Quantifying the Role of Vulnerability in Hurricane Damage via a Machine Learning Case Study

Laura D. Szczyrba

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Geosciences

Robert Weiss, Chair
Cristina Dura
Jennifer L. Irish
Venkataramana R. Sridhar
Yang Zhang

May 12, 2020
Blacksburg, Virginia

Keywords: Vulnerability, Impact, Damage, Machine Learning, Hurricane María

Quantifying the Role of Vulnerability in Hurricane Damage via a Machine Learning Case Study

Laura D. Szczyrba

ABSTRACT

Pre-disaster damage predictions and post-disaster damage assessments are challenging because they result from complicated interactions between multiple drivers, including exposure to various hazards as well as differing levels of community resiliency. Certain societal characteristics, in particular, can greatly magnify the impact of a natural hazard, however they are frequently ignored in disaster management because they are difficult to incorporate into quantitative analyses. In order to more accurately identify areas of greatest need in the wake of a disaster, both the hazards and the vulnerabilities need to be carefully assessed since they have been shown to be positively correlated with damage patterns. This study evaluated the contribution of eight drivers of structural damage from Hurricane María in Puerto Rico, leveraging machine learning algorithms to determine the role that societal factors played. Random Forest and Stochastic Gradient Boosting Trees algorithms analyzed a diverse set of data including wind, flooding, landslide, and vulnerability measures. These data trained models to predict the structural damage caused by Hurricane María in Puerto Rico and the importance of each predictive feature was calculated. Results indicate that vulnerability measures are the leading predictors of damage in this case study, followed by wind, flood, and landslide measures. Each predictive variable exhibits unique, often nonlinear, relationships with damage. These results demonstrate that societal-driven vulnerabilities play critical roles in damage pattern analysis and that targeted, pre-disaster mitigation efforts should be enacted to reinforce household resiliency in socioeconomically vulnerable areas. Recovery programs may need to be reworked to focus on the highly impacted vulnerable populations to avoid the persistence, or potential enhancement, of preexisting social inequalities in the wake of a disaster.

Quantifying the Role of Vulnerability in Hurricane Damage via a Machine Learning Case Study

Laura D. Szczyrba

GENERAL AUDIENCE ABSTRACT

Disasters are not entirely natural phenomena. Rather, they occur when natural hazards interact with the man-made environment and negatively impact society. Most risk and impact assessment studies focus on natural hazards (processes beyond human control) and do not incorporate the role of societal circumstances (within human agency). However, it has been shown that certain socioeconomic, demographic, and structural characteristics increase the severity of disaster impacts. These characteristics define the the susceptibility of a community to negative disaster impacts, known as vulnerability. This study quantifies the role of vulnerability via a case study of Hurricane María. A variety of statistical modeling, known as machine learning, analyzed flood, wind, and landslide hazards along with the aforementioned vulnerabilities. These variables were correlated with a damage assessment database and the model calculated the strength of each variable's relationship with damage. Results indicate that vulnerability measures exhibit the strongest predictive correlations with the damage caused by Hurricane María, followed by wind, flood, and landslide measures, respectively, suggesting that efforts to improve societal equality and improvements to infrastructure in vulnerable areas can mitigate the impacts of future hazardous events. In addition, societal information is critical to include in future risk and impact assessment efforts in order to prioritize areas of greatest need and allocate resources to those who would benefit from them most.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Number 1735139. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The author would like to thank the Disaster Resilience and Risk Management program at Virginia Tech for their valuable discussions and feedback on this research.

List of Abbreviations

ARA	Applied Research Associates
AveDepth	Average flood depth (m)
AveLS	Average landslide density code
CDC	Center for Disease Control
CDCVuln	CDC vulnerability index
DI	Damage Index
DI1	Basline damage index 1
DI2	Damage index 2, excluding highest outlier (0.36)
DI3	Damage index 3, excluding 0 damage
FEMA	Federal Emergency Management Agency
HurTrack	Distance from Hurricane María track
MaxDepth	Maximum depth of flooding (m)
MaxSusWinds	Maximum sustained winds (m/s)
ME	Mean Error
MSE	Mean Squared Error
NFIP	National Flood Insurance Program
NHC	National Hurricane Center
NOAA	National Oceanic and Atmospheric Administration
PCA	Principal Component Analysis
PR-OFSA	Puerto Rican Office of the General Coordinator for Social Financing and Self-Management (English translation)

OSM	OpenStreetMap
PeakGust	Peak gust (m/s)
PropFA	Proportion of flooded area
PropSC	Proportion of area considered to be a special community
PropSFHA	Proportion of area considered to be Special Flood Hazard Area
RF	Random Forest
SeVI	Socioeconomic Vulnerability Index
SFHA	Special Flood Hazard Area
SGBT	Stochastic Gradient Boosting Trees
StrVI	Structural Vulnerability Index
SVI	Social Vulnerability Index
USCB	United States Census Bureau
USGS	United States Geological Survey

Contents

Chapter 1

Introduction

Hurricanes are becoming increasingly hazardous events, resulting in more severe impacts on communities (Walsh et al., 2016). Political and economic tendencies have also led to under-regulated housing development in many at-risk regions, potentially exacerbating risk of hurricane-related damage (Dahl et al., 2018). The effect of unchecked development can be quantified by assessing the increasing number of people or assets exposed to hazards. However, addressing simply the hazards and exposure produces an incomplete assessment of disaster impacts and the traditional approach of quantifying disasters in terms of the physical world has gradually been expanded to account for the inherent social nature of disaster risk. For example, Hurricane Katrina demonstrated the particular importance of race, social class, gender, and age, among a host of other indelible factors, to both the severity of household impacts and as well as recovery rates (Hartman & Squires, 2006). However, the influence of societal inequalities on the intensity of disaster impacts is difficult to evaluate because of the inherent complexity and qualitative nature of the challenges introduced. Unfortunately history has a way of repeating itself, and Hurricane María similarly revealed underlying social issues, manifestations of the overarching failings of leadership produced long before María's

landfall, that may have intensified the impact of the storm for certain populations. In the aftermath of María, few assessments of storm damage and response priorities considered vulnerability because of the challenge in quantifying such topics.

