ABSTRACT

Title of Dissertation:	DATA-DRIVEN ASSESSMENT FOR UNDERSTANDING THE IMPACTS OF LOCALIZED HAZARDS
	Hamed Ghaedi, Doctor of Philosophy, 2022
Dissertation directed by:	Assistant Professor, Allison C. Reilly, Department of Civil & Environmental Engineering, University of Maryland

Both the number of disasters in the U.S. and federal outlays following disasters are rising. Thus, evaluating the impact of varying natural hazards on the built environment and communities rapidly and at various spatial scales is of the utmost importance. Many hazards can cause significant and repetitive economic and social damages.

This dissertation is a collection of studies that broadly evaluates resilience outcomes in urban areas using data-driven approaches. I do this over three chapters, each of which explores a unique aspect of hazards and their impact on society. The first two chapters are devoted to federal disaster programs aimed at supporting recovery and building resilience. I especially seek to understand how characteristics of hazards intersect with aspects of the physical and social environment to drive federal intervention. The final chapter explores the heterogenous impacts of natural

hazards in urban communities and how disparities correlate with various socioeconomic and demographic characteristics.

The first two studies examine two major federal disaster programs in the U.S. – FEMA Public Assistance (PA) and FEMA National Flood Insurance Program (NFIP) – at varying spatial and temporal scales. Both leverage parametric and non-parametric statistical learning algorithms to understand how measures of hazard intensity and local factors drive federal intervention. These studies could be used by federal/state-level resource managers for planning the level of aid that may be required after a disaster. This study can also potentially be useful for decision-makers to identify the potential causes of increased disaster spending over time.

In the final chapter, I evaluate the links between public transit disruptions, socioeconomic characteristics, and precipitation. By analyzing the spatial distribution and clustering of infrastructure disruptions, I identify the area(s) susceptible to a disproportionate amount of disruptions. Additionally, spatial statistical models are developed to investigate the relationship between infrastructure disruptions and the characteristics of the communities by including variables related to socioeconomic, demographics, social vulnerability, traffic volume, transit system, road connectivity, and the built environment characteristics. For the decision-makers with the goal of improving the performance and resilience of the transit system, this study can provide insight to locate critical areas impacted by such disruptions.

DATA-DRIVEN ASSESSMENT FOR UNDERSTANDING THE IMPACTS OF LOCALIZED HAZARDS

by

Hamed Ghaedi

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor Allison C. Reilly, Chair
Professor Gregory B. Baecher
Professor Michelle T. Bensi
Professor Katrina M. Groth, Dean's Representative
Professor Tom Logan

Dedication

To my parents, Karim and Sareh, for their unconditional love and support.

Acknowledgements

I would like to express my gratitude to the following people who helped me along the way on this journey and made this dissertation possible.

First and foremost, I am extremely grateful to my advisor, Dr. Allison C. Reilly for her continuous support, invaluable feedback, patience, and encouragement throughout my Ph.D. studies. As my mentor, she has taught me more than I could ever give her credit for here. I am extremely grateful for our conversations, which were significant in inspiring me to think outside the box and from different perspectives to form a comprehensive analysis.

Each of the members of my Dissertation Committee has provided me extensive assistance and suggestions, which contributed to this work. I would like to thank Dr. Michelle T. Bensi who has been an unofficial advisor to me supporting and encouraging me during my program. I appreciate Dr. Gregory B. Baecher whose wealth of knowledge and experience has been the North Star to me and everyone else in our department to seek guidance. I would further like to thank Dr. Katrina M. Groth for her invaluable insights on how to improve my research as well as communicating and presenting research findings effectively. I would like to thank Dr. Tom Logan for his support and encouragement and to show me that the realm of disaster resilience does not have geographical boundaries.

Special thanks to Dr. Marccus Hendricks and Dr. Hiba Baroud for their research collaboration and support throughout my doctorate studies. I am especially grateful to all my friends and lab mates, past and present, in the Civil and Environmental Engineering department for their help and support. With a special

thanks to Zeinab Yahyazadeh Jasour, Dr. Paul Magoulick, and Dr. Aref Darzi with whom I spent cherished time in the research meetings and social settings. I really appreciate their friendship and I am particularly grateful for our enjoyable and memorable conversations. I would further like to acknowledge Dr. Sachraa Borjigin, Dr. Ashkan Keivan, Dr. Yalda Saadat, and my friends in Maryland and Virginia for their help and support through this journey.

Last but not least, I would like to thank my family for their endless support and encouragement – especially my wife and best friend, Setareh, for being a source of advice, respect, reassurance, and love. Without your unyielding love and support, I never would have made it.

Table of Contents

List of Tables

List of Figures

List of Major Abbreviations

BART	Bayesian Additive Regression Trees
CART	Classification And Regression Tree
CRS	Community Rating System
FEMA	Federal Emergency Management Agency
FPR	Flood Peak Ratio
GBM	Gradient Boosting Machines
GFF	Giovanni Flooded Fraction
GLM	Generalized Linear Model
MAE	Mean-Absolute Error
MARS	Multivariate Adaptive Regression Splines
NCA-LDAS	National Climate Assessment - Land Data Assimilation System
NFIP	National Flood Insurance Program
NLCD	National Land Cover Database
NN	Neural Networks
PA	Public Assistance Program
PDD	Presidentially Declared Disaster
PIF	Policies in Force
RF	Random Forest
RMSE	Root Mean-Squared Error
SFHA	Special Flood Hazard Area
SVR	Support Vector Regression

USGS	United States Geological Survey
VIF	Variance Inflation Factor
ZINB	Zero-Inflated Negative Binomial

1 Chapter 1: Introduction

1.1 Background

In recent years, the United States has experienced disasters, ranging from the wildfires in California, to record floods in the Midwest, to hurricanes in the Southeast, and experts say the frequency and intensity of such extreme events are on the rise (Emanuel, 2020; Taherkhani et al., 2020). Since 2005, the United States has averaged more than 65 disasters each year that rise to the level of requiring federal aid and intervention (U.S. Government Accountability Office, 2016). The frequency and magnitude of these events bolster the necessity of promoting resilience in our infrastructure and communities to cope with the devastating consequences of such hazards.

A report by U.N. World Meteorological Organization estimated 11,000 weather-related disasters globally in the past 50 years, and the United States has incurred the most losses at 38% of global economic losses (World Meteorological Organization, 2021). Scientists in this report argue that climate change, more extreme weather, and better reporting are the main contributors to the rise. Other researchers assert that the main reasons behind the rising disaster losses are the increasing hazard exposure due to urbanization, and population and economic growth, particularly in hazard-prone regions such as coastal areas (Botzen et al., 2019; Jongman et al., 2012; Neumann et al., 2015). The impact on infrastructure and societies from natural hazards are categorized as direct losses, which result from damages to the structures

from immediate contact with natural hazards (e.g. roof destruction by high wind), and indirect losses which consist of losses not directly induced by the disaster itself, but by its consequences (e.g., revenue or output reduction, and other disruptions to or burdens on daily life) (Hallegatte & Przyluski, 2010). Natural hazards can also have detrimental impacts on public health in terms of the well-being of the communities (Kovats & Hajat, 2008) or even fatalities (Borden & Cutter, 2008; Sharif et al., 2015), where the impact is often higher for socially vulnerable groups (Cutter & Finch, 2008).

The unprecedented occurrences of natural hazards in recent decades have given rise to the notion of building infrastructure with the capability of resisting and quickly recovering from those hazards. The concept of resilience has received great attention from modern societies as they come to acknowledge that they cannot avert all threats or disasters (Renschler et al., 2010). Resilience was first introduced by Holling (1973) in the area of ecological systems as "a measure of the persistence of systems and of their ability to absorb change and disturbance and still maintain the same relationships between populations or state variables." Since then, significant research has been done in the area of infrastructure and community resilience (Bocchini et al., 2014; Cutter et al., 2008; Koliou et al., 2020; Sharifi, 2016). For example, Folke et al. (2002) argued that in addition to resisting and rapidly recovering the shocks, the ability of learning to adapt to future shocks and vulnerabilities are important components of resilience. A well-cited resilience framework presented by Bruneau et al. (2003) identifies four properties to characterize resilience: robustness, redundancy, resourcefulness, and rapidity.

However, a consensus over a unified and consistent framework to evaluate resilience does not exist (Meerow et al., 2016), leading to challenges in how to adequately measure resilience and effectively invest resources to build resilience. Resilience in the context of hazards usually refers to the pre-event measures to prevent disaster damage and losses and post-event strategies to overcome and minimize disaster impacts (Tierney, 2003).

In the United States, the Federal Emergency Management Agency (FEMA) plays a large role in managing risk from and building resilience to natural hazards in the built environment. Unfortunately, given their limited resources and the scale of the disaster problem, their efforts are largely reactionary. Indeed, only 15% of all FEMA funds spend on disasters go towards mitigation through Hazard Mitigation Grant Program (FEMA, 2021), despite estimates that mitigation has a 6-to-1 return-on-investment (Multihazard Mitigation Council, 2017). Instead, most of their efforts and expenditure goes towards aiding communities recover from disasters – usually to their pre-disaster condition, meaning that communities are not necessarily more able to resist and overcome future disasters (Olshansky & Johnson, 2014).

Despite the recent Justice40 Initiative, which seeks to use a whole-of-government approach to bring environmental justice and economic opportunities to disadvantaged communities, FEMA and the federal government broadly have been unable to address the fact that risk to hazards follows lines of race and class (Cutter et al., 2006). Socially vulnerable communities are more likely to be exposed to hazards, more likely to live in housing vulnerable to extreme weather, and more likely to be under- or uninsured (Dash et al., 2009). However, compared to wealthier

communities, socially vulnerable communities may experience less damage in net, simply because they have fewer assets to begin with (SAMHSA, 2017). Because of how the federal government responds to needs – usually based on measured damage – socially vulnerable communities frequently receive less recovery aid and fewer mitigation dollars, preventing investments in long-term resilience and thus exacerbating disparities (Emrich et al., 2022). Further, disparities within the built environment – many of which are a direct legacy of redlining and other acts of systemic racism – mean that horizontal support infrastructure, including electric power and water, is often less reliable during and after extreme events (Mitsova et al., 2018). These outages lead to additional indirect costs, which are rarely measured, especially for poorer communities that lack the resources to adequately adapt.

This dissertation makes progress in the broad area of disaster resilience by using data-driven approaches to understand the ways in which the built environment interacts with extreme events to ultimately impact individuals and communities. I do this over three chapters, each of which explores a unique aspect of hazards and their impact on society. The first two chapters are devoted to federal disaster programs aimed at supporting recovery and building resilience. I especially seek to understand how characteristics of hazards intersect with aspects of the physical and social environment to drive federal intervention. This work is especially novel because it is among the first to leverage machine learning methods to decompose the highly non-linear relationships between hazards and government response. The final chapter explores the impacts that extreme events have on the serviceability of infrastructure

systems, and unpacks the role that social and racial disparities may have on diminished serviceability after extreme events.

While the chapters focus on different problems within a resilience context, they are thematically connected through their data-driven approaches that seek to unpack how hazards impact society. Understanding these relationships can potentially guide new interventions to reduce risk in the built environment. An overview of the chapters follows.

1.2 Chapter 1: The role of repeated hazards and other salient factors influencing growing Public Assistance outlays

Both the number of disasters in the U.S. and federal outlays following disasters are rising. The Public Assistance Program (PA) is a key program administered by the Federal Emergency Management Agency (FEMA) dedicated to rebuilding damaged public infrastructure and aiding local and state governments through their recovery; after most disasters, it is the primary source of recovery funds. While the degree to which PA builds local resilience has repeatedly been questioned (Davlasheridze et al., 2017), as the structure of the program reimburses applicants for damages and does not reward loss avoidance, a goal of FEMA is presumably to administer the program equitably, or at least evenly. In this work, I evaluate the salient features that contribute to the number of county-level PA applicants and projects following disasters using statistical learning theory.

PA specifically reimburses expenses related to emergency protective measures and restoration of public infrastructure for both natural hazards and man-made

disasters. To be eligible for aid, the state and the county must surpass specific damage thresholds; in these instances, the event is designated a Presidentially Declared Disaster (PDD), which, under the Stafford Act, authorizes the use of federal resources for recovery. Despite the PA program being in existence for nearly 50 years in some form, the literature evaluating its effectiveness is notably sparse. Little is empirically known about how it contributes to national resilience, if at all. The current disaster management system focuses heavily on short-term recovery as opposed to strategic planning for long-term nationwide hazard reduction and resilience (U.S. Government Accountability Office, 2019). Communities lack financial incentives to invest in mitigation, though are guaranteed by law (via the Stafford Act) significant funding to repair damaged infrastructure after PDDs - i.e., disasters recognized by the federal government (Kousky & Shabman, 2012).

There are currently many limitations to our understanding of how PA operates in response to hazards and supports local recovery. For instance, while procedural equity concerns are well documented within PA (Domingue & Emrich, 2019), in that more funds tend to go to higher-income communities and that this may widen the wealth gap, the variables that are correlated with these findings are inconsistent. Further, there is reasonable evidence that local, state, and national governments learn from their disaster experiences and that this could influence disaster outlays (Crow et al., 2018; May, 1992). This, in principle, could lead to larger PA outlays as jurisdictions become more adept at the reimbursement process - though this has not been investigated with significant depth. Finally, many researchers hypothesize that local capacity influences disaster preparedness which, in turn, correlates with local

outcomes (Cigler, 2007; Cutter et al., 2010; Norris et al., 2008). The GAO, among others, has struggled to measure local capacity via indicators in a way that portends local ability to respond to a disaster and short- and long-term outcomes (Hall, 2008; U.S. Government Accountability Office, 2012).

Thus, little is known about how evenly public assistance is distributed amongst locals following PDDs. In order to address the above gaps, this work supports the burgeoning inquiry into this area by identifying the salient factors that are associated with receipt of aid in order to determine the degree to which the program is administered consistently despite factors related to the socio-economic condition and local capacity. I specifically focus on repetitive flooding events in the upper Midwest between 2003 and 2018. This work uses a variety of local indicators that are popular for predicting local capacity and local learning over time to plan for and recover from disasters (e.g., educational attainment, income, demographics, disaster experience). The results suggest that while hazard intensity is a key indicator for PA outlays, so are other indicators that are not reflective of flood risk, including the state in which the disaster occurred, the county's prior experience with disasters, the county's median income, and the length of time between the end of the disaster and the date when a disaster is declared. These findings may help explain some increases in outlays over time and ultimately inform discussions on how infrastructure disaster policy could be reformed to induce better local resilience decisions about infrastructure.

This work uses machine learning to support the investigation. While the use of statistical learning theory is becoming increasingly common in the area of

infrastructure and disasters (e.g., (Shashaani et al., 2018)), statistical learning theory is virtually absent in the federal disaster aid space, which tends to favor parametric modeling. I use parametric, semi-parametric, and non-parametric statistical models to explore the salient factors for predicting the level of PA grants in communities. A discovery of wholesale relationships between the number of projects and metrics associated with local capacity and experience with disasters could suggest that causes for increases in public outlays go beyond changes to the hazard environment associated with climate change and changes to the vulnerability of our infrastructure. It may imply that increasing federal outlays are tied also to how communities learn during the recovery process and their local capacity to navigate the "system." In addition to having cost implications, this could have other implications, including those surrounding equity and the long-term viability of how disasters are currently financed in the US.

1.3 Chapter 2: Evaluating the Flood Peak Ratio and Giovanni Flooded Fraction as useful proxies of flood intensity

Similar to the PA program, many other federal programs that financially support recovery after extreme events are faced with an increased financial burden. In the United States, the National Flood Insurance Program (NFIP), which is administered by FEMA, is the main provider of flood insurance and currently insures over $1.3 trillion in assets (Horn & Webel, 2021). The growing economic activity and development in risky areas are directly influencing the increasing insured losses since the early 1990s (S. D. Brody et al., 2008; Highfield & Brody, 2017; Michel-Kerjan et

al., 2012). These increased losses have required the NFIP to borrow money from the U.S. Treasury, which has led to a deficit of nearly $25 billion (Wells, 2017). Therefore, the ability to accurately forecast the amount of losses in the aftermath of such extreme events plays an important role in achieving the goal of promoting resilience and loss mitigation throughout the communities.

Forecasting losses or liability incurred by a hazard requires spatially-resolved knowledge about the hazard's intensity. For flooding events, it has been primarily done through either computationally intensive hydrodynamic modeling or by flood hazard maps. However, both have specific limitations. The former requires both subject matter expertise to create the models and significant computation time. The time limitation prevents comparisons over wide spatial and temporal scales. It is time-intensive, and perhaps impossible, to create spatially-resolved hydrodynamic models for the entire US or state on a daily level. The latter, flood hazard maps, are a good indicator of the area's general threat over a long timescale but are static. That is, this metric provides no variability among specific events and does not reflect, for example, the intensity of particular rainfall or snowmelt events. As a result, flood hazard maps may underestimate the true flood exposure.

Data-driven methods to rapidly estimate flood occurrence and intensity at localized levels are in their infancy, and have not been evaluated at wide spatial or temporal scales in terms of their ability to predict damage or infrastructure reliability. Besides Czajkowski et al. (2013) and Czajkowski et al. (2017), the research to date that uses statistical methods to forecast flood losses generally relies on the 100-year floodplain or hydrodynamic models to measure the flood's intensity, with the

limitations explained above. To address these limitations, I assess the usefulness of two data-driven heuristics to forecast flood losses that both are simple and computationally rapid to compute, benefiting from publicly available data on NFIP claims and other underlying factors such the land cover and demographic characteristics.

I specifically evaluate the usefulness of the Flood Peak Ratio (FPR) and NASA's Giovanni Flooded Fraction (GFF) for forecasting flood damages. The FPR measures the discharge intensity relative to past intense events in order to isolate the flood severity in a particular region. The GFF grids the entire US and relies on satellite imagery to report the fraction of each grid that is flooded. Both the FPR and the GFF are conceptually simple and computationally rapid to compute. This research uses data from the National Flood Insurance Program (NFIP) to proxy flood damage. This study focuses on flooding events that occurred in 2016 in the contiguous United States to forecast NFIP claims. To identify the flood events in 2016, I implemented a novel approach by building a data-driven classifier using NFIP and FPR data between 2005 and 2015 for each county. I built separate models at the nationwide level and as well as the state level. I found that the model performance improved for most of the states compared to the nationwide analysis. Another key finding is that for the vast majority of states, the most important variable in terms of error reduction for models developed for each state was an FPR variable for the FPR approach or a GFF variable for the GFF approach. Overall, both the FPR and GFF approaches present promising performance considering their simplicity and diverse characteristics of each state.

1.4 Chapter 3: Evaluating the links between public transit disruptions, socioeconomic characteristics, and heavy precipitation

Studies have shown that the impact and recovery of the communities in the aftermath of extreme events are greatly influenced by the demographic and socioeconomic characteristics of the communities (Horney et al., 2017; Knighton et al., 2020; Rufat et al., 2015). However, many studies have only focused on the community outcomes after extreme incidents (Cutter et al., 2008; Finch et al., 2010; Wang & Ganapati, 2018). Less is understood about the resilience of urban communities with different characteristics to the low-to-moderate intensity natural hazards and the outcome of such events on everyday public services.

One of the public services vital to urban communities is the public transit system and one of the natural hazards that may cause disruptions to its services is heavy precipitation. Such disruptions may have larger implications for certain communities, particularly disadvantaged communities, as they are highly reliant on public transit and often lack other means of transportation. The majority of the existing literature on public transit during extreme precipitation events has solely focused on either its performance (Hofmann & O'Mahony, 2005; Kulkarni & Shafei, 2018; Suarez et al., 2005) or changes in travel behaviors and ridership (Guo et al., 2007; Kashfi Syeed et al., 2013; Stover & McCormack, 2012). However, little is known about the system performance and the distribution of its disruptions during extreme precipitation events among the neighborhoods with varying socioeconomic and demographic characteristics. To address the above gaps, I evaluate the nexus between public transit disruptions, socioeconomic characteristics, and heavy

precipitation in an urban environment. I hypothesize that the communities with lower socioeconomic and demographic status are susceptible to disproportionate disruptions to public transit, induced by heavy precipitation. In particular, I look at the delays in the public bus system of the city of Baltimore, MD.

A spatial analysis is developed to evaluate the spatial autocorrelation of bus delays and identify the location of neighborhoods with clustered delays. Subsequently, upon confirming the existence of the spatial autocorrelation effect, spatial statistical models are created to investigate the relationship between the bus delays and the characteristics of the communities. Various parameters are used in building the models including those related to socioeconomic, demographics, social vulnerability, traffic volume, transit system, road connectivity, and the built environment. The effect of precipitation on the bus delays is incorporated by examining the delays during wet periods versus the delays during dry periods. The results show a spatial autocorrelation effect for bus delays, with and without accounting for precipitation. However, since the spatial autocorrelation is not strong enough, the non-spatial statistical model investigated outperformed spatial statistical models. Assessing the non-spatial statistical model, various parameters are identified as statistically significant variables such as the proportion of the non-white population and level of English proficiency.

1.5 Overview of Dissertation

The remainder of this research dissertation proceeds as follows. The study on PA and the salient factors influencing its outlays is presented in Chapter 2. In Chapter

3, I evaluate the usefulness of two data-driven methods as proxies of the NFIP reimbursements. The study of public transit disruptions and their nexus to the characteristics of the communities is covered in Chapter 4. The conclusions and future work are presented in Chapter 5.

2 Chapter 2: The role of repeated hazards and other salient factors influencing growing FEMA Public Assistance outlays[1]

Abstract

Both the number of disasters in the U.S. and federal outlays following disasters are rising. FEMA's Public Assistance (PA) is a key program for rebuilding damaged public infrastructure and aiding local and state governments in recovery. After most disasters, it is the primary source of recovery funds. Between 2000 and 2019, more than $125B (adjusted, 2020 dollars) has been awarded through PA. While all who qualify for PA should have equal opportunity to receive aid, no study has systematically reviewed the distribution of these outlays to determine how evenly the program has been administered. In this work, we evaluate the salient features that contribute to the number of county-level PA applicants and projects following disasters using statistical learning theory, using repetitive flooding events in the upper-Midwest between 2001 and 2017 as a case study. The results suggest that while hazard intensity is a key indicator for PA outlays, so are other indicators that are not reflective of the nature of the disaster itself, including the state in which the disaster occurred, the county's prior experience with disasters, the county's median income, and the length of time between the end of the disaster and the date when a disaster is

[1] Ghaedi, H., Reilly, A., and Abrahams, L. (expected 2022). "The role of repeated hazards and other salient factors influencing growing FEMA Public Assistance outlays". Responding to the comments of the reviewers (American Review of Public Administration).

declared. These findings may help inform discussions on how infrastructure disaster policy could be reformed to both increases the equitability of federal PA outlays as well as better incentivize local resilience efforts.

Keywords: *FEMA's Public Assistance, public infrastructure, repetitive flood losses, statistical learning theory*

2.1 Introduction

Since 2005, the United States has averaged more than 65 disasters[2] each year that rise to the level of requiring federal aid and intervention (U.S. Government Accountability Office, 2016). These events, which range from floods to tornadoes to chemical spills, are declared disasters by the president (i.e., presidentially-declared disasters or PDDs), and are deemed so significant that there is little reasonable expectation that either the locals or the state can coordinate or finance the recovery. In 2020, the US witnessed 22 billion-dollar plus weather disasters according to the National Oceanic and Atmospheric Administration, the most on record (National Centers for Environmental Information, 2020). The US government spent over $180 billion across all federal programs for expenses related to Hurricanes Katrina and Sandy alone – a cost of nearly $600 per American (Plyer, 2016; Weidman & Weiss, 2013).

[2] The term "disaster" is used very specifically in this work. We use this term to mean a hazardous event that was declared a disaster by the president. These disasters are deemed so significant that they are beyond the capacity of local and state governments, and warrant federal intervention.

While aid can come from numerous federal programs, following most disasters, the majority of the aid comes from the Federal Emergency Management Agency (FEMA) Public Assistance (PA). PA primarily supports the repair and rebuilding of publicly-owned and, on occasion, private, non-profit infrastructure, and partially reimburses locals for emergency management and debris removal expenses. Between 2000 and 2019, more than $125B (adjusted, 2020 dollars) has been awarded through PA.

In this work, we develop statistical models to explore what factors, beyond hazard severity and local capacity, are influential in the PA outlays received by communities to assess if the aide is distributed equitably, or at least evenly. This can advise FEMA and other federal disaster assistance agencies on if/how to target their programs to address systemic inequities. PA will continue to be a central focus of disaster response, as scientists project climate change will increase the frequency and intensity of natural hazards. It is critical to develop assessment methods by which to ensure this assistance reaches everyone who qualifies, including those in traditionally underserved communities. By creating predictive models, we evaluate the influence of factors that are known to cause communities to receive more aid, such as hazard intensity, and also consider factors that are less explored, like the repetitive nature of hazards and the applicants' abilities to learn over time.

As a case study, we analyze all PA grants for five Midwestern states between 2003 and 2018 to see how the number of PA applicants and the number of PA projects at the county-level are tied to a variety of factors including the intensity of the hazard, local capacity, and prior disaster experience. We use parametric, semi-

parametric, and non-parametric statistical models to make these forecasts. The discovery of wholesale relationships between the number of projects and metrics associated with local capacity and experience with disasters could suggest that causes for public outlays go beyond hazard intensity and community impact. By better understanding the characteristics of communities capable of navigating the "system," we can better identify and address issues surrounding equity and the long-term viability of how disasters are currently financed in the US.

The repetitive nature of disasters in many regions allows individuals and local governments to learn and adapt over time (Crow et al., 2018). In many instances, this is beneficial; for example, while buyouts in flood-prone regions were once generally unpopular, many saw Hurricane Sandy as the watershed moment whereby its importance as a national priority was realized, which may lower long-term losses (Marino, 2018). We hypothesize, however, that adaptation and learning have the potential to increase government spending on disasters as local governments and other organizations eligible for relief become more adept at applying for grants. We hypothesize further that additional factors influence the level of aid that is received locally, including local capacity and regional policies. Many researchers have found that local capacity influences disaster preparedness which, in turn, correlates with local outcomes (Cigler, 2007; Cutter et al., 2010; Norris et al., 2008). Recent work also suggests that more emphasis in higher capacity jurisdictions has been placed on receiving federal grants (Smith et al., 2013). While there are obvious local benefits to receiving more aid, it also creates broader societal questions about the opportunity

costs of this money, which communities are not getting this aid, and how demand for aid could increase in the future as more hazards occur.

We view this work as contributing to the literature in three distinct ways. First, this is among a few studies to explore FEMA's PA program with depth, and the first to our knowledge, that uses information about states and counties and the floods they experience to predict their subsequent level of federal disaster aid, and the degree to which it was administered evenly. Second, along this vein, we pay particular attention to the role that learning and capacity have on disaster aid for infrastructure. While several studies have found a relationship between capacity and federal intervention, the studies are limited to FEMA's Hazard Mitigation Grants, FEMA's Community Rating System, and the National Flood Insurance Program (Landry & Li, 2012; Sadiq & Noonan, 2015a; Smith & Vila, 2020). Finally, this work highlights the usefulness of statistical learning theory in the disaster aid and policy space in making non-linear inferences on how various disaster attributes, including hazard intensity and experience with past disasters, influence disaster aid. While the use of statistical learning theory is becoming increasingly common in the area of infrastructure and disasters (e.g., (Shashaani et al., 2018), statistical learning theory is virtually absent in the federal disaster aid space, which tends to favor parametric modeling.

The next section introduces PA and provides a brief overview of relevant literature that focuses on salient features known to influence disaster aid. The case study data are described in Section 3, along with an overview of the methodology. The results are presented in Section 4, and concluding remarks in Section 5.

2.2 Background

2.2.1 FEMA Public Assistance

Many disasters are beyond the capacity of a local and state government to adequately respond. In these situations, the event is designated as a PDD, which, under the Stafford Act (The Robert T. Stafford Disaster Relief and Emergency Assistance Act of 1988, 1988), authorizes the use of federal resources for recovery. States and counties are then designated disaster areas, allowing FEMA to provide aid in the form of response coordination and financial assistance. The work presented here focuses on financial assistance, and specifically, assistance in the form of PA. PA provides reimbursement of (at least) 75% of expenditures incurred by locals related to emergency protective measures, debris removal, and recovery of public infrastructure (called "permanent work"). This includes the repair of bridges, roads, and schools. Occasionally, depending on the severity of the disaster, some private not-for-profit infrastructure qualifies as well for PA. Locals, which could include school districts, state agencies, and rural utility cooperatives and are henceforth called "applicants," must apply for these grants and are responsible for financing the remaining 25% of expenditures. Applicants usually file multiple project claims to reflect instances of damage to multiple assets. Only direct losses (i.e., losses stemming from damaged infrastructure or assets and administrative costs) are eligible for reimbursement.

There has been significant criticism of this program, in that it does not encourage or reward planning or mitigation (Kousky & Shabman, 2012; Tonn et al., 2021). Further, though with some exceptions, the funding must be used to rebuild the

infrastructure to its pre-hazard condition. Moving a structure to a less hazardous area is generally disallowed though grant recipients are now allowed to rebuild to "relevant consensus-based codes, specifications, and standards that incorporate the latest hazard-resistant designs" (Section 406 of the Stafford Act, as Amended by Section 1235(b) of the Disaster Recovery Reform Act of 2018, 2018).

Unlike the extensive inquiry into other FEMA-run programs (e.g., National Flood Insurance Program and Hazard Mitigation Grant Program (see: (Bin et al., 2017; Fan & Davlasheridze, 2016; Godschalk et al., 2009; Rose et al., 2007; Strother, 2018)), the research body of knowledge on Public Assistance is limited. The programmatic analysis conducted by the US Government Accountability Office (GAO) and the Congressional Research Service has provided the most breadth. 2009 and 2015 Congressional Research Service reports both summarize the disaster declaration process and the PA program (Brown & Richardson, 2015; McCarthy, 2009). In the latter report, descriptive summary statistics on PA are provided - the report is among the first to do so. A 2021 report again summarized the program, and also legislative changes made since its initial 2009 report. It notes concerns about project execution delays, whether limited capacity applicants are able to navigate the PA system, and the program's widening scope - PA was a primary mechanism for early COVID-19 relief (Lee, 2021). The GAO has focused primarily on the program's escalating costs, stemming from worsening hazards, program structure, and administrative costs, and has made recommendations to make the program more fiscally sustainable (U.S. Government Accountability Office, 2012, 2014). One recommendation was to increase the threshold that warrants federal disaster

intervention; FEMA provided notification of their plans to adopt this recommendation in late 2020 (*Cost of Assistance Estimates in the Disaster Declaration Process for the Public Assistance Program*, 2020). While this plan has some potential to reduce FEMA's expenditures, albeit minorly (Kousky et al., 2016), there has been no evaluation on whether it will encourage new resilience actions by local and state governments or how it may affect funding distribution among communities.

While more emphasis in recent has been placed on procedural equity following disasters and how this intersects with US federal aid, this body of work primarily focuses on FEMA's Individual and Housing Assistance program and Hazard Mitigation Grant Program. Indeed, higher capacity and resourced communities are consistently found to access more federal resources and fare better during the recovery process (Finch et al., 2010; Muñoz & Tate, 2016). For some federal programs, especially mitigation programs, this may depend on the local's ability to come up with a cost-share - a known impediment for low-resourced communities (Martín & Williams, 2021). Domingue and Emrich (2019) focus solely on PA, the first to do so, and resource equity following the 2012-2015 major US disasters. While they find inequities in funding distributions, they were unable to identify consistent vulnerability indicators to explain these distributions across the years. Thus, significant questions remain about the multibillion-dollar PA program and its contribution to national resilience, from how pre-disaster mitigation affects post-disaster aid to how to measure the local capacity to know which regions are in actual need of federal assistance to rebuild public infrastructure.

2.2.2 Salient factors influencing disaster losses and aid

There have been significant efforts in developing methodology for predicting disaster losses given a hazard's intensity, measuring the accuracy of loss data put forth by federal agencies (Downton & Pielke, 2005), and quantifying secondary effects of natural disasters (Cochrane, 2004). Botzen et al. (2019) present a timely review of this work. Methods for predicting hazard losses tend to be either bottom-up approaches (e.g., (Kreibich et al., 2005; A. C. Reilly et al., 2017)), which typically rely on surveys or probabilistic methods that overlay hazard intensity with building fragility, or top-down approaches (e.g., (S. D. Brody et al., 2008; Cutter & Emrich, 2005)) which use prior data, economic models, and analytical tools to forecast damage and losses. These methods are all critical for understanding the economic and personal toll that disasters have on communities and how this may rise under climate change and widening economic stratification.

While not a perfect proxy for direct losses, there is a body of work on predicting National Flood Insurance Program (NFIP) claims following floods. (The NFIP is an insurance program managed by FEMA, though policies are purchased through private companies. Not everyone who is required to have flood insurance has a policy, and in some regions, the penetration rate is quite low (Michel-Kerjan et al., 2012)). Czajkowski et al. (2013), Czajkowski et al. (2017), and Tonn and Czajkowski (2018) all primarily rely on flood intensity and risk indicators, and NFIP penetration rates as predictors of claims. Mobley et al. (2021) use machine learning approaches that combine NFIP claims and topographic and hydrological indicators to better forecast flood hazard maps. Other recent work found that race and income tend to be

key predictors of whether a household would be likely to have a policy in place (Atreya et al., 2015; Dixon et al., 2017; Knighton et al., 2020).

Compared with the literature that focuses on translating the impacts of a hazard on direct (and indirect) losses, the research that forecasts federal aid as a function of hazard intensity and other salient factors is notably sparse. When it comes to forecasting the level of government aid, it is not as obvious that the same top-down and bottom-up appraisals discussed above are valid, meaning disaster aid may not be proportional to losses - though this has not been investigated with significant depth. A lack of data transparency from all government agencies involved in the recovery process (e.g., HUD's Community Development Block Grants - Disaster Recovery, USACE), makes this process in some ways virtually impossible (A. Reilly et al., 2020).

In a landmark study, Howell and Elliot (2019) use federal aid to forecast wealth over time (and not the relationship between damage and aid), finding that regions that receive more aid are more prone to wealth inequality. The discoveries in this body of work, however, do not truly provide direct insight into the relationship between regional losses and PA grants, given how PA grants are for municipal infrastructure, that grants are given to organizations and not individuals, and that few denials are ever made (though it is not clear whether all entities eligible for assistance ultimately apply).

As with other federal grants, it is highly possible that PA levels depend on local capacity. The past decades of disaster science research has shown that factors related to local capacity play a key role in how communities fare after disasters and

adapt to hazards (Norris et al., 2008). Here, capacity refers to an ability to anticipate hazards, make informed decisions and implement strategies to address hazards, and then later evaluate whether those decisions and strategies were effective (Honadle, 1981). Generally, individuals and organizations (e.g., local governments) with more capacity engage in more risk-reducing activities (Sadiq & Noonan, 2015b). This would suggest fewer losses. At a regional level, higher capacity also portends higher NFIP penetration rates and a greater likelihood of participation in other federal programs focused on hazard reduction, such as the Community Rating System (Landry & Li, 2012; Li & Landry, 2018). Local governments with greater capacity also tend to place greater emphasis on procurement of disaster aid, making the influence of capacity on PA challenging to forecast without numerical analysis (Smith et al., 2013).

