes spatial. European handbook of crowdsourced geographic information 45.

Carter, J.G., Cavan, G., Connelly, A., Guy, S., Handley, J., Kazmierczak, A., 2015. Climate change and the city: Building capacity for urban adaptation. Prog Plann 95, 1–66.

Cervone, G., Sava, E., Huang, Q., Schnebele, E., Harrison, J., Waters, N., 2016. Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study. Int J Remote Sens 37, 100–124.

Chegini, T., Li, H.-Y., 2022. An Algorithm for Deriving the Topology of Below-ground Urban Stormwater Networks. EGUsphere 1–30.

Chen, H., Chandrasekar, V., 2015. The quantitative precipitation estimation system for Dallas–Fort Worth (DFW) urban remote sensing network. J Hydrol (Amst) 531, 259–271.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. Association for Computing Machinery, New York, NY, USA, pp. 785–794.

Chu, X., Yang, J., Chi, Y., Zhang, J., 2013. Dynamic puddle delineation and modeling of puddle-to-puddle filling-spilling-merging-splitting overland flow processes. Water Resour Res 49, 3825–3829.

Chung, Y., 2012. Assessment of non-recurrent congestion caused by precipitation using archived weather and traffic flow data. Transp Policy (Oxf) 19, 167–173.

Coles, A.R., Hirschboeck, K.K., 2020. Driving into Danger: Perception and Communication of Flash-Flood Risk. Weather, Climate, and Society 12, 387–404.

Coles, D., Yu, D., Wilby, R.L., Green, D., Herring, Z., 2017. Beyond 'flood hotspots': Modelling emergency service accessibility during flooding in York, UK. J Hydrol (Amst) 546, 419–436.

Craglia, M., Ostermann, F., Spinsanti, L., 2012. Digital Earth from vision to practice: making sense of citizen-generated content. Int J Digit Earth 5, 398–416.

Deng, D., Li, G., Hao, S., Wang, J., Feng, J., 2014. MassJoin: A mapreduce-based method for scalable string similarity joins. In: 2014 IEEE 30th International Conference on Data Engineering. pp. 340–351.

Djokic, D., Ye, Z., Dartiguenave, C., 2011. Arc hydro tools overview. Redland, Canada, ESRI 5.

Douglas, I., Garvin, S., Lawson, N., Richards, J., Tippett, J., White, I., 2010. Urban pluvial flooding: a qualitative case study of cause, effect and nonstructural mitigation. J Flood Risk Manag 3, 112–125.

Edelbrock, C., 1979. Mixture model tests of hierarchical clustering algorithms: The problem of classifying everybody. Multivariate Behav Res 14, 367–384.

Estellés-Arolas, E., González-Ladrón-de-Guevara, F., 2012. Towards an integrated crowdsourcing definition. J Inf Sci 38, 189–200.

Fill, H.D., Stedinger, J.R., 1998. Using regional regression within index flood procedures and an empirical Bayesian estimator. J Hydrol (Amst) 210, 128–145.

Gaitan, S., ten Veldhuis, M. claire, van de Giesen, N., 2015. Spatial Distribution of Flood Incidents Along Urban Overland Flow-Paths. Water Resources Management 29, 3387–3399.

Goodrich, K.A., Basolo, V., Feldman, D.L., Matthew, R.A., Schubert, J.E., Luke, A., Eguiarte, A., Boudreau, D., Serrano, K., Reyes, A.S., Contreras, S., Houston, D., Cheung, W., AghaKouchak, A., Sanders, B.F., 2020. Addressing Pluvial Flash Flooding through Community-Based Collaborative Research in Tijuana, Mexico. Water (Basel) 1–17.

Gouldby, B., Samuels, P., Klijn, F., Messner, F., 2005. Language of risk-project definitions. Floodsite project report T32-04-01.

Guidolin, M., Chen, A.S., Ghimire, B., Keedwell, E.C., Djordjević, S., Savić, D.A., 2016. A weighted cellular automata 2D inundation model for rapid flood analysis. Environmental Modelling and Software 84, 378–394.

Haghighatafshar, S., Becker, P., Moddemeyer, S., Persson, A., Sörensen, J., Aspegren, H., Jönsson, K., 2020. Paradigm shift in engineering of pluvial floods: From historical recurrence intervals to risk-based design for an uncertain future. Sustain Cities Soc 61, 102317.

Han, D., Chan, L., Zhu, N., 2007. Flood forecasting using support vector machines. Journal of Hydroinformatics 9, 267–276.

Hauer, E., Harwood, D.W., Council, F.M., Griffith, M.S., 2002. Estimating safety by the empirical bayes method: A tutorial. Transp Res Rec 126–131.