1.1 Purpose and Outline

As the third costliest storm in U.S. history, with damages estimated between $65 - $115 billion (Pasch, Penny, & Berg, 2018), Hurricane María offers an opportunity to gain an understanding of the underlying social factors that contributed to its destruction. This study provides a quantitative measure of the influence of the vulnerabilities, originating from prevailing societal conventions, on the damage caused by María, complimenting surrounding qualitative discussions. This case study exemplifies a unique application of machine learning algorithms to illustrate the importance of holistic data analyses that incorporate human variables and is outlined as follows:

- Chapter 2 grounds the current study with a review of current literature. This will provide the reader with the necessary background information and context to interpret any findings by explaining the influence of vulnerability on disaster impacts and providing relevant applications of ensemble decision tree algorithms. This chapter closes by introducing Hurricane María as the selected case study.

- Chapter 3 explains the methods applied to the study. It identifies and describes all the raw data, how they were processed, and analysis methods. It elucidates the way in which the algorithms were trained, tuned, assessed, and how they quantified variable importance in order to assess the role of vulnerability in the case study. The chapter also provides an overall workflow summary.

- Chapter 4 presents the raw results of the analyses including the findings from data processing, model performance, and the relationships between the various predictive features and the damage produced by Hurricane María.

- Chapter 5 provides a contextual discussion of the study's findings in terms of both the role of vulnerability as well as model performance. The suitability of machine learning in disaster management studies is also discussed. The chapter ends with thoughts on future work as well as a brief summary of the key findings of the study.

Chapter 2

Literature Review

2.1 The Role of Vulnerability in Disaster Impacts

Disasters affect different demographic groups with immense disparity. While physical hazards often remain the leading contributors to damage intensity, studies have demonstrated that social factors contribute significantly as well. Namely, vulnerability, or the collective factors influencing a community's susceptibility to damage, can intensify disaster impacts and inhibit post-disaster recovery (Fothergill & Peek, 2004; Flanagan, Gregory, Hallisey, Heitgerd, & Lewis, 2011). Vulnerability can be quantified using either inductive or deductive statistical methods to create indices focused on socioeconomic, structural (i.e. built environment), and/or comprehensive (e.g. Cutter's Social Vulnerability Index) measures (Cutter, Boruff, & Shirley, 2003; Flanagan et al., 2011; Holand, Lujala, & Rød, 2011).

Evident across many scales, vulnerability and disaster resilience have been shown to be correlated from a nationwide (Ward & Shively, 2017) to household level (Highfield, Peacock, & Van Zandt, 2014). A review by Fothergill and Peek (2004) concluded that socioeconomic

status consistently emerges as a contributing factor in physical damage for a variety of disasters. Furthermore, they found that vulnerable populations often reside in homes with structural qualities that cause them to be more susceptible to wind and flood forcings resulting in higher levels of damage incurred during a hurricane. Van Zandt and Rohe (2011) hypothesized that this may be because the houses that are affordable to lower income groups are often older and poorly maintained due to the lack of spare financial resources to fund needed repairs and updates, thus increasing their risk of damage. A study of Hurricane Ike by Highfield et al. (2014) confirmed that, while controlling for other factors, cheaper and/or older homes received higher levels of damage than newer, more expensive homes. Another study of Hurricane Ike and Hurricane Andrew by Peacock, Van Zandt, Zhang, and Highfield (2014) concluded that social factors, such as income, were important determinants of both residential building damage and recovery since houses in wealthier neighborhoods retained a higher relative percentage of their home value post-disaster and recuperated their value more quickly.

Vulnerability not only increases the level of disaster impact a community experiences, it also affects post-disaster recovery, with more vulnerable groups recovering at slower rates than less vulnerable groups (Bolin, 1985). Impoverished populations are less likely to be insured or accumulate private savings and thus they commonly rely on government assistance for recovery resources (Bolin & Stanford, 1991). However, access to public funding is often unequal (Peacock, Dash, & Zhang, 2007) and government assistance programs have been shown to systematically exclude the most vulnerable groups from qualifying for aid, namely renters, multi-family households, minorities, and families lacking legal documentation of ownership or residence (Fothergill, Maestas, & Darlington, 1999; Kitzbichler, 2011; Mueller, Bell, Chang, & Henneberger, 2011; Chhotray & Few, 2012; Van Zandt et al., 2012). Therefore, vulnerable populations more likely to experience intense structural damage and are also

less likely to fully recover prior to the next disaster, resulting in a positive feedback loop that amplifies vulnerability over time. To address this inequity, the role that vulnerability plays in structural damage must first be quantified. Because of the quantity and complexity of contributing factors to structural damage, a large amount of diverse data is required to build a representative statistical model. The presently developing field of machine learning for data analytics is well suited to accommodate complex, multivariate datasets and parse out variables' relationships.

2.2 Ensemble Decision Trees

Random Forest (RF) and Stochastic Gradient Boosting Trees (SGBT) are part of a family of highly flexible, nonparametric machine learning algorithms known as ensemble decision trees. They accommodate nonlinear, multidimensional datasets from a variety of sources and formats - allowing for both continuous numeric as well as categorical data within the same model. Because the algorithms leverage learned, as opposed to assumed (such as in linear regression), patterns in large datasets, no assumptions regarding the statistical relationships in the data are needed to run the models and data do not need to be transformed before the analysis. Ensemble decision tree algorithms provide useful and interpretable machine learning approaches to disaster data analytics.