Finally, there is reasonable evidence that local, state, and national governments learn from their disaster experiences and that this could influence disaster outlays. May (1992) and Crow et al. (2018) deconstruct the collective and organizational learning exhibited by governments into a total of four orthogonal classes. Social and government learning - the highest orders of learning - focus on revisiting assumptions and policy goals, leading to more effective governing solutions to address chronic problems. Birkland (2006) and Crow et al. (2018) show that these types of learning are uncommon after extreme events, but possible. Instrumental and political learning - the lowest orders of learning - are more common. These classes of learning focus on understanding intergovernmental bureaucracy and policy rules and could result in new internal processes or strategies for working within the federalist

system. Crow et al. (2018) found these to be far more common in local governments after flood disasters. In fact, all jurisdictions surveyed after major flooding reported learning about PA processes and how to more effectively apply for reimbursements. This, in principle, could lead to larger PA outlays as jurisdictions become more adept at the reimbursement process.

2.3 Location, Data, and Methods

2.3.1 Location

In this study, we focus on presidentially-declared disasters in five Midwestern states: North Dakota, South Dakota, Iowa, Wisconsin, and Minnesota. This area experiences frequent flooding. Between 2003 and 2018, counties in North Dakota had received the most disaster declarations for flooding with six, though only four counties have been designated disaster zones in all 6 PDDs (Figure 2.1). Other non-flooding PPDs occurred in this region during this period for hazards including severe storms and tornadoes. Again, counties in North Dakota had received the most disaster declarations in the study's time frame with 15. Numerous counties in Iowa had more than 10 PDDs, when considering severe storms and tornadoes in addition to floods. We focus on flooding disasters only to measure hazard intensity uniformly across counties. However, later, when we measure a county's experience with disasters, all PDDs that the county experiences are included.

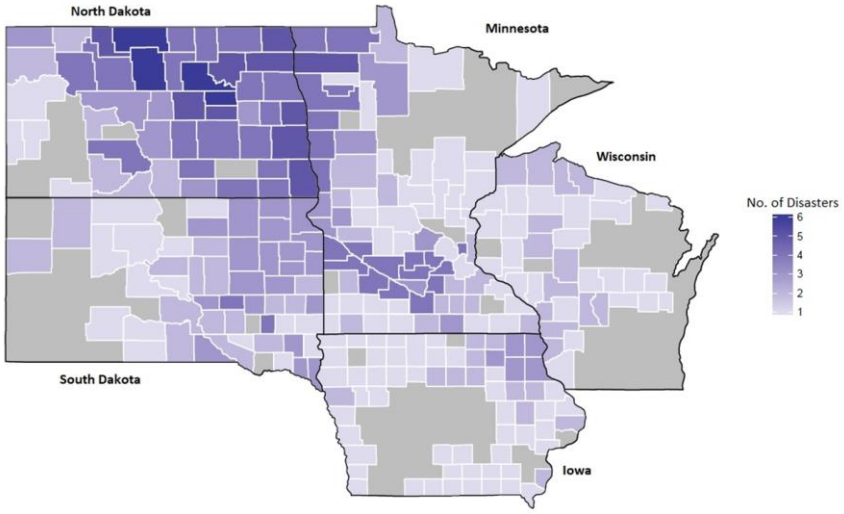

Figure 2.1. The number of PDD flood disasters in each county between 2003 and 2018.

2.3.2 Data

We use the FEMA PA database to conduct our analysis (Federal Emergency Management Agency, 2021). This database provides project-level data, and contains information on the type of project (e.g., roads and bridges), the applicant's address and county, the amount of money obligated to the applicant along with the applicant's contribution, and a brief description of the need. For each PDD, we count the number of applicants and projects in each county and these serve as our dependent variables, as an applicant could apply for multiple projects. In total, there are 566 PDD-county observations in our dataset.

Figure 2.2 shows the empirical cumulative distribution function plots for the number of applicants and projects per county per flood PDD, respectively factored by states. Iowa generally has the fewest applicants and projects for flood PDDs. North Dakota similarly has among the fewest applicants per county per disaster.

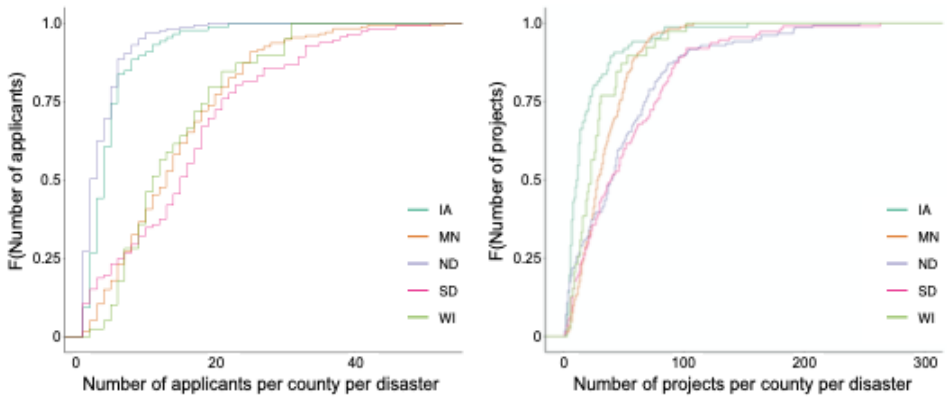

Figure 2.2. a. The empirical cumulative distribution function for the number of applicants per county, divided by state. b. The empirical cumulative distribution function for the number of projects per county, divided by state.

We consider a variety of explanatory variables that could help to explain the number of county-level applicants and projects. These variables are divided into five broad indicator categories: Hazard and geographic indicators, Consequence indicators, Human capital and socioeconomic indicators, Policy indicators, and Experience indicators. The subsections below explain the data that are used. Some of the variables may be relevant in multiple categories, but are still assigned to only one category for simplicity. A list of the variables, along with their summary statistics is provided in Supplementary Table A.1.

***Hazard and geographic indicators*.** The severity of the hazard, the land cover of a region, and the level of moisture in the soil are all known to be associated with damage (Burby et al., 2000; Nateghi et al., 2014). We use the 10-year flood peak ratio (10-year FPR) and precipitation to measure hazard severity (Villarini, Smith, Baeck, Marchok, et al., 2011). The 10-year FPR for a specific streamgage is the ratio between its maximum instantaneous discharge during the flood and the 90th percentile of annual maximum instantaneous discharge. It measures the intensity of conditions during the flood relative to other unusually bad conditions. Czajkowski et al. (2013)

found this ratio to be highly predictive of the number of NFIP claims. Because some of the PDDs have a duration of a few weeks to a couple of months, we use the maximum discharge during the PDD incident window. We then spatially interpolate these ratios over the region using the Inverse Distance Weighted interpolation method (Lam, 1983). After mapping, the fraction of each county (by size) with 10-year FPR greater than 0.5 and 1.0 is calculated along with the maximum value of flood ratios observed in the county during the disaster incident period.

To account for precipitation, we first geolocate the National Weather Service weather stations in each county. There is often more than one in each county. We then collect 24-hour cumulative rain totals for each weather station for each day in the disaster incident period, and for each weather station, we compute the 3-day and 5-day cumulative maximum (Chamberlain et al., 2016). Then, four metrics are recorded: two that average the maximum 3-day and 5-day rainfall over all stations in a given county, and two that find the maximum of the maximum 3-day and 5-day rainfall over all stations in a given county.

Soil moisture influences how much water the ground can absorb during a flood. We examine the soil moisture at the start of each disaster's incident period (for each county) to provide a consistent baseline across events at depths of 10cm and 100cm. These data are collected from Giovanni, NASA's open data portal (Acker & Leptoukh, 2007), and a single value is reported for each 0.125 degree grid (approximately 12.5 km) in the US for each depth. To get a single value for a county for each incident and depth, we overlay the grids with county boundaries and conduct a weighted average over the county's area to find county-level soil moisture.

Land cover has an important role in determining flood intensity and losses. Highly developed areas with little pervious surface are more prone to flood losses than regions with more pervious surface (S. D. Brody et al., 2007). We use the 2011 USGS National Land Cover Database (Multi-Resolution Land Characteristics Consortium, 2021a), which divides land cover into 21 classes, and record the fraction of each county that is in each land cover class. Consistent with Nateghi et al. (2011), we consolidate land cover classes to be the following eight aggregated classes: water, developed (including residential, commercial, and industrial), barren, forest, scrub, grass, pasture, and wetland.

***Consequence indicators*.** Severe PDDs are likely to mobilize more federal resources than PDDs that are less severe. As a proxy for severity, we compute the time in days between the disaster incident end date and the date when the president declared a disaster. Sometimes, this occurs during the incident period, which often signifies that the disaster was especially severe. These instances are assigned a value of 0. Otherwise, the PDD comes after the incident window and the value is positive. Langabeer et al. (2012), to some degree, explore the importance of this variable, but more from a political posturing sense. Disasters that result in fewer losses may take longer to assess whether the event rose above the threshold to be considered a PDD, in part due to fewer personnel assigned to the event. When delays occur, it could prompt local mobilization to tally losses if the capacity exists. The longest lag in the data set is 74 days. We considered including county-level loss data from the Special Hazard Events and Losses Database for the U.S. (SHELDUS), though multiple flood

PDDs reported no losses, though FEMA's Preliminary Damage Assessment reported higher impacts.

Human capital and socioeconomic indicators. County-level losses are expected to be somewhat dependent on population; larger populations tend to have more infrastructure and assets, and thus more potential for exposure. To account for this, we include the county-level population in 2010 and the number of county-level businesses. However, infrastructure needs vary based on both population size and whether the region is urbanized or not. For example, a large, rural county may require the same number of fire stations as an urbanized county to account for size disparities. To capture this phenomenon, we include the number of fire stations in the county (U.S. Fire Administration, 2021), the number of schools (The National Center for Education Statistics, 2021), and the county's Urban-Rural Continuum code (Economic Research Service, 2020). Fire departments and school districts are eligible for PA. These help to proxy the total number of organizations that could apply for PA in a county. The Office of Management and Budget Urban-Rural Continuum code classifies each county in the US into one of nine categories, based on its population, whether it is included in a metro area, and its proximity to other rural or urban counties.

Disaster theory has long demonstrated that communities with fewer resources fare worse than communities with more (Cutter et al., 2008; Freeman, 2000; Noonan & Sadiq, 2018). We control for this by using county-level median household income, the fraction of the county at or below the poverty line, the fraction of the county that has at least a bachelor's degree, and the fraction of the county population that is non-

white (U.S. Census Bureau, 2021). We use Census data from 2010 which is close to the midpoint of our study period and make an implicit assumption that these indicators do not vary significantly from year to year.

As mentioned earlier, local governments are responsible for financing up to 25% (i.e., the cost-share) of expenditures through PA, which can be a significant hurdle for low-resourced communities (Martín & Williams, 2021) to receive much-needed disaster aid. Lalancette and Charles (2022) reinforced the importance of sufficient financial resources in hazard response by identifying the variables measuring the revenue of municipalities as significant determinants in whether hazard management appeared in the priorities and the number of responses of the local government. To capture the fiscal health of the counties before a disaster occurs, we include the total revenue per capita of the prior year. The data on the total revenue is obtained from U.S. Census Annual State and Local Government Finances and the total revenue consists of intergovernmental revenue, taxes, charges, liquor store revenue, utility revenue, and social insurance trust revenue (U.S. Census Bureau, 2021). For years where revenue data is not available, we use linear interpolation between years with available revenue data.

Policy indicators. Local, regional, and state policies and procedures may influence who applies for aid, especially if some states have additional personnel in place to guide applicants, or if additional resources are made available to applicants during the process (Smith et al., 2013). To account for possible state and regional discrepancies in the level of support that is offered to locals, we include binary variables to account for both the state and the FEMA regions in which the county

resides. FEMA divides its local administrative duties into ten regions across the US, which work with locals to prepare for and recover from disasters. The study area includes three FEMA regions: Regions 5, 7, and 8. Schmidtlein et al. (2008) found the FEMA region to be a potentially important indicator of how adept a region was at receiving a PDD. In addition, we include binary variables indicating the year of the disaster occurrence to capture policy changes through the years (although we found them not to be significant indicators in our modeling).

Mitigation and preparedness reduce losses, and would therefore reduce the number of applicants and projects following a PDD. To measure preparedness, we include the miles of levees in the county in 2019, whether the county participated in the Community Rating System (CRS) between 2011 and 2015, and whether at least one jurisdiction in the county participated in the CRS between 2011 and 2015. We use the National Levee Database to identify cumulative levee length in a county (U.S. Army Corps of Engineers, 2016). There are more than 3,000 levee structures in the five states in the investigation. Unfortunately, the data are reported for 2019 and some of the structures may have not existed during the time period of the PDD. However, the average age of levees in the United States is over 50 years old, which lessens this concern (American Society of Civil Engineers (ASCE), 2017).

The CRS is a voluntary incentives-based program administered by FEMA in which communities participate to reduce flood insurance rates. Highfield and Brody (2017) discovered that participation in the CRS led to a 41% reduction in flood insurance claims. An entire county can participate in the CRS or any of its jurisdictions. While ideally, the data would ascertain whether the county or at least

one jurisdiction participated in the CRS during the PDD, data limitations prevent this. Of the 270 counties that we consider in our study, only 11 counties have at least one jurisdiction participating in the CRS between 2011 and 2015.

Experience indicators. We hypothesize that counties that experience repetitive disasters gain experience when applying for PA. This experience could include knowledge of the application process and who to contact for application assistance. On the other hand, one could hypothesize that when counties experience repetitive disaster losses, the quality (of some) of their infrastructure is higher as it was recently repaired or replaced. (Applicants are generally allowed to use PA to repair or replace, as needed, infrastructure up to current code.) To capture experience, for each PDD at the county-level, we compute the number of years between the PDD and the county's previous disaster and flood disaster (two metrics) and the total number of disasters they experienced in the previous five years and the previous twenty years (an additional two metrics).

2.3.3 Methods

We use statistical learning theory to construct fully-validated models that predict the number of PA applicants and projects after a disaster. Statistical learning models include a wide set of unsupervised and supervised learning models, and each has advantages in terms of complexity and interpretability (Obringer & Nateghi, 2018). Unsupervised methods (e.g., clustering) find patterns in unlabeled data. Because we are trying to predict the number of claims and applicants, which makes

our data labeled, our work leverages supervised models. Supervised learning models or predictive models,

$$Y = f(X) + e$$

make inferences about a function, f, that makes predictions about a response variable, Y, using a vector of predictor variables, X. X consists of the data reported in Section 3.2, with each observation consisting of one county's experience during one PDD. The objective is to find the functional relationship that minimizes the difference, e, between the predicted values, Y, and the observed values (James et al., 2013). In this work, we evaluate this functional relationship using a variety of parametric, semi-parametric, and non-parametric methods, and then identify which method results in the best predictions. Parametric models (e.g., generalized linear models) are highly interpretable, but their strong parametric assumptions are often unsuitable for highly non-linear data, which we have (Hastie et al., 2009). Non-parametric models, on the other hand, are flexible and nicely accommodate non-linear data, but are often hard to interpret (Hastie et al., 2009).

The variance inflation factor (VIF) is computed to evaluate multicollinearity among the covariates. While highly correlated features do not influence the predictive power of the model, they do potentially impede model interpretation. To eliminate this issue, features with high VIF scores are removed until all features have a VIF value below 10 (Alin, 2010).

To estimate the number of applicants and projects and to identify their most important predictors, we tested and evaluated the predictive performance of eight types of models. These models are: Bayesian additive regression trees (BART),

classification and regression tree (CART), generalized linear model (GLM), gradient boosting machines (GBM), multivariate adaptive regression splines (MARS), neural networks (NN), random forest (RF), and support vector regression (SVR) (James et al., 2013). These were selected because they are most suitable for interpreting the relationship between the variables and the response variable (J. Bennett et al., 2020). A brief summary of these methods is provided in Appendix A. Of these eight models, RF proved to be the best model for predicting both the number of applicants and the number of projects, so a detailed description is provided in the following section.

Random Forest (RF). RF is a nonparametric model that improves accuracy and stability over traditional tree methods by combining multiple trees (Breiman, 2001). More specifically, an RF model is an ensemble of regression trees. Regression trees recursively partition data based on a threshold that is selected to enhance the model's predictive power. While regression trees usually capture the structure of the data easily (i.e., low bias) - especially if the tree is allowed to have many branches - they are sensitive to outliers (i.e., high variance). A model with a lot of branches is also prone to overfitting the data. RF models address both these issues. RF models work by sampling N bootstrapped observations from the dataset to create a regression tree and using the remaining data to evaluate the model in terms of its predictive power. The training data are split at $m << M$ nodes, where m is predefined and M is the number of covariates. This prevents the tree from growing too large, and thus overfitting. The tree is then evaluated using the withheld data. This is repeated K times and as a result builds K regression trees. The model's prediction is the unweighted average prediction over all K regression trees.

Model Assessment. We determine the best model by measuring their out-of-sample predictive accuracy through a validation process. The models are validated using a *k*-fold cross-validation approach, with $k = 10$ in our case. The idea is to divide the data into *k* partitions or folds of approximately equal size. Ultimately, *k* models are iteratively created by withholding one partition in each iteration and building a model with the remainder of the data. This process is repeated until each of the partitions has been withheld once. The models are evaluated on the withheld data using an out-of-sample mean-absolute error (MAE) metric and a root mean-squared error (RMSE) metric. This entire process was repeated three times (i.e., with new partitioning). The metrics are averaged over all iterations for each model for comparison and also ultimately compared to the MAE and RMSE of the null model. The null model is simply the average over the response variables. A lack of improvement over the null model implies that the statistical models provide no more insight than simple averaging. Additional cross-validation approaches are investigated including event-based and year-based cross-validations. The results show lower model performance than the above repeated k-fold cross-validation, therefore they are not used in this study (see Table A.2).

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

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Variable selection. As the results section later reports, the RF model performed the best. However, this model includes numerous predictor variables, which can lead to inaccurate inferences and overfitting (J. B. Bennett, 2019). To address this, we use variable selection to develop a parsimonious model that still has high out-of-sample predictive accuracy. The idea behind the variable selection is that we want to keep the predictor variables that contribute most to the model's out-of-sample predictive accuracy and cull the rest. In some instances, this is as straightforward as evaluating and ranking each predictor variable's importance in terms of how much worse the model accuracy is without it, and then removing variables that do not contribute much to the model's predictive accuracy. However, with RF and other stochastic models, the process of variable selection is more involved. The issue lies with model stability; two tree ensembles grown using the same data will likely differ in the end, and as a result, the variables' rankings may not be consistent.

To address this, we used the VSURF heuristic technique for variable selection on RF models (Genuer et al., 2010). This heuristic works by repeatedly building RF models with all the variables and for each model, ranking the importance of each predictor variable in terms of its contribution to out-of-sample predictive accuracy. The standard deviation of each variable's rankings is then computed, with the idea being that noisy variables will have standard deviations close to zero (i.e., they are usually not important). A predictive CART model is then built to forecast the standard deviations of the variable importance, and the model's minimum prediction becomes a cut-off threshold. Variables with the highest average ranking and standard

deviations above this threshold are included for further consideration. The heuristic then makes nested models, starting with the most important variable (on average), and continues adding variables one at a time based on rank until the effect on out-of-sample predictive accuracy is negligible. Our final RF model uses only variables identified by VSURF to be useful.

Statistical inference. The ultimate goal of the model is to be able to use it to make inferences between the response and predictor variables. As opposed to parametric models, nonparametric models are more difficult to interpret. They do not have a well-defined functional form with interpretable parameters. To overcome this hurdle, we rely on partial dependence plots to characterize the relationship between single predictor variables and the response variable. A plot is made for each predictor variable used in the final model and shows the effect of the predictor variable on the response variable while the remaining predictor variables are averaged:

$$\bar{f_s}(X_s) = \frac{1}{N}\sum_{i=1}^{N} f(X_s, x_{iC})$$

where $X_s$is the predictor variable that we are interested in, $x_{iC}$are the other predictor variables, $\bar{f_s}(X_s)$ is the final RF model (Hastie et al., 2009). We also include the 95% confidence intervals on the partial dependence plots.

2.4 Results

Model Performance Assessment. After feature reduction to reduce multicollinearity, the final model included 40 covariates, down from 56 originally.

The highest VIF after feature selection was 9.5, though 90% of variables had a VIF below 5. Notable eliminated variables included the maximum average 3-day county-level precipitation, which was highly correlated with the maximum average 5-day county-level precipitation, and the number of schools, which was highly correlated with county-level population. For the models that predict the number of projects, we trained the models, identified outliers, and retrained the models with the outliers removed. There was one outlier; the City of Minot in Ward County, North Dakota, received 83M USD for flood relief in 2011 to cover 824 projects - about four times the next highest observation.

Table 2.1 shows the average, across all hold-out analyses, of the in-sample and out-of-sample MAE and RMSE and R^2 for the nine learning models, including a null model, for predicting the number of county-level *applicants* for a PDD. The random forest (RF) model outperformed other models in terms of its average R^2, MAE, and RMSE values. It outperformed the null model by 50% for the MAE. Table 2.2 shows the average, across all hold-out analyses, of the in-sample and out-of-sample MAE and RMSE and R^2 for the nine learning models, including a null model, for predicting the number of county-level *projects* for a PDD. Again, the RF model outperformed other models and led to a 32% improvement over the null model. However, in this case, the predictive performance is less than when predicting the number of applicants, suggesting that there is either more variability in the number of projects than the number of applicants or there are salient features that help explain the number of projects that are not captured in our models. These results are statistically significant at the 5% significance level except for MAE between the RF

model and the SVR model when predicting the number of county-level projects. However, the correlation between the predicted and the observed values of the RF model (97%) is much higher than for the SVR model (79%), showing that the RF model has superior performance. It is perhaps not surprising that RF outperformed other models. RF is a highly flexible model that is able to capture the structure of non-linear data. It tends to be robust to outliers because of how it averages predictions over its trees (Hastie et al., 2009).

Table 2.1. Model results forecasting the number of applicants.

	In-sample		Out-of-sample			
Model	**MAE**	**RMSE**	**MAE**	**RMSE**	R^2	**% Improvement over null model (MAE)**
BART	4.4	6.0	7.0	8.9	0.32	7%
CART	3.3	4.6	4.2	6.0	0.61	44%
GBM	2.7	3.6	4.2	5.7	0.63	45%
GLM	4.7	6.4	5.1	7.0	0.47	32%
MARS	4.4	5.9	4.9	6.7	0.51	35%
NN	2.8	4.3	5.3	7.3	0.45	30%
RF	**1.5**	**2.2**	**3.8**	**5.5**	**0.67**	**50%**
SVR	3.1	5.4	4.3	6.5	0.56	42%
Null model			7.5	9.5	N/A	-

***RF Variable selection*.** VSURF was used to improve RF model stability and its predictive ability by eliminating variables that contribute less to error reduction. VSURF found 21 variables to be important for predicting the number of county-level applicants and 12 variables to be important for predicting the number of county-level projects. For both models, the predictive power of the RF model improved slightly (Table 2.3).

Table 2.2. Model results forecasting the number of projects.

	In-sample		Out-of-sample			
Model	**MAE**	**RMSE**	**MAE**	**RMSE**	R^2	**% Improvement over null model (MAE)**
BART	17.7	23.8	22.6	30.6	0.37	16%
CART	15.5	21.9	19.8	28.2	0.44	27%
GBM	11.9	16.7	19.0	27.0	0.48	30%
GLM	20.1	28.8	22.1	31.3	0.32	18%
MARS	18.3	25.7	21.1	29.3	0.39	22%
NN	14.2	20.7	23.8	32.6	0.30	12%
RF	**7.4**	**11.1**	**18.3**	**26.2**	**0.53**	**32%**
SVR	13.8	25.2	18.4	28.7	0.45	32%
Null model			27.0	37.4	N/A	-

Table 2.3. Model improvements from VSURF.

	MAE		RMSE	
	Before	After	Before	After
Number of applicants	3.77	3.78	5.48	5.45
Number of projects	18.34	18.33	26.20	26.00

***Model diagnostics*:** Figure 2.3 shows the normal quantile plot (i.e., QQ plot) and the predicted versus actual plots for the number of applicants and the number of projects for the final RF models. For both RF models, residuals generally fall along the 45-degree line, suggesting that the models are capable of capturing the variance in the data. The Pearson's correlation coefficient for both models is high (97% and 97%). However, the plots show more error in the tails; they tend to under-predict instances where the actual number of applicants and projects are higher and overpredict instances where the number of applicants and projects are low. This suggests that some salient features may be missing. For example, in some instances,

though it is not clear whether it has occurred here, FEMA occasionally awards additional mitigation funds under PA. It is possible that some projects here are not for recovery but rather mitigation, and thus actuals are higher than we predict.

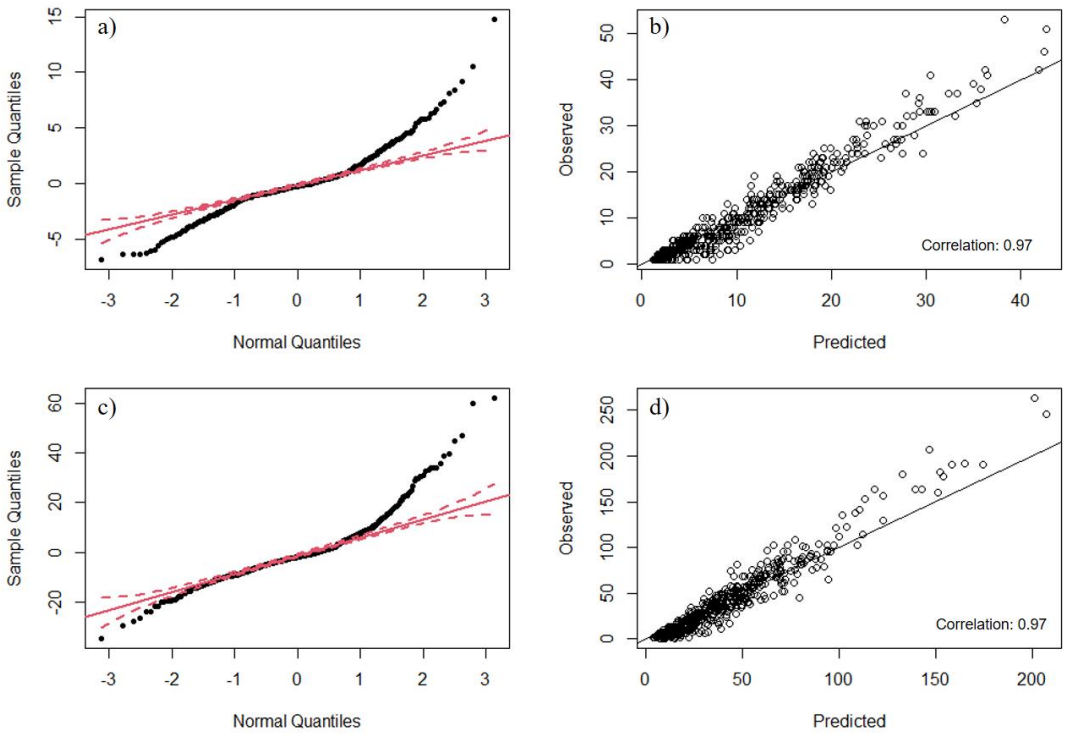

Figure 2.3. a. Q-Q plot for the final RF model for predicting the number of applicants. b. The predicted versus actual plot for the number of applicants. c. Q-Q plot for the final RF model for predicting the number of projects. d. The predicted versus actual plot for the number of projects. The red dashed lines on the Q-Q plot represent the 95% confidence interval.

Variable Importance and Model Inference. Figure 2.4 shows the variable ranking for both the number of applicants and the number of projects. The higher the variable importance, the greater the average decrease between the predictive accuracy of the model and a model that excludes that variable. This figure is the average over all trees in the final model.

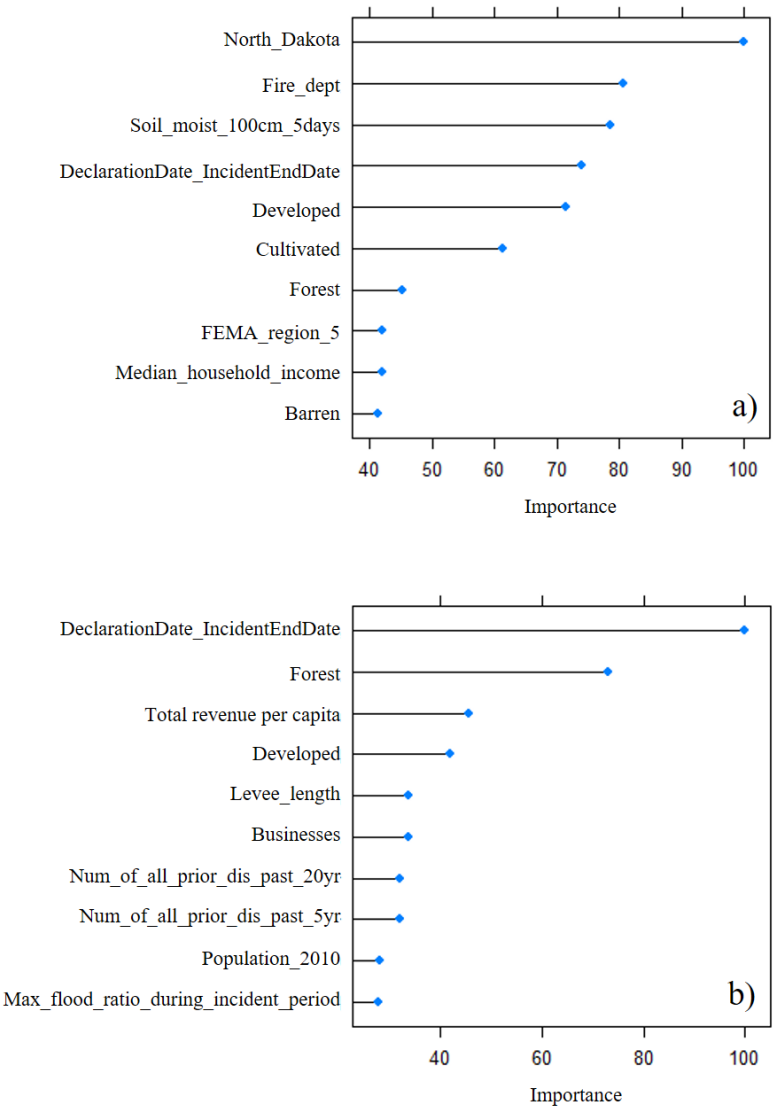

Figure 2.4. a. Variable importance plot for the number of applicants. b. Variable importance plot for the number of projects (only the top ten important variables are shown).

For the model that forecasts the number of applicants, the four most important variables are whether the flood occurred in North Dakota, the number of fire departments in the county, average soil moisture in the past 5 days, and the time between the end of the disaster and the time in which a declaration was declared. (A value of zero implies that the disaster declaration occurred while the flood was

ongoing.) For the model that forecasts the number of projects, the four most important variables are the time between the end of the disaster and the time in which a declaration was declared, the fraction of the county covered in forest, total revenue per capita, and the development. The time between the end of the disaster and the time in which a declaration was declared, the variable used to proxy consequences, is highly important for predicting both response variables, potentially suggesting it indeed proxies consequences. The flood hazard intensity metrics did not rank among the top ten important variables in either model, suggesting perhaps, that the flood metrics we used are unable to capture the complexity of floods over large geographic regions and time scales. We note that heuristics, as opposed to accurate yet complex hydrological models, that reflect actual flood hazard intensity over wide geographic and temporal scales are in their infancy (Czajkowski et al., 2017).

Figure 2.5 shows the relative influence that each of the ten most important variables has on the response variable representing the number of applicants - i.e., the partial dependence plots. The y-axis shows the average degree to which the response variable is influenced by the feature assuming all other features are held constant.

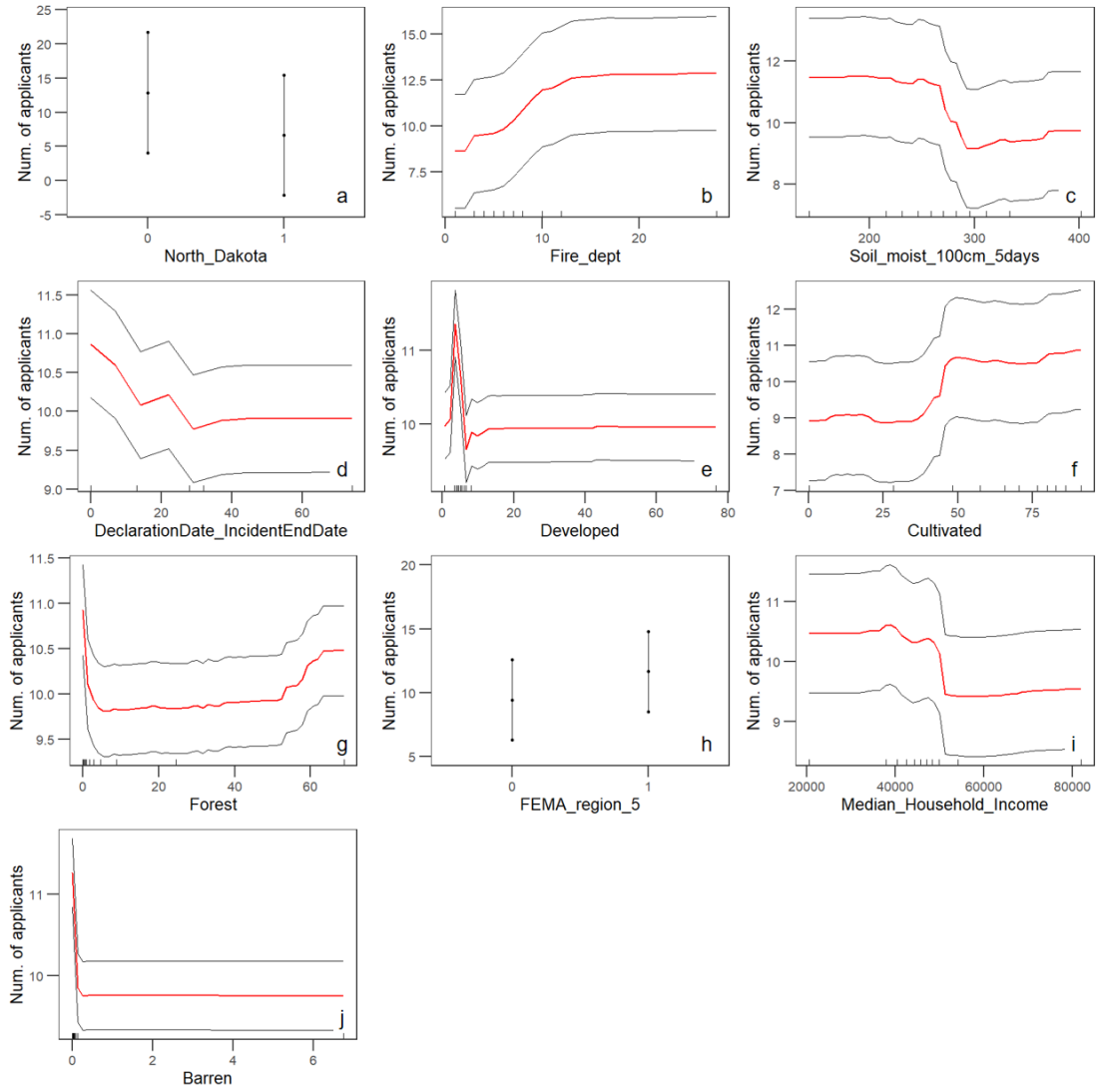

Figure 2.5. Partial dependence plots for the RF model forecasting number of applicants for the following variables: a. North Dakota, b. The number of fire departments, c. Soil moisture over the last 5-days before the start of the disaster at 100cm, d. The time between incident end date and declaration date, e. Fraction of the county land developed, f. Fraction of the county land that is cultivated, g. Fraction of the county land that is forested, h. FEMA Region 5, i. Median household income, j. Fraction of the county land that is barren.