Helmrich, A.M., Ruddell, B.L., Bessem, K., Chester, M. v, Chohan, N., Doerry, E., Eppinger, J., Garcia, M., Goodall, J.L., Lowry, C., Zahura, F.T., 2021. Opportunities for crowdsourcing in urban flood monitoring. Environmental Modelling and Software 143, 105124.

Hemmati, M., Ellingwood, B.R., Mahmoud, H.N., 2020. The role of urban growth in resilience of communities under flood risk. Earths Future 8, e2019EF001382.

Hemmati, M., Mahmoud, H.N., Ellingwood, B.R., Crooks, A.T., 2021. Shaping urbanization to achieve communities resilient to floods. Environmental Research Letters 16, 094033.

Homer, C., Fry, J., 2012. The National Land Cover Database. US Geological Survey Fact Sheet 1–4.

Homer, C.H., Fry, J.A., Barnes, C.A., 2012. The national land cover database. US geological survey fact sheet 3020, 1–4.

Hooper, E., Chapman, L., Quinn, A., 2014. Investigating the impact of precipitation on vehicle speeds on UK motorways. Meteorological Applications 21, 194–201.

Huang, R., Ma, C., Ma, J., Huangfu, X., He, Q., 2021. Machine learning in natural and engineered water systems. Water Res 205, 117666.

Jack, K., Jaber, F., Heidari, B., Prideaux, V., 2021. Green Stormwater Infrastructure for Urban Flood Resilience: Opportunity Analysis for Dallas, Texas, The Nature Conservancy.

Janizadeh, S., Vafakhah, M., Kapelan, Z., Mobarghaee Dinan, N., 2021. Hybrid XGboost model with various Bayesian hyperparameter optimization algorithms for flood hazard susceptibility modeling. Geocarto Int 1–20.

Kalantari, Z., Cavalli, M., Cantone, C., Crema, S., Destouni, G., 2017. Flood probability quantification for road infrastructure: Data-driven spatial-statistical approach and case study applications. Science of the Total Environment 581–582, 386–398.

Kalantari, Z., Nickman, A., Lyon, S.W., Olofsson, B., Folkeson, L., 2014. A method for mapping fl ood hazard along roads. J Environ Manage 133, 69–77.

Ke, Q., Tian, X., Bricker, J., Tian, Z., Guan, G., Cai, H., Huang, X., Yang, H., Liu, J., 2020. Urban pluvial flooding prediction by machine learning approaches – a case study of Shenzhen city, China. Adv Water Resour 145.

Kecman, V. (Vojislav), 2001. Learning and soft computing : support vector machines, neural networks, and fuzzy logic models 541.

Kousky, C., 2018. Financing flood losses: A discussion of the national flood insurance program. Risk Management and Insurance Review 21, 11–32.

Kuczera, G., 1982. Combining site-specific and regional information: An empirical Bayes Approach. Water Resour Res 18, 306–314.

le Coz, J., Patalano, A., Collins, D., Guillén, N.F., García, C.M., Smart, G.M., Bind, J., Chiaverini, A., le Boursicaud, R., Dramais, G., 2016. Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand. J Hydrol (Amst) 541, 766–777.

Le, P.V. v, Kumar, P., 2014. Power law scaling of topographic depressions and their hydrologic connectivity. Geophys Res Lett 41, 1553–1559.

Lhomme, J., Sayers, P., Gouldby, B., Wills, M., Mulet-Marti, J., 2008a. Recent development and application of a rapid flood spreading method. Flood Risk Management: Research and Practice 15–24.

Lhomme, J., Sayers, P., Gouldby, B.P., Samuels, P.G., Wills, M., Mulet-Marti, J., 2008b. Recent development and application of a rapid flood spreading method. Flood Risk Management.

Li, M., Huang, Q., Wang, L., Yin, J., Wang, J., 2018. Modeling the traffic disruption caused by pluvial flash flood on intra-urban road network. Transactions in GIS 22, 311–322.

Li, X., Willems, P., 2019. Probabilistic flood prediction for urban sub-catchments using sewer models combined with logistic regression models. Urban Water J 16, 687–697.

Lindsay, J.B., Dhun, K., 2015. Modelling surface drainage patterns in altered landscapes using LiDAR. International Journal of Geographical Information Science 29, 397–411.

Liu, J., Wang, J., Xiong, J., Cheng, W., Li, Y., Cao, Y., He, Y., Duan, Y., He, W., Yang, G., 2022. Assessment of flood susceptibility mapping using support vector machine, logistic regression and their ensemble techniques in the Belt and Road region. Geocarto Int 1–30.