RF is a common ensemble decision tree algorithm that constructs a group of independent classification or regression trees and leverages the majority vote of trees to determine the resultant prediction or the average prediction per data sample, respectively (Breiman, 1996, 2001). A random subset of both the training data and the predictive features construct each decision tree (Ho, 1995) which ensures that the decision trees are, ideally, unique and uncorrelated. Each tree begins at a parent node and then the algorithm recursively splits

the data using Boolean Logic towards increasingly homogeneous groups of child nodes. At each node, the model analyzes a random subset of the data (referred to as in-bag data), and does not train on an internal testing set of data (referred to as out-of-bag data), and then tests different split options while monitoring the effect on the out-of-bag (OOB) data. On average, about 38.6% of the randomly sampled evaluation data are set aside as out-of-bag data (Grömping, 2009). The model selects the optimal division by measuring the changes in splitting criterion on the out-of-bag data. In this study, mean squared error (MSE) was selected as the splitting criterion, from Grömping (2009):

$$OOB - MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{\hat{y}}_{iOOB})^2 \tag{2.1}$$

where $\overline{\hat{y}}_{iOOB}$ is the average prediction of the ith observation on the out-of-bag data across all regression trees, y_i are the true, observed values across n data instances. Data continue to be divided at splits until the tree reaches a stopping criterion (e.g. maximum depth) at the terminal nodes, where the predictions are produced.

While RF creates a group of randomized decision trees and selects the average prediction per data sample, SGBT iteratively generates a series of classification or regression trees, each improving upon the performance of the previous (Friedman, 2002). This stagewise improvement of model predictions is known as boosting. After the initial tree is constructed from a random subset of the data, the residual errors are passed to the next tree. The new tree reduces the error loss with the gradient descent cost function, then new residual errors are calculated. Trees are sequentially constructed until the loss function reaches a threshold or another stopping criterion is met.

2.2.1 Applications in Geosciences

In the past 5 - 10 years, geoscience studies have begun incorporating decision tree algorithms into their analyses. Owing to the complex nature of the phenomena, many flood studies in particular have embraced this strategy. Wang et al. (2015) first applied RF for regional flood risk assessment with a case study of the Dongjiang River Basin in China. Five thousand data samples and eleven predictive indices trained and tested the performance of RF on 4 historical flood events. The model classified areas within the basin as either lowest, lower, medium, higher, or highest risk levels and concluded that RF is a reasonable machine learning model for flood risk quantification, with a misclassification error rate of 8.76% on the testing dataset. Model interpretation revealed that maximum three-day precipitation, runoff depth, and typhoon frequency were the leading predictive indices. In 2017, Chapi et al. similarly quantified flood hazard risk in the Haraz Watershed, Iran on a smaller dataset of 201 data samples. They trained five separate decision tree classifiers, including RF and an algorithm they produced - Bagging Logistic Model Tree. The algorithm they produced performed best, however it only marginally outperformed RF. Slope angle, distance to river, and river density were found to have the highest predictive power.

Several flood forecast studies have recently used RF for flood impact prediction. Sadler, Goodall, Morsy, and Spencer (2018) applied a RF regression model to predict the number of crowd-sourced flood reports (a quantitative proxy for flood impact) as a result of 45 extreme events between 2010 and 2016, including several hurricanes, in Norfolk, Virginia. They compared RF with traditional Poisson regression and found that RF more accurately predicted the number of expected flood reports. Shafizadeh-Moghadam, Valavi, Shahabi, Chapi, and Shirzadi (2018) analyzed the same dataset used by Chapi et al. (2017) with a variety of machine learning models including an ensemble of boosted trees. Amongst the eight models tested, boosted regression trees performed best. Overall, they concluded that

ensemble machine learning methods perform well for flood impact prediction, however they acknowledged that there is a dearth of similar studies to compare their results to.

In addition to flooding, the broader field of natural hazards has taken advantage of the capabilities of ensemble decision tree algorithms. One of the earliest hazards studies that applied RF, Tesfamariam and Liu (2010) identified buildings susceptible to earthquake damage with eight machine learning models. Six characteristic measures of 484 concrete buildings were used to predict field observations of earthquake damage from a 1999 event in the Duzce-Bolu region of Turkey. Landslides have also been predicted with these strategies. In 2015, Trigila, Iadanza, Esposito, and Scarascia-Mugnozza predicted landslides in Italy from a 2009 event that was induced by heavy rainfall. They analyzed aerial imagery to manually identify 1,490 rapid, shallow landslides and collected data on several predictive factors including slope angle, distance to stream, land cover, etc. All data were then aggregated to 4 m x 4 m grid cells. They found that RF performed well as a predictive model and was most effective at minimizing classification errors. Another landslide study by Hong, Pourghasemi, and Pourtaghi (2016) analyzed 163 individual landslides in Lianhua County, China. They compared two methods of calculating the predictive power of features: permutation and mean decrease in impurity. Both methods of calculating feature importance produced similar results. They also concluded that, in their case, the RF algorithm performed adequately enough to be used in landslide susceptibility mapping.

Beyond natural hazards, ensemble decision trees have been applied in a variety of other geoscience subfields, including the prediction of riverine discharge (Prieto, Le Vine, Kavetski, García, & Medina, 2019), groundwater potential mapping (Rahmati, Pourghasemi, & Melesse, 2016; Naghibi, Moghaddam, Kalantar, Pradhan, & Kisi, 2017), depth of the limit of soil redox processes (Koch et al., 2019), and gold deposit locations (Rodriguez-Galiano, Chica-Olmo, & Chica-Rivas, 2014; McKay & Harris, 2016). Intriguingly, the McKay and

Harris study found that the RF algorithm is better at locating gold deposits than expert, knowledge-driven models. However, none of these aforementioned studies analyzed how vulnerability factors affected these events.

2.2.2 Applications in Disaster Vulnerability Assessment

Ensemble decision tree strategies can provide more holistic approaches to disaster data analytics and a few studies have recently emerged that apply this method to explore the societal role in disasters. A RF assessment of wildfire damage in Portugal by Oliveira, Zêzere, Queirós, and Pereira (2017) concluded that purchasing power and housing quality were significantly correlated with the extent of wildfire damage and that certain demographic groups, such as the elderly and households with lower education levels, were relatively more vulnerable to wildfire impacts. Merz, Kreibich, and Lall (2013) studied the correlation between flood damage and voluntary precautionary measures across German households, finding that households with resources to implement mitigation actions sustained lower structural losses. They also concluded that ensemble decision trees more accurately predicted damage than traditional impact models. By leveraging household-level damage assessments in Bangladesh, another flood study comparing linear regression, RF, and Artificial Neural Networks concluded that larger households and higher education levels were associated with lower flood damage (Ganguly, Nahar, & Hossain, 2019). This study employs ensemble decision tree algorithms, specifically RF and SGBT, to quantitatively explore the relative role that societal factors played in the structural damage caused by Hurricane María.