When predicting the number of applicants, many of the key variables behave as expected. As the number of fire stations and businesses increases, so do the number of applicants. (Most businesses cannot apply for PA, but the metric captures a combination of population and density of infrastructure). The most important variable in terms of error reduction is whether the event occurred in North Dakota, which

appears to have about half the number of applicants as other states, holding other factors constant. This could suggest that despite having numerous floods, ND applicants have challenges working with the state and FEMA administrators in submitting applications. Alternatively, it could suggest that there is more centralization of public entities in North Dakota compared to other states (e.g., perhaps school districts are a part of local government and not separate public institutions).

The longer the period between the end of the disaster and when the disaster was declared, the fewer the applicants. This could reflect the fact that a longer period usually coincides with a less intense disaster (e.g., fewer FEMA personnel may be sent to do damage assessments, so the assessment takes longer, so the declaration is delayed). If this is true, this duration variable likely proxies consequences. Alternatively, this could reflect a possibility that in these situations of long durations, some entities that would have been eligible for aid did not retain necessary documentation for reimbursement because they did not know a federal aid would later be made available. This should be investigated further.

Interestingly, in terms of the importance of local capacity, county-level medium income was among the more important variables and there is a precipitous *drop* in the number of applicants once the median income reaches approximately $50,000. This could indicate that either states or FEMA send additional personnel to resource-poor counties to support PA applications. Alternatively, it could indicate that infrastructure in resource-rich counties is less vulnerable to floods, perhaps because the infrastructure was better maintained. While the marginal effect is small, there is a

slight increase in the number of applicants for the top decile of income counties, suggesting that wealthier jurisdictions do place additional emphasis on procuring grants.

It is challenging to interpret many of the land use variables. For example, the greater the fraction of the county land cover that is cultivated, the more applicants. There is a significant spike once the county reaches approximately 45% cultivated, and remains high once the fraction increases. Because the study area is in the upper-Midwest, where much of the land is cultivated, counties with low fractions of cultivated land may be particularly unpopulated (as opposed to being more urbanized, as you might expect in other areas of the US), suggesting less infrastructure and less need for aid after floods.

When predicting the number of projects, many of the same findings from above persist, including the time between the end of a disaster and the date the disaster was declared, the fraction of the land that is developed, and the fraction of the land that is forest (see Figure 2.4b and Figure 2.6). In fact, all other measured factors held constant, a two-month period between the end of the disaster and a disaster declaration resulted in a 33% reduction in the number of projects. However, unlike before, the variable “North Dakota” was not important - in fact, it was eliminated during the VSURF feature selection. This could suggest that of the entities in North Dakota that do apply for aid, they apply for more projects compared to applicants in other states. Alternatively, this may imply that North Dakota’s governance structure is indeed more centralized and that the number of projects may be a more representative metric than the number of applicants for understanding outlays. If the

centralization theory is not true however, there are potentially eligible applicants who receive no aid in North Dakota and FEMA may need to especially target some eligible applicants.

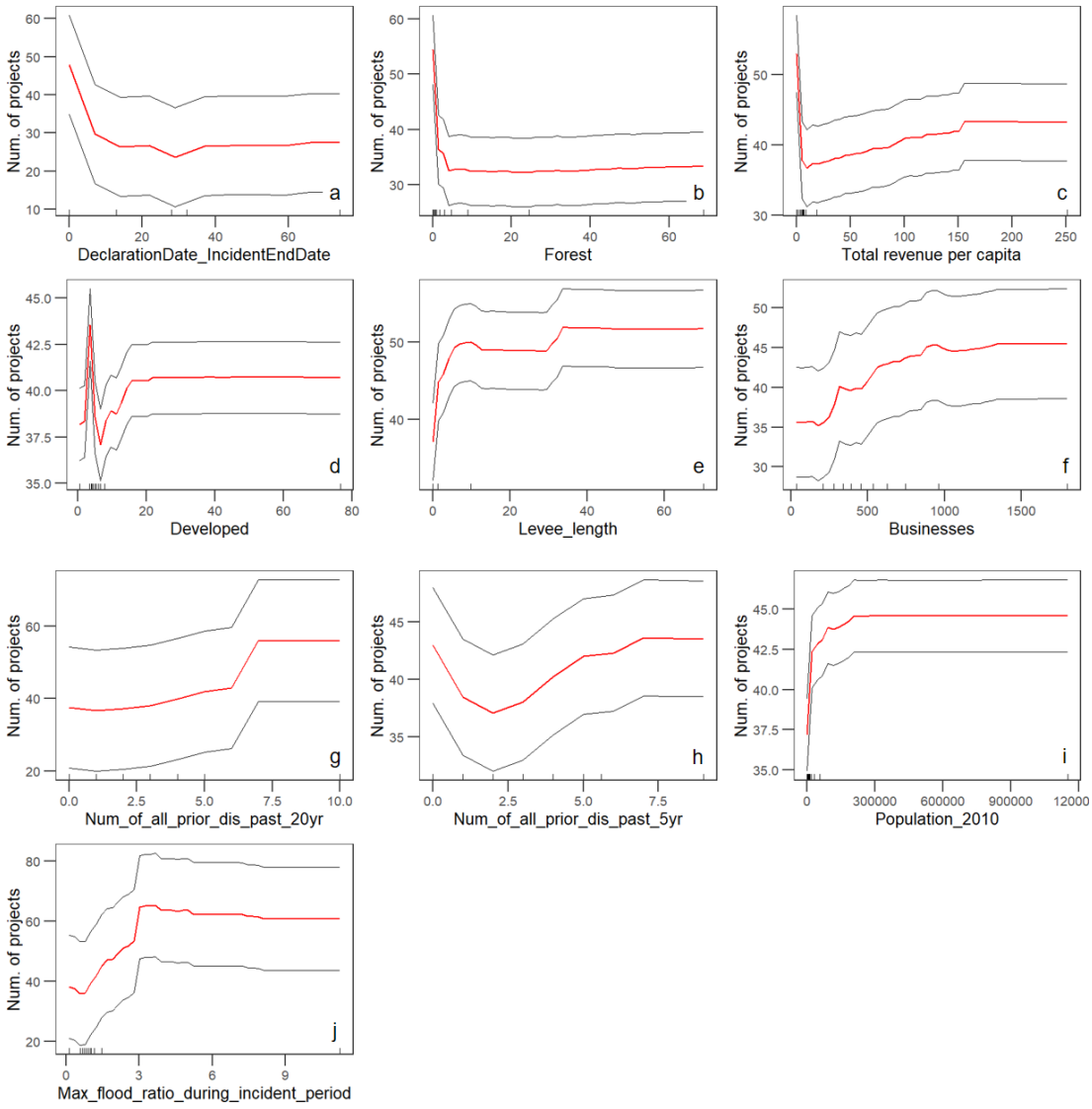

Figure 2.6. Partial dependence plots for the RF model forecasting number of projects for the following variables: a. The time between incident end date and declaration date, b. Fraction of the county land that is forested, c. Total revenue per capita, d. Fraction of the county land that is developed, e. Levee length, f. The number of businesses, g. The number of prior disasters in the last 20 years in the county, h. The number of prior disasters in the last 5 years in the county, i. County population in 2010, j. Maximum flood ratio during the incident period.

There are two variables that are among the top ten most important factors for predicting the number of projects but are eliminated during feature selection when predicting the number of applicants: the number of presidentially-declared disasters over the previous five years, and the number of presidentially-declared disasters over the previous twenty years. In the case of the number of disasters in the past five years, as the number increases from zero to two, the number of projects decreases. This could suggest that key infrastructure is relatively new and able to withstand new flooding. Interestingly, however, after two disasters, the number of projects that are expected rises quickly. For example, a county that has experienced seven disasters in the past five years is expected to have the same number of projects as a county that has experienced no disasters and 15% more projects than a county that has experienced two disasters. Similar patterns emerge when considering the number of disasters over the prior twenty years. This could suggest that as applicants learn over time how to apply for PA, and potentially get better at finding cost-share, and thus increase project submissions. Alternatively, though potentially less plausible, it may reflect that some communities rely on general funds to finance some projects after disasters as opposed to applying for PA; they may have found the process cumbersome in the past, reflecting the initial drop. However, multiple disasters in a row could drain a general fund and require them to apply for PA. Regardless, this relationship between past experience with disasters and projects should be explored in more depth, and could be a factor explaining the rising number of federal outlays that have not been identified previously.

2.5 Conclusion

Presently, little is known about how equitably or evenly public assistance is distributed amongst locals following PDDs. As the number of natural disasters is only expected to increase, FEMA will need to restructure the PA process to ensure that the limited quantity of federal outlays is effectively distributed amongst those in need. There are currently no metrics or indicators used to predict what factors are associated with historical receipt of aid, which limits the ability to objectively evaluate PA programs.

This work supports the burgeoning inquiry into this area by identifying the salient factors that are associated with receipt of aid in order to determine the degree to which the program is administered equitably, or at least consistently. It is clear that factors associated with population, such as the number of fire departments, and land use influence aid, as expected. This work found additional factors that are less tied to exposure and more tied to bureaucratic attributes and human behavior. For example, we found that North Dakota has fewer applicants, *ceteris paribus*, which could stem from having fewer state-level resources to support the acquisition of aid, less ability to find cost-share, or differences in the centralization of public entities in the state.

A longer time duration between the end of the disaster and the date on which a disaster is declared seems to lead to both marginally fewer applicants and projects. While this variable may be a proxy for a less intense disaster, it also could suggest that entities that are otherwise eligible for aid are not made aware of the availability of aid once a declaration is declared, that they have already recovered and lost

necessary receipts and paperwork to apply for aid, or that fewer resources are sent to these areas to support recovery and aid applications.

County-median income and total revenue per capita were the only capacity indicators that were identified as important. Specifically, counties with a median income below $50,000 have more applicants than more well-resourced counties. One explanation could be that FEMA or state officials provide additional personnel and resources to support PA applications in low-resourced counties. An alternative explanation is that low-resourced communities tend to be located in more hazardous areas, have fewer resources before the disaster to engage in mitigation, and have lower-quality infrastructure due to chronic underinvestment - and thus have more destruction and more need for aid after disasters. Similarly, there is a decrease in the number of projects as the total revenue per capita increases, though the number of projects increases for the counties with the highest revenue. This may suggest that the lowest-resourced counties require the most intervention from the federal government while the counties with higher revenue have the capacity to cover some of the losses by the local government. However, there is a significant increase in the number of projects for the counties with the highest revenues, which might indicate that the losses are significantly higher due to being the most developed counties in the study region.

An interesting next step could be to explore the relationship between median income and receipt of a PDD itself. Counties in the highest income decile tend to have marginally more applicants than counties in the middle deciles, perhaps suggesting that well-resourced counties do systematically focus on procurement of

aid (Smith et al., 2013). It also could suggest that well-resourced counties have more private non-profit entities which are eligible for disaster aid after some disasters (e.g., private hospitals and universities). Regardless, these findings bring up pressing questions about disaster aid and procedural equity.

Finally, it is worth discussing that the number of disasters that a county has experienced in the past is a strong predictor of how many projects they will have in future disasters, but is not a predictor for the number of applicants. The lack of influence from past disasters on the number of PA applicants suggests that entities are generally made equally aware of their eligibility regardless of their past experience, and that additional county-level experience is not advantageous here. There is a possibility, however, that additional county-level experience is advantageous in terms of the number of expected projects undertaken. After one or two disasters, the number of projects decreases and subsequently increases with more disasters. It is possible that initially, infrastructure is relatively new and robust, suggesting less damage and need for repair. After more disasters, locals may become better adept at applying for aid and finding cost-share. Alternatively, it could reflect FEMA's desire to support counties with excessive disaster experience by identifying additional projects that could be undertaken, perhaps mitigation projects, that could make them less susceptible to damage in the future.

If these patterns persist on a larger spatial scale across other disaster types, when paired with in-depth interviews of applicants to verify these findings, this type of analysis could support FEMA in better understanding the drivers of PA outlays that are beyond damage intensity and support their quest of ensuring resources are

delivered where they are most needed. These salient attributes are critical when envisioning how PA could be reformed to better build future resilience.

3 Chapter 3: Predicting flood damage using the Flood Peak Ratio and Giovanni Flooded Fraction[3]

Abstract

Currently, researchers who wish to predict flood losses or predict infrastructure reliability following a flood usually rely on computationally intensive hydrodynamic modeling or on flood hazard maps (e.g., the 100-year floodplain) to build a spatially-resolved understanding of the flood's intensity. However, both have specific limitations. The former requires both subject matter expertise to create the models and significant computation time, while the latter is a static metric that provides no variation among specific events. This work takes a data-driven approach to estimate the flood damages by implementing two emerging data-driven flood-intensity heuristics, namely the flood peak ratio (FPR) and NASA's Giovanni Flooded Fraction (GFF). The objective of this work is to develop an integrated data-driven method to rapidly predict continental-scale flood damage using FPR and GFF as flood intensity heuristics, along with other well-established flood exposure variables, such as regional slope and population. This study uses data on flood claims from the National Flood Insurance Program (NFIP) to proxy flood damage. We perform the analysis using predictive models at two spatial levels nationwide and statewide for the entire contiguous United States. A variable importance analysis demonstrates the

[3] Ghaedi, H., Reilly, A., Baroud, H., Perrucci, D., Ferreira, C. (expected 2022). "A rapid, data-driven approach for forecasting flood damage using the Flood Peak Ratio and Giovanni Flooded Fraction". Under Review (PLOS ONE).

significance of FPR and GFF data in predicting flood damage. Also, the model performance at the state level was much higher than the nationwide level analysis which indicates the effectiveness of both FPR and GFF models at the regional level.

3.1 Introduction

Accurately predicting infrastructure reliability or losses incurred by a hazard requires spatially-resolved knowledge about the hazard's intensity. For example, to predict electric-power reliability resulting from hurricane winds at a census tract level, researchers first parametrically downscale a hurricane's 6-hr track to a 3-sec peak wind gust for each census tract centroid (Holland, 2008). This can be done in a matter of seconds. Peak ground acceleration models similarly exist to rapidly predict localized peak ground acceleration for earthquakes, which can then be used to predict building damage, traffic disruption, etc. (Mangalathu et al., 2020; Thomas et al., 2017). The state of the practice is quite different for floods. Currently, researchers who wish to predict flood losses or predict infrastructure reliability following a flood usually rely on computationally intensive hydrodynamic modeling (Horritt & Bates, 2002; Merz et al., 2010) or on flood hazard maps (e.g., the 100-year floodplain) (Luke et al., 2018) to build a spatially-resolved understanding of the flood's intensity. Both have specific limitations. The former requires both subject matter expertise to create the models and significant computation time. The time limitation prevents comparisons over wide spatial and temporal scales (i.e., it is time-intensive to create highly spatially-resolved hydrodynamic models for the entire U.S. on an hourly, or even daily basis, though NOAA's National Water Model arguably is coming closest

to this promise) (*Office of Water Prediction*, 2016). The latter, flood hazard maps, are a good indicator of the area's general threat over a long-time scale, but are static. That is, while these metrics provide insight into hazard occurrence probability, they provide no variability among specific events and do not reflect, for example, the intensity of particular rainfall or snowmelt events. As a result, special flood hazard area (SFHA) flood hazard maps may underestimate the depth and extent of some actual floods. Blessing et al. (2017), for example, reports that, in at least two subbasins in Texas, the 100-year FEMA floodplain fails to adequately predict flood losses compared to more advanced hydrological models (e.g., distributed hydrologic models) for finding the 1% flood. Further, during Hurricane Sandy in New York, nearly 290,000 residents experienced unexpected inundation because they were located outside the SFHA boundary (Knowlton & Rotkin-Ellman, 2014).

Data-driven methods to *rapidly* estimate flood occurrence and intensity at localized levels are in their infancy, and have not been evaluated at wide spatial or temporal scales in terms of their ability to predict damage or infrastructure reliability. Two emerging data-driven methods are the flood peak ratio (FPR) (Villarini, Smith, Baeck, Marchok, et al., 2011) and NASA's Giovanni Flooded Fraction (GFF) (Rui, H., Teng, W., Vollmer, B., Jasinski, M., Mocko, D., & Kempler, S., 2014). The FPR is a ratio of the largest height of a streamgage's hydrograph (i.e., its instantaneous peak discharge) during a flood to the same streamgage's 90th percentile annual maximum instantaneous peak discharge (other reference percentiles may be used e.g., see (Czajkowski et al., 2017)). It measures the discharge intensity relative to past intense events in order to isolate the relative flood severity. The metric is found for all

streamgages in a region, and then a spatial interpolation method (i.e., an Inverse Distance Weighted (IDW) interpolation) is used to find the FPR for all locations in the study region. On the other hand, the GFF grids the entire US and relies on satellite imagery to report the fraction of each grid that is flooded. From this data, one can compute the fraction of spatial area (e.g., a county) that is flooded. The data to compute both the FPR and the GFF are publicly available and both metrics can be computed in the order of minutes for the entire US. The temporal resolution for the FPR varies between 15 and 60 minutes (depending on how frequently gages report their discharge value) and daily for the GFF.

The objective of this work is to evaluate the usefulness of both the FPR and GFF as data-driven heuristics for predicting flood damages. More specifically, we evaluate the adequacy of the FPR and GFF in terms of being able to predict National Flood Insurance Program (NFIP) claims for every flooding event in 2016 for the entire contiguous United States at the county level using statistical learning methods. We expect that in some locations, one metric will outperform the other. If one of these metrics proves useful, even for particular areas, it enables a new branch of flood hazard research in which researchers can use simplified, data-driven flood metrics to predict localized losses and infrastructure reliability over wide spatial and temporal scales without first developing complex hydrodynamic models. The researcher in this case, however, would be trading hydrodynamic accuracy for time savings.

This work builds on Czajkowski et al. (2017) which uses FPR to predict freshwater flood risk, proxied by NFIP claims, following 28 U.S. landfalling tropical cyclones between 2001 and 2014. The authors powerfully illustrate the potential uses

of the FPR as a flood-intensity indicator, and found a strong positive relationship between a FEMA community's FPR and the number of NFIP claims in that community. Czajkowski et al. (2017) use a zero-inflated negative binomial regression model to make predictions. The work presented here expands upon Czajkowski et al. (2017) in five ways: First, we develop a data-driven classifier in Stage 1 of our model that predicts whether a county had a flooding event intense enough to produce at least one NFIP claim. Thus, our work is not event-driven, meaning that we are not only attempting to find the number of NFIP claims conditioned on knowing an event, say a tropical cyclone, has occurred. Rather, our work is data-driven and "discovers" when a flooding event is significant enough to produce at least one NFIP claim at a county-level by using a fully-validated classifier. Stage 2 of our model then predicts the number of claims, given an event has been predicted. Second, we consider additional statistical learning techniques to better capture the relationship between our independent variables, including the FPR and GFF, and the number of claims. We evaluate models specifically in terms of their predictive accuracy, which is accomplished through cross-validation and evaluation of out-of-sample error reduction. Third, to expand the data-driven approach, we include additional independent variables that are known to influence flood intensity, including land cover and land slope (Bhattarai et al., 2016; S. Brody et al., 2014; Highfield & Brody, 2013). The data for these land variables is publicly available and the variables compute for any spatial scale. We note, however, that by doing this, we do not precisely mirror the variables considered in Czajkowski et al. (2017). Fourth, we build separate models for both the FPR and GFF metrics to identify which is most

influential in improving predictive accuracy, and in which regions of the county each is potentially more useful. Finally, we expand the spatial and temporal resolution to consider all flooding events in the contiguous U.S. for all of 2016. The start of a flooding event in a county is discovered using a classifier we construct using daily NFIP and FPR data between 2005 and 2015 to make a county-level binary prediction for whether at least one claim is made. It is then applied to 2016 data.

Neither the FPR nor the GFF has been extensively used in the literature. The FPR was initially developed by Villarini et al. (2011) as a way to make comparisons of inland flooding resulting from three tropical cyclones. The purpose of normalizing the highest instantaneous peak discharge during a flood by the 90th percentile (i.e., 10-year) annual maximum instantaneous peak discharge is to account for the fact that larger watersheds have larger drainage areas which are usually linked to larger discharge values (Gupta et al., 1994). Czajkowski et al. (2013) and Czajkowski et al. (2017) then use the FPR to predict NFIP claims to evaluate the extent of freshwater flood risk due to tropical cyclones both now and into the future. They found, for example, through regression models, that if the rate of urbanization were to keep pace with 2001-2011 levels, future flood insurance claims would increase by 2.4% - roughly keeping pace with the rate of urbanization.

The Giovanni Flooded Fraction (GFF) variable is available through the National Climate Assessment - Land Data Assimilation System (NCA-LDAS ver. 2.0) (Rui, H., Teng, W., Vollmer, B., Jasinski, M., Mocko, D., & Kempler, S., 2014). NCA-LDAS includes 42 variables including routing variables (e.g., flooded area), land-surface fluxes (e.g., precipitation), stores (e.g., soil moisture and snow), and

states (e.g., surface temperature). Jasinski et al. (2019) evaluated many NCA-LDAS variables and by comparing them to other independent datasets concluded that, in general, NCA-LDAS is an effective tool for merging diverse satellite data products that quantify climate trends for scientific understanding and decision support. This work leverages the flooded area variable which reports the area within a grid (size: 0.125 deg. by 0.125 deg.) that is flooded. The grid is extended over the entire U.S. To the best of our knowledge, this study is the first to use the flooded fraction variable from NCA-LDAS to evaluate the relationship between flooding and damage.

Both the FPR and the GFF are conceptually simple and computationally rapid to compute. The FPR proxies flood stage as it is based on the height of a streamgage, while the GFF is a better proxy for the spatial extent of the flood. Obviously, both the FPR and GFF have significant limitations, which ultimately may not be mitigated by their computational ease. The FPR relies on a spatial interpolation which does not take into consideration topography that governs flooding patterns (e.g., levees, mountains) and is especially unrepresentative of the actual flood depth in regions with variable terrain. Further, while streamgages are fairly abundant east of the Mississippi River, they are comparatively sparse to its west (See Supplementary Figure B.1). In some instances, spatial interpolation of over 100 miles is required. GFF data are only available through 2016, but this could soon change. Furthermore, its spatial resolution is 0.125 deg. by 0.125 deg. grids - extremely coarse relative to more sophisticated hydrodynamic models.

This research uses data from the National Flood Insurance Program (NFIP) to proxy flood damage. In the United States, the NFIP is the main provider of flood

insurance and currently insures over $1.3 trillion of assets (Horn & Webel, 2021). The growing economic activity and development in risky areas are directly influencing the increasing insured losses since the early 1990s (S. D. Brody et al., 2008; Highfield & Brody, 2017; Michel-Kerjan et al., 2012). These increased losses have required the NFIP to borrow money from the U.S. Treasury, which has led to a deficit of nearly $25 billion (Wells, 2017). Many papers, including Mobley et al. (2021) and Knighton et al. (2020), have developed statistical models to predict NFIP claims as a function of hydrologic, socioeconomic, and development variables. An anonymized NFIP dataset has been made publicly available, and includes all U.S. records of flood claims, along with the level of the damage and spatial information about the claim. It also includes information on the start and termination dates of (anonymized) policyholders. Specific limitations, however, exist when using NFIP claims as a proxy for damage. While the NFIP has the bulk of the market share for flood insurance products, only about 50% of households within the 100-year flood zone (i.e., the Special Flood Hazard Area) have flood insurance (Dixon et al., 2006; Kousky, 2018). Further, 25% of NFIP claims are outside the 100-year floodplain boundaries - i.e., areas not required to have flood insurance to obtain a federally-backed mortgage - showing that flooding does occur outside the 100-year floodplain (S. D. Brody et al., 2013). Thus, this dataset is highly likely to exclude a significant number of uninsured homes that have been flooded. Future work could consider separate flood loss datasets, such as the NOAA Storm Events database (National Centers for Environmental Information, 2021).

3.2 Data and Methods

3.2.1 Data Description

The data for this study is obtained from several publicly available U.S. government sources, including the Department of Homeland Security (DHS), the U.S. Geological Survey (USGS), and the U.S. National Aeronautics and Space Administration (NASA). Supplementary Table B.1 summarizes the data and includes summary statistics. In addition, Figure 3.1 provides density plots of key variables. A description of the data follows:

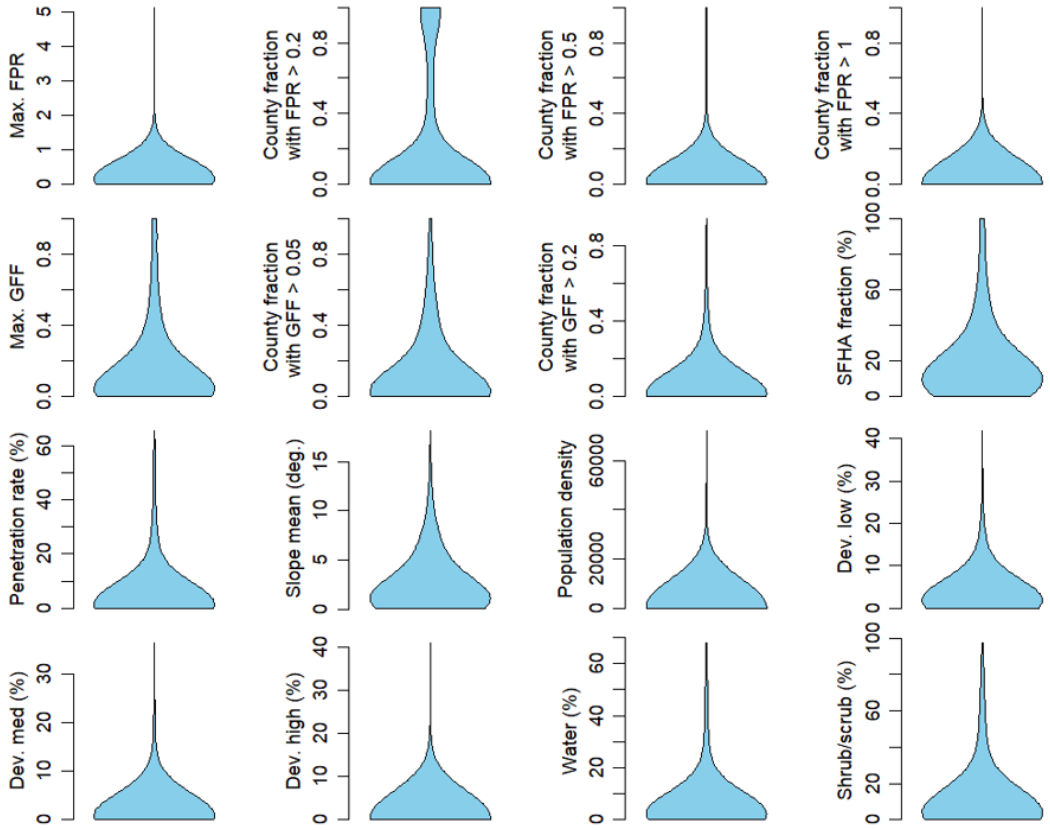

Figure 3.1. Density plots of key variables.

Spatial and Temporal Extent: This study focuses on flooding events that occurred in 2016 in the contiguous United States to predict NFIP claims. The year 2016 was selected because (i) it had the highest number of billion-dollar flooding events (National Centers for Environmental Information, 2020), and (ii) it is the last year for which GFF data are available.

However, the Stage 1 classifier is built using FPR data (thus streamgage data) and NFIP claims data between 2005 and 2015 in the contiguous United States. It is then applied to 2016 data to predict flooding events. The classifier is described in the methods section. The FRP data are used over GFF data in Stage 1 because a comparative sampling of the two datasets in a few states suggests that the FPR data would have better predictive performance overall; future work could expand this analysis to see if this is true nationwide.

NFIP claims: The response variable in this study is the number of residential NFIP claims at a county-level during a *predicted* flooding event (determined by the Stage 1 classifier). Anonymized NFIP data is freely available through OpenFEMA, with records dating back to 1970. The records of claims include the date of the flood and information about the county code of the parcel (among other data, including information about losses). While flood insurance is available for businesses, these data are excluded from the analysis to focus on residential claims since businesses are a small fraction of the claims data. Residential insurance is available for both single-family and multi-family residential structures, though the majority of the coverage is for single-family homes. The start of a predicted flooding event (in 2016) is identified

by the Stage 1 classifier. The duration of a flooding event, we assume, extends 3 days beyond the start of the event, thus a 4-day event in total (see Figure 3.2). During each predicted event, we count the number of NFIP claims. If the Stage 1 classifier fails to identify an actual flooding event in a county, and thus its claims, those claims are not included (though are considered when evaluating cumulative model error over Stage 1 and 2 later). We evaluated different definitions of events, and this definition offered the best in terms of its predictive ability.

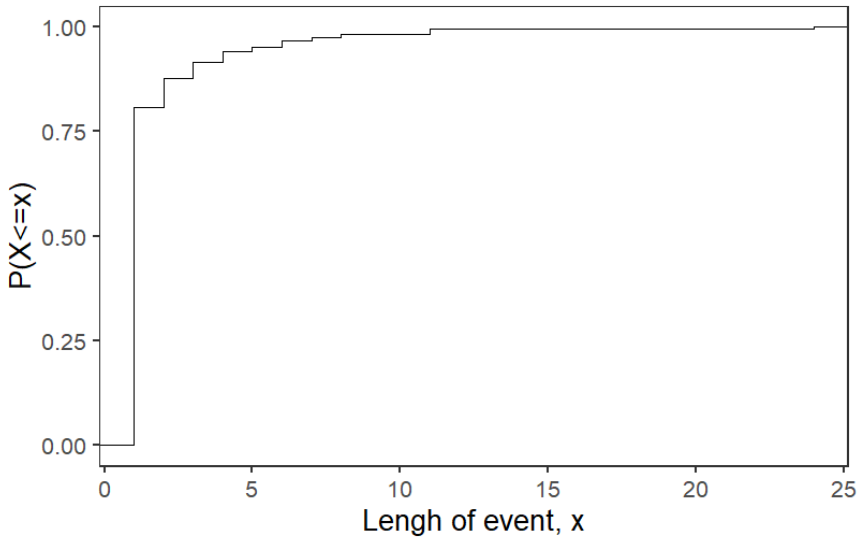

Figure 3.2. Ogive of the length of all flood events between 2005 and 2015. An event starts on the day of the first NFIP claim in a county and then continues for all the (semi-)consecutive days in which at least one NFIP claim is made. We allowed for at most two days without an NFIP claim to pass to consider those observations to be within the same event. The empirical data show that 95% of flood events are four days or shorter.

The NFIP had 97,253 residential claims in 2016 in over 1,239 counties, which resulted in over $3.4 billion (USD) in residential building payouts, along with another $570 million (USD) in residential content payouts (Federal Emergency Management Agency, 2021). The average and median combined payout were $43,000 and $25,200 (USD) respectively, though this ranges from $0 to $250,000 (USD), which is the

maximum allotted by the NFIP. Figure 3.3 shows the number of county-level claims in 2016. Louisiana had the highest number of claims, with 34% of all records. 69% of all claims were within the 100-year floodplain in 2016.

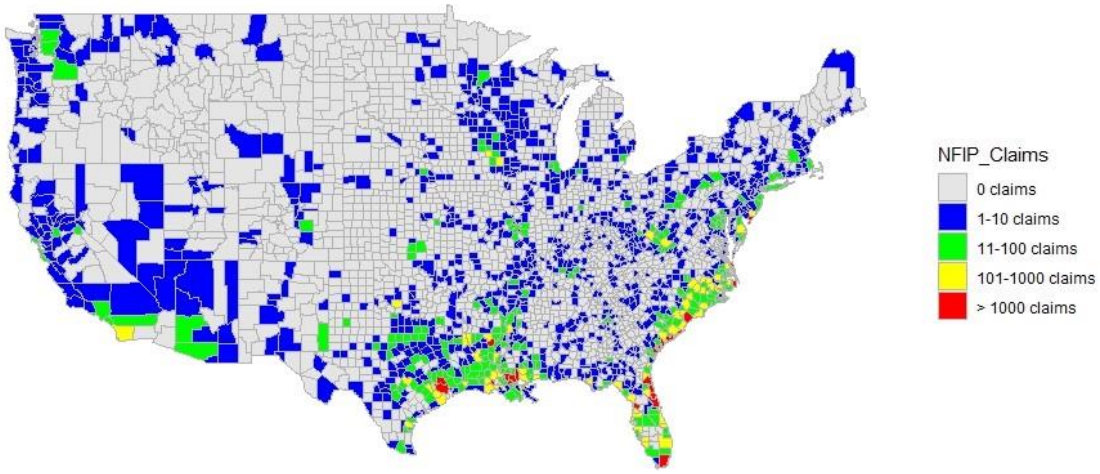

Figure 3.3. The distribution of NFIP claims in 2016 aggregated to the county level.

NFIP policies-in-force (i.e., the penetration rate): NFIP coverage is known to vary throughout the U.S., and penetration rates strongly depend on prior flooding (Atreya et al., 2015), the fraction of the region inside the 100-year floodplain, and regional wealth (Zahran et al., 2009). The best approach for computing NFIP penetration rates is to divide the number of policies-in-force by the number of residential buildings in a given area. However, data on the number of policies-in-force are not publicly available for years other than 2020. Thus, we use the number of policies-in-force in 2020 and divide it by the number of households in that county. At the time of the analysis, the number of households in a county was available through 2019, and thus 2019 data were used. We do not expect significant changes in county-level data

between 2016-2020, but acknowledge this limitation. A nationwide map of computed penetration rates is shown in Figure 3.4.

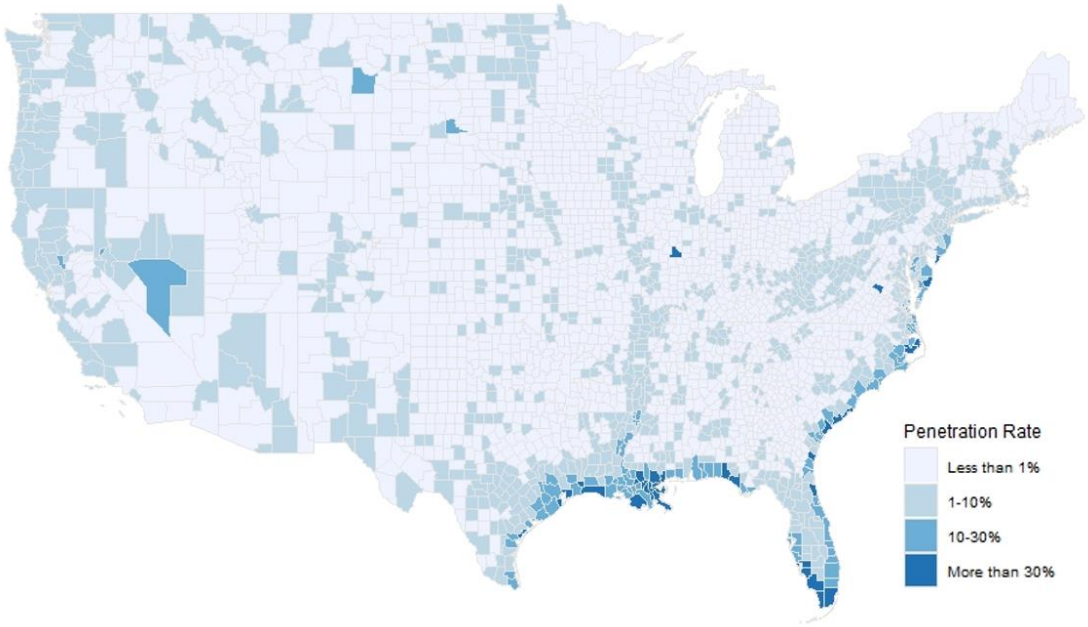

Figure 3.4. Computed NFIP penetration rates at the county level.