Liu, L., Liu, Y., Wang, X., Yu, D., Liu, K., Huang, H., Hu, G., 2015. Developing an effective 2-D urban flood inundation model for city emergency management based on cellular automata. Natural Hazards and Earth System Sciences 15, 381–391.

Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. Accid Anal Prev 37, 35–46.

Manfreda, S., Samela, C., 2019. A digital elevation model based method for a rapid estimation of flood inundation depth. J Flood Risk Manag 12, e12541.

Mark, A., Marek, P.E., 2011. Hydraulic design manual. Austin, Texas Department of Transportation.

Mignot, E., Li, X., Dewals, B., 2019. Experimental modelling of urban flooding: A review. J Hydrol (Amst) 568, 334–342.

Mignot, E., Paquier, A., Haider, S., 2006. Modeling floods in a dense urban area using 2D shallow water equations. J Hydrol (Amst) 327, 186–199.

Molinari, D., de Bruijn, K.M., Castillo-Rodríguez, J.T., Aronica, G.T., Bouwer, L.M., 2019. Validation of flood risk models: Current practice and possible improvements. International journal of disaster risk reduction 33, 441–448.

Mooney, P., Minghini, M., 2017. A review of OpenStreetMap data.

Moy De Vitry, M., Kramer, S., Dirk Wegner, J., Leitao, J.P., 2019. Scalable flood level trend monitoring with surveillance cameras using a deep convolutional neural network. Hydrol Earth Syst Sci 23, 4621–4634.

National Weather Service, 2022. Turn Around Don't Drown [WWW Document]. URL https://www.weather.gov/tsa/hydro_tadd (accessed 8.2.22).

'Nelson, B.R.' 'Zhang, J. "Stevens, S." 'Wilkens, S. 'NOAA C.P., 2017. NOAA Climate Data Record (CDR) of NEXRAD Quantitative Precipitation Estimates (QPE), Version 1.0. NOAA National Centers for Environmental Information.

Nobre, A.D., Cuartas, L.A., Hodnett, M., Rennó, C.D., Rodrigues, G., Silveira, A., Saleska, S., 2011. Height Above the Nearest Drainage–a hydrologically relevant new terrain model. J Hydrol (Amst) 404, 13–29.

Noh, S.J., Lee, J.-H., Lee, S., Kawaike, K., Seo, D.-J., 2018. Hyper-resolution 1D-2D urban flood modelling using LiDAR data and hybrid parallelization. Environmental Modelling & Software 103, 131–145.

Paquier, A., Mignot, E., Bazin, P.-H., 2015. From hydraulic modelling to urban flood risk. Procedia Eng 115, 37–44.

Pedregosa, Fabian, Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. the Journal of machine Learning research 12, 2825–2830.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Perica, S., Pavlovic, S., Laurent, M. st., Trypaluk, C., Unruh, D., Wilhite, O., 2018. NOAA Atlas 14, Precipitation-Frequency Atlas of the United States, Volume 11 Version 2.0: Texas.

Praharaj, S., Chen, T.D., Zahura, F.T., Behl, M., 2021. Estimating impacts of recurring flooding on roadway networks : a Norfolk , Virginia case study. Natural Hazards.

Pregnolato, M., Ford, A., Glenis, V., Wilkinson, S., Dawson, R., 2017. Impact of Climate Change on Disruption to Urban Transport Networks from Pluvial Flooding. Journal of Infrastructure Systems 23, 04017015.

Rafieeinasab, A., Norouzi, A., Kim, S., Habibi, H., Nazari, B., Seo, D.J., Lee, H., Cosgrove, B., Cui, Z., 2015. Toward high-resolution flash flood prediction in large urban areas - Analysis of sensitivity to spatiotemporal resolution of rainfall input and hydrologic modeling. J Hydrol (Amst) 531, 370–388.

RD Manual, 2004. Texas Department of Transportation. Austin, Texas, February.

Risk factor, 2021. Risk Factor. Flood Risk Overview, Does Dallas have risk? [WWW Document]. URL https://riskfactor.com/state/texas/48_fsid/flood (accessed 8.6.21).

Rosenzweig, B.R., McPhillips, L., Chang, H., Cheng, C., Welty, C., Matsler, M., Iwaniec, D., Davidson, C.I., 2018. Pluvial flood risk and opportunities for resilience. Wiley Interdisciplinary Reviews: Water 5, e1302.