2.3 Hurricane María

Hurricane María was a Category 5 storm that made landfall in Puerto Rico on September 20, 2017 and devastated the island for months to follow. María is the third costliest storm in U.S. history, after Hurricanes Katrina (2005) and Harvey (2017), with total approximate damages in the U.S. Virgin Islands and Puerto Rico of $90 billion (Pasch et al., 2018). Two weeks prior to María's landfall, Hurricane Irma skirted the island 50 km north, weakening the island's infrastructure. An independent study commissioned by Puerto Rico's government estimated a death toll of 2,975 individuals, although the number of fatalities remains a contentious and politicized topic (Federal Emergency Management Agency, 2018b).

In addition to the storm's intensity, the severity of María's impact could have been influenced by the preexisting social disparities that have been exacerbated in the past two decades by the economic downturn in Puerto Rico (Santiago-Bartolomei, 2018). With a poverty rate of 44%, Puerto Rico has a much higher low income population than the national average (poverty rate of 13%) as well as a higher percentage of the population is age 65 and over (18%) compared to the national average (15%) (U.S. Census Bureau, 2017). As economic challenges often are, this was the result of years of harmful political and economic decisions, often made by leaders that the people of Puerto Rico did not elect (Klein, 2018). The influence of these disparities is reflected in mortality statistics, Santos-Burgoa et al. (2018) determined that the 2,975 fatalities caused by María were concentrated in areas of low socioeconomic status as well as areas with the highest ratios of men over age 65.

While María's effect on infrastructure in Puerto Rico was widespread, one of the most critical sectors hit was housing and the aforementioned challenges resulted in a shortage of resilient low-income housing (Santiago-Bartolomei, 2018). The housing stock on the island is 40-50 years old on average and new home construction has been stagnant due to the high cost of

construction relative to median income (U. S. Department of Housing and Urban Development, 2018). Therefore, much of the housing stock is incompatible with current building codes, which applied to houses constructed after 2011, and outdated due to the high relative cost of proper maintenance (U. S. Department of Housing and Urban Development, 2018). Affluent residents of Puerto Rico often reside in concrete structures built to code by licensed contractors, however many lower income populations only have access to substandard houses that have not been updated or are informally self-built, resulting in an uneven distribution of structural resiliency across demographic groups (U. S. Department of Housing and Urban Development, 2018). Informal housing refers to structures that are self-built without proper titles or permitting and are often out of compliance with zoning and building regulations. A 2018 report by the Puerto Rico Home Builders Association estimated that 45% of structures on the island are informal (Asociación de Constructores de Puerto Rico, 2018). However, these areas of vulnerable housing were difficult to prioritize in post-disaster response and recovery efforts because these structures lack legal documentation and government housing databases do not include them (Federal Emergency Management Agency, 2018b). Given the widespread prevalence of these structures, existing government structural databases are incomplete and could not be solely utilized to accurately model traditional structure-damage functions across the island.

Damage patterns produced by wind events specifically are indicators of structural and socioeconomic vulnerability (Eaton, 1980). Since Hurricane María passed diagonally across the center of Puerto Rico, all structures across the island were exposed to some degree of wind and/or flood forcing. As opposed to flooding, which has a discrete spatial extent, high velocity winds spanned the entirety of the island and revealed areas of poor infrastructure investment and high structural vulnerability. Ma and Smith (2019) analyzed Individual Assistance data from the Federal Emergency Management Agency (FEMA) and determined

that María's wind was the cause of 99% of the destroyed homes in Puerto Rico. They also concluded that renters and lower income populations sustained higher levels of damage than homeowners or higher income households.

Although Cutter et al. (2003) established that lack of wealth and housing quality are primary contributors to hazard vulnerability, societal factors have not yet been used as a proxy to analyze correlations between areas with high concentrations of community vulnerability with damage due to Hurricane María. This is especially vital given the high poverty rate and housing challenges experienced by Puerto Rico and while it has been the focus of many qualitative discussions, the role of vulnerability as it pertains to Hurricane María's impact on Puerto Rico remains to be quantitatively constrained in a robust manner. In this study, resultant damage patterns are hypothesized to be a function of both hazardous forcings and preexisting vulnerabilities.

Chapter 3

Methods

3.1 Data

The widely recognized disaster risk assessment model defines risk as a function of hazards, exposure, and vulnerability. This framework was adapted for a post-event application by proposing that impact is a function of hazards, exposure, and vulnerability (Figure 3.1). Best available data represented the components of this conceptual framework, including impact (number of buildings damaged), exposure (total number of buildings), natural hazards (wind, flooding, landslides), and vulnerabilities (socioeconomic and/or structural). Table 3.1 provides a summary of all variables produced by data gathering and processing, indicating their relation to the conceptual framework and abbreviations.

An aerial damage assessment database from the Federal Emergency Management Agency (FEMA) documented the structural damage of individual buildings after María (Federal Emergency Management Agency, 2018c). Visual comparisons of pre- and post-event nadir imagery from National Oceanic and Atmospheric Administration's (NOAA) National Geode-

tic Survey informed categorical determinations of the relative level of damage each structure on the island received (either "Affected" or "Destroyed"). Oblique imagery, where available, supplemented the creation of this dataset. A total of 53,664 structures were visually designated as "Affected" (49,972) or "Destroyed" (3,692). Because nadir imagery is collected perpendicular to the terrain, these data did not capture damage to the sides of structures and residential versus nonresidential structures are often indistinguishable.