100-year floodplain: Past studies found a strong influence between the percentage of a region in the 100-year floodplain and flood-induced property damage (S. D. Brody et al., 2008, 2009). Thus, we consider the fraction of the county in the floodplain. As mentioned earlier, a limitation of using this metric alone (without e.g., GFF or FPR) is that this metric is static; it does not vary from event to event and thus cannot differentiate between moderate and extreme floods.

Land cover: Past studies have shown a significant relationship between the land cover of the study region and the amount of flood losses (S. Brody et al., 2014). Land cover data are publicly available through the National Land Cover Database (NLCD)

(Multi-Resolution Land Characteristics Consortium, 2021b). The database has a spatial resolution of 30m by 30m and is sorted into 20 classes. For simplicity and to enhance interpretability, classes with similar characteristics are aggregated into 11 categories consistent with Nateghi et al. (2011). For each county, the fraction of the area covered by each land cover category is computed. The four land cover variables related to development are indicators of the level of impervious surface.

Land slope: Multiple studies have shown a significant positive correlation between slope and flood losses (S. D. Brody & Highfield, 2013; Frimpong et al., 2020; Highfield & Brody, 2013). The argument is that higher slopes result in greater rainfall concentration and subsequently result in faster and higher stream peaks and mean annual flows (Stuckey, 2006). To calculate the slope, we use a digital elevation model (DEM) obtained from the Amazon Web Services (AWS) Terrain Tiles using the 'elevatr' package in R (*Amazon Web Services (AWS) Terrain Tiles*, 2017; Hollister et al., 2020). The slope has a spatial resolution of 500m by 500m. The average of slope grids in each county is calculated to indicate the mean slope.

Coastal indicator: We include a binary variable indicating whether the county is coastal. The rationale is that instantaneous discharge data from streamgages, used to compute the FPR, is an indicator of freshwater (riverine) flooding and is less likely to indicate flooding due to coastal inundation.

Population per area: Because exposure can vary significantly between counties, and counties with more exposure are expected to receive more damage, we control for exposure by including population per unit area, similar to Highfield and Brody (2017) and Czajkowski et al. (2013). This metric is calculated by dividing the estimated population of a county by the land area (in square miles) of the county (U.S. Census Bureau, 2021).

Flood Peak Ratio (*FPR*): The FPR was first introduced by Villarini et al. (2011) as an indicator of flood intensity. It is defined as the ratio between the maximum instantaneous discharge during a flooding event at a given streamgage and the 90^{th} percentile of annual maximum instantaneous peak discharge for that streamgage. The 90th percentile of annual maximum instantaneous peak discharge is equivalent to 10-yr flood peak value (Czajkowski et al., 2013; Villarini, Smith, Baeck, & Krajewski, 2011; Villarini, Smith, Baeck, Marchok, et al., 2011). An FPR value is calculated for all streamgages in the region of study. These FPR records are geographic point data. Next, consistent with Czajkowski et al. (2013), an Inverse Distance Weighted (IDW) interpolation method is applied to assign an FPR value across the entire region (Pebesma, 2004). In principle, the FPR is a continuous variable over space. The FPR is computed daily for the entire US between 2005 and 2016 (using the peak daily discharge for each gauge) to build the classifier and predict flooding events in Stage 1. An example of daily FPR data, interpolated across a large spatial area, is shown in Supplementary Figure B.2. In order to obtain county-level intensity measures, we

compute five FPR metrics. The first is the maximum flood peak ratio in a county and the others are the fraction of the county with an FPR above 0.2, 0.5, 1.0, 1.5, and 2.0.

Giovanni Flooded Fraction (*GFF*): Historically, NASA's NCA-LDAS data, from which we extract the flooded fraction of a region, and other land data assimilation systems that collect their data through satellite imagery, have been used for climatic assessment and trend studies (Jasinski et al., 2019; Jasinski, M.F. et al., 2018; Mucia et al., 2020). We use NCA-LDAS Version 2.0, which includes 42 climate variables measured daily at a 0.125 deg. x 0.125 deg. grid-size across the U.S. from January 1, 1979 to December 31, 2016, to evaluate the fraction of the land area that floods (Jasinski, M.F. et al., 2018). For the day of the event that had the greatest flooded fraction for a county (recall an event is defined to last 4 days), we overlaid the spatial extent of the flood (0.125 deg. by 0.125 deg. gridded data) atop the counties to compute the fraction of the county that was flooded. For grid cells that encompass two or more counties, we proportionally divided the flooded area in the grid cell with the fraction of the county in the grid cell. Initially, eight county-level metrics are computed: the maximum flood fraction of the county, which is the maximum flooded fraction of any grid cells within a county, and average flooded fraction, which is the average flooded fraction over all grid cells within a county, and the fraction of the county covered by flooded fractions greater than 0.05, 0.1, 0.2, 0.4, 0.6, and 0.8.

3.2.2 Methods

This work uses statistical learning theory to build statistical models that first predict whether a flood occurred in a county with intensity significant enough to trigger at least one NFIP claim (i.e., our Stage 1 "classifier") and then to predict the number of NFIP claims during that event (i.e., our Stage 2 NFIP regression model). Separate statistical models are built for both the FPR variables and the GFF variables when predicting the number of NFIP claims in Stage 2. An overview of the approach is presented in Figure 3.5.

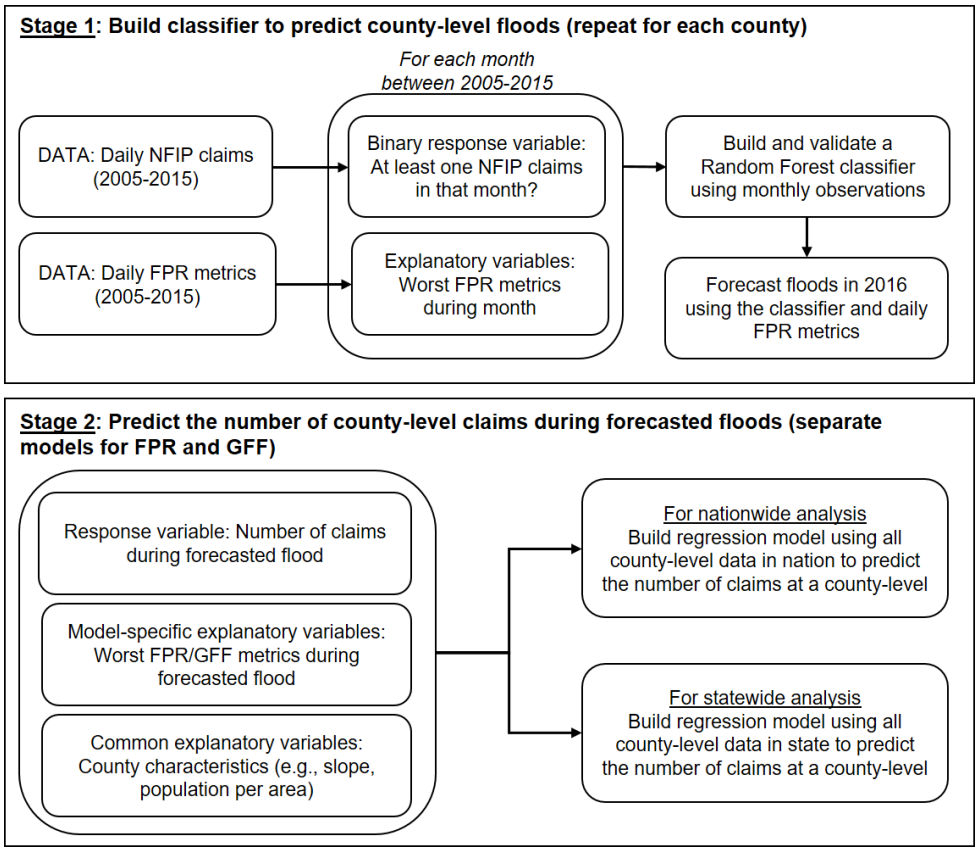

Figure 3.5. The framework of the methodology

Statistical learning theory: Supervised learning theory is becoming increasingly popular in the hazards literature, especially for identifying the relationship between

hazards intensity and infrastructure serviceability (e.g., predicting electric power outages (Guikema et al., 2014), predicting pipe breaks (Chen et al., 2019)). Broadly, supervised learning models are categorized into parametric, semi-parametric, and non-parametric models and seek to find the structural relationship between independent variables and a response variable that minimizes the cumulative delta between observations and predicted outcomes. The delta function could represent a straight difference function, e.g., Euclidean distance, or something more complicated. Parametric models - in which parameters are computed that best project the data onto a prescribed functional model form (often linear) - are popular, especially in hazards research and policy analysis, because of their stability and interpretability. Non-parametric models do not assume a relationship between the predictor variables and the response variables, and rather extract patterns from the data that elicit the functional relationship. Thus, they may better approximate the true relationship between the predictor and response variables, but this comes at the cost of interpretability, (potentially) model stability, and (potentially) requirements for a lot of data (Nateghi, 2018).

Stage 1 Classifier - Identification of flooding events: The ultimate objective of this work is to predict the number of NFIP claims during a flood "event" in a county, where an event is a date range during which flooding occurs. Prior work pre-defines the time window of an event by considering only specific incidents. For example, when Czajkowski et al. (2017) make predictions on the number of NFIP flood claims, they do so for a (large) set of predefined tropical cyclones to impact the US. The

approach in the current work is data-driven, meaning that we leverage the data to "discover" if an event likely occurred. Thus, we devise an approach to predict whether a county experiences a flood resulting in at least one NFIP claim regardless of the scale of the flood or source of the water.

Perhaps an intuitive way to approach this is to focus on periods of anomalous precipitation. This approach, however, could exclude events that resulted in minor flooding and would exclude events that were not precipitation induced. Another plausible approach would be to leverage the NFIP dataset, and consider that there is a flood event whenever there is a claim. However, if we were to define a flood event using the NFIP claim data, we would effectively introduce endogeneity into our model by preconditioning our model to predict the number of claims only for events that we knew *a priori* to have at least one claim. Thus, we develop a novel approach by building a statistical classifier using 2005 through 2015 county-level data to find the functional relationship between FPR indicators and the occurrence of at least one NFIP claim. This classifier is then applied to the 2016 FPR data to define county-level events.

First, we compute the FPR for each streamgage at a daily level between 2005 and 2015 using its maximum daily discharge and then compute the FPR metrics described above for each county. We also identify whether at least one NFIP claim was made on that day. Unfortunately, a lag could exist between the maximum FPR and when an NFIP claim is made (i.e., the conditions that cause a flood may not align with the date on which the claim is reported). Therefore, we aggregate the data to monthly level (e.g., 1-Jan to 31-Jan) by finding the maximum over all the FPR

indicators (i.e., maximum county FPR, and the fraction of county with an FPR greater than 0.2, 0.5, 1, and 2) over the month and creating a binary response variable that indicates whether at least one claim was made during that month in a given county. The idea is to take a snapshot of the worst hydraulic conditions over that month in that county and see whether it induced at least one flood claim at some point. This assumes that no county had more than one flooding event in a month - a potentially erroneous assumption that we found to occur exceedingly infrequently. Finer temporal divisions (e.g., a week as opposed to a month) increase the likelihood of not capturing potential lags between worst-case hydraulic conditions and the reporting of an NFIP claim, which would divide the actual event into two: one with severe hydraulic conditions and no claims and one with typical hydraulic conditions and an NFIP claim. A wider temporal division (e.g., two months) increases the likelihood of having multiple actual flooding events in the interval.

Using this county-level data, we built a cross-validated random forest statistical model (i.e., a "classifier") to identify the FPR indicators that trigger at least one NFIP claim in each county. Random forests (RF) are a statistical tree-based ensemble model that bootstraps data to build multiple decision trees and then averages the predictions over all trees (Breiman, 2001). We choose an RF model because it is a low-bias, low-variance model given how it is able to capture the structure of the data and how it averages over uncorrelated trees. Because of its averaging approach, it is highly effective at variance reduction. We test the classifier using a 5-fold cross-validation approach, whereby a model is built using 80% of the data and the model is then tested using the withheld data. This is repeated five times

so that all data are withheld once and the entire process is repeated three times. Because Stage 2 focuses on predicting the number of NFIP claims should a flood event occur (and it can predict zero), we tune the model to minimize false negatives as opposed to false positives.

As expected, most counties have a data imbalance, as a majority of counties had no claims for many of the months. To address the imbalance in the data, we compared the performance of the classifier when combined with downsampling and upsampling algorithms. While the details of each algorithm differ, the general concept is similar - when the data are bootstrapped, more emphasis is placed on observations with the minority class (i.e., the observations when the response variable is greater than zero) (H. He & Garcia, 2009). The downsampling algorithm, which randomly subset the majority class in the training set to match the frequency of the minority class, resulted in a better performance of the classifier.

The final classifier, in essence, identifies the unique FPR conditions that portend a flood event in each county based upon its historical FPR and reported NFIP claims. We then use this validated classifier on the 2016 FPR data to identify when a flood began in a given county. Each flood event needs an end date. We add an end date that is four days after the beginning date (rationale below). Then, for each county, we tabulate the number of NFIP records during that predicted flood event. This becomes the response variable for the model that predicts flood claims in Stage 2. If the classifier erroneously predicts a flood, but there is not one, zero (0) NFIP claims are recorded.

The duration of four (4) days was determined as follows. For each county, starting in 2005, we identified the date of its first actual NFIP claim, and then counted the number of (semi-)consecutive days in which at least one NFIP claim was made in that county. We allowed for at most two days without an NFIP claim to pass to consider those observations to be within the same event. We repeat this process through 2015. By examining the ogive over the lengths of all events, 95% of them were less than 4 days (see Figure 3.2).

Stage 2 Model - Prediction of the number of claims during each flooding event: Once the flooding events have been predicted for 2016, we build statistical learning models to predict the number of claims for each event in each county. The "best" model is identified for the FPR data (along with the other covariate data) and another model for the GFF data, so that we can compare the two flood heuristics. Note that if the classifier from Stage 1 failed to predict an actual flooding event, the claims from that event would not be captured in the data (which is why the model is tuned to minimize false negatives at the expense of including more false positives).

The correlation matrices of the predictors are shown in Figure 3.6 and Figure 3.7. As expected, some land cover variables are highly correlated, which is also reflected in high Variable Inflation Factors (VIFs). The VIF is a measure of multicollinearity. To address this, we removed the following variables: the fraction of the county that is forested, the fraction of the county that is highly developed, the fraction of the county with an FPR above 1.5, and the average flooded fraction and the fraction of the county covered by flooded fractions greater than 0.1, 0.4, and 0.8.

Elimination of highly correlated variables above results in a maximum VIF of 4.79 for the FPR model and a maximum VIF of 5.95 for the GFF model. The predictors are then standardized to support model prediction. Note that population per area is a continuous unbounded variable. The data are zero-inflated as there are many predicted flood events that do not actually have claims (i.e., the response variable is 0).

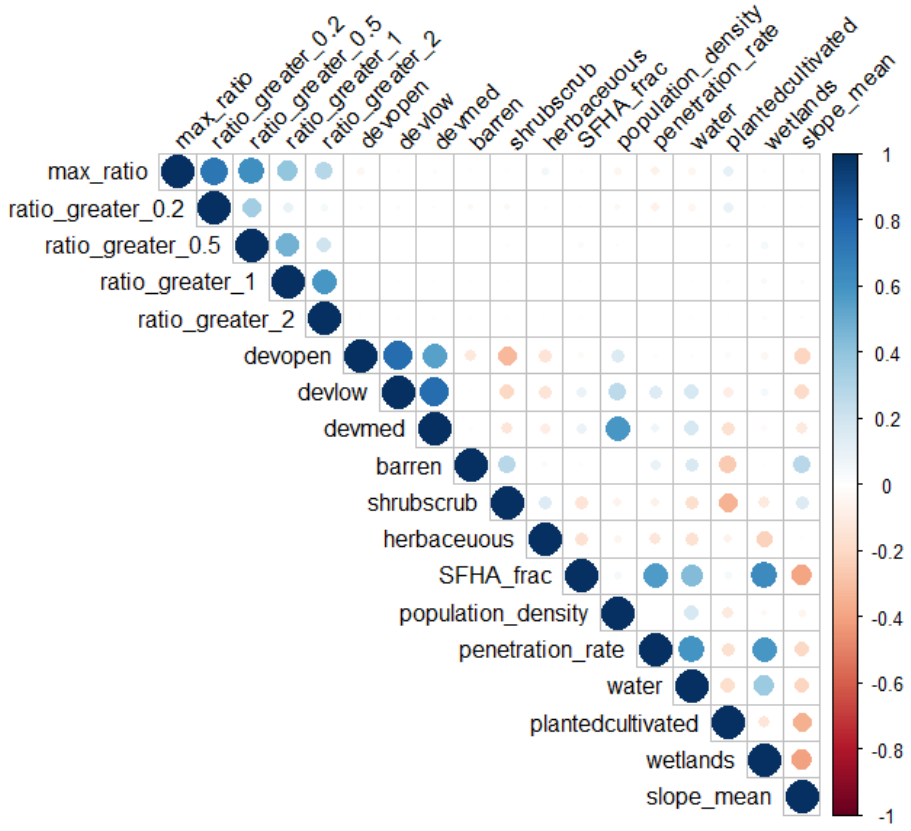

Figure 3.6. A correlation matrix of the predictor variables for the FPR model.

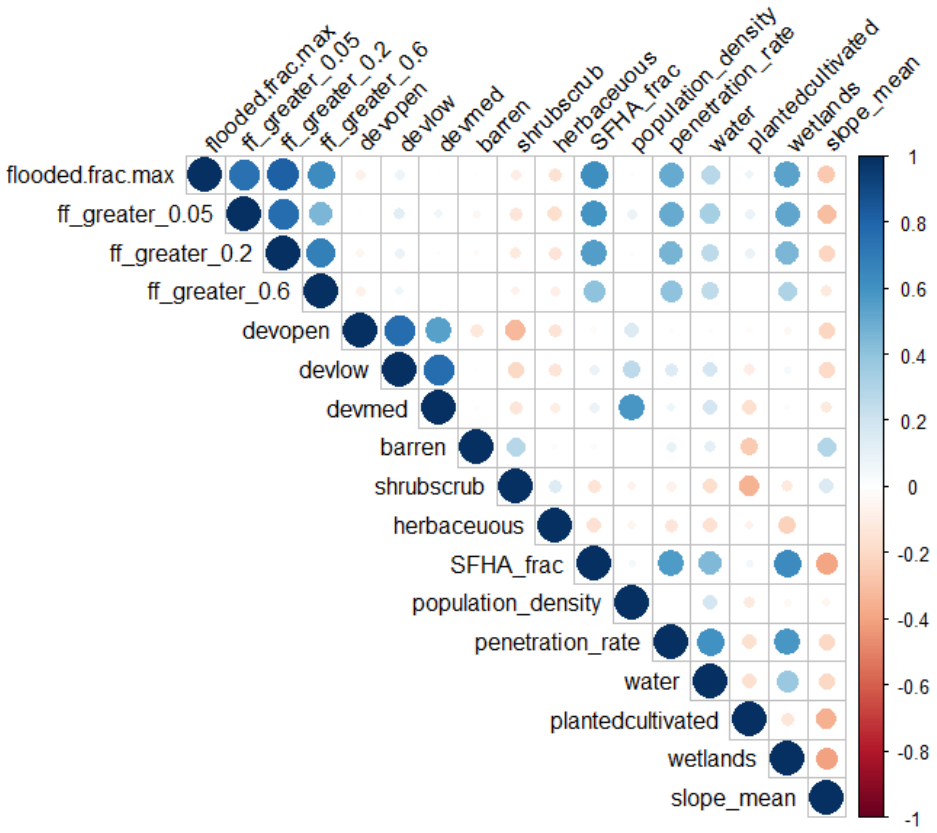

Figure 3.7. A correlation matrix of the predictor variables for the GFF model.

We then evaluate the performance of several statistical models in their ability to predict the number of claims in a county during a flooding event as a function of the independent variables. We build both nationwide models and statewide models using county-level observational data to determine whether models with more data (the nationwide models) are better predictors of claims despite containing data with significantly more variance. The statistical models considered are: (1) a zero-inflated negative binomial (ZINB) regression (i.e., the same model used in (Czajkowski et al., 2017); (2) a random forest (RF) model; (3) a support vector regression (SVR) model; and (4) a classification and regression tree (CART) model. More details on these models can be found in Appendix B. We specifically include a zero-inflated negative-

binomial model as our work builds from Czajkowski et al. (2017) which uses a negative-binomial model - though direct model comparisons between our work and Czajkowski et al. (2017) cannot be made because our work is data-driven, not event-driven, variable selection is slightly different, and we are operating in a different spatial-scale.

For the zero-inflated negative binomial (ZINB) model, we use a combination of forward and backward feature selection to identify the subset of variables with the best performance model. The approach starts with no predictor variables and then iteratively builds a model by adding one predictor at a time that most contributes to the model performance. In the same step, it eliminates any variables that do not provide any improvements. The stepwise feature selection resulted in a significant error reduction and model performance improvement. Similar stepwise feature selection was used in conjunction with other statistical models (i.e., CART, RF, and SVR) but the performance improvement was negligible.

We validated each model using a 10-fold cross-validation approach, whereby a model is built using 90% of the data and the model is then tested using the 10% of the data that is withheld. This is repeated 3 times. We then compute model performance on out-of-sample data (i.e., test data). In-sample predictions are vulnerable to bias (i.e., overfitting) and thus the in-sample metrics are less informative. The error metrics that we use are the root mean squared error (RMSE) (i.e., the sum of the squared differences between the predicted value and the observed value) in order to penalize larger deviations from observed values more than smaller deviations and the mean absolute error (MAE) (i.e., the average difference between

the observed value and predicted value). All models are compared relative to a null model, which is simply the average number of NFIP claims during predicted events statewide or nationwide (depending on the spatial level of the model).

After model selection for both the FPR and GFF approaches, we built variable importance plots, which rank each variable's importance in terms of its contribution toward error reduction. The variable that contributes most to out-of-sample predictive accuracy (based on models with and without that variable) is deemed most important, and so on (Hastie et al., 2009). Finally, partial dependence plots are built to map the relationship between each covariate and the response variable, without the influence of other variables (i.e., they are held constant).

3.3 Results

Stage 1 - Flood-event classifier: The classifier is built for each county using 2005-2015 data using a Random Forest (RF) model and then is validated using 2016 data from the same county. Note that we exclude counties for which there are no claims between 2005 and 2015, which leaves 2,555 counties remaining. For each of the 2,555 counties, the classifier predicts which months in 2016 had at least one claim. A confusion matrix is then computed for each of the 2,555 counties using the 2016 data, and these confusion matrices are summed together to produce an aggregated out-of-sample confusion matrix for the 2016 data (Table 3.1).

Table 3.1. Out-of-sample confusion matrix for monthly-county records in 2016. This matrix sums together each of the 2,555 counties' confusion matrices.

Scale: county-month		**Predicted**	
		No claim	*At least one claim*
Actual	*No claim*	20,188 (66%)	7,587 (25%)
	At least one claim	1,246 (4%)	1,639 (5%)

Table 3.1 summarizes the results of 30,660 county-months in 2016 (i.e., 12 months for each of 2,555 counties). The classifier successfully predicted 71% of county-months; 66% of county-months had no claims in actuality, and the model predicted no claims while 5% of county-months had at least one claim in actuality, and the model predicted there to be at least one claim. As mentioned above, because Stage 2 predicts the number of NFIP claims for a predicted flood event, which can be zero, we are interested in minimizing the number of county-months that falsely predict no claim at the expense of falsely predicting a county-month with a claim. This ensures that Stage 2 includes as many actual claims as possible. To this end, the Stage 1 model has a 4% false negative rate as opposed to a 25% false positive rate. However, while the model did miss 4% of county-months with actual claims, 81% of all actual claims in 2016 are contained within the 1,639 county-months that were predicted correctly. In other words, while the model misclassified many months with actual claims, these months generally had relatively few claims.

The classifier then is applied to daily FPR data at the county-level in 2016 to determine *when* an event began (i.e., on which dates are the conditions beyond the county-level threshold, created by the classifier, that would suggest at least one NFIP claim). We assume a flood event lasts three days beyond the initiating day. As

expected, due to the high false positive rate in the classifier, 94.4% of the predicted events have no actual claims. Ultimately, however, 81% of the 97,253 claims are contained within events that were correctly predicted. (We account for the 19% of actual claims that we fail to predict when evaluating cumulative model error later.)

Stage 2 - Predicting the number of claims: The second stage predicts the number of claims given that the classifier in Stage 1 predicted a flooding event. Separate models are built using the FPR and GFF statistics. The remaining covariates, described in the data section, are identical among the models. Two types of models using county-level observations are built - one for the entire U.S. and one for each state.

3.3.1 Nationwide Model

The nationwide model includes county-level data from the 48 contiguous states. It benefits from a much larger dataset, but in totality, these data are also high in variance which potentially reduces the predictive power. The results are shown in Table 3.2 and Table 3.3. We ran pairwise t-tests among the models built for the FPR data and the GFF data to confirm statistical performance differences exist among the errors; the results are different at a 5% level of significance.

For the nationwide FPR model, the RF model performed the best in terms of out-of-sample MAE but slightly worse in terms of out-of-sample RMSE when compared to the CART model. However, because the final model developed by RF provides the best fit to the data as determined by the correlation between the observed and predicted values, the RF model is selected as the best model overall. The RF

model shows a 23% improvement and 9% improvement over the null model in terms of MAE and RMSE, respectively. For the nationwide approach using GFF, the RF model outperformed all other models in terms of all performance metrics and thus was selected as the best model. In this case, the RF model results in a 22% improvement in terms of MAE and a 7% improvement in terms of RMSE over the null model.

Table 3.2. Summary of model performance for FPR model.

Model	**In-Sample**		**Out-of-sample**			***y-ŷ corr.***
	MAE	***RMSE***	***MAE***	***RMSE***	***R^2***	
CART	0.64	6.11	0.67	6.70	0.16	0.58
RF	0.51	5.41	0.64	6.75	0.15	0.73
SVR	0.98	7.09	1.01	7.09	0.14	0.48
ZINB	0.79	13.88	0.81	12.45	0.13	0.28
Null Model			0.83	7.39		

Table 3.3. Summary of model performance for GFF model.

Model	**In-Sample**		**Out-of-sample**			***y-ŷ corr.***
	MAE	***RMSE***	***MAE***	***RMSE***	***R^2***	
CART	0.67	6.36	0.73	6.97	0.08	0.57
RF	0.42	4.33	0.65	6.91	0.11	0.86
SVR	1.04	7.26	1.05	7.11	0.10	0.42
ZINB	0.81	7.86	0.83	7.89	0.02	0.10
Null Model			0.83	7.39		

Overall, the performance of the best models that use FPR and GFF data is marginal, given their low predictive power. These findings somewhat align with Czajkowski et al. (2017) where the authors found an R^2 of 0.13 (compared with 0.15 and 0.11 for our FPR and GFF models, respectively) though the scale and

methodology of that study differ from ours. The nationwide GFF model had higher errors in comparison to the FPR model among all the four models evaluated in the nationwide analysis (except the RMSE for the ZINB model). Ultimately, at a nationwide level, the RF model using the FPR data had better performance.

Figure 3.8 shows county-level error using the nationwide RF models for the FPR and GFF data. The R^2 values for the GFF model are greater than those for the FPR model for a majority of counties using the nationwide RF model (i.e., 54% of the counties considered). In addition, for counties with more than 100 actual claims in 2016, for which there are 63 counties, the R^2 values for the nationwide GFF model were higher 71% of the time. The MAE and RMSE values for the GFF model are less than those for the FPR model for a majority of counties using the nationwide RF model (i.e., 75% and 72% of the counties considered, respectively). Overall, the GFF model provides lower county-level errors more often and a slightly better model fit in comparison to the FPR model. However, the GFF model is worse when considering aggregate nationwide error (Table 3.2 and Table 3.3), meaning that in counties where GFF performed poorly, it likely performed especially poorly.

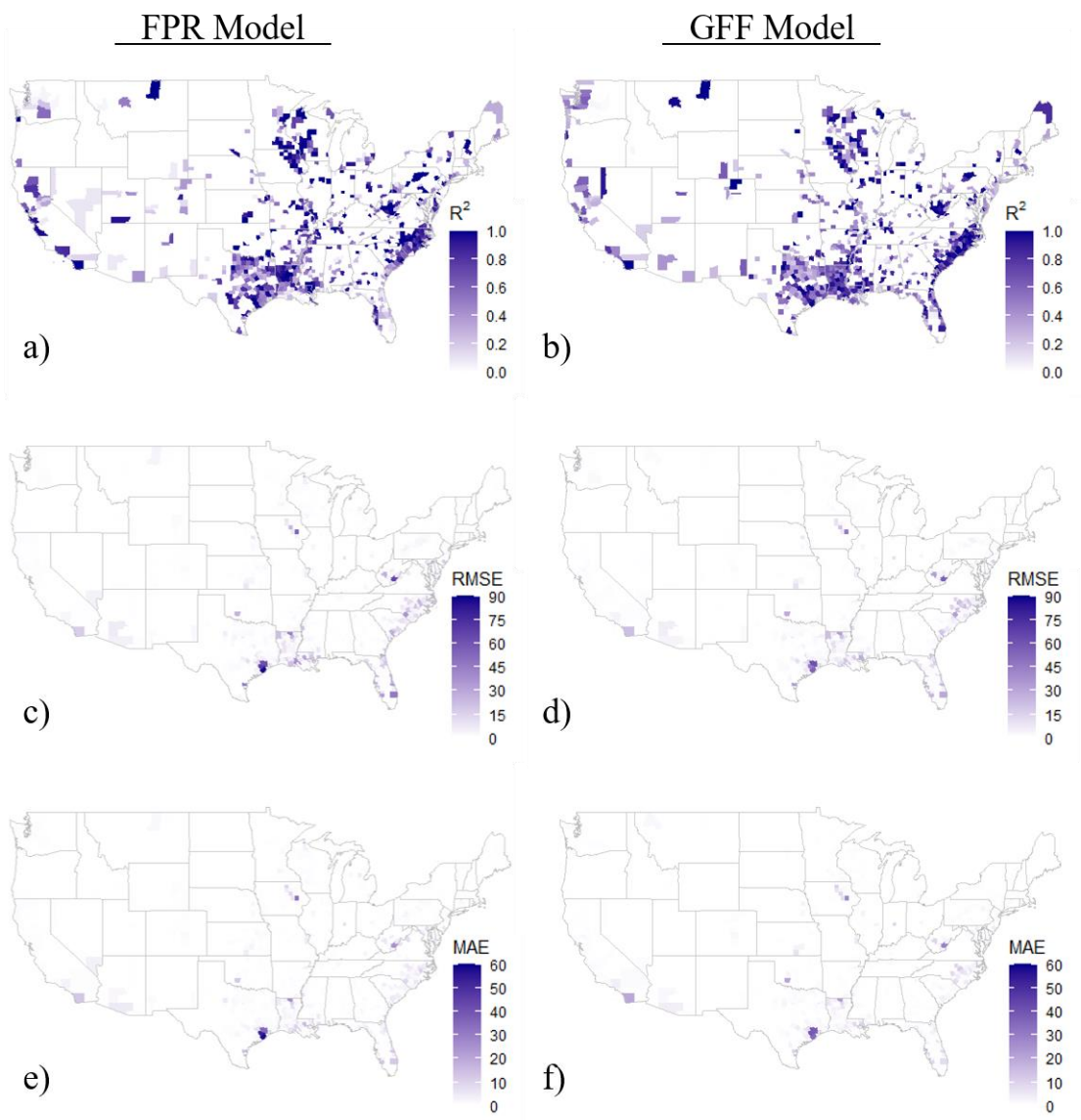

Figure 3.8. County-level error using the nationwide RF models for the FPR and GFF data: a. R2 (FPR model), b. R2 (GFF model), c. RMSE (FPR model), d. RMSE (GFF model), e. MAE (FPR model), f. MAE (GFF model). Note: These maps exclude counties that have zero claims in predicted events.

Figure 3.9 and Figure 3.10 show the variable importance plot for the nationwide RF models using the FPR and GFF data respectively. For the FPR model, the three most important variables are the fraction of a county with FPR greater than 1, the maximum FPR in a county, and the fraction of a county with FPR greater than 0.5. On the other hand, for the GFF model, the three most important variables are the

fraction of a county with FF greater than 0.05, the maximum FF in a county, and the fraction of a county with FF greater than 0.2. In both approaches, variables that reflect the characteristics of the flood are among the top important variables.

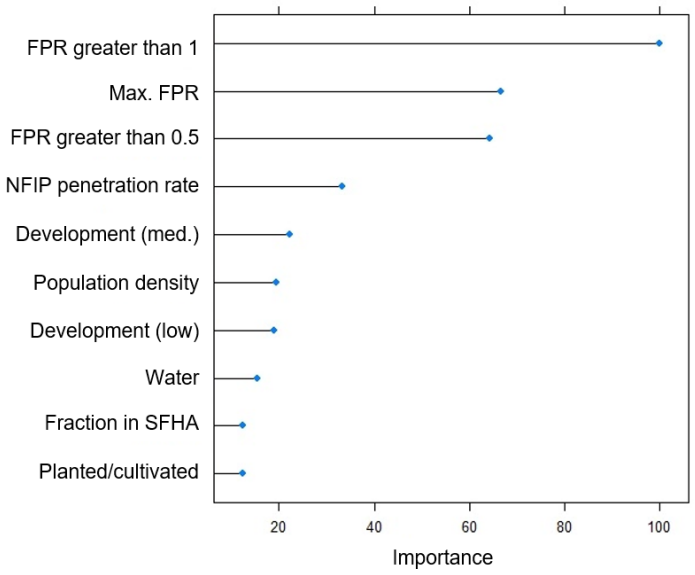

Figure 3.9. Variable importance plot for the FPR model.

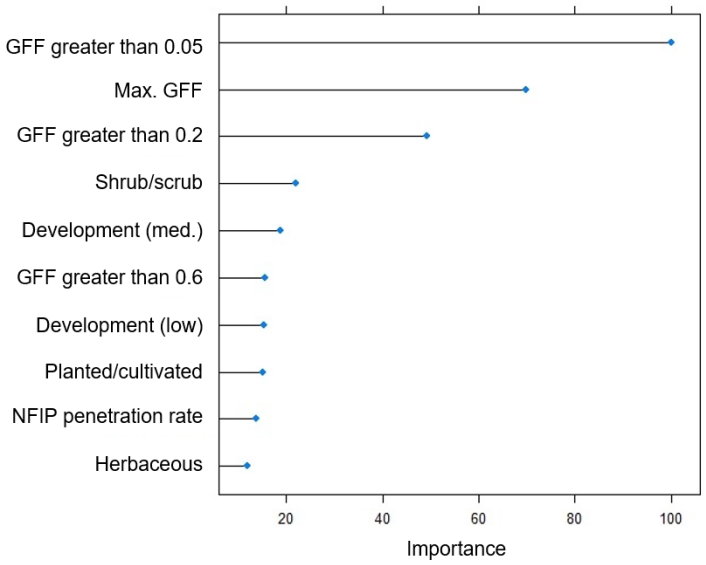

Figure 3.10. Variable importance plot for the GFF model.