Ross, C.W., Prihodko, L., Anchang, J., Kumar, S., Ji, W., Hanan, N.P., 2018. HYSOGs250m, global gridded hydrologic soil groups for curve-number-based runoff modeling. Sci Data 5, 1–9.

Rossman, L.A., 2010. Storm water management model user's manual.

Sadler, J.M., Goodall, J.L., Morsy, M.M., Spencer, K., 2018. Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest. J Hydrol (Amst) 559, 43–55.

Safaei-Moghadam, A., Tarboton, D., Minsker, B., 2022. Estimating the likelihood of roadway pluvial flood based on crowdsourced traffic data and depression-based DEM analysis. Natural Hazards and Earth System Sciences Discussions 2022, 1–29.

Saha, A., Pal, S.C., Arabameri, A., Blaschke, T., Panahi, S., Chowdhuri, I., Chakrabortty, R., Costache, R., Arora, A., 2021. Flood Susceptibility Assessment Using Novel Ensemble of Hyperpipes and Support Vector Regression Algorithms. Water 2021, Vol. 13, Page 241 13, 241.

Saito, T., Rehmsmeier, M., 2015a. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets.

Saito, T., Rehmsmeier, M., 2015b. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. PLoS One 10.

Samela, C., Persiano, S., Bagli, S., Luzzi, V., Mazzoli, P., Humer, G., Reithofer, A., Essenfelder, A., Amadio, M., Mysiak, J., 2020. Safer_RAIN: a DEM-based hierarchical filling-&-Spilling algorithm for pluvial flood hazard assessment and mapping across large urban areas. Water (Basel) 12, 1514.

Sanders, W., Li, D., Li, W., Fang, Z.N., 2022. Data-Driven Flood Alert System (FAS) Using Extreme Gradient Boosting (XGBoost) to Forecast Flood Stages. Water (Switzerland) 14.

Schnebele, E., Cervone, G., Waters, N., 2014. Road assessment after flood events using non-authoritative data. Natural Hazards and Earth System Sciences 14, 1007–1015.

Shen, J., Tong, Z., Zhu, J., Liu, X., Yan, F., 2016. A new rapid simplified model for urban rainstorm inundation with low data requirements. Water (Switzerland) 8, 1–20.

Smith, B.L., Byrne, K.G., Copperman, R.B., Hennessy, S.M., Goodall, N.J., 2004. An investigation into the impact of rainfall on freeway traffic flow. 83rd annual meeting of the Transportation Research Board, Washington DC.

Smith, L., Liang, Q., James, P., Lin, W., 2017. Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. J Flood Risk Manag 10, 370–380.

Smith, L.S., Liang, Q., Quinn, P.F., 2015. Towards a hydrodynamic modelling framework appropriate for applications in urban flood assessment and mitigation using heterogeneous computing. Urban Water J 12, 67–78.

Smith, T., Marshall, L., Sharma, A., 2014. catchment knowledgebase : A Bayes empirical Bayes approach. Water Resour Res 1189–1204.

Sohrabi, S., Lord, D., 2022. Navigating to safety: Necessity, requirements, and barriers to considering safety in route finding. Transp Res Part C Emerg Technol 137, 103542.

Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. Journal of hydroinformatics 10, 3–22.

Strupczewski, W.G., Singh, V.P., Feluch, W., 2001. Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation. J Hydrol (Amst) 248, 123–142.

Suarez, P., Anderson, W., Mahal, V., Lakshmanan, T.R., 2005. Impacts of flooding and climate
    change on urban transportation: A systemwide performance assessment of the Boston Metro
    Area. Transp Res D Transp Environ 10, 231–244.

Tehrany, M.S., Pradhan, B., Jebur, M.N., 2013. Spatial prediction of flood susceptible areas
    using rule based decision tree (DT) and a novel ensemble bivariate and multivariate
    statistical models in GIS. J Hydrol (Amst) 504, 69–79.

Teng, J., Jakeman, A.J., Vaze, J., Croke, B.F.W., Dutta, D., Kim, S., 2017. Flood inundation
    modelling: A review of methods, recent advances and uncertainty analysis. Environmental
    Modelling and Software 90, 201–216.

Thrysøe, C., Balstrøm, T., Borup, M., Löwe, R., Jamali, B., Arnbjerg-Nielsen, K., 2021.
    FloodStroem: A fast dynamic GIS-based urban flood and damage model. J Hydrol (Amst)
    600, 126521.

Tien Bui, D., Hoang, N.D., 2017. A Bayesian framework based on a Gaussian mixture model
    and radial-basis-function Fisher discriminant analysis (BayGmmKda V1.1) for spatial
    prediction of floods. Geosci Model Dev 10, 3391–3409.