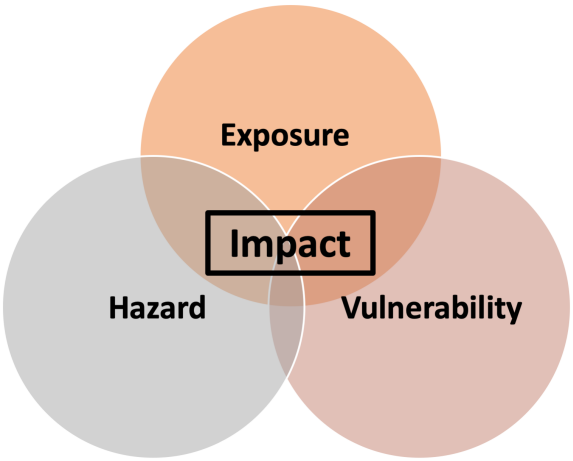

Figure 3.1: Conceptual framework illustrating that disaster impacts result from the nexus between natural hazards, exposure, and vulnerabilities. This principle guides the analysis, with damage indices representing impact normalized by exposure and predictive features represented by hazard and vulnerability data.

Two other sources of damage data were also considered for this study. FEMA's National Flood Insurance Program (NFIP) conducted detailed damage estimates in the field after María, however these data only account for structures within the Special Flood Hazard Area (SFHA) (Federal Emergency Management Agency, 2018a). FEMA also deployed damage inspectors to households that applied for Individual Assistance in order to evaluate the amount of aid that each applicant qualified for (Federal Emergency Management Agency,

Table 3.1: Summary of all data included in this analysis, including all predictive (P) and target (T) variables

Category	Measure	Type	Source	Abbreviation
Wind	Distance from hurricane center (deg)	P	NHC	HurTrack
	Peak gust (m/s)	P	ARA	PeakGust
	Max Sustained Winds (m/s)	P	ARA	MaxSusWinds
Flood	Proportion of flooded area	P	FEMA	PropFA
	Average depth of flooding (m)	P	FEMA	AveDepth
	Max depth of flooding (m)	P	FEMA	MaxDepth
	Proportion of SFHA	P	FEMA	PropSFHA
Landslide	Average landslide density code	P	USGS	AveLS
Vulnerability	Proportion of special communities	P	PR-OFSA	PropSC
	Social vulnerability	P	CDC	CDCVuln
	Structural vulnerability	P	Eroglu et al.	StrVI
	Socioeconomic vulnerability	P	Eroglu et al.	SeVI
Damage / Exposure	Basline Damage Index	T	FEMA	DI1
	DI1, excluding highest outlier	T	FEMA	DI2
	DI1, excluding 0 damage tracts	T	FEMA	DI3

2019b). Again, this dataset is limited to those homes that both applied and met the initial qualifications for disaster assistance, and therefore may exclude a significant number of socially vulnerable households (Fothergill et al., 1999; Ross, 2013). Furthermore, these data are only publicly accessible at the zip code summary level. While the FEMA geospatial damage assessment database may not be as detailed as other assessments, it is the only database that includes every structure on the island of Puerto Rico and remains the best source of damage data appropriate for this study.

Building footprint polygons, created by OpenStreetMap (OSM) and hosted on the Humanitarian Data Exchange platform, delineated all structures across Puerto Rico (OpenStreetMap, 2019). To develop these data, crowd-sourced volunteers analyzed aerial imagery to hand-delineate visible structural boundaries. A total 1,500,308 residential and nonresidential struc-

tures on the island were digitized, representing structural exposure in Puerto Rico.

Best available hazard information, including wind, flooding, and landslides, represented the threats posed by Hurricane María. For wind hazards, FEMA's Natural Hazards Risk Assessment Program provided measures of 3-second peak gusts (m/s) and maximum sustained winds (m/s) at 10 m elevation over flat, open terrain (Applied Research Associates, 2017). Applied Research Associates (ARA) modeled these hazards with their hurricane wind field model (Vickery, Skerlj, Steckley, & Twisdale, 2000) along with National Hurricane Center (NHC) data. NHC best track data for María charted the center path of the storm (National Hurricane Center, 2017). FEMA's flood event depth grids measured the severity of flooding (Federal Emergency Management Agency, 2017a), created by subtracting the terrain base elevation from a water surface elevation model at a cell size of 7.62 m, representing inundation extent and intensity (Federal Emergency Management Agency, 2017b). Puerto Rico's NFIP SFHA database (Federal Emergency Management Agency, 2018d) provided an additional measure of general flood risk, depicted by the one-percent annual chance flood polygons (Federal Emergency Management Agency, 2019a). A United States Geological Survey (USGS) dataset documented the spatial density of landslides triggered by Hurricane María (United States Geological Survey, 2019). This dataset was rapidly created to provide situational awareness and does not record individual landslide occurrences, but rather aggregates landslide measures to a uniform grid comprised of 2,506 2 km x 2 km cells throughout the island (Bessette-Kirton et al., 2019). Each grid cell represents an area with: (1) no landslides, (2) fewer than 25 landslides/4 km^2 , or (3) more than 25 landslides/4 km^2. Areas that were not studied were attributed as 0.

Four datasets represented vulnerability in this study: two represented comprehensive measures (with both soceioeconomic and structural factors), one focused on socioeconomic factors, and one focused on structural (built environment) factors. The Center for Disease

Control (CDC) created a comprehensives social vulnerability index (SVI) for Puerto Rico in 2017 (Center for Disease Control, 2017) that incorporates multiple themes including socioeconomic status,language, housing, and transportation, using deductive methods from Flanagan et al. (2011). The Puerto Rican Geographic Information Systems catalogue provided a shapefile of "special communities" that delineates the spatial extent of 713 identified disadvantaged communities throughout the island (Oficina del Coordinador General para el Financiamiento Social y la Autogestión, 2008). The Puerto Rican Office of the General Coordinator for Social Financing and Self-Management (PR-OFSA) compiled the data however, neither well-documented metadata nor methods accompany this dataset. Eroglu, Pamukçu, Szczyrba, and Zhang (2020) (collaborators at Virginia Tech) created a socioeconomic vulnerability index and a structural vulnerability index for Puerto Rico by adapting methods from Cutter et al. (2003) using JMP statistical software (SAS Institute Inc., 1989-2019). The contributions of socioeconomic and structural vulnerabilities were examined separately by categorizing variables into two groups, where the socioeconomic vulnerability index was measured with variables representing living conditions and population characteristics, and the structural vulnerability index was measured with variables representing housing characteristics and the structural quality (Holand et al., 2011). The most representative variables in the StrVI include: the number of structural units per structure, number of units without heating fuel, median number of rooms, vacancy rate, and age of construction. The most representative variables in the SeVI include: per capita income, age of population, ethnicity, mode of transportation, and number of non-English speaking households. 8 components make up the StrVI and 12 components makeup the SeVI, accounting for 67% of the variance in each.