Partial dependence plots for the three most important variables for both nationwide models are provided in Figure 3.11. As expected, for both the FPR and GFF models, the expected number of NFIP claims increases as flood intensity metrics increase. For the FPR model, an FPR greater than one at a county-level is the most important predictor variable in terms of its contribution to error reduction; when this fraction is close to 0, the number of claims is similarly expected to be around 0, *ceteris paribus*. As this fraction approaches 1, the expected number of claims rises to over 20, *ceteris paribus*. A nearly identical pattern emerges for the most important variable in the GFF model (fraction of grids with GFF greater than 0.05).

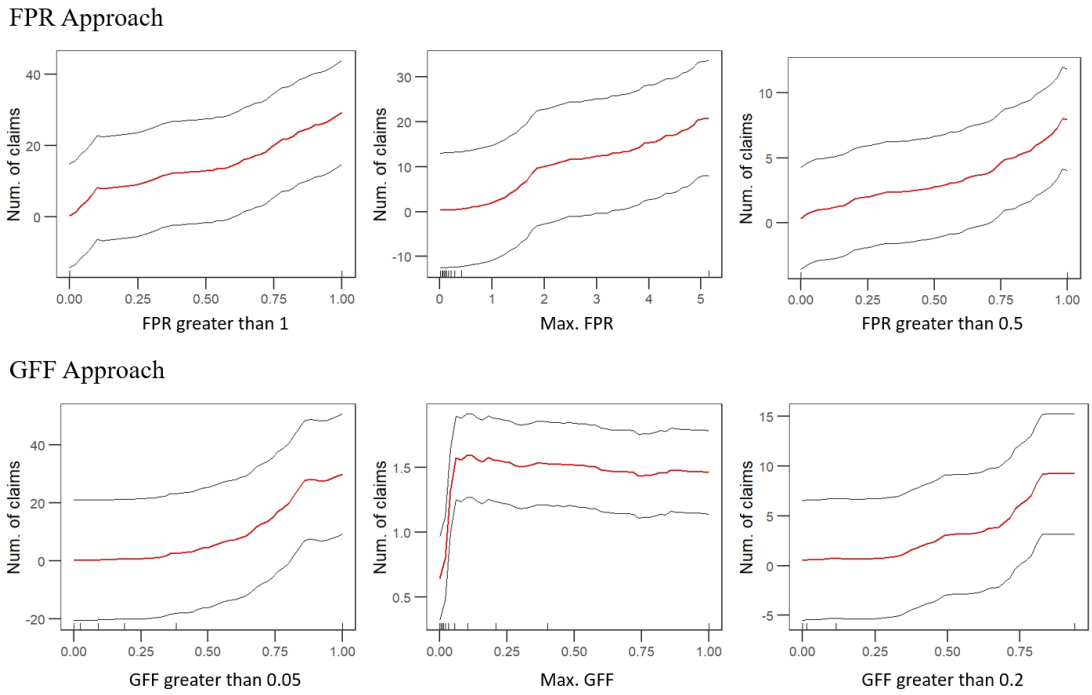

Figure 3.11. Partial dependence plots of the top three important variables for both the FPR model and the GFF model.

Figure 3.12a,b plot the fitted and the observed values for the FPR and the GFF models respectively ($\rho_{FPR} = 0.73$ and $\rho_{GFF} = 0.86$). 99.2% of flood events predicted by the classifier (i.e., the observations shown in these figures) are captured within the red boxes; Figure 3.12c,d plot the fitted and the observed values in these red boxes for the FPR and the GFF models respectively.

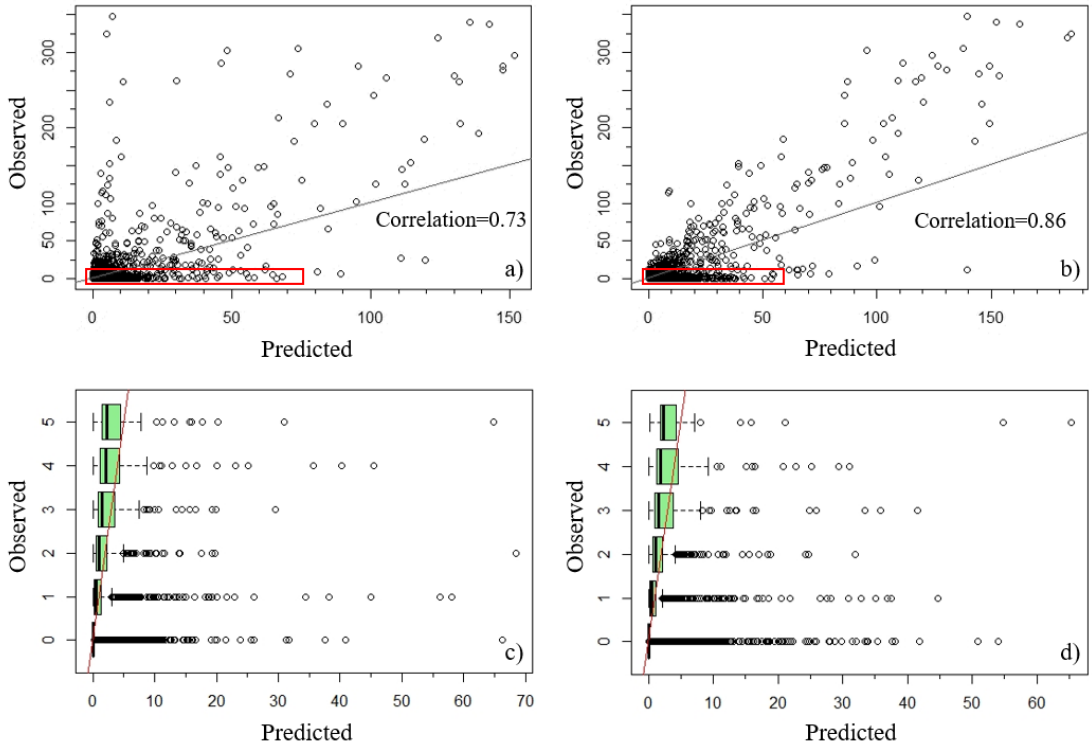

Figure 3.12. The plot of fitted vs observed values of the dependent variable for the RF model. 99.2% of flood events predicted by the classifier (i.e., the observations shown in these figures) are captured within the red boxes. a. Plot of fitted vs observed values for FPR approach, b. The plot of fitted vs observed values for the GFF approach, c. The plot of fitted and observed values in the red box for the FPR approach, d. The plot of fitted and observed values in the red box for the GFF approach.

3.3.2 Statewide Models

We use the same dataset of predicted flood events to perform state-level analyses. More specifically, for each state, we predict the number of NFIP claims

using county-level predicted flood events, FPR or GFF data, and the other shared county-level covariate data. RF model is selected as the best model because, in combination, it outperformed all other models in the nationwide analysis, and when we tested different models using state-specific data, the RF model consistently outperformed others. Predictive accuracy measures for each state using RF models are shown in Figure 3.13.

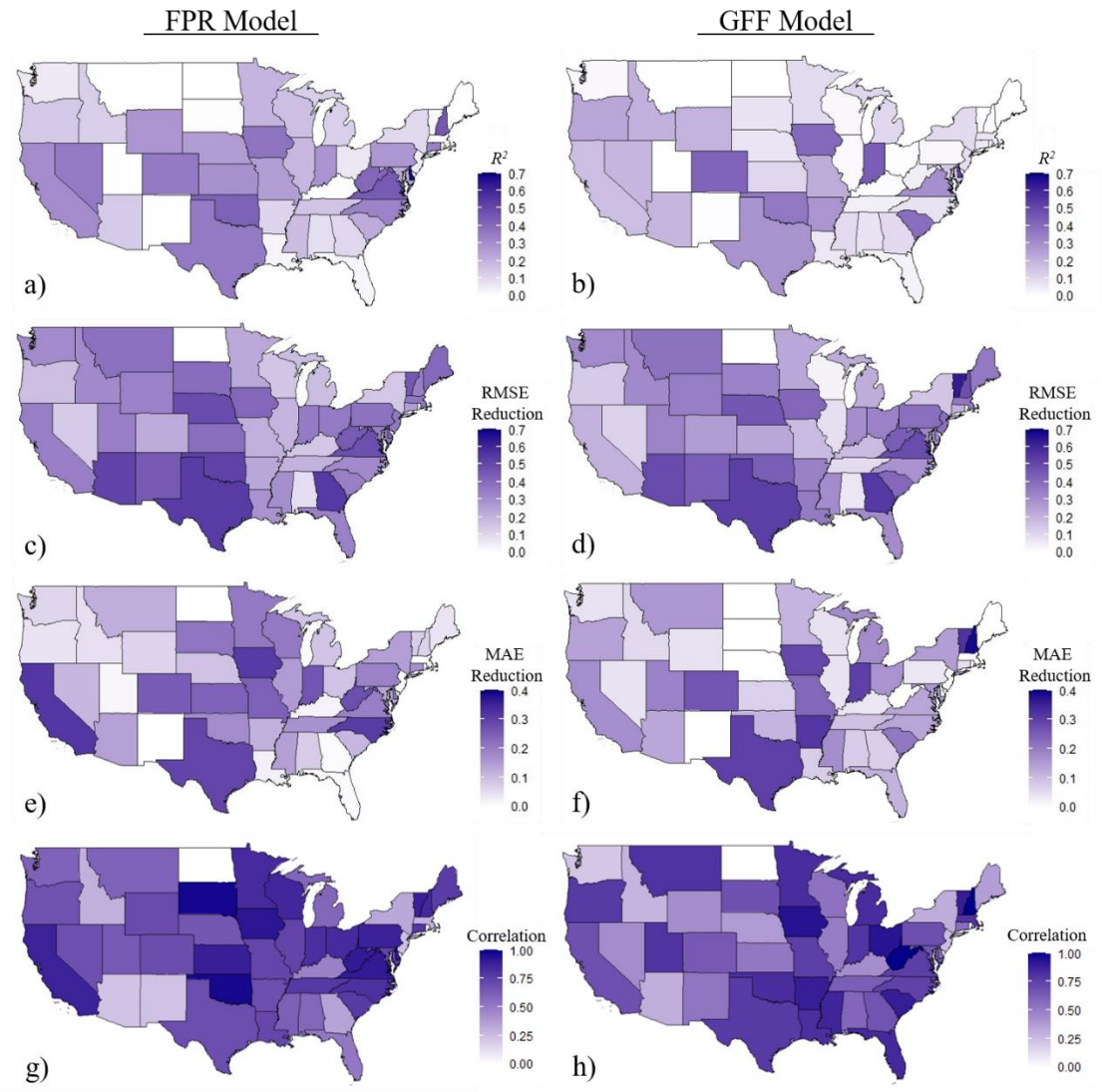

Figure 3.13. Predictive accuracy measures for each state using the RF model: a. R2 (FPR model), b. R2 (GFF model), c. RMSE reduction compared to null model (FPR model, out-of-sample), d. RMSE reduction compared to null model (GFF model, out-of-sample), e. MAE reduction compared to null model (FPR model, out-of-sample), f. MAE reduction compared to null model (GFF model, out-of-sample), g. Correlation between predicted and actual values (FRP model), h. Correlation between predicted and actual values (GFF model).

The R^2 for FPR models generally performed about 10% better than the GFF models; the R^2 for FPR models ranged from 0.2% to 67.9% with a mean of 19.6%, while the R^2 for GFF models ranged from 0.1% to 52.8% with a mean of 14.3%. Note that, on average, this is a significant improvement over the nationwide model (R^2 of

0.15 for the FPR model and R^2 of 0.11 for the GFF model) and Czajkowski et al. (2017) (R^2 of 0.13, though differences in approaches exist). The FPR model outperformed the GFF model in 28 out of the 46 states with predicted claims in terms of RMSE reduction. The mean of RMSE reduction among all states over the null model was 30% and 28% for FPR and GFF models, respectively. For FPR and GFF models, 45 and 42 states (out of the 46 states) had an improvement of more than 10% in RMSE reduction, respectively. Note that in two states (North Dakota and Rhode Island), all predicted events had zero actual claims, and therefore we did not build an RF model for these states. North Dakota and Rhode Island had only 7 and 3 actual claims in 2016, respectively, which occurred during events that were not predicted in Stage 1.

The RMSE reduction over the null model for both approaches is significant in some states, including Texas which had 14,751 actual claims in 2016, and Georgia, which had 1,582 actual claims in 2016 (See Supplementary Table B.2). The predictive performance was poorest in Alabama, which had 66 claims in 2016. However, few claims alone do not destine a state to poor model accuracy. Vermont had only 5 claims in 2016 and yet had a 45% RMSE improvement for the FPR model over the null model and a 63% RMSE improvement for the GFF model over the null model. This shows how other factors, such as insurance penetration rates and land cover factors play a significant part in predicting claims. For the MAE metric, the FPR model provides significantly higher error reduction in many states compared to the GFF model, including California, North Carolina, and much of the central U.S.

However, the MAE metric in the GFF model is higher for some states, including Arkansas, New Hampshire, and Vermont.

Finally, comparing the correlation between prediction and observation values using the RF model at the state level, the GFF model provides a higher correlation in states with the most actual claims, including Louisiana, Florida, Texas, and South Carolina. This may suggest that the GFF model delivers a better model fit for coastal states, which are more likely to experience extreme precipitation, surge, and more claims. On the other hand, the FPR model performs better in terms of correlation in most midwest states and west-coast states which may experience more riverine flooding. Over all states, the average correlation for the FPR models is 0.69, which is slightly higher than that for the GFF models at 0.66.

Table 3.4 shows the most important variable in terms of error reduction for the RF models (the FPR and GFF models) developed for each state. For the FPR statewide models, in all except eight states, the most important variable was an FPR indicator. The FPR indicator with the highest occurrence as the top most important variable was the maximum FPR. For the GFF statewide models, only seven states had a most important variable other than a GFF indicator. Similar to the FPR statewide models, in GFF statewide models, the maximum GFF was the indicator with the highest occurrence as the top indicator.

Table 3.4. The most important variable, in terms of error reduction, for each state in statewide analysis.

State	Top Important Variable (FPR)	Top Important Variable (GFF)	State	Top Important Variable (FPR)	Top Important Variable (GFF)
AL	ratio_greater_0.2	flooded.frac.max	MT	ratio_greater_1	flooded.frac.max
AZ	population_density	flooded.frac.max	NE	ratio_greater_0.5	flooded.frac.max
AR	max_ratio	gff_greater_0.05	NV	ratio_greater_0.5	population_density
CA	max_ratio	barren	NH	ratio_greater_0.5	flooded.frac.max
CO	devmed	herbaceuous	NJ	max_ratio	flooded.frac.max
CT	ratio_greater_0.2	water	NM	penetration_rate	flooded.frac.max
DE	ratio_greater_1	gff_greater_0.05	NY	penetration_rate	penetration_rate
FL	ratio_greater_0.5	gff_greater_0.05	NC	ratio_greater_1	gff_greater_0.05
GA	max_ratio	gff_greater_0.05	OH	ratio_greater_0.5	flooded.frac.max
ID	population_density	devmed	OK	ratio_greater_0.5	flooded.frac.max
IL	ratio_greater_1	gff_greater_0.05	OR	max_ratio	flooded.frac.max
IN	ratio_greater_2	flooded.frac.max	PA	ratio_greater_1	flooded.frac.max
IA	ratio_greater_1	gff_greater_0.05	SC	max_ratio	gff_greater_0.05
KS	ratio_greater_0.5	flooded.frac.max	SD	max_ratio	gff_greater_0.05
KY	water	flooded.frac.max	TN	max_ratio	gff_greater_0.05
LA	max_ratio	gff_greater_0.2	TX	max_ratio	gff_greater_0.05
ME	max_ratio	flooded.frac.max	UT	max_ratio	flooded.frac.max
MD	ratio_greater_0.5	gff_greater_0.05	VT	ratio_greater_0.5	herbaceuous
MA	planted/cultivated	flooded.frac.max	VA	ratio_greater_1	gff_greater_0.05
MI	population_density	flooded.frac.max	WA	max_ratio	flooded.frac.max
MN	ratio_greater_0.5	gff_greater_0.05	WV	max_ratio	flooded.frac.max
MS	ratio_greater_1	gff_greater_0.05	WI	max_ratio	gff_greater_0.05
MO	ratio_greater_0.2	flooded.frac.max	WY	max_ratio	flooded.frac.max

3.3.3 Cumulative Model Performance Over Stages 1 and 2

The validation approaches in Stages 1 and 2 are stage-specific, and unable to capture cumulative model error over both stages. What makes computing cumulative model error challenging is that if the classifier in Stage 1 misses an actual flood event,

or misses days from an actual flood event (say because the 4-day window does not perfectly align with what occurred), there is no record of those claims to validate the model built in Stage 2. Thus, to evaluate the aggregated performance, we compute three different county-level residuals using 2016 data and the state-level models: (1) the difference between the number of actual claims for the actual event with the most claims over the year and the number of predicted claims for the predicted event with the most claims over the year in each county; (2) the difference between the total number of actual claims and the total number of predicted claims summed over the year in each county and (3) the difference between the number of claims for predicted and actual events in each county with any temporal overlap. The residuals are then used to compute MAE, RMSE, and R^2 for each of the three approaches above.

The first metric determines the degree to which Stages 1 and 2 together are able to predict the peak number of claims in a county over all of its events. The metric is unable to capture whether the event that produced the peak number of actual claims aligns in time with the predicted event that produced the peak number of predicted claims. Results are shown in Table 3.5. The null model is the average of the maximum number of actual claims for all actual events over all counties. The FPR and GFF models resulted in similar performance. The approach made no improvement over the null model for RMSE and the R^2 is low. The high RMSE is driven by the small number of observations that are significantly underpredicted. However, the MAE presents a significant error reduction in comparison to the null model, suggesting most predictions, on average, were better than the null.

Table 3.5. Comparison of the maximum actual number of claims from actual events versus the maximum predicted the number of claims from predicted events using the state-level analysis in 2016 at the county level.

Error metric	Null model	Maximum actual vs. maximum predicted claims		Error reduction compared to null model (%)	
		FPR model	GFF model	FPR model	GFF model
MAE	112.77	62.28	62.60	44.77	44.49
RMSE	471.65	471.57	472.07	0.02	-0.09
R^2		0.05	0.06		

For the second metric, we tally all predicted claims among all predicted events for each county and compare this to the actual number of claims in that county in 2016. The null model is the average of the actual number of claims in each county. Table 3.6 shows the error performance. Comparable to the first metric, the results are similar for the FPR and GFF models. However, the results show significantly better performance ($R^2 \sim 0.22$) and greater error reduction compared to the first performance metric. Thus, while our two-stage modeling approach has some ability to accurately predict the worst events in a county (as measured by the peak of claims), it has a stronger ability to predict the cumulative number of claims that a county experienced over a year. Similar to the first metric, the higher MAE reduction and lower RMSE reduction relative to the null model show that the FPR and GFF models generally perform better for the events with fewer claims.

Table 3.6. Comparison of the total actual number of claims from actual events versus the total predicted number of claims from predicted events using the state-level analysis in 2016 at the county level.

Error metric	Null model	Total actual vs. total predicted claims		Error reduction compared to null model (%)	
		FPR model	GFF model	FPR model	GFF model
MAE	138.76	67.32	66.23	51.48	52.27
RMSE	581.69	560.83	560.15	3.59	3.70
R^2		0.22	0.22		

In the third metric, we identify the predicted and actual events in each county with any temporal overlap, and for all those events in a given county, compare the sum of the number of predicted and actual claims. The goal of this metric is to assess generally how well the two-stage approach is at predicting the correct number of claims at the approximate right point in time. However, it is unable to capture error produced by erroneously predicting a flood event that did not actually occur (and thus potentially predicting erroneous claims), and error produced by predicting an event that did occur, but at a different point in time. The null model is the average of the actual number of claims from the actual events with temporal overlap with predicted events in each county. Table 3.7 shows the model performance. In contrast to the prior metric, the GFF model provides slightly lower error and better model fit in comparison to the FPR model. Although the model fit (in terms of R^2) is lower than the prior metric, the error reduction relative to the null model is higher, particularly for the RMSE. This shows that by focusing on the predicted and actual events with any temporal overlap, the model performance relative to the null model improves. Particularly, the higher reduction of RMSE relative to the null model demonstrates there are fewer predicted outliers (albeit, based on the RSME, the number of outliers is still large).

Table 3.7. Comparison of the total actual number of claims from actual events versus the total predicted the number of claims from only predicted events with some time overlap with the actual events using the state-level analysis in 2016 at the county level.

Error metric	Null model	Actual claims vs. overlapped predicted claims		Error reduction compared to null model (%)	
		FPR model	GFF model	FPR model	GFF model
MAE	65.79	28.82	27.50	56.19	58.19
RMSE	195.17	180.27	178.21	7.63	8.69
R^2		0.16	0.18		

3.4 Discussion and Conclusion

Making accurate and rapid predictions of flood losses is challenging; hydrodynamic models can provide detailed representations of floods but can be difficult to produce with limited time. Flood hazard maps, such as FEMA's flood insurance rate maps, are static and do not reflect current hydraulic conditions. Further, mapping hazard intensity to losses provides additional challenges because of localized differences in exposure. Over time, data-driven approaches, such as the one proposed in this work, may provide long-term value for being able to rapidly and accurately predict local losses, and how to direct state and federal mitigation and recovery resources efficiently.

The present work evaluated the usefulness of two flood heuristics - the FPR and GFF - in accurately predicting localized flood losses, as proxied by the number of National Flood Insurance Program claims. The two-stage approach is novel in that it is the first work, to the authors' knowledge, to develop a data-driven approach to determining when floods severe enough to cause at least one NFIP claim have occurred on a national scale. It is also the first to use the GFF to investigate flood

losses. While the work is not exactly comparable to prior studies that use the FPR to predict insurance claims - due to differences in spatial-scale (ours is coarser), differences in being data-driven versus event-driven (ours is data-driven), and differences in the independent variables that were used - our state-level models found slightly better accuracy in terms of the correlation between the actual and observed variables. There are likely two explanations for this. The first is that we consider non-parametric machine learning models that do not presuppose a specific shape for the response surface, which in this case, leads to better predictive accuracy compared to, say, a negative binomial model. The second explanation is that the state-level models (using county data) are not mired by high variance data that exist in a singular national-level model. Local factors, such as exposure, risk tolerance, and local adaptation to flooding, that drive NFIP claims likely vary significantly across the U.S. The complexity of these factors is challenging to capture in a few simple variables, such as population and NFIP penetration rates. The state-level models, to some degree, are able to weed out the variance stemming from local factors, though this comes at the cost of less data availability.

For the national-level models, the top three important covariates impacting the error reduction of the models are FPR or GFF metrics. Similarly, for most statewide models, the most important covariate in terms of error reduction is one of the FPR or GFF metrics; this is reassuring as these covariates are supposed to proxy flood intensity - the driver of the losses. There are moderate spatial differences where the statewide FPR and the GFF models performed better. The FPR approach showed slightly better predictive performance for most states. Interestingly, the GFF model

provided a higher correlation in states with the most actual claims (i.e., Louisiana, Florida, Texas, and South Carolina), which also happen to be coastal states, while the correlation is higher in midwestern states for the FPR models, where riverine flooding tends to be the driver. Thus, there could be important considerations for where particular metrics are more useful.

There are specific limitations with the metrics and modeling approach that could be addressed. First, there are serious concerns about whether NFIP claims are truly indicative of direct flood losses (Tonn & Czajkowski, 2018). Penetration rates are especially low in some areas, and just because there is not a claim does not mean that there was no flood damage to a house. Stage 1 of our model predicted many counties had at least one NFIP claim during a given month but did not in actuality (i.e., false positives). It is possible that in many of these counties, there was housing damage, but the houses were uninsured. Also, there are likely community characteristics that impact the number of claims, such as construction practices, building codes, local regulations, past flood experience, socioeconomic characteristics, etc., that are not included in our models. This may lead to lower predictive accuracy. Finally, it is likely that more spatially-granular models (e.g., using ZIP-code data as opposed to countywide data) would lead to better predictive accuracy. However, this would be more computationally intensive and we found that the majority of ZIP-codes did not have a sufficient number of historical NFIP claims to develop a reliable Stage 1 classifier.

With the advancements in continental scale hydrological modeling and real-time flood mapping, this line of research could support real-time flood damage

prediction by anticipating flood intensity ahead of extreme events. This could ultimately support the rapid and appropriate deployment of recovery resources in flood-impacted areas. Overall, the models using the FPR and GFF data offer promise considering their relative simplicity, their reliance on publicly accessible data, and their comparatively fast computational speed.

4 Chapter 4: Evaluating the links between public transit disruptions, socioeconomic characteristics, and heavy precipitation

Abstract

This study explores the impacts that extreme events have on the serviceability of infrastructure systems and unpacks the role that social and racial disparities may have on diminished serviceability after extreme events. In particular, this study explores disruptions to public transit systems in response to heavy precipitation events. The public bus system in Baltimore, Maryland is selected as a case study to examine the nexus between bus delays and the city's socioeconomic and demographic characteristics, with and without regard for the precipitation condition. The work considers various spatial autocorrelation analyses and spatial statistical models, and includes various variables including those related to socioeconomic, demographic and social vulnerability, traffic volume, transit system, road connectivity, and the built environment. The results demonstrate a statistically significant spatial autocorrelation between the delays, aggregated at the census tract level, both with and without consideration for precipitation.

4.1 Introduction

Urban floods occur when extreme rainfall events overwhelm the capacity of natural and human-constructed drainage systems and inundate properties in a built environment. Although most do not rise in economic costs comparable to major disasters, they can cause significant economic and social damages (Galloway et al., 2018). In addition to direct damages to the buildings and infrastructure from flooding, such events also cause indirect losses. Heavy rainfalls in urban communities, for instance, can disrupt public transit services that are vital to the well-being of those communities (He et al., 2021).

Rainfall events have a detrimental impact on the performance of transit systems and subsequently can affect the everyday travels of the population. The rainfall events can increase the travel time (Ma et al., 2009) and time travel variability (i.e., cause the service to be less reliable) in comparison to good weather conditions (Tu et al., 2007). Such disruptions may have larger implications for disadvantaged communities, as they are highly reliant on public transit and often lack other means of transportation (Anderson, 2016). Thus, evaluating the impact of rainfall on the transit system and assessing the affected areas are of significant importance.

Many studies have focused on the impact of extreme climatic conditions on solely the performance of public transportation systems without regard to the impacted population and their characteristics (Hofmann & O'Mahony, 2005; Kulkarni & Shafei, 2018; Suarez et al., 2005) or assessed the changes in travel behaviors and ridership in response to rain (Guo et al., 2007; Kashfi Syeed et al., 2013; Stover & McCormack, 2012). However, little is known about the system

performance and the distribution of its disruptions during extreme precipitation events among the neighborhoods with varying socioeconomic and demographic characteristics. To address the above gaps, I evaluate the links between public transit disruptions, socioeconomic characteristics, and heavy precipitation in an urban environment. I hypothesize that the communities that are more socially-vulnerable are susceptible to higher disruptions, induced by heavy precipitation, for public transit. Driving this hypothesis is that infrastructure assets in communities with lower socioeconomic status are often of lower quality and less reliable due to underinvestment and a smaller tax base ("The Unequal Commute," 2020). This has been exacerbated by decades of racially motivated policies including segregation and redlining (Archer, 2021).

In this work, I evaluate the impact of precipitation on on-time performance for a large bus system and the spatial variations of the delays. I then correlate delays to characteristics in the built and social environment. In particular, I look at the delays in the public bus system in the City of Baltimore, Maryland.

This type of inquiry is important because past work has shown that the impact and recovery of the communities in the aftermath of extreme events are greatly influenced by the demographic and socioeconomic characteristics of the communities (Horney et al., 2017; Knighton et al., 2020; Rufat et al., 2015). Most studies, though, only focused on the community outcomes after extreme events (Cutter et al., 2008; Finch et al., 2010; Wang & Ganapati, 2018). However, less is understood about the resilience of urban communities with different characteristics to the low-to-moderate

intensity natural hazards and the outcome of such events on the transit infrastructure and the population affected by potential service disruptions.

This study presents research novelty in a few ways: First, past studies have evaluated the bus disruptions mainly either at the route level or system level. But in this study, I perform the analysis at the bus-stop level and include several bus routes to evaluate the spatial distribution of such disruptions more effectively. Secondly, past studies have focused on evaluating solely the performance of the transit system or the changes in the travel behavior. However, in this study, I am interested in investigating the spatial distribution of transit disruptions among neighborhoods with vulnerable socioeconomic characteristics. Finally, I incorporate precipitation data at fine temporal (hourly) and spatial (4km*4km) resolutions to more accurately associate extreme weather conditions with bus-stop level disruptions.

4.2 Data and Methods

Evaluating the impact of natural hazards on the transit system and assessing the affected areas are of significant importance. Many studies have focused on the impact of extreme climatic conditions on solely the performance of public transportation systems or assessed the changes in travel behaviors and ridership (Kulkarni & Shafei, 2018; Tao et al., 2016). However, little is known about the relationship between socioeconomic and demographic characteristics of the communities and transit disruptions during precipitation events. In order to address this gap, I evaluate the impact of rainfall events on a bus system in terms of delays at the bus stop and variations of such delays among the city's neighborhoods. This is

done by first using spatial clustering that shows the existance of spatial autocorrelation in bus delays. Then, I develop spatial autoregressive models to conduct a statistical analysis that includes demographic and socioeconomic characteristics of impacted communities.

4.2.1 Case Study

The City of Baltimore, Maryland is used as a case study. Baltimore is the largest city in the State of Maryland (and 30th in the USA) with a total population of 585,708 (62% Black, 31% White, and 7% other based on US Census, 2020). Baltimore has a land area of 80.94 mi^2, which is divided into 200 census tracts and 653 census block groups (Figure 4.1). Baltimore is selected as the case study for two reasons: first, there is a significant transit-related inequality where low-income and people of color (also the majority of transit riders) face a higher than average commute times and also worse air pollution from the vehicle emissions (*Transit Equity & Environmental Health in Baltimore*, 2021). Secondly, it was selected due to the availability of on-time performance data at the stop level, provided by the University of Maryland (UMD) Center For Advanced Transportation Technology (CATT).

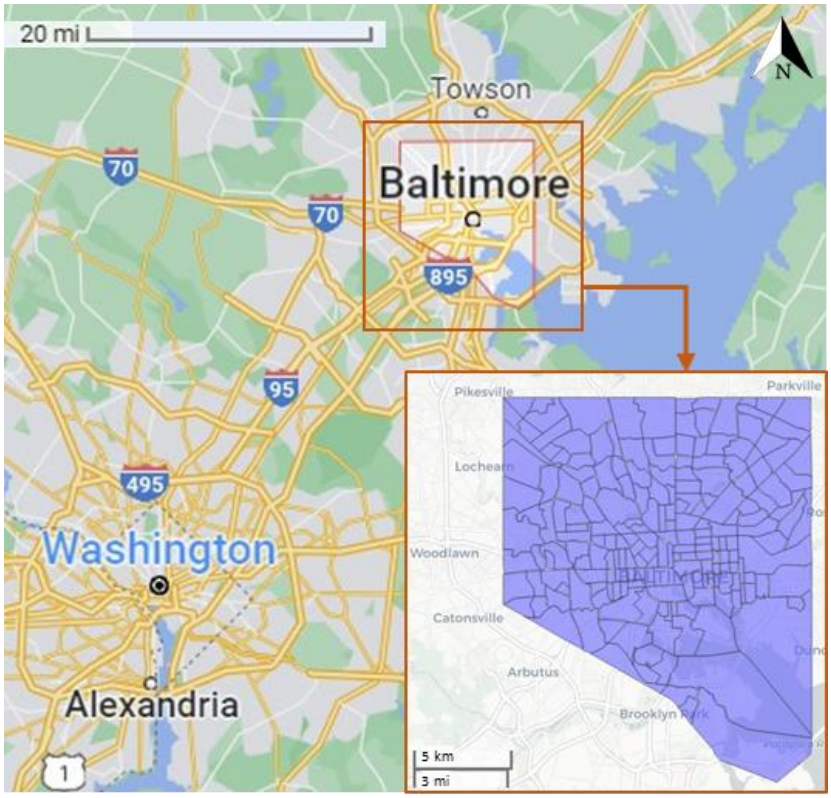

Figure 4.1. Location of the study area and the boundaries of census tracts

4.2.2 Data

To investigate the proposed case study application, several datasets are collected including historical on-time performance data for the bus system, census data, precipitation data, and traffic volume data. Table 4.1 outlines the summary statistics of these characteristics and additional details regarding each data follow.

Table 4.1. Summary statistics of data

Variable Type	Description	Min.	Max.	Mean	Temporal Resolution	Unit	Source
Dependent Variable	Median of bus delays at the census tract level	0	0.97	0.387	each bus arrival	sec./sec.*	MTA
	Difference between the median of delays during wet and dry periods	-0.14	0.19	0.005	each bus arrival	sec./sec.*	MTA
Independent Variable	Max non-highway traffic	275	130,685	16,770	year 2018	count	NaNDA
	Stops per capita	0.537	30.046	5.210	year 2018	count/person	NaNDA
	Census tract area	0.067	3.863	0.383	year 2010	mi^2	NaNDA
	Street network density	18,176	91,145	56,958	year 2010	m/mi^2	NaNDA
	Imperviousness	0.223	0.966	0.645	year 2016	fraction	NLCD
	Public transp. to work	0.028	0.616	0.222	year 2018	fraction	CDC SVI
	Below poverty	2.100	61.400	24.630	year 2018	%	CDC SVI
	Unemployed	0.500	32.500	10.570	year 2018	%	CDC SVI
	No diploma	0.000	43.500	17.110	year 2018	%	CDC SVI
	Aging (65+)	1.200	26.900	12.000	year 2018	%	CDC SVI
	Aging (17-)	1.100	50.700	21.180	year 2018	%	CDC SVI
	Disability	2.500	36.500	15.880	year 2018	%	CDC SVI
	English (less than well)	0.000	24.500	2.156	year 2018	%	CDC SVI
	Race (non-white)	0.052	1.000	0.733	year 2018	fraction	CDC SVI
	Population density	1,195	33,791	12,472	year 2018	$person/mi^2$	CDC SVI

** The bus delay is defined as the ratio of the delay in the preceding route segment (in sec.) prior to the stop divided by the scheduled travel time (in sec.) of the same segment. Thus, the unit of the bus delay is sec./sec.*

Historical bus data: The historical on-time performance data from the Maryland Transit Administration (MTA), which manages and oversees Baltimore's transit system, is provided by the University of Maryland (UMD) Center For Advanced Transportation Technology (CATT). The data is for the entire CityLink

bus system and for the October 1 through December 31, 2019 timespan. Figure 4.2 shows the spatial network of the CityLink bus system. Within the CityLink bus network, there are 12 routes containing a total of 1,428 bus stops. The data reports the time at which buses *should* have arrived at each bus stop and when buses actually arrive.