Tyralis, H., Papacharalampous, G., Langousis, A., 2019. A brief review of random forests for
    water scientists and practitioners and their recent history in water resources. Water
    (Switzerland) 11.

United  Nations, D. of E. and S.A., 2018. WorldUrbanizationProspectsThe 2018 Revision
    (ST/ESA/SER.A/420), Demographic Research.
    United Nations, Department of Economic and Social Affairs, Population Division .

United States Department of Agriculture, 1986. Urban Hydrology for Small Watersheds.

U.S. Department of Transportation FHWA, 1979. DESIGN OF URBAN HIGHWAY DRAINAGE, THE STATE OF THE ART. FHWA-TS-79-225.

van der Knijff, J.M., Younis, J., de Roo, A.P.J., 2010. LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. International Journal of Geographical Information Science 24, 189–212.

Vapnik, V., 1999. The nature of statistical learning theory. Springer science & business media.

Versini, P.A., Gaume, E., Andrieu, H., 2010. Assessment of the susceptibility of roads to flooding based on geographical information - Test in a flash flood prone area (the Gard region, France). Natural Hazards and Earth System Science 10, 793–803.

Wu, Q., Lane, C.R., 2017. Delineating wetland catchments and modeling hydrologic connectivity using lidar data and aerial imagery. Hydrol Earth Syst Sci 21, 3579–3595.

Wu, Q., Lane, C.R., Wang, L., Vanderhoof, M.K., Christensen, J.R., Liu, H., 2019a. Efficient delineation of nested depression hierarchy in digital elevation models for hydrological analysis using level-set method. JAWRA Journal of the American Water Resources Association 55, 354–368.

Wu, Q., Lane, C.R., Wang, L., Vanderhoof, M.K., Christensen, J.R., Liu, H., 2019b. Efficient delineation of nested depression hierarchy in digital elevation models for hydrological analysis using level-set method. JAWRA Journal of the American Water Resources Association 55, 354–368.

Wu, Q., Lane, C.R., Wang, L., Vanderhoof, M.K., Christensen, J.R., Liu, H., 2019c. Efficient Delineation of Nested Depression Hierarchy in Digital Elevation Models for Hydrological Analysis Using Level-Set Method. J Am Water Resour Assoc 55, 354–368.

Xiong, J., Li, J., Cheng, W., Wang, N., Guo, L., 2019. A GIS-Based Support Vector Machine
Model for Flash Flood Vulnerability Assessment and Mapping in China. ISPRS
International Journal of Geo-Information 2019, Vol. 8, Page 297 8, 297.

Yin, J., Yu, D., Yin, Z., Liu, M., He, Q., 2016. Evaluating the impact and risk of pluvial flash
flood on intra-urban road network: A case study in the city center of Shanghai, China. J
Hydrol (Amst) 537, 138–145.

Yong-He, L., Wan-Chang, Z., Jing-Wen, X., 2015. Another Fast and Simple DEM Depression-
Filling Algorithm Based on Priority Queue Structure. Atmospheric and Oceanic Science
Letters 2, 214–219.

Zahura, F.T., Goodall, J.L., Sadler, J.M., Shen, Y., Morsy, M.M., Behl, M., 2020. Training
Machine Learning Surrogate Models From a High-Fidelity Physics-Based Model:
Application for Real-Time Street-Scale Flood Prediction in an Urban Coastal Community.
Water Resour Res 56.

Zhang, S., Pan, B., 2014. An urban storm-inundation simulation method based on GIS. J Hydrol
(Amst) 517, 260–268.

Zhao, T., Minsker, B., Salas, F., Maidment, D., Diev, V., Spoelstra, J., Dhingra, P., 2018.
Statistical and Hybrid Methods Implemented in a Web Application for Predicting Reservoir
Inflows during Flood Events. JAWRA Journal of the American Water Resources
Association 54, 69–89.

Zheng, X., Maidment, D.R., Tarboton, D.G., Liu, Y.Y., Passalacqua, P., 2018. GeoFlood: Large-
Scale Flood Inundation Mapping Based on High-Resolution Terrain Analysis. Water
Resour Res 54, 10, 10–13, 33.

Zhu, H., Obeng Oforiwaa, P., Su, G., 2022. Real-time urban rainstorm and waterlogging disasters detection by Weibo users. Natural Hazards and Earth System Sciences Discussions 1–15.

Zou, Y., Ash, J.E., Park, B.J., Lord, D., Wu, L., 2018. Empirical Bayes estimates of finite mixture of negative binomial regression models and its application to highway safety. J Appl Stat 45, 1652–1669.