A preexisting dataset of Puerto Rican structural vulnerability does not exist and, furthermore, the island lacks accurate permitting and construction data. The few available datasets

are incomplete and fail to capture the widespread abundance informal homes, estimated to comprise up to 45% of the homes on the island (Asociación de Constructores de Puerto Rico, 2018), and were therefore not used in this analysis.

3.1.1 Data Processing

Census tracts are widely used for public policy and urban planning that specifically promote socioeconomic well-being and equality and, therefore, provide an appropriate level of analysis for stakeholders interested in this study (Krieger, 2006). To summarize the structural damage database at the census tact level, first, the building footprint polygons sourced from OSM were used to provide a count of undamaged structures. The damage points and building polygons were collected using various sources of nadir imagery which could result disparate aerial perspectives. To overcome point placement mismatches due to imagery differences, a 3 foot buffer was applied to select the building footprints not associated with damage points, which were then converted to center points attributed as "Not Affected". Structure points were spatially joined with the 905 land-based census tracts in Puerto Rico, giving a summary of the total number of structures per tract and the categorical level of damage that each structure received (i.e. "Not Affected", "Affected", "Destroyed"). Figure 3.2 summarizes the creation of the damage index. The baseline damage index (DI1) created for this study measured the total number of "Affected" ($N_{Affected}$) or "Destroyed" ($N_{Destroyed}$) structures (impact) normalized by total number of structures (N_{Total}) in each census tract (exposure),:

$$DI1 = \frac{N_{Affected} + N_{Destroyed}}{N_{Total}} \tag{3.1}$$

following similar methods as Burton (2010) and Ganguly et al. (2019). Three additional variations from DI1 were calculated to reduce noise: the second index (DI2) removed the

highest outlying data point - value of 0.36, the third index (DI3) excluded 117 census tracts that contained no damage. Table 3.1 summarizes the damage indices used.

Figure 3.2: Hurricane María damage index development, illustrating that (a) FEMA damage assessment points (Federal Emergency Management Agency, 2018c) were combined with (b) OSM building footprints (OpenStreetMap, 2019) resulting in (c) a categorical assessment of damaged (red) and undamaged (green) structures across the island.

The wind, flood, and landslide data required spatial processing to represent census tract-level hazard summaries. A within-polygon center point (i.e. centroid) was extracted from each census tract. Then, the proximity tool Near calculated the nearest distance (in degrees) between each center point and the NHC Hurricane María best track line. ARA windfield data are partitioned to the census tract scale and did not require further processing. FEMA flood event depth grids provided maximum and mean depth of flooding (m) measurements and were used to calculate flooded area. Zonal statistics calculated the maximum and mean raster values per census tract. The Raster to Polygon conversion tool recast the raster grids to polygons, allowing measurement of flooded area per census tract. Depth grid polygons were intersected with the census tracts to calculate overlapping area (m^2) divided by the total area of the census tract to determine the proportion of flooded area. Similar calculations provided the proportional SFHA area as well. The USGS dataset was converted from a 2 km x 2 km grid to census tract measures. After intersecting the grid with the tracts, spatial statistics summarized the average landslide code value within each tract. Table 3.1 provides a summary of all variables produced by data processing and aggregation and the abbreviations

with which the variables will be referred to as henceforth.

Because census tracts are common for social analyses, the Center for Disease Control and Eroglu et al. vulnerability measures were already defined at this scale. However, the shapefile of special communities from PR-OFSA required processing in order to be incorporated into this study. The same methods used to calculate the proportional flooded area and proportional area of SFHA were also applied: the area of each census tract was intersected with the special communities data and then divided by the total area.

After all predictive variables were processed and aggregated, a correlation analysis eliminated colinear variables (i.e. variables that were highly correlated and thus contained redundant statistical relationships). A matrix of Spearman rank correlation coefficients (Moran, 1948) were calculated to compare the correlations between all predictive features. This coefficient examines monotonic relationships between variables, either linear or nonlinear, and ranges from -1 (negative correlation) to + 1 (positive correlation). If variable pairings exhibited very strong correlations, i.e., Spearman correlation value higher than 0.9 (Schober, Boer, & Schwarte, 2018), one variable would be carried forth into the machine learning analysis in order to maintain the interpretability of the model results (Karagiannopoulos, Anyfantis, Kotsiantis, & Pintelas, 2007).

3.2 Ensemble Decision Trees

After finalizing the dataset, RF and SGBT regression algorithms, sourced from Python's SciKit Learn machine learning package (Pedregosa et al., 2011), were trained to assess the relative influences of hazards and vulnerabilities on structural damage due to Hurricane María in Puerto Rico. These algorithms were selected for their high interpretability compared to other machine learning algorithms and their ability to quantify relative importances of

predictive features.

Before applying the algorithms on the data, randomized division created a training set containing 80% of the data and an evaluation set containing the remaining 20% of data, following commonly accepted procedures, e.g., Suthaharan (2016). Data stratification upon division ensured that the distribution of the training and testing data were similar. Automated optimization techniques, including randomized search cross validation and grid search cross validation, tuned the model to the ideal hyperparameters and explained variance (R^2), mean absolute error (MAE), and mean error (ME), calculated on the evaluation set, assessed model performance,:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \tag{3.2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.3}$$

$$ME = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{y}_i \tag{3.4}$$

where y_i represents the true, observed values, $\hat{y}_i$ is the corresponding model predictions, and $\overline{y}_i$ is the mean value over n data instances. ME can provide information on average model bias, such as whether the model is consistently under-predicting or over-predicting, but should be interpreted with the understanding that positive and negative errors can cancel each other out.