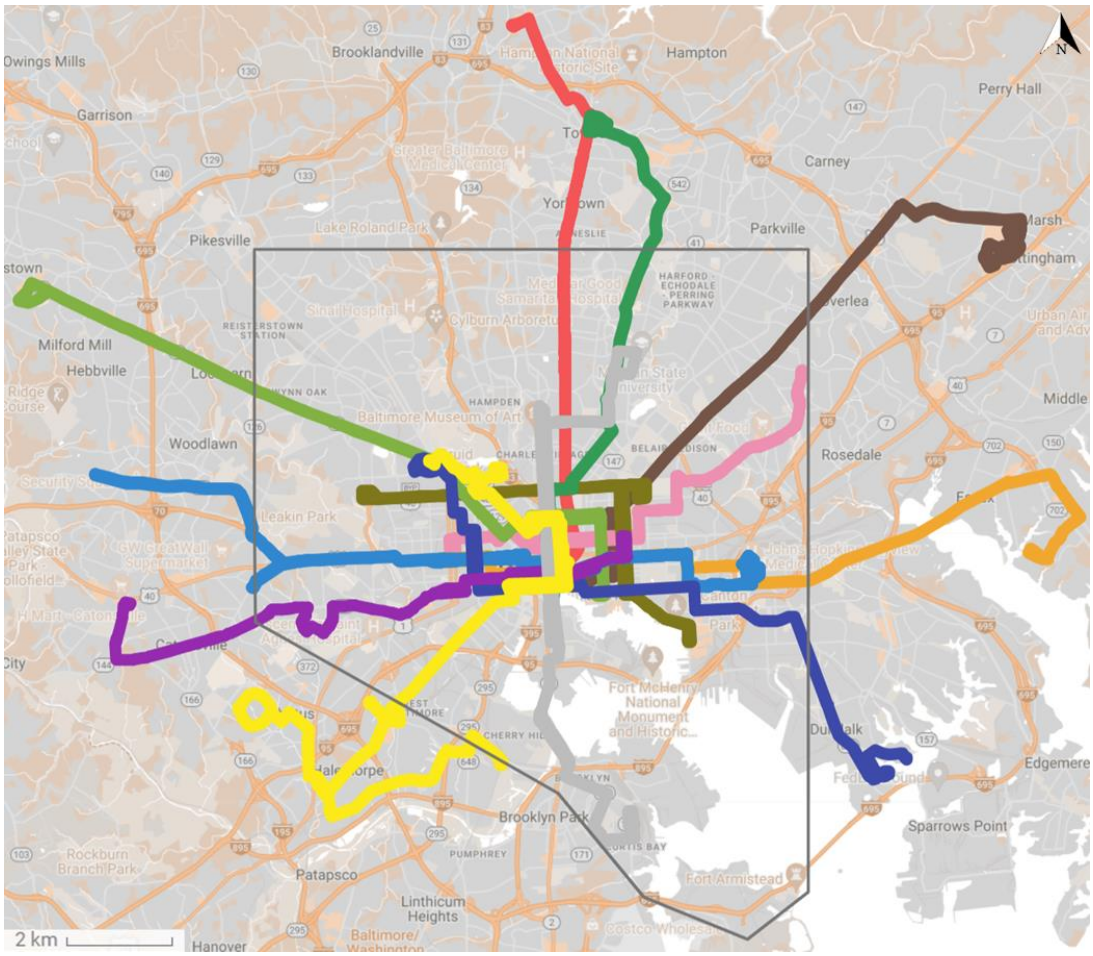

Figure 4.2. Spatial network of the CityLink bus system in Baltimore City.

In this study, I use the delays at the bus stop level as a mean of assessing the performance of the transit system. One way to define a bus delay is to identify the actual arrival time and the scheduled arrival time of a bus arrival at a bus stop, and

then compute its time difference (absolute bus delay). However, I am interested in localized delays at the stop level. So, this definition is not suitable because it is not fair to penalize the local area for incidents that might have occurred farther away and caused a delay down the road, for example a delay at the origin. Consistent with Chu (2014), I define the bus delay as the ratio of the delay in the preceding route segment prior to the stop divided by the scheduled travel time of the same segment.

$$\text{Bus Delay} = \frac{\text{actual travel time} - \text{scheduled travel time}}{\text{scheduled travel time}}$$

I categorize the delays into three categories depending on the time of the day that they occurred, namely morning peak (6-10 am), mid-day off-peak (10 am-2 pm), and evening peak (2-6 pm). Each category is analyzed separately. Figure 4.3 and Figure 4.4 outline the distribution of scheduled travel times (in seconds) and the distribution of segment lengths (in miles) of all segments in the network of the CityLink bus system, respectively. From these figures, for example, 90% of all segments have travel times less than 101 seconds and segment lengths less than 0.54 miles.

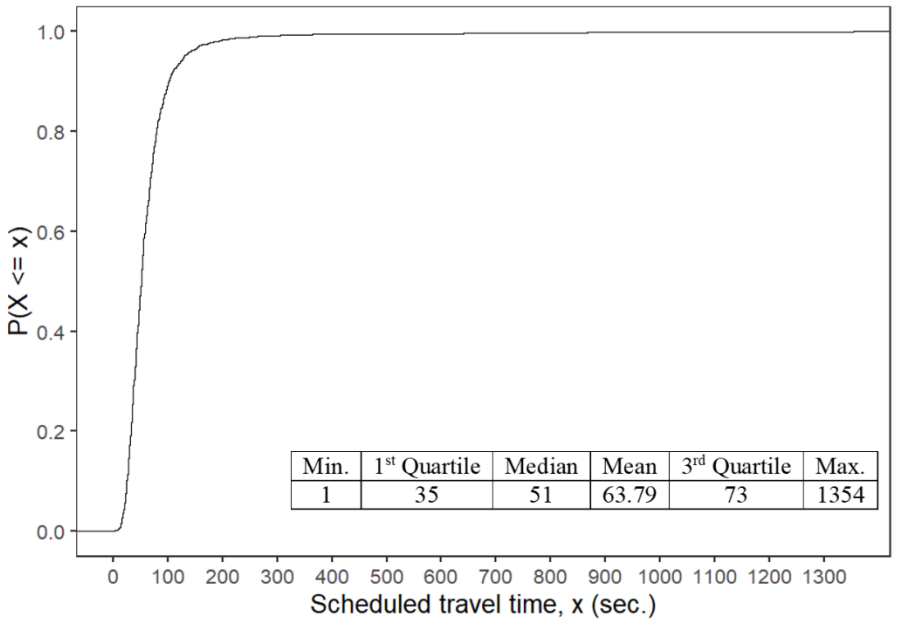

Figure 4.3. Empirical cumulative distribution function (eCDF) and distribution of scheduled trip times of all the segments.

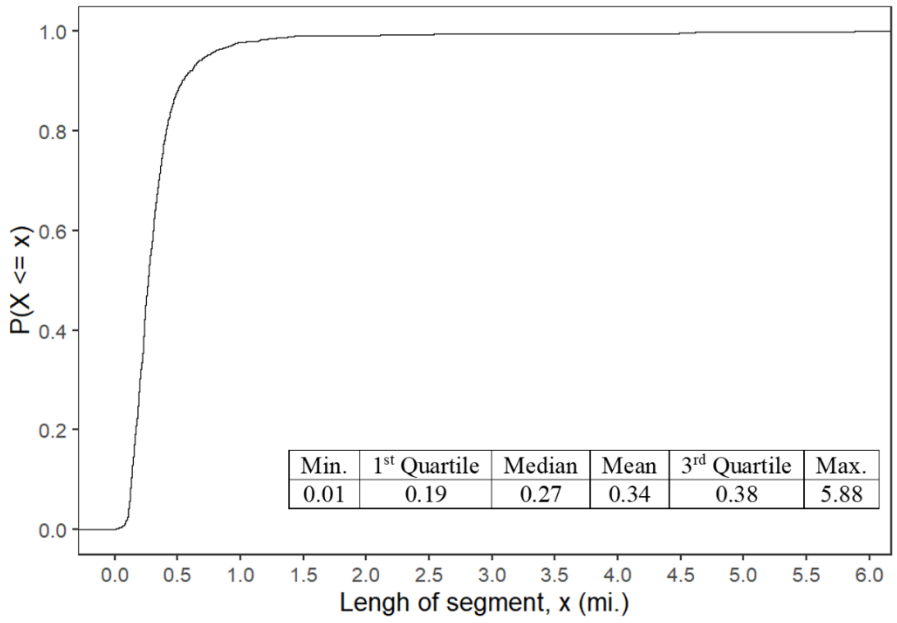

Figure 4.4. Empirical cumulative distribution function (eCDF) and distribution of lengths of all the segments.

Precipitation: Precipitation data are collected from the Stage-IV product, which is produced by National Centers for Environmental Prediction (NCEP). The NCEP Stage-IV is based on the multi-sensor precipitations developed by 12 River Forecast Centers (RFCs) for the entire contiguous US. RFCs are responsible for providing critical hydrologic and hydrometeorological forecasts and data to both technical and public partners. The NCEP Stage-IV product is mosaicked on the 4km (local polar-stereographic) grids and the precipitations are available hourly, 6-hourly, and 24-hourly. I used the hourly precipitation data from October 1 through December 31, 2019, and crop the collected data to Baltimore City.

Traffic volume: The data on traffic volume is obtained from National Neighborhood Data Archive (NaNDA, 2018). It provides measures of traffic volume at the census tract level from 1963 to 2019. The available measures are average, maximum, and minimum traffic volume for all streets, highways, and non-highways per tract per year. I use the maximum and minimum average daily non-highway traffic count in a census tract because transit buses usually use non-highway roads. Also, the count of transit stops per capita and per square miles are included to capture the density of the transit system in census tracts.

Demographic and socioeconomic: The impact of hazards on the communities, whether natural or man-made, can highly depend on the characteristics of the communities (Cutter & Finch, 2000). Several characteristics such as poverty, unemployment, low income, and lack of transportation can affect the ability of a

community to prepare, respond, recover, and adapt (i.e., the resilience) to the hazardous event, thus making them socially vulnerable (Flanagan et al., 2018; Lehnert et al., 2020). As part of the development of the U.S. Centers for Disease Control and Prevention (CDC)'s Social Vulnerability Index (SVI), 15 factors are listed as the characteristics defining the overall vulnerability of the communities which will most likely need support before, during, and after hazardous events (CDC SVI, 2018). They can be categorized under four themes namely socioeconomic status, household composition and disability, minority status and language, and housing and transportation. To achieve a more parsimonious dataset, a subset of nine factors are selected as they are deemed to be more relevant and also with some commonality of variables with another Social Vulnerability Index (SOVI) developed by Cutter, Boruff, and Shirley (Cutter et al., 2003). These nine factors are: fraction of non-white population, below poverty (%), unemployment (%), no high school diploma (%), aging (65+) (%), aging (17-) (%), disability (%), English (less than well) (%), and use of public transportation to get to work (%).

4.2.3 Methods

I investigate two approaches separately to evaluate the distribution of bus delays and the impact of rainfall on bus delays. In the first approach (Approach-I), I use the median bus delay of all bus stops in a census tract without regard for precipitation, as the variable of interest. In the second approach (Approach-II), I incorporate the impact of rainfall by assessing the difference between the median bus delays during “wet periods” and the “dry periods” in a census tract. A “wet period” is

a one-hour period where any portion of the study area measured precipitation greater than 0.025 inches (equivalent to 0.635 mm); otherwise, the hour is considered "dry." Several other precipitation thresholds were evaluated, and 0.025 inches showed high statistical significance in spatial analysis. The following analyses are conducted on both approaches.

4.2.3.1 Spatial Analysis

Moran's I spatial autocorrelation: In order to evaluate the distribution of bus delays spatially, I implemented a global spatial autocorrelation technique namely Global Moran's I (Moran, 1948). This technique identifies whether there is a Global autocorrelation effect. A positive indicator shows that the variable of interest is spatially clustered (i.e., a positive spatial autocorrelation) or conversely, a negative indicator explains that the variable of interest is spatially dispersed (i.e., a negative spatial autocorrelation). Similar to other correlation techniques, the range of Moran's I indicator is from -1 (highest negative spatial autocorrelation) to +1 (highest positive spatial autocorrelation). Moran's I indicator close to zero indicates that the variable is spatially random (i.e., no spatial autocorrelation). The Global Moran's I can be computed by (Anselin, 1995).

$$I = (n/S_0) \sum_i \sum_j w_{ij} z_i z_j \Big/ \sum_i z_i^2$$

Where n is the number of observations, w_{ij} are weights, μ is the mean of the variable, X_i and X_j are the variable value at location i and j respectively, $z_i = X_i - \mu, z_j = X_j - \mu$, and $S_0 = \sum_i \sum_j w_{ij}$.

The Global Moran's I defines one indicator of spatial autocorrelation for the entire study area. To further investigate the spatial autocorrelation and identify the location of spatial clustered, I used a local spatial autocorrelation technique namely Local Moran's I which is defined as (Anselin, 1995).

$$I_i = z_i \sum_j w_{ij} z_j$$

Local Moran's I is capable of identifying the clustered areas with positive spatial autocorrelation (i.e., high-high or low-low values), negative spatial autocorrelation (i.e., high-low or low-high), and areas with no statistically significant spatial autocorrelation (Anselin, 2005). In order to create a spatial weight matrix, I use Queen contiguity with row-standardized weights. Queen contiguity refers to census tracts that share either a border or a point and row-standardized weight indicates that the weight for each spatial unit (i.e., a census tract) is standardized to 1. Thus, if a census tract has n neighbors, each of the neighbors will weight of 1/n.

4.2.3.2 Spatial Statistical Analysis

To investigate the relationship between the bus delays and the characteristics of the communities, I develop a statistical analysis. In case I encounter a spatial autocorrelation effect of the dependent variables, then performing a spatial statistical analysis is deemed more appropriate than a non-spatial statistical analysis. In this section, I am interested to answer the question: What is the relationship between the demographic and socioeconomic characteristics of the neighborhoods and the bus delays with and without the influence of precipitation events? Also, in the case of

existing spatial autocorrelation, will the spatial statistical models perform better than the non-spatial models? To answer these questions, I use two spatial statistical models the spatial autoregressive model (SAR), and the spatial errors model (SEM), along with a non-spatial ordinary least squares model (OLS). In the following, I briefly describe each model.

Ordinary Least Squares (OLS): OLS method can be used to estimate the unknown parameters in a linear regression model (Hastie et al., 2009). It uses the observed outcomes in the dataset and the predicted outcomes from a linear approximation and minimizes the sum of squared vertical distances between them. Then the estimated parameters (a.k.a. coefficients) can be used to describe the relationship between the dependent variable and independent variable(s).

Spatial autoregressive model (SAR): spatial autocorrelation can be modeled using a functional relationship between a variable (y) or error term (ε), and its associated spatial lag which is W_y for a spatially lagged dependent variable (Anselin & Bera, 1998). Spatial lag dependence in a regression model (a.k.a. mixed regressive) can be defined as

$$y = \rho W_y + X\beta + \varepsilon$$

Where y is an N by1 vector of observations on the dependent variable, X is an N by K matrix of observations on the independent variables, W_y is the corresponding spatially lagged dependent for weight matrix W, ε is an N by 1 vector of error terms, β is a K by 1 vector regression coefficients, and ρ is the spatial autoregressive parameter. The inclusion of the spatial lag term W_y will provide a non-zero correlation with the error term (ε) which is similar to the presence of an endogenous

variable. If one develops an OLS model and a spatial lag process is presented, the estimated coefficients will be biased and inefficient. This leads to underestimating the standard errors and the coefficient sizes and their associated signs not being close to their true value (Anselin & Bera, 1998).

Spatial errors model (SEM): another way for the spatial autocorrelation to be accounted for in a regression model is to specify a spatial process for the disturbance terms (Anselin & Bera, 1998). In this case, the OLS estimates are unbiased and inefficient because the resulting error covariance will be nonspherical. Spatial autocorrelation can be modeled using a functional relationship between a variable (y) or error term (ε), and its associated spatial lag which is W_ε for a spatially lagged error term. The spatial error dependence can be defined as

$$y = X\beta + \varepsilon$$

where $\varepsilon = \lambda W_\varepsilon + \zeta$ and λ is the spatial autoregressive coefficient for the error lag W_ε, and ζ is an uncorrelated error term.

Independent variables: An initial set of 30 variables are selected for the statistical modeling related to socioeconomic, demographics, social vulnerability, traffic volume, transit system, road connectivity, and the built environment. In order to eliminate the potential redundancy between the independent models, a correlation analysis is performed over all independent variables and the variables with a correlation higher than 75% are eliminated (Supplementary Table C.1). In addition, the variance inflation factor (VIF) is computed to evaluate multicollinearity among the covariates. While highly correlated features do not influence the predictive power of the model, they do potentially impede model interpretation. To eliminate this issue,

features with high VIF scores are removed until all features have a VIF value below 10 (Alin, 2010). In order to compare the coefficients of the independent variables in a model's results, each variable is standardized by subtracting the mean of the variable from each observed value (centering) and dividing by the standard deviation of the same variable (scaling).

4.3 Results

4.3.1 Approach I: Entire Bus Delays With Precipitation Influence

4.3.1.1 Spatial Analysis

I break the analysis into different times of the day namely Morning, Midday, and Evening over all weekdays during the study period. Table 4.2 summarizes the results of the clustering analysis for each analysis. It can be seen that the category of Evening on weekdays demonstrates a spatial clustering effect among census tracts based on the Global Moran's I spatial autocorrelation analysis. Thus, the case of Evening was selected for further investigation due to the higher Moran's I value (i.e. more clustering effect). The Global Moran's I analysis resulted in Moran's I value of 0.11 (P-value < 0.05), indicating that it rejects the null hypothesis of no spatial correlation. Thus, non-spatial models that disregard the spatial correlation between the observations (for example linear model) are not appropriate here because they cannot capture the implicit spatial relationship between the observations. For that reason, the implementation of spatial models such as SAR and SEM is needed (additional details in Section 4.2.3.2).

Table 4.2. Spatial autocorrelation analysis using Global Moran's I

Time	Moran's I value	P-value	Clustering?
Morning	0.063	0.0812	Random
Midday	-0.006	0.4933	Random
Evening	0.110	0.0096	Clustered

An investigation into the spatial distribution of the delays outlines the critical areas susceptible to the highest amount of delays during the Evening on weekdays. Figure 4.5a demonstrates the median delays of bus arrivals among all bus stops in a census tract. It shows that the bus delays are fairly distributed around Baltimore City and the majority of the city is susceptible to bus delays. Nonetheless, there are distinct areas with high amounts of delays, particularly downtown and areas further away from the downtown area. Although a visual evaluation of Figure 4.5 may not show a clustering pattern, however, as shown previously, indeed there is a spatial clustering pattern of delays around the city based on the Global Moran's I.

In order to identify the areas with spatial clusters of delays at the census tract level, I use the Local Moran's I technique (Figure 4.5b). It reveals that there is a high spatial clustering of bus delays in the northwest, southeast, and a smaller area in the south of the city. Identifying the areas with spatial clusters of delays using visualization in Figure 4.5a,b can be useful for transit officials to locate the critical regions of delays in the bus system.

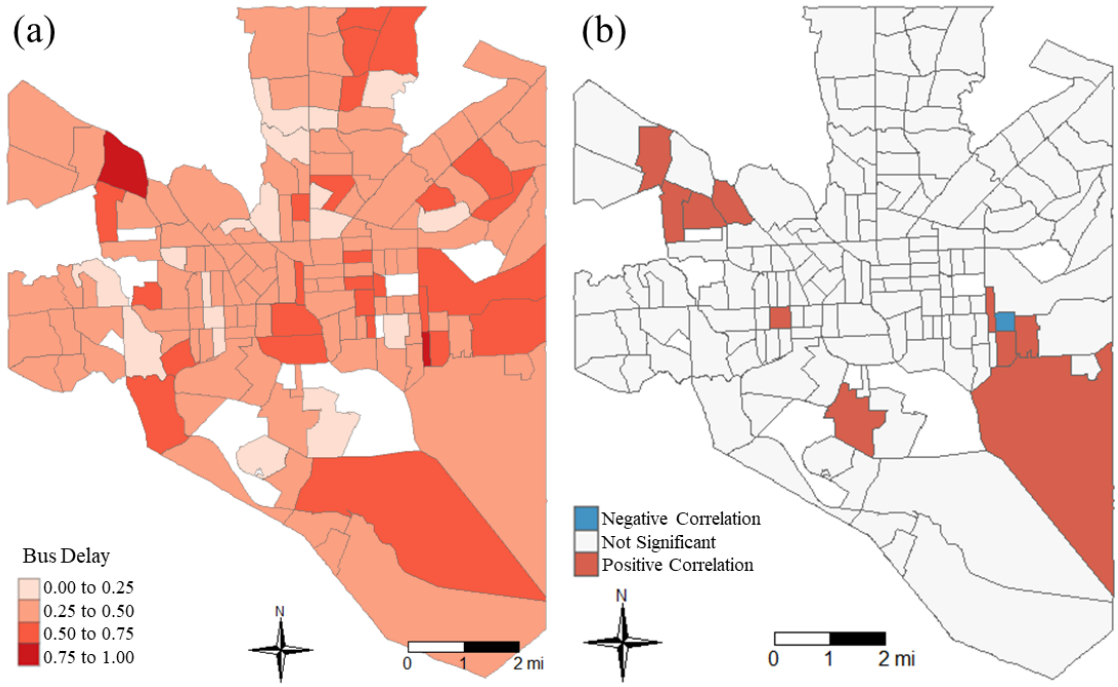

Figure 4.5. a) Median of bus delays at census tract level during evening period on weekdays. b) Clusters of bus delays identified using Local Moran's I.

4.3.1.2 Spatial Statistical Analysis

The spatial analysis in the previous section demonstrated an inherent spatial relationship between the observations. This indicates that the non-spatial models (e.g. linear model) may not be appropriate since they do not capture the spatial relationships. To further investigate the association between the bus delays and socioeconomic and demographic characteristics of the neighborhoods, I perform spatial statistical analysis using SAR and SEM models and compare them with the OLS model. Table 4.3 details the significance of the model parameters used. ρ and λ are the parameters in spatial autoregressive models. The lag parameter in the SAR model (ρ) has the value of 0.139 and is not statistically significant. This may suggest that the explanatory variables have accounted for the spatial lag in the response

variable and a spatial lag on the response variable is not needed. Similarly, the lag error parameter in the SEM model (λ) has a value of 0.093 and is not statistically significant, suggesting that I do not need to account for the spatial autocorrelation in the error. On the other hand, the OLS model is statistically significant at a significance level of 5% and an adjusted R^2 of 0.11. This indicated that the OLS model is more appropriate in this case.

Table 4.3. Significance of model parameters (dependent variable: median of delays during the entire study period at the census tract level)

Parameters	SAR (P-value)	SEM (P-Value)	Parameters	OLS
ρ	0.139 (0.26)	-	Adj-R^2	0.11
λ	-	0.093 (0.55)	P-value	0.01

Table 4.4 presents the results of regression models. Although I found the SAR and SEM models not to be statistically significant, I included their results for comparison to the OLS. In the OLS model, it is found that 5 out of 15 variables show statistically significant effects (at a significance level of 10%) on the amount of bus delays. The variable "Imperviousness" is statistically significant with a positive coefficient implying that the higher amount of built environment and development, the higher amount of delays. "Street network density" shows that the higher the density of streets per square mile in a census tract, the higher amount of the delays. Also, the amount of delays' increase with the increasing "census tract area" suggesting that the larger census tracts are susceptible to higher delays. Among the variables accounting for the vulnerability of neighborhoods, the two variables of "Race (non-white)" and "English (less than well)" have a positive and significant

relationship with the amount of delays. This finding implies that the census tracts with a higher percentage of the non-white population and lower English proficiency are reported to have an increasing amount of delays.

Table 4.4. Results of spatial and non-spatial statistical models

Variables	OLS		SAR		SEM	
	β	*P-value*	*β*	*P-value*	*β*	*P-value*
Intercept	0.387	0.00 *	0.333	0.00 *	0.387	0.00 *
Max non-highway traffic	-0.011	0.41	-0.012	0.34	-0.010	0.42
Public transp. to work	-0.006	0.71	-0.005	0.77	-0.005	0.77
Stops per capita	0.023	0.11	0.022	0.11	0.022	0.12
Census tract area	0.036	0.02 *	0.035	0.02 *	0.034	0.02 *
Imperviousness (%)	0.032	0.09 *	0.030	0.08 *	0.031	0.08 *
Street network density	0.040	0.03 *	0.038	0.03 *	0.039	0.03 *
Below poverty (%)	-0.020	0.32	-0.019	0.33	-0.018	0.34
Unemployed (%)	-0.026	0.12	-0.027	0.08 *	-0.027	0.09 *
No diploma (%)	-0.024	0.28	-0.023	0.27	-0.025	0.24
Aging (65+) (%)	0.014	0.34	0.013	0.32	0.013	0.32
Aging (17-) (%)	0.004	0.81	0.003	0.86	0.002	0.90
Disability (%)	-0.006	0.73	-0.007	0.70	-0.007	0.69
English (less than well) (%)	0.031	0.03 *	0.029	0.03 *	0.030	0.02 *
Race (non-white)	0.037	0.08 *	0.037	0.06 *	0.037	0.06 *
Population density	-0.018	0.40	-0.017	0.39	-0.018	0.37
Adj-R^2	0.110	0.01	-	-	-	-
ρ	-	-	0.139	0.26	-	-
λ	-	-	-	-	0.093	0.55

β Estimated coefficient, * *P-value* < 0.1

4.3.2 Approach II: Difference in Bus Delays between Wet and Dry Periods

4.3.2.1 Spatial analysis

In order to investigate the impact of precipitation events on the amount of delays, we compute the difference between the median of delays during wet periods and the median of delays during dry periods. I start by evaluating the Global Moran's I index for different times of the day. Similar to section 3.1, I found that the Evening category presents the highest spatial clustering effect among the census tracts using the Global Moran's I technique. Thus, the case of Evening was selected for further investigation due to the higher Moran's I value (i.e. higher clustering effect). Table 4.5 outlines the details of the spatial autocorrelation analysis of the difference in the delays. The result for spatial clustering analysis shows a positive spatial autocorrelation of delays with Moran's I value of 0.093 (P-value < 0.05) which means that the null hypothesis of no spatial correlation is rejected. Therefore, one cannot disregard the spatial correlation between the observations and use non-spatial models such as the Linear model. In order to capture the inherent spatial relationship between the observations, I utilize two spatial models namely SAR and SEM (additional details in Section 4.2.3.2).

Table 4.5. Spatial autocorrelation analysis using Global Moran's I

Time	Moran's I value	P-value	Clustering?
Morning	-0.031	0.6897	Random
Midday	-0.011	0.5448	Random
Evening	0.093	0.0213	Clustered

Figure 4.6a outlines the spatial distribution of the difference between the median of delays during wet periods and the median of delays during dry periods at the census tract level. It reveals that some census tracts (i.e. 61 census tracts) have a lower median of delays during the wet periods than the median of delays during the dry periods. However, the majority of census tracts (i.e. 89 census tracts) have a median of delays during wet periods of higher or equal to the median of delays during dry periods. Also, the majority of census tracts (i.e. 134 census tracts) have a difference of delay medians between -0.05 and 0.05 which indicates that the precipitation events do not have a significant impact on the bus delays at the census tract level. Nonetheless, there are distinct areas with either much higher or much lower amounts of delays during wet periods in comparison to dry periods, particularly the mideast and the outskirts of the city. Although a visual inspection of Figure 4.6a may not reveal a spatial clustering, however using the Global Moran's I above, I showed the existence of spatial autocorrelation.

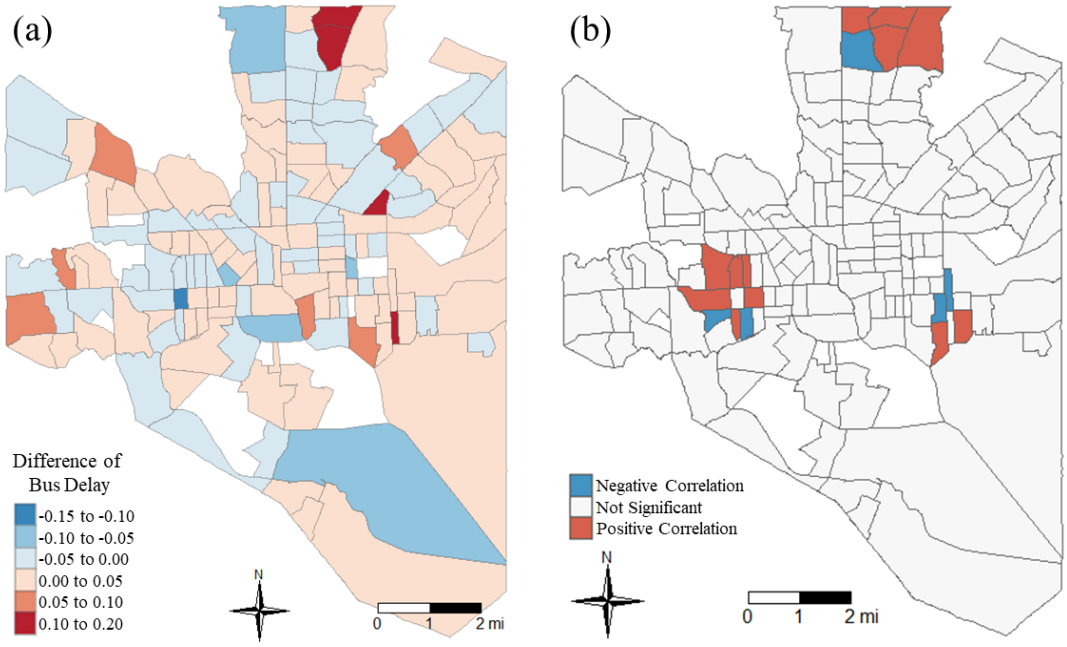

Figure 4.6. a) Difference between the median of delays during wet periods and the median of delays during dry periods at the census tract level. b) Clusters between the median of delays during wet periods and the median of delays during dry periods were identified using Local Moran's I.

Figure 4.6b demonstrates the areas with spatial clustering of differences between the median of delays during wet periods and the median of delays during dry periods at the census tract level using Local Moran's I technique. It shows high clustering in three areas including the midwest, midwest, and north of the city. Most of the census tracts display a positive spatial autocorrelation meaning that in these census tracts, similar values to their neighboring census tracts. This can be helpful to transit officials to identify the areas where the bus delays are exacerbated by the occurrence of precipitation events.

4.3.2.2 Spatial statistical analysis

Based on Moran's I in the previous section, I discovered the spatial autocorrelation of the dependent variable (i.e. differences between the median of

delays during wet periods and the median of delays during dry periods at the census tract level) is statistically significant (but not strong). This justifies the necessity of investigating the spatial statistical models (such as SAR and SEM) as the non-spatial models do not capture the spatial relationships. To evaluate the association between the socioeconomic and demographic characteristics and the dependent variable, I developed three models including SAR, SEM (spatial model), and OLS (non-spatial model). Table 4.6 details the significance of the model parameters used. I find that neither the spatial models (SAR and SEM) nor the non-spatial model (OLS) resulted in statistically significant models. The lag parameter in the SAR model (ρ) and the lag error parameter in the SEM model (λ) are 0.140 and 0.106, respectively, suggesting that the explanatory variables have accounted for the spatial lag and the spatial autocorrelation in the error. The adjusted R^2 in the OLS model has the value of -0.03 and the model is not statistically significant, indicating the insignificance of explanatory variables to explain differences in the response variable.

Table 4.6. Significance of model parameters (dependent variable: differences between the median of delays during wet periods and the median of delays during dry periods at the census tract level)

Parameters	SAR (P-value)	SEM (P-Value)	Parameters	OLS
ρ	0.140 (0.27)	-	Adj-R^2	-0.03
λ	-	0.106 (0.47)	P-value	0.81

4.4 Discussion and Conclusion

In this study, the distribution of bus delays was investigated spatially and statistically, with and without including the influence of precipitation events, in order to understand the regional variations of the bus delays in the transit system and in

different socioeconomic and demographic communities. It is accomplished by analyzing the delays in bus arrivals in Baltimore City at the census tract level for three months. Two dependent variables are considered in separate approaches including the median of delays during the entire study and the difference in median of delays between wet and dry periods. Starting with spatial analysis, the areas with the highest amount of delays and their spatial clusters were identified, with and without precipitation impact. Then, spatial statistical models were used to investigate the relationship between the aforementioned dependent variables and variables related to socioeconomic, demographics, social vulnerability, traffic volume, transit system, road connectivity, and the built environment.

I found that although there were spatial autocorrelation effects in evaluating the distribution of both dependent variables, however, only the OLS model for the median of delays resulted in a statistically significant model. This may suggest that the spatial autocorrelation effect was not strong enough to impact statistical analysis and justify the implementation of spatial statistical models. Nonetheless, I discovered that two characteristics of the socially vulnerable communities have a significant relationship with the amount of delays. It implies that the census tracts with a higher percentage of non-white and lower English proficiency are reported to have an increasing amount of delays.

By analyzing the relationship between bus delays, with and without accounting for precipitation, and the characteristics of the communities, I identified communities with disproportionate delays and their unique characteristics. For the decision-makers with the goal of improving the performance and accessibility of

public transit for the disadvantaged groups, this study can be useful to identify the socioeconomic and demographic characteristics that might be associated with the disproportionate impact of transit disruptions. In addition, improving the performance of the transit system by reducing the delays may result in higher ridership (higher revenue), less use of personal vehicles (less traffic), and less environmental pollution.

It is noteworthy to mention some of the limitations of this study. The CityLink system is not the only bus system operating in Baltimore City. There are other local bus routes as well as commuting bus routes providing service in the city. Also, the CityLink system does not cover all census tracts in Baltimore City. However, only the historical data of the CityLink bus system was provided for this study. Future work may include the data on the other bus systems in the study. As mentioned earlier, I did not investigate the reasons behind the bus delays, other than considering the impact of precipitation, but rather evaluated the spatial distribution of the delays and the impacted neighborhoods. Subsequent research may consider additional variables relevant to the bus delays including road conditions, bus conditions, the existence of road constructions, and stormwater drainage systems. Such data was not publicly available to be collected at the time of this study.

5 Chapter 5: Conclusions and Future Work

5.1 Conclusions

Implementing data-driven methods is a promising approach to rapidly elucidate the post-disaster losses and the resilience of impacted communities and infrastructure. By proving the usefulness of these methods, even for particular areas or infrastructure systems, it enables a new branch of flood hazard research in which researchers can use simplified, data-driven flood metrics to forecast localized impacts. This would provide the decision-makers at entities such as FEMA and public transportation agencies with the tools to rapidly and effectively allocate the resources necessary to reduce the impact of natural hazards on the communities, and infrastructure systems and services.

Chapter 2 systematically evaluated the distribution of FEMA Public Assistance outlays following flooding disasters to assess the evenness of program administration. This is the first work to understand the salient features that contribute to aid and I identify previously unidentified reasons for why outlays may be increasing – including jurisdictions that experience repetitive hazards becoming more skilled at grant application. This work could ultimately support programmatic improvements and policy reforms for aid for infrastructure following disasters.

Chapter 3 developed an integrated data-driven approach for rapidly predicting continental-scale flood losses using two flood intensity heuristics – the flood peak ratio and the Giovanni flooded fraction. Prior to this work, these heuristics, have not been evaluated in terms of their ability to predict whether a flood with significant

enough intensity to result in losses has occurred, nor has a comparison been conducted to understand the value of these heuristics. This research could ultimately support rapid and appropriate deployment of recovery resources in flood-impacted areas.

In Chapter 4, the spatial distribution of transit disruptions (in this case bus delays) was investigated using various spatial and spatial statistical analyses by incorporating the influence of precipitation. To the best of my knowledge, this is the first study to evaluate the bus delays at stop level and investigate the relationship of such disruptions with the characteristics of the neighborhoods with and without incorporation of precipitation impact. In addition, in contrast to prior works, the finer spatial and temporal resolution of the collected data in conjunction with the selection of a long study period enables a more accurate assessment of such relationship between the delays and the neighborhoods' characteristics. This study could support the decision-makers with the goal of improving transit performance, the tool to pinpoint the critical areas of high or clustered delays, and identify the characteristics that might be associated with disproportionate disruptions.