The tuning process evaluated the number of trees, tree depth, number of features considered at each split, minimum number of data samples required to incite a split, and the minimum

number of samples required at each node. For SGBT, the tuning process also included learning rate and the proportion of randomly selected data instances used to construct each tree. The accuracy of the average prediction per data instance converges as parameters are tuned.

3.2.1 Assessing the Role of Vulnerability

If the models heavily relied upon vulnerability measures to determine damage patterns, it would indicate that vulnerability played an important role in this case study. Ensemble decision tree algorithms provide straightforward means to interpret the importance of predictive features (Breiman, Friedman, Olshen, & Stone, 1984). Because the trees built in this study relied upon MSE to split each node, they simultaneously quantified the quality of each split as the effectiveness of each predictive feature in reducing predicted target variable variance. Therefore, the default importance calculation provided by SciKit Learn - mean decrease in MSE - represented one measure of the importance of each predictive feature (Pedregosa et al., 2011).

The default importance measure can potentially be biased towards favoring features with high-cardinality (Strobl, Boulesteix, Zeileis, & Hothorn, 2007), therefore the importance of each predictive feature was also calculated by permuting, or randomly shuffling, each feature's values while measuring changes in variance (R^2) before and after permutation (Breiman, 2001). For RF, the out-of-bag data monitored variance changes while for SGBT the entire training dataset monitored variance changes. If shuffling one feature resulted in a sharp increase in model variance (i.e. sharp decrease in model performance), the feature is considered important. However, if the permuted feature correlated with another, the relationship would be retained, thus reducing the perceived measure of importance. To

mitigate this effect, related groups of features were also permuted in tandem to determine how correlated variable categories affected the model (Koch et al., 2019). Variable groupings are indicated in Table 3.1.

While feature importance measures indicate which features are most-valuable to model performance, they provide little information in terms of how or why features are important. Learned partial dependencies demonstrate the marginal effect each predictive feature exhibits on the damage index (Friedman, 2001). In other words, it shows how the model's predictions partially depend on each predictive feature, represented by the following equation from Liaw and Wiener (2018),:

$$\tilde{f}(x) = \frac{1}{n}\sum_{i=1}^{N} f(x, X_{i_C}) \tag{3.5}$$

where x is predictive feature of interest among the other predictive features $\{X_{1_C}, X_{2_C}, X_{3_C}...X_{N_C}\}$ used in the machine learning model $\tilde{f}$ of n samples. The function explains, for a given value x, the marginal effect it has on the prediction by creating an average prediction for each value of x over the distribution of X_{i_C}. Thus, this analysis is computationally expensive. The partial dependence for each predictive feature along with the feature distribution were plotted to determine the marginal relationship that each predictive feature exhibited with damage.

3.3 Work Flow Summary

Three unique damage indices were used as target variables while the same set of features were used as predictors in a machine learning analysis. Predictive features represented wind (HurTrack, PeakGust, MaxSuSWinds), flood (PropFA, AveDepth, MaxDepth, PropSFHA),

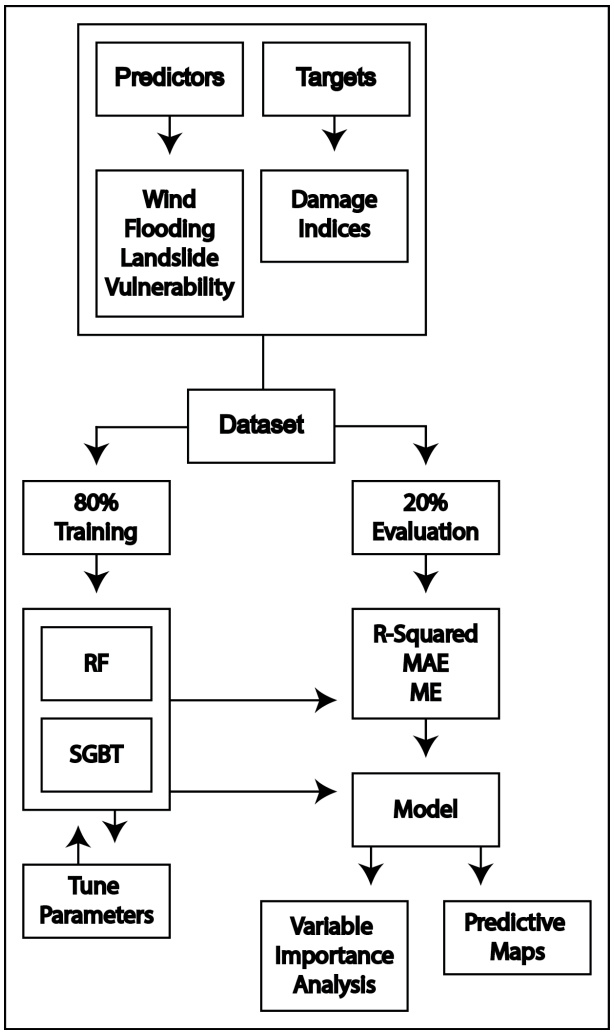

Figure 3.3: Workflow summary.

and landslide (AveLS) hazards, as well as vulnerability (SeVI, StrVI, CDCVuln, PropSC). RF and SGBT regression algorithms were optimized and trained on a dataset, then validated with a separate evaluation dataset. Three measures of feature importance were extracted from each trained model as well as spatial visualizations of the predictions. Figure 3.3 summarizes this study's workflow.