5.3 Future Work

Some recommendations for future studies are detailed below based on the results of the work presented in this dissertation:

Future Work for Chapter 2

Under the first study, county-median income was the only capacity indicator that was identified as important. Specifically, counties with a median income below $50,000 have more applicants than more well-resourced counties. One explanation could be that FEMA or state officials provide additional personnel and resources to support PA applications in low-resourced counties. An alternative explanation is that low-resourced communities tend to be located in more hazardous areas, have fewer resources before the disaster to engage in mitigation, and have lower-quality infrastructure due to chronic underinvestment - and thus have more destruction and more need for aid after disasters. An interesting next step could be to explore the relationship between median income and receipt of a presidentially declared disaster (PDD) itself. Counties in the highest income decile tend to have marginally more applicants than counties in the middle deciles, perhaps suggesting that well-resourced counties do systematically focus on procurement of aid (Smith et al., 2013). Regardless, these findings bring up pressing questions about disaster aid and procedural equity.

Further, I hypothesized that factors that partially influence the level of aid that is received locally include local capacity and regional policies. Many researchers have found that local capacity influences disaster preparedness which, in turn, correlates with local outcomes (Cigler, 2007; Cutter et al., 2010; Norris et al., 2008). In light of that, this work uses a variety of local indicators that are popular for predicting local capacity and local learning over time to plan for and recover from disasters (e.g., educational attainment, income, demographics, disaster experience).

However, additional factors related to local capacity such as the availability of permanent FEMA emergency staff and the existence of a comprehensive disaster plan to respond at the local level, which were not readily available at the time of this study, may shed light on the experience of local governments in receiving disaster recovery.

Future Work for Chapter 3

Chapter 3 evaluated the usefulness of two data-driven approaches in rapidly and accurately predict local losses from flooding events, to deliver more effective estimates than using the existing approaches (with many limitations) such as hydrodynamic models or flood hazard maps. Subsequent work may consider devising a comparative analysis between the application of these new data-driven approaches and the hydrodynamic models developed for a region.

In this study, several variables are considered as the predictors in regression models mainly related to the characteristics of land cover, population, and NFIP penetration rates. However, there are likely community characteristics that impact the number of claims, such as construction practices, building codes, local regulations, past flood experience, socioeconomic characteristics, etc., that are not included in our models which may cause lower predictive accuracy. In future work, additional parameters such as the above characteristics may be included to further examine the local factors such as adaptation practices, exposure, risk perception, etc.

Future Work for Chapter 4

In chapter 4, spatial and spatial statistical methods were used to investigate the relationship between the public transit disruptions and variables related to socioeconomic, demographics, social vulnerability, traffic volume, transit system, road connectivity, and the built environment. This study did not investigate the salient reasons behind the transit disruptions comprehensively, other than considering the impact of precipitation, but rather evaluated the spatial distribution of the delays and the impacted neighborhoods. Subsequent research may consider additional variables, which might attribute to such disruptions including variables related to road conditions, bus conditions, the existence of road construction, and stormwater drainage systems. Such data were not publicly available at the time of this study.

This dissertation aims at examining the interactions between the natural hazardous events, which impact the individuals and societies, and the ways that such events affect the built environment and infrastructure systems in order to contribute to the broader area of disaster resilience. Understanding these interactions allows us to evolve the existing practices and develop new strategies in the built environment to diminish the risk of natural hazards. With technological advancements in computational tools and the introduction of robust statistical models, the data-driven approaches have the potentials to address many of our nowadays issues, which were previously considered infeasible to be solved.

A Appendix: Supplementary Material Chapter 2

Summary of data

Table A.1. List of 56 variables initially considered, along with their abbreviations. The five columns on the right provide variable summary statistics and data sources. The asterisks and carat at the end of the variable name indicate whether they were included after feature reduction. ("" for the number of applicants, "^" for the number of projects).*

Variable	Description	Min.	Median	Mean	Max	Data Source
Soil_moist_10cm_1day*	Average soil moisture content 1-day preceding disaster arrival (soil depth: 0-10cm)	17.1	28.9	29.6	42.9	Acker & Leptoukh, 2007
Soil_moist_100cm_1day	Average soil moisture content 1-day preceding disaster arrival (soil depth: 0-100cm)	142.4	273.3	273.6	401.0	
Soil_moist_100cm_5days*	Average soil moisture content 5-days preceding disaster arrival (soil depth: 0-100cm)	142.5	270.5	272.2	401.6	
Max_3d_avg_prcp	Max. 3-day precipitation averaged among stations in a county	0.0	64.7	72.9	249.8	Chamberlain et al., 2021
Max_5d_avg_prcp	Max. 5-day precipitation averaged among stations in a county	0.0	72.9	82.1	278.5	
Max_3d_max_prcp	Max. 3-day precipitation using among the stations in a county	0.0	101.5	109.4	500.4	
Max_5d_max_prcp	Max. 5-day precipitation using among the stations in a county	0.0	114.7	124.7	579.9	
Max_flood_ratio_during_incident period*^	Max. value of flood ratios observed during the incident period in a county	0.13	0.91	1.04	11.2	
Area_perc_ratio_greater_0.2_incidentperiod	County area percentage with flood ratio greater than 0.2 during the incident period	0.00	1.00	0.98	1.0	
Area_perc_ratio_greater_0.5_incidentperiod	County area percentage with flood ratio greater than 0.5 during the incident period	0.00	1.00	0.85	1.0	
Area_perc_ratio_greater_1_incident period^	County area percentage with flood ratio greater than 1.0 during the incident period	0.00	0.00	0.18	1.0	

Pop_2010*^	Population of county in 2010	783	11.1k	29.2k	1,152k	U.S. Census Bureau, 2021
NonWhite_perc	Non-white population percentage in a county	1.1	4.5	8.4	87.4	
Bachelors_or_ Higher	Population percentage with bachelor's degree or higher	4.2	13.2	13.9	30.3	
Poverty_Perc_All_ Ages*	Population percentage in poverty in a county	4.9	11.6	12.9	50.1	
Median_household _income*	Median household income in a county	20.6k	45.8k	46.0k	81.9k	
NonProfit_2010	Number of non-profit organizations in a county (year: 2010)	7	115	223.4	8005	
Businesses*^	Number of businesses in a county	35	456	530.8	1800	
Total_rev_ per_capita*^	Total revenue (in thousand dollars) per capita in the prior year	0.03	5.78	11.15	251.4	
Fire_dept*	Number of fire departments in a county (year: 2017)	1	6	6.98	28	U.S. Fire Administration , 2021
Schools	Number of public schools in a county	1	10	13.9	271	National Center for Education Statistics, 2021
School_districts*	Number of school districts in a county	0	4	4.6	73	
Levee_length^	Cumulative levee length of a county	0	0	3.2	70.1	U.S Army Corps of Engineers, 2016
Metro_des	Metro designation (a value of 1 is most urban and a value of 9 is most rural)	1	7	6.6	9	Economic Research Service, 2020
Yrs_since_prev_ flood_dis	Number of years since previous flood PDD in a county	0	4	8.1	50	Federal Emergency Management Agency, 2021
Yrs_since_prev_dis	Number of years since previous PDD in a county	0	1	2.1	25	
Num_of_all_prior_ dis_past_5yr^	Number of PDDs in the past 5 years in a county	0	3	3.2	9	
Num_of_all_prior_ dis_past_20yr^	Number of PDDs in the past 20 years in a county	0	3	3.0	10	
Num_of_prior_floo d_dis_past_5yr	Number of flood PDDs in the past 5 years in a county	0	1	1.0	5	
Num_of_prior_floo d_dis_past_20yr	Number of flood PDDs in the past 20 years in a county	0	3	3.0	10	

DeclarationDate_ IncidentEndDate*^	Time between the disaster incident end date and the disaster declaration (in days)	0	0	10.2	74	
Water*^	Fraction of a county covered by each type of land cover	0.1	2.6	3.8	26.5	Multi-Resolution Land Characteristics Consortium, 2021
Developed*^		0.6	5.1	6.1	76.6	
Barren*		0.0	0.0	0.1	6.8	
Forest*^		0.0	1.7	7.2	68.8	
Shrub		0.0	0.0	0.5	9.9	
Herbaceous		0.1	4.5	12.3	86.7	
Cultivated*^		0.2	70.4	64.3	91.2	
Wetlands*		0.2	3.7	5.8	74.5	
FEMA_region_5*^	FEMA region of PDD	Binary Variable				Federal Emergency Management Agency, 2021
FEMA_region_7*						
FEMA_region_8*^						
MN*	State of PDD					
WI						
IA*						
ND*						
SD*						
County_CRS	CRS program participation by a county					
City_CRS	CRS program participation by at least one jurisdiction in a county					
2001	Disaster year					
2002						
2006						
2010						
2011						
2013						
2014						
2016						
2017						

Table A.2. Additional cross-validation (CV) approaches investigated

Type of CV	MAE	RMSE	R^2
Dependent Variable: No. of applicants			
Current CV (10-fold, repeated 3 times)	3.8	5.5	0.67
Event-based CV	4.4	5.5	0.41
Year-based CV	4.5	5.8	0.47
Dependent Variable: No. of projects			
Current CV (10-fold, repeated 3 times)	18.3	26.0	0.51
Event-based CV	21.3	26.4	0.25
Year-based CV	22.5	28.9	0.11

Bayesian Additive Regression Trees (BART): BART is a Bayesian sum-of-trees ensemble where each tree consists of independently-constructed binary trees that are ultimately summed over all *m* branches. Each binary branch is built using a fully Bayesian probability model (Chipman et al., 2010), which thus limits each tree's contribution to the model through its priors (Kapelner & Bleich, 2013). In BART, a Markov Chain Monte Carlo (MCMC) algorithm is used as the backfitting algorithm to fit and infer the model (Chipman et al., 2010). The BART model is represented by:

$$Y = \sum_{j=1}^{m} g(x; T_j, M_j) + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2)$$

where T_j denotes a set of terminal nodes, M_j denotes a set of parameters associated with each node T_j, and g denotes a single tree model. ε represents model noise, and is assumed to be normally distributed with the mean.

One advantage of the BART model over other ensemble-of-trees models is its probability model. Similar to other Bayesian models, it includes a set of priors for the structure and the leaf parameters and a likelihood for data in the terminal nodes. The

purpose of including a set of priors is to provide regularization (Kapelner & Bleich, 2013). Regularization prevents the total fit to be dominated by any single regression tree and each of the trees accounts for only part of the overall fit (Chipman et al., 2010).

Classification and Regression Tree (CART): A CART model is a basic non-parametric decision tree-base where data are "forked" into binary branches based on the predictor variables. The model is built using a bootstrapped sampling approach whereby at each branch, the split is made to minimize error. This is repeated until a pre-specified terminal node is met, at which time, the prediction at the terminal nodes is the average value. The tree may then be pruned based on a cost complexity function to balance the size of the tree and fit (Breiman, 1996).

Gradient Boosting Machines (GBM): GBM is a non-parametric model where decision trees are grown using the information from previously grown trees (Hastie et al., 2009). Instead of fitting the model to the response variable, a GBM uses a Greedy algorithm to build trees that reduce residual variance from prior models. The Greedy algorithm forces minimization of a loss function or other user-specified cost function (Kuhn & Johnson, 2013).

Generalized Linear Model (GLM): A GLM is a parametric model, similar in spirit to an ordinary linear model, except that the response variable need not be normally

distributed and is generally assumed to be a member of the family of exponential distributions. In order to achieve this, the response variable is mapped to predictor variables using a link function (Zuur et al., 2009).

Multivariate adaptive regression splines (MARS): This semi-parametric method is based on the recursive partitioning approach in regression analysis but in contrast, MARS produces continuous models with continuous derivatives (Friedman, 1991). Its ability to allow for non-linearity behavior and interaction effects makes MARS a powerful technique for high-dimensional datasets (Nateghi et al., 2011). MARS is a generalization of stepwise linear regression and is built by summing a series of linear splines that enables the dependent variable to vary non-linearly with the independent variables. More specifically, MARS minimizes the sum of squared errors of the following model:

$$f(X) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(x)$$

where β_0 and $\beta_m$are coefficients and $h_m(x)$ are linear splines (Hastie et al., 2009).

Support Vector Regression (SVR): This non-parametric method is a robust modeling approach that defines decision boundaries (i.e., hyperplanes) to separate continuous data. The approach employs optimization to find the hyperplane that maximizes the number of observations between a margin defined by the hyperplane and two boundary lines, ε, away from the hyperplane. When data are highly non-

linear, the data cannot be linearly separated. To address this problem, SVR projects the data into higher dimensions using kernel functions (often a radial kernel function). The kernel function reduces both the computational complexity and interpretability of the model. The hyperplane is found using

$$\hat{f}_x = \sum_{i=1}^{n} \hat{a}_x K(x, x_i)$$

Where $\hat{a}_x = (HH^T + \lambda I)^{-1} y$, H is the basis function, and $K(x, x_i)$ is the kernel function (linear, polynomial, or radial) (Hastie et al., 2009).

Neural Networks (NN): The concept of NN is based on how thought in the human brain is structured. In NNs, artificial neurons connect in series or parallel to other neurons over many layers to develop a complex network. A neuron and its connections are called perceptrons. Progression through the network is achieved through activation functions which are non-linear transformations of input into perceptrons. More specifically, one layer of the model is represented by

$$f_{i+1}(X) \; = \; a(W_{i+1} \times f_i(X) + b_{i+1})$$

where *a* is the activation function, W_i is the weight tensor of layer *i*, b_i is the bias of layer *i*, and $f_i(X)$ is the output of layer *i*. This process is repeated for each subsequent layer (Kuhn & Johnson, 2013).

B Appendix: Supplementary Material Chapter 3

Below is a description of the statistical models used in the research.

Classification and Regression Tree (CART): CART is a nonparametric decision-tree model where each node is split based on a threshold value of a predictor. CART uses a bootstrapped sampling approach whereby each branch splits with the goal of minimizing the error. This process is repeated until pre-specified criteria are met and the prediction at the terminal nodes is the averaged value. In order to balance the size of the tree and avoid overfitting, the tree may then be pruned using a complexity function (Breiman, 1996). In this study, we developed regression decision trees using the CART model.

Random Forest (RF): RF is a nonparametric ensemble learning technique based on the concept of decision trees. Decision trees tend to produce high variance (Hastie et al., 2009). Decision tree algorithms are greedy because they consider all possible predictors at each split point of each tree. As such, averaging over an ensemble of decision trees can lead to highly correlated predictions. To address this issue, RF models consider a random sample of the predictors at each split. This is then repeated K times in order to build K regression trees. The prediction is then the average prediction over all K trees.

Support Vector Regression (SVR): SVR is a popular machine learning tool that is able to capture non-linearity in the data. It tries to find the best fit line (hyperplane) within a threshold value (i.e., the distance between the hyperplane and boundary line). Advantages of SVR include high prediction accuracy, good generalization capability, and computational complexity that is not affected by the input dimensionality (Awad & Khanna, 2015).

Zero-Inflated Negative Binomial Regression Model (ZINB): Negative binomial is a generalized linear model used for count data with excess zeros and over-dispersion (variation is higher than would be expected) (Liu et al., 2005).

Supplementary Data

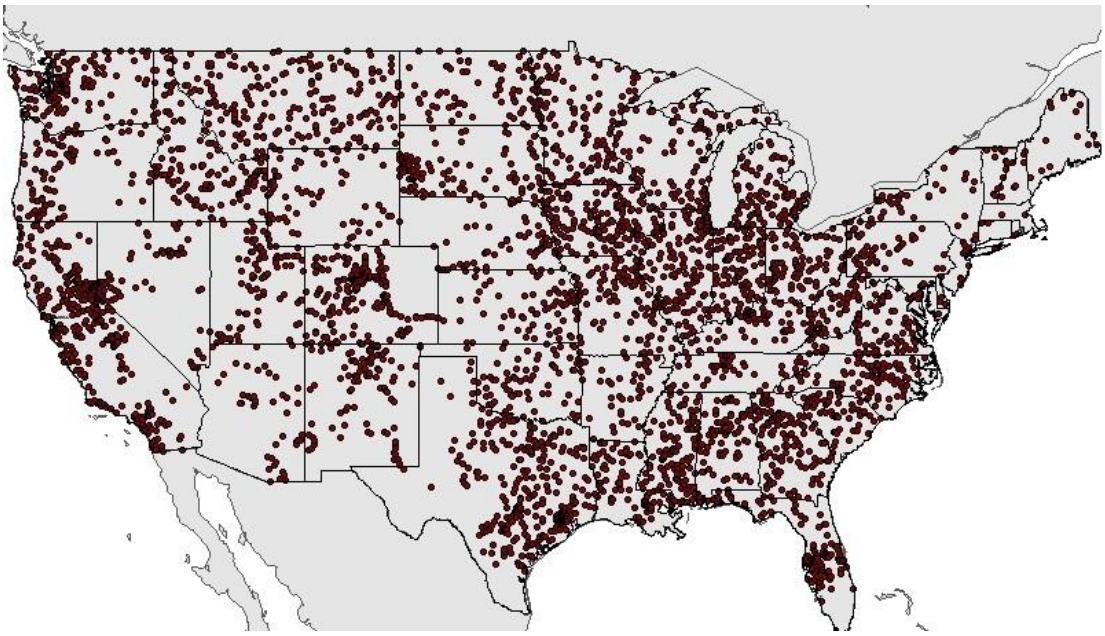

Figure B.1. The location of USGS streamgages throughout the continental U.S.

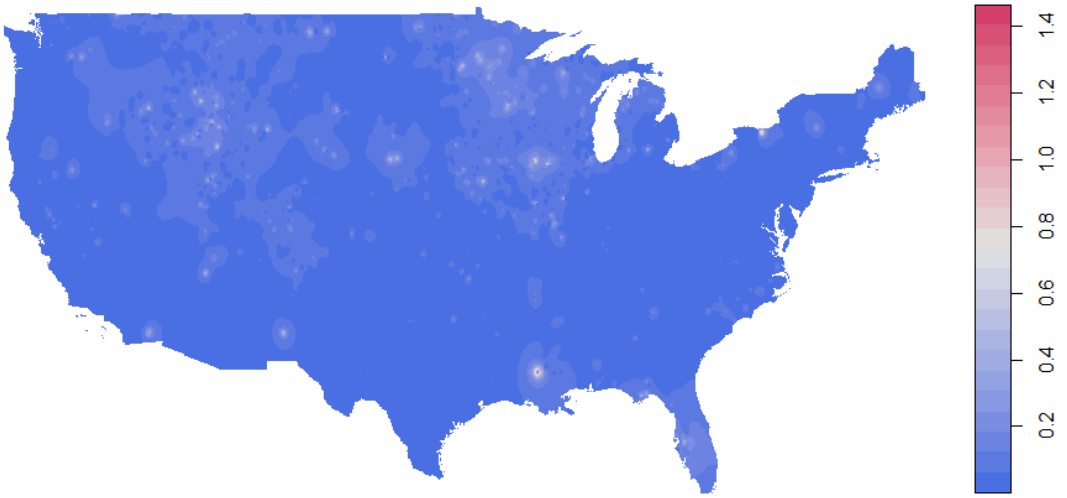

Figure B.2. An example of daily maximum FPR spatially interpolated across the entire U.S. (Louisiana flooding event, Aug. 12, 2016).

Table B.1. Flood metrics and summary statistics of 2016 data.

Type	Variable	Description	Unit	Minimum	Maximum	Mean	Source	Resolution
Dependent	NumClaims_all	Number of NFIP claims	Count	0	8433	1.5	FEMA	County
Independent variables used solely in the FPR model	max_ratio	Maximum value of observed flood peak ratios in a county	Unitless	0.01	5.15	0.19	USGS	Streamgage
	ratio_greater_0.2	Fraction of county with flood peak ratio greater than 0.2	Fraction	0	1	0.17	USGS	Streamgage
	ratio_greater_0.5	Fraction of county with flood peak ratio greater than 0.5	Fraction	0	1	0.01	USGS	Streamgage
	ratio_greater_1	Fraction of county with flood peak ratio greater than 1.0	Fraction	0	1	0.001	USGS	Streamgage
	ratio_greater_2	Fraction of county with flood peak ratio greater than 2.0	Fraction	0	0.33	0.000051	USGS	Streamgage
Independent variables used solely in GFF model	ff_max	Maximum value of flooded fraction observed in a county	Fraction	0.001	1	0.13	Giovanni	0.125*0.125 (deg^2)
	ff_greater_0.05	Fraction of county with flooded fraction greater than 0.05	Fraction	0	1	0.11	Giovanni	0.125*0.125 (deg^2)
	ff_greater_0.2	Fraction of county with flooded fraction greater than 0.2	Fraction	0	0.94	0.03	Giovanni	0.125*0.125 (deg^2)
	ff_greater_0.6	Fraction of county with flooded fraction greater than 0.6	Fraction	0	0.6	0.01	Giovanni	0.125*0.125 (deg^2)

Independent variables common to both FPR and GFF models	SFHA_frac	Fraction of county within a 100-year floodplain	Fraction	0	1	0.18	FEMA	N/A
	coastal	Is the county a coastal county?	Binary	No	Yes	N/A	Census	County
	slope_mean	Mean land slope	Degree	0.1	18.1	2.2	AWS Terrain Tiles	500x500 m^2
	Population_density	Ratio between the total population and land area	Population per mi^2	0.5	72035	423	Census	County
	devopen (developed, open space)	Fraction of county covered by land cover variable	Fraction	0.1	30.4	5.4	NLCD	30x30 m^2
	devlow (developed, low)			0.04	41.8	2.9		
	devmed (developed, medium)			0.002	36.4	1.5		
	devhigh, (developed, high)			0	41.1	0.7		
	barren			0	16.1	0.6		
	shrub/scrub			0	97.4	9.1		
	herbaceous			0	82.6	6.7		
	water			0.003	68.0	4.6		
	planted/cultivated			0	90.6	25.3		
	wetlands			0	72.8	9.6		

Table B.2. Total count of residential claims and policies in force (PIF) in 2016.

State	Claims	PIF	State	Claims	PIF	State	Claims	PIF
AK	7	2,246	LA	35,492	504,331	OK	178	12,112
AL	66	52,912	MA	65	57,421	OR	35	24,995
AR	274	14,060	MD	316	65,281	PA	260	50,879
AZ	114	28,218	ME	10	7,728	PR	17	4,889
CA	503	210,081	MI	116	20,683	RI	3	11,675
CO	53	19,322	MN	108	10,197	SC	13,501	211,064
CT	62	33,938	MO	241	19,086	SD	6	3,607
DC	13	2,069	MS	859	61,831	TN	167	27,242
DE	230	26,361	MT	8	4,328	TX	14,751	793,083
FL	18,525	1,729,744	NC	6,485	142,451	UT	12	3,829
GA	1,582	81,589	ND	7	13,159	VA	2,796	103,639
HI	1,294	60,964	NE	100	8,960	VT	5	3,298
IA	707	12,290	NH	6	7,620	WA	198	32,884
ID	5	5,941	NJ	2,732	214,906	WI	90	12,732
IL	208	36,177	NM	27	11,270	WV	1,026	13,011
IN	104	19,359	NV	14	10,464	WY	10	1,678
KS	157	8,705	NY	221	169,059			
KY	132	19,011	OH	81	28,044			

C Appendix: Supplementary Material Chapter 4

Table C.1. Description of initial independent variables.

Variable	Description	Variable	Description
max_traffic	Highest average daily traffic count in tract	pop_density	Population density
min_traffic	Lowest average daily traffic count in tract	EP_POV	Percentage of persons below poverty estimate
max_nonhw_traffic	Highest average daily non-highway traffic count in tract	EP_UNEMP	Percentage of civilian (age 16+) unemployed estimate
min_nonhw_traffic	Lowest average daily non-highway traffic count in tract	EP_PCI	Per capita income estimate, 2012-2016 ACS
median_income	Median household income	EP_NOHSDP	Percentage of persons with no high school diploma (age 25+) estimate
median_age	Median age of persons	EP_AGE65	Percentage of persons aged 65 and older estimate, 2012-2016 ACS
race_nonwhite_frac	Fraction of persons of non-white race	EP_AGE17	Percentage of persons aged 17 and younger estimate, 2012-2016 ACS
edu_college_frac	Fraction of persons with college degree	EP_DISABL	Percentage of civilian noninstitutionalized population with a disability estimate, 2012-2016 ACS
transp_public_frac	Fraction of persons using public transportation to work	EP_SNGPNT	Percentage of single-parent households with children under 18 estimate, 2012-2016 ACS
census_tract_area	Area of census tract	EP_MINRTY	Percentage minority (all persons except white, non-Hispanic) estimate, 2012-2016 ACS
stops_per_capita	Transit stops per 1000 people	EP_LIMENG	Percentage of persons (age 5+) who speak English "less than well" estimate, 2012-2016 ACS
stops_per_sqmile	Transit stops per square mile	EP_MUNIT	Percentage of housing in structures with 10 or more units estimate
impervious_frac	Fraction of impervious area	EP_CROWD	Percentage of occupied housing units with more people than rooms estimate
intDensity	Intersection density	EP_NOVEH	Percentage of households with no vehicle available estimate
strNetDensity	Street network density	EP_GROUPQ	Percentage of persons in institutionalized group quarters estimate, 2012-2016 ACS

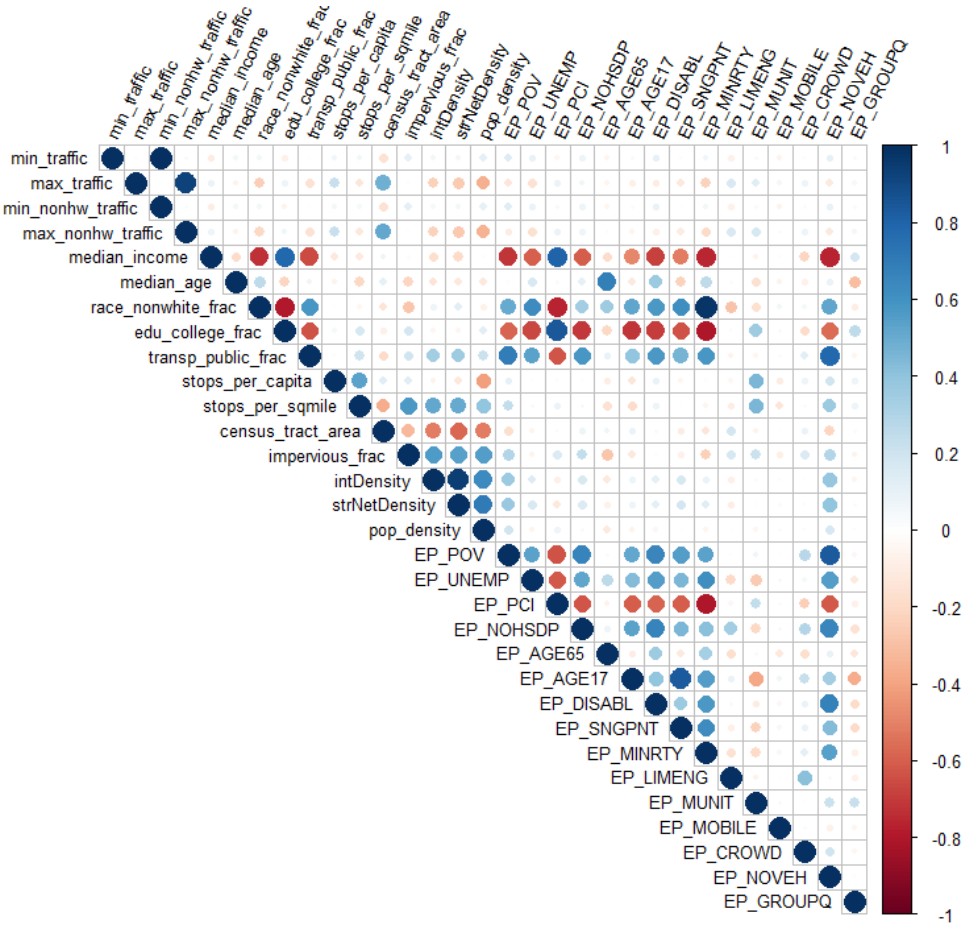

Figure C.1. A correlation matrix of initial independent variables.

Bibliography

Acker, J. G., & Leptoukh, G. (2007). Online analysis enhances use of NASA Earth science data. *Eos, Transactions American Geophysical Union*, *88*(2), 14–17. https://doi.org/10.1029/2007EO020003

Alin, A. (2010). Multicollinearity. *WIREs Computational Statistics*, *2*(3), 370–374. https://doi.org/10.1002/wics.84

Amazon Web Services (AWS) Terrain Tiles. (2017). [Database]. https://registry.opendata.aws/terrain-tiles

American Society of Civil Engineers (ASCE). (2017). *America's Infrastructure Report Card (Levees)*. American Society of Civil Engineers (ASCE). https://www.infrastructurereportcard.org/wp-content/uploads/2017/01/Levees-Final.pdf

Anderson, M. (2016). Who relies on public transit in the U.S. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2016/04/07/who-relies-on-public-transit-in-the-u-s/

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Anselin, L., & Bera, A. (1998). Spatial Dependence in linear Regression Models with an Introduction to Spatial Econometrics. In *Handbook of Applied Economic Statistics* (pp. 237–290). CRC Press.

Archer, D. N. (2021). Transportation Policy and the Underdevelopment of Black Communities. *Iowa Law Review*, *106*(5), 2125–2152.

Atreya, A., Ferreira, S., & Michel-Kerjan, E. (2015). What drives households to buy flood insurance? New evidence from Georgia. *Ecological Economics, 117,* 153–161. https://doi.org/10.1016/j.ecolecon.2015.06.024

Awad, M., & Khanna, R. (2015). Support Vector Regression. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers* (pp. 67–80). Apress. https://doi.org/10.1007/978-1-4302-5990-9_4

Bennett, J. B. (2019). *Attitude and Adoption: Understanding Climate Change Through Predictive Modeling* [Thesis, Purdue University Graduate School]. https://doi.org/10.25394/PGS.9108197.v1

Bennett, J., Baker, A., Johncox, E., & Nateghi, R. (2020). Characterizing the Key Predictors of Renewable Energy Penetration for Sustainable and Resilient Communities. *Journal of Management in Engineering*, *36*(4), 04020016. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000767

Bhattarai, R., Yoshimura, K., Seto, S., Nakamura, S., & Oki, T. (2016). Statistical model for economic damage from pluvial floods in Japan using rainfall data and socioeconomic parameters. *Natural Hazards and Earth System Sciences*, *16*(5), 1063–1077. https://doi.org/10.5194/nhess-16-1063-2016

Bin, O., Bishop, J., & Kousky, C. (2017). Does the National Flood Insurance Program Have Redistributional Effects? *The B.E. Journal of Economic Analysis & Policy*, *17*(4). https://doi.org/10.1515/bejeap-2016-0321

Birkland, T. A. (2006). *Lessons of Disaster: Policy Change after Catastrophic Events*. Georgetown University Press.

Blessing, R., Sebastian, A., & Brody, S. D. (2017). Flood Risk Delineation in the United States: How Much Loss Are We Capturing? *Natural Hazards Review*, *18*(3), 04017002. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000242

Bocchini, P., Frangopol, D. M., Ummenhofer, T., & Zinke, T. (2014). Resilience and Sustainability of Civil Infrastructure: Toward a Unified Approach. *Journal of Infrastructure Systems*, *20*(2), 04014004. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000177

Borden, K. A., & Cutter, S. L. (2008). Spatial patterns of natural hazards mortality in the United States. *International Journal of Health Geographics*, *7*(1), 64. https://doi.org/10.1186/1476-072X-7-64

Botzen, W. J. W., Deschenes, O., & Sanders, M. (2019). The Economic Impacts of Natural Disasters: A Review of Models and Empirical Studies. *Review of Environmental Economics and Policy*, *13*(2), 167–188. https://doi.org/10.1093/reep/rez004

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brody, S., Blessing, R., Sebastian, A., & Bedient, P. (2014). Examining the impact of land use/land cover characteristics on flood losses. *Journal of Environmental Planning and Management*, *57*(8), 1252–1265. https://doi.org/10.1080/09640568.2013.802228

Brody, S. D., Blessing, R., Sebastian, A., & Bedient, P. (2013). Delineating the Reality of Flood Risk and Loss in Southeast Texas. *Natural Hazards Review*, *14*(2), 89–97. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000091

Brody, S. D., & Highfield, W. E. (2013). Open space protection and flood mitigation: A national study. *Land Use Policy*, *32*, 89–95. https://doi.org/10.1016/j.landusepol.2012.10.017

Brody, S. D., Zahran, S., Highfield, W. E., Bernhardt, S. P., & Vedlitz, A. (2009). Policy Learning for Flood Mitigation: A Longitudinal Assessment of the Community Rating System in Florida. *Risk Analysis*, *29*(6), 912–929. https://doi.org/10.1111/j.1539-6924.2009.01210.x

Brody, S. D., Zahran, S., Highfield, W. E., Grover, H., & Vedlitz, A. (2008). Identifying the impact of the built environment on flood damage in Texas. *Disasters*, *32*(1), 1–18. https://doi.org/10.1111/j.1467-7717.2007.01024.x

Brody, S. D., Zahran, S., Maghelal, P., Grover, H., & Highfield, W. E. (2007). The Rising Costs of Floods: Examining the Impact of Planning and Development Decisions on Property Damage in Florida. *Journal of the American Planning Association*, *73*(3), 330–345. https://doi.org/10.1080/01944360708977981

Brown, J., & Richardson, D. (2015). *FEMA's Public Assistance Grant Program: Background and Considerations for Congress* (No. R43990). Congressional Research Service (CRS).

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., & von Winterfeldt, D. (2003). A Framework to Quantitatively Assess and Enhance the Seismic

Resilience of Communities. *Earthquake Spectra*, *19*(4), 733–752. https://doi.org/10.1193/1.1623497

Burby, R. J., Deyle, R. E., Godschalk, D. R., & Olshansky, R. B. (2000). Creating Hazard Resilient Communities through Land-Use Planning. *Natural Hazards Review*, *1*(2), 99–106. https://doi.org/10.1061/(ASCE)1527-6988(2000)1:2(99)

Chamberlain, S., Anderson, B., Salmon, M., Erickson, A., Potter, N., Stachelek, J., Simmons, A., Ram, K., & Edmund, H. (2016). *Package 'rnoaa.'*

Chen, T. Y.-J., Beekman, J. A., David Guikema, S., & Shashaani, S. (2019). Statistical Modeling in Absence of System Specific Data: Exploratory Empirical Analysis for Prediction of Water Main Breaks. *Journal of Infrastructure Systems*, *25*(2), 04019009. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000482

Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, *4*(1), 266–298. https://doi.org/10.1214/09-AOAS285

Chu, H.-C. (2014). Modeling Variations in Bus Schedule Adherence at the Stop Level on an Urban Arterial Road. *Journal of Transportation Engineering*, *140*(7), 05014003. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000688

Cigler, B. A. (2007). The "Big Questions" of Katrina and the 2005 Great Flood of New Orleans. *Public Administration Review*, *67*(s1), 64–76. https://doi.org/10.1111/j.1540-6210.2007.00814.x

Cochrane, H. C. (2004). Indirect Losses from Natural Disasters: Measurement and Myth. In Y. Okuyama & S. E. Chang (Eds.), *Modeling Spatial and Economic Impacts of Disasters* (pp. 37–52). Springer. https://doi.org/10.1007/978-3-540-24787-6_3

Cost of Assistance Estimates in the Disaster Declaration Process for the Public Assistance Program (to be codified at 44 C.F.R. § 206; p. 85 FR 80719). (2020). https://www.federalregister.gov/documents/2020/12/14/2020-27094/cost-of-assistance-estimates-in-the-disaster-declaration-process-for-the-public-assistance-program

Crow, D. A., Albright, E. A., Ely, T., Koebele, E., & Lawhon, L. (2018). Do Disasters Lead to Learning? Financial Policy Change in Local Government. *Review of Policy Research*, *35*(4), 564–589. https://doi.org/10.1111/ropr.12297

Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., & Webb, J. (2008). A place-based model for understanding community resilience to natural disasters. *Global Environmental Change*, *18*(4), 598–606. https://doi.org/10.1016/j.gloenvcha.2008.07.013

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social Vulnerability to Environmental Hazards*. *Social Science Quarterly*, *84*(2), 242–261. https://doi.org/10.1111/1540-6237.8402002

Cutter, S. L., Burton, C. G., & Emrich, C. T. (2010). Disaster Resilience Indicators for Benchmarking Baseline Conditions. *Journal of Homeland Security and Emergency Management*, *7*(1). https://doi.org/10.2202/1547-7355.1732

Cutter, S. L., & Emrich, C. (2005). Are natural hazards and disaster losses in the U.S. increasing? *Eos, Transactions American Geophysical Union*, *86*(41), 381–389. https://doi.org/10.1029/2005EO410001

Cutter, S. L., Emrich, C. T., Mitchell, J. T., Boruff, B. J., Gall, M., Schmidtlein, M. C., Burton, C. G., & Melton, G. (2006). The Long Road Home: Race, Class, and Recovery from Hurricane Katrina. *Environment: Science and Policy for Sustainable Development*, *48*(2), 8–20. https://doi.org/10.3200/ENVT.48.2.8-20

Cutter, S. L., & Finch, C. (2000). *Temporal and spatial changes in social vulnerability to natural hazards*. 6.

Cutter, S. L., & Finch, C. (2008). Temporal and spatial changes in social vulnerability to natural hazards. *Proceedings of the National Academy of Sciences*, *105*(7), 2301–2306. https://doi.org/10.1073/pnas.0710375105

Czajkowski, J., Villarini, G., Michel-Kerjan, E., & Smith, J. A. (2013). Determining tropical cyclone inland flooding loss on a large scale through a new flood peak ratio-based methodology. *Environmental Research Letters*, *8*(4), 044056. https://doi.org/10.1088/1748-9326/8/4/044056

Czajkowski, J., Villarini, G., Montgomery, M., Michel-Kerjan, E., & Goska, R. (2017). Assessing Current and Future Freshwater Flood Risk from North Atlantic Tropical Cyclones via Insurance Claims. *Scientific Reports*, *7*(1), 41609. https://doi.org/10.1038/srep41609

Dash, N., McCoy, B. G., & Herring, A. (2009). Class. In *Social Vulnerability to Disasters* (1st ed., pp. 75–100). CRC Press.

Davlasheridze, M., Fisher-Vanden, K., & Allen Klaiber, H. (2017). The effects of adaptation measures on hurricane induced property losses: Which FEMA investments have the highest returns? *Journal of Environmental Economics and Management*, *81*, 93–114. https://doi.org/10.1016/j.jeem.2016.09.005

Dixon, L., Clancy, N., Miller, B. M., Hoegberg, S., Lewis, M. M., Bender, B., Ebinger, S., Hodges, M., Syck, G. M., Nagy, C., & Choquette, S. R. (2017). *The Cost and Affordability of Flood Insurance in New York City: Economic Impacts of Rising Premiums and Policy Options for One- to Four-Family Homes*. RAND Corporation. https://www.rand.org/pubs/research_reports/RR1776.html

Dixon, L., Clancy, N., Seabury, S. A., & Overton, A. (2006). *The National Flood Insurance Program's Market Penetration Rate: Estimates and Policy Implications*. RAND Corporation. https://www.rand.org/pubs/technical_reports/TR300.html

Domingue, S. J., & Emrich, C. T. (2019). Social Vulnerability and Procedural Equity: Exploring the Distribution of Disaster Aid Across Counties in the United States. *The American Review of Public Administration*, *49*(8), 897–913. https://doi.org/10.1177/0275074019856122

Downton, M., & Pielke, R. (2005). How Accurate are Disaster Loss Data? The Case of U.S. Flood Damage. *Natural Hazards*, *35*(2), 211–228. https://doi.org/10.1007/s11069-004-4808-4

Economic Research Service. (2020, December 10). *Rural-Urban Continuum Codes*. https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/

Emanuel, K. (2020). Evidence that hurricanes are getting stronger. *Proceedings of the National Academy of Sciences, 117*(24), 13194–13195. https://doi.org/10.1073/pnas.2007742117

Emrich, C. T., Aksha, S. K., & Zhou, Y. (2022). Assessing distributive inequities in FEMA's Disaster recovery assistance fund allocation. *International Journal of Disaster Risk Reduction, 74,* 102855. https://doi.org/10.1016/j.ijdrr.2022.102855

Fan, Q., & Davlasheridze, M. (2016). Flood Risk, Flood Mitigation, and Location Choice: Evaluating the National Flood Insurance Program's Community Rating System. *Risk Analysis, 36*(6), 1125–1147. https://doi.org/10.1111/risa.12505

Federal Emergency Management Agency. (2021, August 19). *OpenFEMA Data Sets*. https://www.fema.gov/about/openfema/data-sets

FEMA. (2021). *Summary of FEMA Hazard Mitigation Assistance Grant Programs* [Fact Sheet]. FEMA. https://www.fema.gov/sites/default/files/documents/fema_summary-fema-hazard-mitigation-assistance-grant-programs_032321.pdf

Finch, C., Emrich, C. T., & Cutter, S. L. (2010). Disaster disparities and differential recovery in New Orleans. *Population and Environment*, *31*(4), 179–202. https://doi.org/10.1007/s11111-009-0099-8

Flanagan, B. E., Hallisey, E. J., Adams, E., & Lavery, A. (2018). Measuring Community Vulnerability to Natural and Anthropogenic Hazards: The Centers

for Disease Control and Prevention's Social Vulnerability Index. *Journal of Environmental Health*, *80*(10), 34–36.

Folke, C., Carpenter, S., Elmqvist, T., Gunderson, L., Holling, C. S., & Walker, B. (2002). Resilience and Sustainable Development: Building Adaptive Capacity in a World of Transformations. *AMBIO: A Journal of the Human Environment*, *31*(5), 437–440. https://doi.org/10.1579/0044-7447-31.5.437

Freeman, P. K. (2000). Infrastructure, natural disasters, and poverty. In *Managing Disaster Risk in Emerging Economies* (pp. 55–61). World Bank Publications.

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, *19*(1), 1–67. https://doi.org/10.1214/aos/1176347963

Frimpong, E., Petrolia, D. R., Harri, A., & Cartwright, J. H. (2020). Flood Insurance and Claims: The Impact of the Community Rating System. *Applied Economic Perspectives and Policy*, *42*(2), 245–262. https://doi.org/10.1093/aepp/ppz013

Galloway, G. E., Reilly, A., Ryoo, S., Riley, A., Haslam, M., Brody, S., Highfield, W., Gunn, J., Rainey, J., & Parker, S. (2018). The growing threat of urban flooding: A national challenge. *College Park and Galveston: University of Maryland and Texas A&M University*.

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, *31*(14), 2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014

Godschalk, D. R., Rose, A., Mittler, E., Porter, K., & West, C. T. (2009). Estimating the value of foresight: Aggregate analysis of natural hazard mitigation benefits

and costs. *Journal of Environmental Planning and Management*, *52*(6), 739–756. https://doi.org/10.1080/09640560903083715

Guikema, S. D., Nateghi, R., Quiring, S. M., Staid, A., Reilly, A. C., & Gao, M. (2014). Predicting Hurricane Power Outages to Support Storm Response Planning. *IEEE Access*, *2*, 1364–1373. https://doi.org/10.1109/ACCESS.2014.2365716

Guo, Z., Wilson, N. H. M., & Rahbee, A. (2007). Impact of Weather on Transit Ridership in Chicago, Illinois. *Transportation Research Record*, *2034*(1), 3–10. https://doi.org/10.3141/2034-01

Gupta, V. K., Mesa, O. J., & Dawdy, D. R. (1994). Multiscaling theory of flood peaks: Regional quantile analysis. *Water Resources Research*, *30*(12), 3405–3421. https://doi.org/10.1029/94WR01791

Hall, J. L. (2008). Assessing Local Capacity for Federal Grant-Getting. *The American Review of Public Administration*, *38*(4), 463–479. https://doi.org/10.1177/0275074007311385

Hallegatte, S., & Przyluski, V. (2010). *The Economics of Natural Disasters: Concepts and Methods* (SSRN Scholarly Paper No. 1732386). Social Science Research Network. https://papers.ssrn.com/abstract=1732386

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd ed.). Springer, New York, NY.

He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

He, Y., Thies, S., Avner, P., & Rentschler, J. (2021). Flood impacts on urban transit and accessibility—A case study of Kinshasa. *Transportation Research Part D: Transport and Environment*, *96*, 102889. https://doi.org/10.1016/j.trd.2021.102889

Highfield, W. E., & Brody, S. D. (2013). Evaluating the Effectiveness of Local Mitigation Activities in Reducing Flood Losses. *Natural Hazards Review*, *14*(4), 229–236. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000114

Highfield, W. E., & Brody, S. D. (2017). Determining the effects of the FEMA Community Rating System program on flood losses in the United States. *International Journal of Disaster Risk Reduction*, *21*, 396–404. https://doi.org/10.1016/j.ijdrr.2017.01.013

Hofmann, M., & O'Mahony, M. (2005). The impact of adverse weather conditions on urban bus performance measures. *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, 84–89. https://doi.org/10.1109/ITSC.2005.1520087

Holland, G. (2008). A Revised Hurricane Pressure–Wind Model. *Monthly Weather Review*, *136*(9), 3432–3445. https://doi.org/10.1175/2008MWR2395.1

Holling, C. S. (1973). Resilience and Stability of Ecological Systems. *Annual Review of Ecology and Systematics*, *4*(1), 1–23. https://doi.org/10.1146/annurev.es.04.110173.000245

Hollister, J. W., Robitaille, A. L., Beck, M. W., Johnson, M., & Tarak Shah. (2020). *jhollist/elevatr: New CRAN Release v0.3.0* (v0.3.0) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.4282962

Honadle, B. W. (1981). A Capacity-Building Framework: A Search for Concept and Purpose. *Public Administration Review*, *41*(5), 575–580. https://doi.org/10.2307/976270

Horn, D. P., & Webel, B. (2021). *Introduction to the National Flood Insurance Program (NFIP)* (No. R44593). Congressional Research Service (CRS).

Horney, J., Dwyer, C., Aminto, M., Berke, P., & Smith, G. (2017). Developing indicators to measure post-disaster community recovery in the United States. *Disasters*, *41*(1), 124–149. https://doi.org/10.1111/disa.12190

Horritt, M. S., & Bates, P. D. (2002). Evaluation of 1D and 2D numerical models for predicting river flood inundation. *Journal of Hydrology*, *268*(1), 87–99. https://doi.org/10.1016/S0022-1694(02)00121-X

Howell, J., & Elliott, J. R. (2019). Damages Done: The Longitudinal Impacts of Natural Hazards on Wealth Inequality in the United States. *Social Problems*, *66*(3), 448–467. https://doi.org/10.1093/socpro/spy016

James, G., Witten, D., Hastie, T., & Tibshirani, R. (Eds.). (2013). *An introduction to statistical learning: With applications in R*. Springer.

Jasinski, M. F., Borak, J. S., Kumar, S. V., Mocko, D. M., Peters-Lidard, C. D., Rodell, M., Rui, H., Beaudoing, H. K., Vollmer, B. E., Arsenault, K. R., Li, B., Bolten, J. D., & Tangdamrongsub, N. (2019). NCA-LDAS: Overview and

Analysis of Hydrologic Trends for the National Climate Assessment. *Journal of Hydrometeorology*, *20*(8). https://doi.org/10.1175/JHM-D-17-0234.1

Jasinski, M.F., Kumar, S.V., Borak, J.S., Mocko, D.M., Peters-Lidard, C.D., Rodell, M., Rui, H., Beaudoing, H. Kato, Vollmer, B.E., Arsenault, K.R., Li, B., & Bolten, J.D. (2018). *NCA-LDAS Noah-3.3 Land Surface Model L4 daily 0.125 x 0.125 degree, Version 2.0* [Data set]. NASA Goddard Earth Sciences Data and Information Services Center. https://doi.org/10.5067/7V3N5DO04MAS

Jongman, B., Ward, P. J., & Aerts, J. C. J. H. (2012). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*, *22*(4), 823–835. https://doi.org/10.1016/j.gloenvcha.2012.07.004

Kapelner, A., & Bleich, J. (2013). *bartMachine: Machine Learning with Bayesian Additive Regression Trees*. https://arxiv.org/abs/1312.2171v3

Kashfi Syeed, Bunker, J., & Jinwoo (Brian) Lee. (2013). Impact of rain on Daily Bus ridership: A Brisbane Case Study. *Australasian Transport Research Forum 2013 Proceedings*, 1–18. https://doi.org/10.13140/RG.2.1.2256.3289

Knighton, J., Buchanan, B., Guzman, C., Elliott, R., White, E., & Rahm, B. (2020). Predicting flood insurance claims with hydrologic and socioeconomic demographics via machine learning: Exploring the roles of topography, minority populations, and political dissimilarity. *Journal of Environmental Management*, *272*, 111051. https://doi.org/10.1016/j.jenvman.2020.111051

Knowlton, K., & Rotkin-Ellman, M. (2014). *Preparing for Climate Change: Lessons for Coastal Cities from Hurricane Sandy* (No. 14-04-A). Natural Resources Defense Council. https://www.nrdc.org/file/3911/download?token=ft_PoPCN

Koliou, M., van de Lindt, J. W., McAllister, T. P., Ellingwood, B. R., Dillard, M., & Cutler, H. (2020). State of the research in community resilience: Progress and challenges. *Sustainable and Resilient Infrastructure*, *5*(3), 131–151. https://doi.org/10.1080/23789689.2017.1418547

Kousky, C. (2018). Financing Flood Losses: A Discussion of the National Flood Insurance Program. *Risk Management and Insurance Review*, *21*(1), 11–32. https://doi.org/10.1111/rmir.12090

Kousky, C., Lingle, B., & Shabman, L. (2016). *FEMA Public Assistance Grants: Implications of a Disaster Deductible* (No. 16–04). Resources for the Future. https://media.rff.org/documents/RFF-PB-16-04.pdf

Kousky, C., & Shabman, L. (2012). *The Realities of Federal Disaster Aid: The Case of Floods*. Resources for the Future. https://media.rff.org/documents/RFF-IB-12-02.pdf

Kovats, R. S., & Hajat, S. (2008). Heat Stress and Public Health: A Critical Review. *Annual Review of Public Health*, *29*(1), 41–55. https://doi.org/10.1146/annurev.publhealth.29.020907.090843

Kreibich, H., Thieken, A. H., Petrow, T., Müller, M., & Merz, B. (2005). Flood loss reduction of private households due to building precautionary measures – lessons learned from the Elbe flood in August 2002. *Natural Hazards and Earth System Sciences*, *5*(1), 117–126. https://doi.org/10.5194/nhess-5-117-2005

Kuhn, M., & Johnson, K. (2013). Nonlinear Regression Models. In M. Kuhn & K. Johnson (Eds.), *Applied Predictive Modeling* (pp. 141–171). Springer. https://doi.org/10.1007/978-1-4614-6849-3_7

Kulkarni, A. R., & Shafei, B. (2018). Impact of Extreme Events on Transportation Infrastructure in Iowa: A Bayesian Network Approach. *Transportation Research Record*, *2672*(48), 45–57. https://doi.org/10.1177/0361198118795006

Lalancette, A., & Charles, A. (2022). Factors influencing hazard management by municipalities: The case of coastal communities. *Global Environmental Change*, *73*, 102451. https://doi.org/10.1016/j.gloenvcha.2021.102451

Lam, N. S.-N. (1983). Spatial Interpolation Methods: A Review. *The American Cartographer*, *10*(2), 129–150. https://doi.org/10.1559/152304083783914958

Landry, C. E., & Li, J. (2012). Participation in the Community Rating System of NFIP: Empirical Analysis of North Carolina Counties. *Natural Hazards Review*, *13*(3), 205–220. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000073

Langabeer, J. R., DelliFraine, J., & Alqusairi, D. (2012). The Influence of Politics on Federal Disaster Declaration Decision Delays. *Journal of Homeland Security and Emergency Management*, *9*(1). https://doi.org/10.1515/1547-7355.1960

Lee, E. (2021). *FEMA's Public Assistance Program: A Primer and Considerations for Congress* (No. R46749). Congressional Research Service (CRS).

Lehnert, E. A., Wilt, G., Flanagan, B., & Hallisey, E. (2020). Spatial exploration of the CDC's Social Vulnerability Index and heat-related health outcomes in

Georgia. *International Journal of Disaster Risk Reduction*, *46*, 101517. https://doi.org/10.1016/j.ijdrr.2020.101517

Li, J., & Landry, C. E. (2018). Flood Risk, Local Hazard Mitigation, and the Community Rating System of the National Flood Insurance Program. *Land Economics*, *94*(2), 175–198. https://doi.org/10.3368/le.94.2.175

Liu, H., Davidson, R. A., Rosowsky, D. V., & Stedinger, J. R. (2005). Negative Binomial Regression of Electric Power Outages in Hurricanes. *Journal of Infrastructure Systems*, *11*(4), 258–267. https://doi.org/10.1061/(ASCE)1076-0342(2005)11:4(258)

Luke, A., Sanders, B. F., Goodrich, K. A., Feldman, D. L., Boudreau, D., Eguiarte, A., Serrano, K., Reyes, A., Schubert, J. E., AghaKouchak, A., Basolo, V., & Matthew, R. A. (2018). Going beyond the flood insurance rate map: Insights from flood hazard map co-production. *Natural Hazards and Earth System Sciences*, *18*(4), 1097–1120. https://doi.org/10.5194/nhess-18-1097-2018

Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z., & Burton, H. V. (2020). Classifying earthquake damage to buildings using machine learning. *Earthquake Spectra*, *36*(1), 183–208. https://doi.org/10.1177/8755293019878137

Marino, E. (2018). Adaptation privilege and Voluntary Buyouts: Perspectives on ethnocentrism in sea level rise relocation and retreat policies in the US. *Global Environmental Change*, *49*, 10–13. https://doi.org/10.1016/j.gloenvcha.2018.01.002

Martín, C., & Williams, A. (2021). *A Federal Policy and Climate Migration Briefing for State, Tribal, and Local Governments*. Urban Institute.

May, P. J. (1992). Policy Learning and Failure. *Journal of Public Policy*, *12*(4), 331–354. JSTOR.

McCarthy, F. X. (2009). *FEMA's Disaster Declaration Process: A Primer* (No. RL34146). Congressional Research Service (CRS).

Meerow, S., Newell, J. P., & Stults, M. (2016). Defining urban resilience: A review. *Landscape and Urban Planning*, *147*, 38–49. https://doi.org/10.1016/j.landurbplan.2015.11.011

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article "Assessment of economic flood damage." *Natural Hazards and Earth System Sciences*, *10*(8), 1697–1724. https://doi.org/10.5194/nhess-10-1697-2010

Michel-Kerjan, E., Lemoyne de Forges, S., & Kunreuther, H. (2012). Policy Tenure Under the U.S. National Flood Insurance Program (NFIP). *Risk Analysis*, *32*(4), 644–658. https://doi.org/10.1111/j.1539-6924.2011.01671.x

Mitsova, D., Esnard, A.-M., Sapat, A., & Lai, B. S. (2018). Socioeconomic vulnerability and electric power restoration timelines in Florida: The case of Hurricane Irma. *Natural Hazards*, *94*(2), 689–709. https://doi.org/10.1007/s11069-018-3413-x

Mobley, W., Sebastian, A., Blessing, R., Highfield, W. E., Stearns, L., & Brody, S. D. (2021). Quantification of continuous flood hazard using random forest classification and flood insurance claims at large spatial scales: A pilot study in southeast Texas. *Natural Hazards and Earth System Sciences*, *21*(2), 807–822. https://doi.org/10.5194/nhess-21-807-2021

Moran, P. A. P. (1948). The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, *10*(2), 243–251.

Mucia, A., Bonan, B., Zheng, Y., Albergel, C., & Calvet, J.-C. (2020). From Monitoring to Forecasting Land Surface Conditions Using a Land Data Assimilation System: Application over the Contiguous United States. *Remote Sensing*, *12*(12), 2020. https://doi.org/10.3390/rs12122020

Multihazard Mitigation Council. (2017). *Natural Hazard Mitigation Saves: 2017 Interim Report* [Interim Report]. National Institute of Building Sciences. https://www.fema.gov/sites/default/files/2020-07/fema_ms2_interim_report_2017.pdf

Multi-Resolution Land Characteristics Consortium. (2021a). *NLCD 2011 Land Cover (CONUS)*. https://www.mrlc.gov/data/nlcd-2011-land-cover-conus

Multi-Resolution Land Characteristics Consortium. (2021b). *NLCD 2016 Land Cover (CONUS)* [Database]. https://www.mrlc.gov/data/nlcd-2011-land-cover-conus

Muñoz, C. E., & Tate, E. (2016). Unequal Recovery? Federal Resource Distribution after a Midwest Flood Disaster. *International Journal of Environmental Research and Public Health*, *13*(5), 507. https://doi.org/10.3390/ijerph13050507

Nateghi, R. (2018). Multi-Dimensional Infrastructure Resilience Modeling: An Application to Hurricane-Prone Electric Power Distribution Systems. *IEEE Access*, *6*, 13478–13489. https://doi.org/10.1109/ACCESS.2018.2792680

Nateghi, R., Guikema, S. D., & Quiring, S. M. (2011). Comparison and Validation of Statistical Methods for Predicting Power Outage Durations in the Event of

Hurricanes. *Risk Analysis*, *31*(12), 1897–1906. https://doi.org/10.1111/j.1539-6924.2011.01618.x

Nateghi, R., Guikema, S. D., & Quiring, S. M. (2014). Forecasting hurricane-induced power outage durations. *Natural Hazards*, *74*(3), 1795–1811. https://doi.org/10.1007/s11069-014-1270-9

National Centers for Environmental Information. (2020). *U.S. Billion-dollar Weather and Climate Disasters, 1980—Present* [Data set]. https://doi.org/10.25921/STKW-7W73

National Centers for Environmental Information. (2021, July). *Storm Events Database* [Database]. https://www.ncdc.noaa.gov/stormevents/

Neumann, B., Vafeidis, A. T., Zimmermann, J., & Nicholls, R. J. (2015). Future Coastal Population Growth and Exposure to Sea-Level Rise and Coastal Flooding—A Global Assessment. *PLOS ONE*, *10*(3), e0118571. https://doi.org/10.1371/journal.pone.0118571

Noonan, D. S., & Sadiq, A.-A. A. (2018). Flood Risk Management: Exploring the Impacts of the Community Rating System Program on Poverty and Income Inequality. *Risk Analysis*, *38*(3), 489–503. https://doi.org/10.1111/risa.12853

Norris, F. H., Stevens, S. P., Pfefferbaum, B., Wyche, K. F., & Pfefferbaum, R. L. (2008). Community Resilience as a Metaphor, Theory, Set of Capacities, and Strategy for Disaster Readiness. *American Journal of Community Psychology*, *41*(1), 127–150. https://doi.org/10.1007/s10464-007-9156-6

Obringer, R., & Nateghi, R. (2018). Predicting Urban Reservoir Levels Using Statistical Learning Techniques. *Scientific Reports*, *8*(1), 5164. https://doi.org/10.1038/s41598-018-23509-w

Office of Water Prediction. (2016). https://water.noaa.gov/about/nwm

Olshansky, R. B., & Johnson, L. A. (2014). The Evolution of the Federal Role in Supporting Community Recovery After U.S. Disasters. *Journal of the American Planning Association*, *80*(4), 293–304. https://doi.org/10.1080/01944363.2014.967710

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences*, *30*(7), 683–691. https://doi.org/10.1016/j.cageo.2004.03.012

Plyer, A. (2016, August 26). *Facts for Features: Katrina Impact*. The Data Center. https://www.datacenterresearch.org/data-resources/katrina/facts-for-impact/

Reilly, A. C., Guikema, S. D., Zhu, L., & Igusa, T. (2017). Evolution of vulnerability of communities facing repeated hazards. *PLOS ONE*, *12*(9), e0182719. https://doi.org/10.1371/journal.pone.0182719

Reilly, A., Siders, A. R., & Niemeier, D. (2020). *Eliminate Billion-Dollar Disasters: Equitable Science-Based U.S. Disaster Policy for a Resilient Future*. The Day One Project. https://www.dayoneproject.org/post/eliminate-billion-dollar-disasters-equitable-science-based-disaster-policy-for-a-resilient-future

Renschler, C. S., Frazier, A. E., Arendt, L. A., Cimellaro, G. P., Reinhorn, A. M., & Bruneau, M. (2010). *DEVELOPING THE 'PEOPLES' RESILIENCE FRAMEWORK FOR DEFINING AND MEASURING DISASTER*

RESILIENCE AT THE COMMUNITY SCALE (NIST GCR 10-930; p. 73). National Institute of Standards and Technology (NIST).

Rose, A., Porter, K., Dash, N., Bouabid, J., Huyck, C., Whitehead, J., Shaw, D., Eguchi, R., Taylor, C., McLane, T., Tobin, L. T., Ganderton, P. T., Godschalk, D., Kiremidjian, A. S., Tierney, K., & West, C. T. (2007). Benefit-Cost Analysis of FEMA Hazard Mitigation Grants. *Natural Hazards Review*, *8*(4), 97–111. https://doi.org/10.1061/(ASCE)1527-6988(2007)8:4(97)

Rufat, S., Tate, E., Burton, C. G., & Maroof, A. S. (2015). Social vulnerability to floods: Review of case studies and implications for measurement. *International Journal of Disaster Risk Reduction*, *14*, 470–486. https://doi.org/10.1016/j.ijdrr.2015.09.013

Rui, H., Teng, W., Vollmer, B., Jasinski, M., Mocko, D., & Kempler, S. (2014). National Climate Assessment—Land Data Assimilation System (NCA-LDAS) Data at NASA GES DISC. *In American Geophysical Union Fall Meeting, San Francisco, p GC51B–0405. 2014.* https://core.ac.uk/reader/42697560

Sadiq, A.-A., & Noonan, D. (2015a). Local capacity and resilience to flooding: Community responsiveness to the community ratings system program incentives. *Natural Hazards*, *78*(2), 1413–1428. https://doi.org/10.1007/s11069-015-1776-9

Sadiq, A.-A., & Noonan, D. S. (2015b). Flood disaster management policy: An analysis of the United States Community Ratings System. *Journal of Natural*

Resources Policy Research, *7*(1), 5–22. https://doi.org/10.1080/19390459.2014.963373

SAMHSA. (2017). *Greater Impact: How Disasters Affect People of Low Socioeconomic Status* (Disaster Technical Assistance Center Supplemental Research Bulletin). Substance Abuse and Mental Health Services Administration (SAMHSA). https://www.samhsa.gov/sites/default/files/dtac/srb-low-ses_2.pdf

Schmidtlein, M. C., Finch, C., & Cutter, S. L. (2008). Disaster Declarations and Major Hazard Occurrences in the United States. *The Professional Geographer*, *60*(1), 1–14. https://doi.org/10.1080/00330120701715143

Section 406 of the Stafford Act, as amended by Section 1235(b) of the Disaster Recovery Reform Act of 2018, (2018).

Sharif, H. O., Jackson, T. L., Hossain, M. M., & Zane, D. (2015). Analysis of Flood Fatalities in Texas. *Natural Hazards Review*, *16*(1), 04014016. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000145

Sharifi, A. (2016). A critical review of selected tools for assessing community resilience. *Ecological Indicators*, *69*, 629–647. https://doi.org/10.1016/j.ecolind.2016.05.023

Shashaani, S., Guikema, S. D., Zhai, C., Pino, J. V., & Quiring, S. M. (2018). Multi-Stage Prediction for Zero-Inflated Hurricane Induced Power Outages. *IEEE Access*, *6*, 62432–62449. https://doi.org/10.1109/ACCESS.2018.2877078

Smith, G., Lyles, W., & Berke, P. (2013). The Role of the State in Building Local Capacity and Commitment for Hazard Mitigation Planning. *International*

Journal of Mass Emergencies and Disasters, *31*(2). http://ijmed.org/articles/621/

Smith, G., & Vila, O. (2020). A National Evaluation of State and Territory Roles in Hazard Mitigation: Building Local Capacity to Implement FEMA Hazard Mitigation Assistance Grants. *Sustainability*, *12*(23), 10013. https://doi.org/10.3390/su122310013

Stover, V., & McCormack, E. (2012). The Impact of Weather on Bus Ridership in Pierce County, Washington. *Journal of Public Transportation*, *15*(1), 95–110. https://doi.org/10.5038/2375-0901.15.1.6

Strother, L. (2018). The National Flood Insurance Program: A Case Study in Policy Failure, Reform, and Retrenchment. *Policy Studies Journal*, *46*(2), 452–480. https://doi.org/10.1111/psj.12189

Stuckey, M. (2006). *Low-Flow, Base-Flow, and Mean-Flow Regression Equations for Pennsylvania Streams* (Scientific Investigations Report No. 2006–5130). U.S. Geological Survey.

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. *Transportation Research Part D: Transport and Environment*, *10*(3), 231–244. https://doi.org/10.1016/j.trd.2005.04.007

Taherkhani, M., Vitousek, S., Barnard, P. L., Frazer, N., Anderson, T. R., & Fletcher, C. H. (2020). Sea-level rise exponentially increases coastal flood frequency. *Scientific Reports*, *10*(1), 6466. https://doi.org/10.1038/s41598-020-62188-4

Tao, S., Corcoran, J., Hickman, M., & Stimson, R. (2016). The influence of weather on local geographical patterns of bus usage. *Journal of Transport Geography, 54*, 66–80. https://doi.org/10.1016/j.jtrangeo.2016.05.009

The National Center for Education Statistics. (2021). *Search for Public Schools Districts*. National Center for Education Statistics. https://nces.ed.gov/ccd/districtsearch/

The Robert T. Stafford Disaster Relief and Emergency Assistance Act of 1988, Pub. L. No. 100–707, 42 U.S.C. (1988).

The Unequal Commute. (2020). *Urban Institute*. https://www.urban.org/features/unequal-commute

Thomas, S., Pillai, G. N., & Pal, K. (2017). Prediction of peak ground acceleration using ϵ-SVR, ν-SVR and Ls-SVR algorithm. *Geomatics, Natural Hazards and Risk*, *8*(2), 177–193. https://doi.org/10.1080/19475705.2016.1176604

Tierney, K. J. (2003). *Conceptualizing and Measuring Organizational and Community Resilience: Lessons From The Emergency Response Following The September 11, 2001 Attack on The World Trade Center*. https://udspace.udel.edu/handle/19716/735

Tonn, G., & Czajkowski, J. (2018). *An Analysis of U.S. Tropical Cyclone Flood Insurance Claim Losses: Storm Surge vs. Freshwater* (Working Paper No. 2018–12). University of Pennsylvania. https://riskcenter.wharton.upenn.edu/wp-content/uploads/2018/12/WP2018-12_Cyclone-Flood-Claims_TonnCzajkowski.pdf

Tonn, G., Reilly, A., Czajkowski, J., Ghaedi, H., & Kunreuther, H. (2021). U.S. transportation infrastructure resilience: Influences of insurance, incentives, and public assistance. *Transport Policy, 100*, 108–119. https://doi.org/10.1016/j.tranpol.2020.10.011

Transit Equity & Environmental Health in Baltimore. (2021). Johns Hopkins University. https://americanhealth.jhu.edu/news/transit-equity-environmental-health-baltimore

Tu, H., van Lint, H. W. C., & van Zuylen, H. J. (2007). *Impact of Adverse Weather on Travel Time Variability of Freeway Corridors* (No. 07–1642). Article 07–1642. Transportation Research Board 86th Annual MeetingTransportation Research Board. https://trid-trb-org.proxy-um.researchport.umd.edu/view/801786

U.S. Army Corps of Engineers. (2016). *National Levee Database*. https://levees.sec.usace.army.mil/

U.S. Census Bureau. (2021). *Explore Census Data*. https://data.census.gov/cedsci/

U.S. Fire Administration. (2021). *National Fire Department Registry*. https://apps.usfa.fema.gov/registry/

U.S. Government Accountability Office. (2012). *Federal Disaster Assistance: Improved Criteria Needed to Assess a Jurisdiction's Capability to Respond and Recover on Its Own* (GAO-12-838). U.S. GAO. https://www.gao.gov/products/gao-12-838

U.S. Government Accountability Office. (2014). *Opportunities Exist to Strengthen Oversight of Administrative Costs for Major Disasters* (GAO-15-65). U.S. GAO. https://www.gao.gov/assets/gao-15-65.pdf

U.S. Government Accountability Office. (2016). *Federal Departments and Agencies Obligated at Least $277.6 Billion during Fiscal Years 2005 through 2014* (GAO-16-797). U.S. GAO. https://www.gao.gov/products/gao-16-797

U.S. Government Accountability Office. (2019). *Disaster Resilience Framework: Principles for Analyzing Federal Efforts to Facilitate and Promote Resilience to Natural Disasters* (GAO-20-100SP). U.S. GAO. https://www.gao.gov/products/gao-20-100sp

Villarini, G., Smith, J. A., Baeck, M. L., & Krajewski, W. F. (2011). Examining Flood Frequency Distributions in the Midwest U.S.1: Examining Flood Frequency Distributions in the Midwest U.S. *JAWRA Journal of the American Water Resources Association*, *47*(3), 447–463. https://doi.org/10.1111/j.1752-1688.2011.00540.x

Villarini, G., Smith, J. A., Baeck, M. L., Marchok, T., & Vecchi, G. A. (2011). Characterization of rainfall distribution and flooding associated with U.S. landfalling tropical cyclones: Analyses of Hurricanes Frances, Ivan, and Jeanne (2004). *Journal of Geophysical Research: Atmospheres*, *116*(D23). https://doi.org/10.1029/2011JD016175

Wang, L., & Ganapati, N. E. (2018). Disasters and Social Capital: Exploring the Impact of Hurricane Katrina on Gulf Coast Counties. *Social Science Quarterly*, *99*(1), 296–312. https://doi.org/10.1111/ssqu.12392

Weidman, J., & Weiss, D. (2013, April 29). *Disastrous Spending: Federal Disaster-Relief Expenditures Rise amid More Extreme Weather*. Center for American Progress. https://www.americanprogress.org/issues/green/reports/2013/04/29/61633/disastrous-spending-federal-disaster-relief-expenditures-rise-amid-more-extreme-weather/

Wells, A. (2017, July 10). *Private Insurers Ready to Plunge into Flood Market*. Insurance Journal. https://www.insurancejournal.com/magazines/mag-features/2017/07/10/456390.htm

World Meteorological Organization. (2021). *WMO ATLAS OF MORTALITY AND ECONOMIC LOSSES FROM WEATHER, CLIMATE AND WATER EXTREMES (1970–2019)* (WMO-No. 1267). World Meteorological Organization. https://public.wmo.int/en/media/press-release/weather-related-disasters-increase-over-past-50-years-causing-more-damage-fewer

Zahran, S., Weiler, S., Brody, S. D., Lindell, M. K., & Highfield, W. E. (2009). Modeling national flood insurance policy holding at the county scale in Florida, 1999–2005. *Ecological Economics*, *68*(10), 2627–2636. https://doi.org/10.1016/j.ecolecon.2009.04.021

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer Science & Business Media.