Chapter 4

Results

4.1 Data Processing

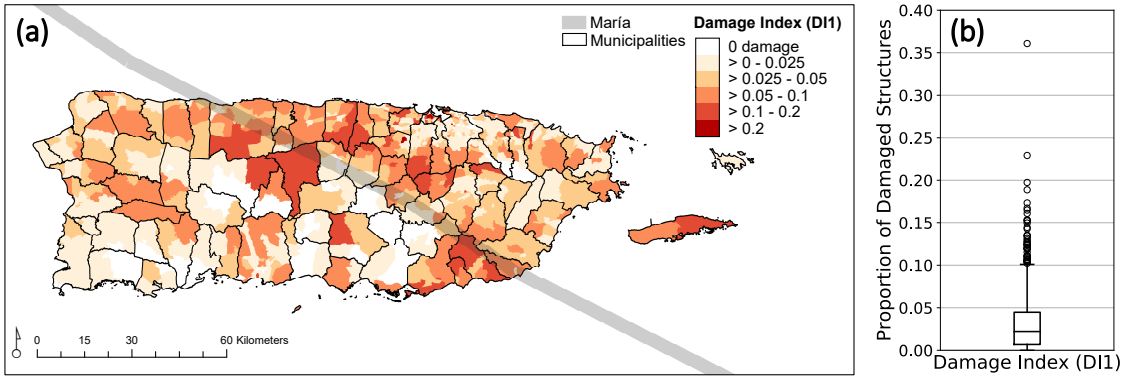

Figure 4.1: Summary of the (a) spatial and (b) statistical distribution of baseline damage index (DI1) values.

Results from DI1 indicated a distribution of damage and areas of highest damage appeared to loosely follow the center path of Hurricane María, illustrated in Figure 4.1a. DI1 ranged from 0 to 0.36 and is skewed towards lower values, with a mean of 0.032 (Figure 4.1b). The census tracts within the municipality of Ciales had the highest mean DI1 value, followed by

Vieques and Arroyo (Figure 4.2a). San Juan contained the census tract with the maximum DI1 value, followed by Toa Alta and Bayamon. Ciales, Arroyo, and Catano were found in both the top 10 mean and maximum census tract baseline damage index value ranks (Figure 4.2b).

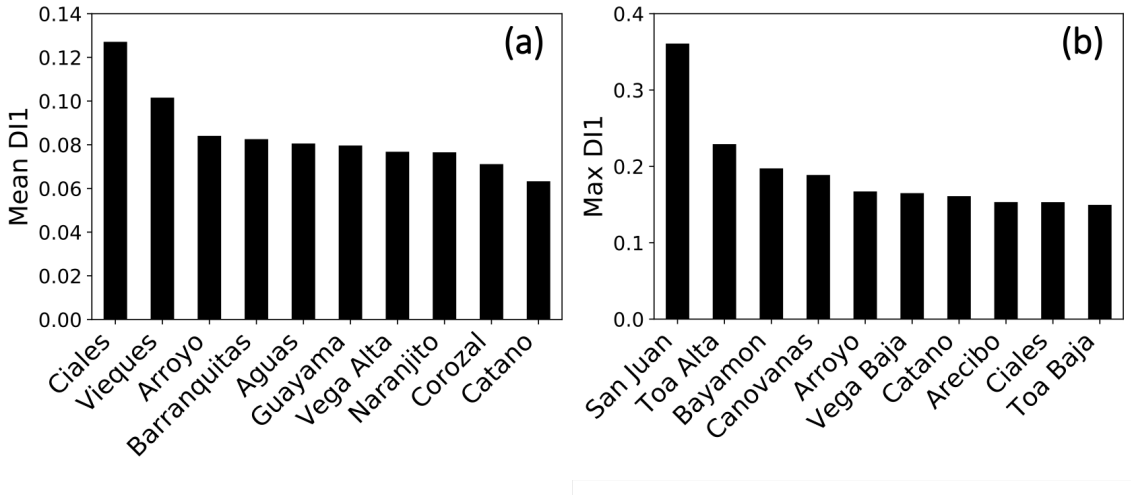

Figure 4.2: Results from damage index development. Mean values (a) result from a municipality- wide average of the census tract damage index values. Maximum values (b) refer to the maximum census tract damage index value within each municipality.

Although there are 905 census tracts in Puerto Rico, 23 tracts did not contain census data because they are unpopulated (e.g. parks, military bases). Feature processing resulted in a total of 12 predictive features and 882 populated census tracts to be analyzed in the study (Figure 4.3). Table 4.1 displays all Spearman correlation values (variable name abbreviations can be found in Table 3.1). Correlations with DI1 were strongest with StrVI (0.23), HurTrack (-0.22), and PropSC (0.21). DI3 exhibited the strongest correlations with all predictive features. The correlation analysis also found that two pairs of features – MaxDepth and AveDepth as well as MaxSusWinds and PeakGust – had Spearman correlation values higher than 0.9. Therefore, one feature from each pair was selected, resulting in a total of 10 predictive features for the machine learning analysis (Figure 4.3).

Table 4.1: Spearman correlation matrix of all predictive and target variables used in the analysis.

	HurTrack	PeakGust	MaxSusWinds	PropFA	AveDepth	MaxDepth	PropSFHA	AveLS	PropSC	CDCVuln	StrVI	SeVI	DI1	DI2	DI3
HurTrack	1														
PeakGust	-0.18	1													
MaxSusWinds	-0.17	0.99	1												
PropFA	-0.11	-0.13	-0.11	1											
AveDepth	-0.17	-0.19	-0.18	0.82	1										
MaxDepth	-0.18	-0.20	-0.19	0.84	0.98	1									
PropSFHA	0.25	-0.04	-0.02	0.41	0.28	0.28	1								
AveLS	-0.22	-0.24	-0.26	-0.02	0.14	0.12	-0.26	1							
PropSC	0.04	-0.09	-0.09	0.01	0.03	0.02	0.04	0.06	1						
CDCVuln	0.01	-0.24	-0.24	0.01	0.01	0.01	0.05	0.06	0.30	1					
StrVI	-0.23	-0.39	-0.40	0.14	0.24	0.25	-0.11	0.24	0.20	0.32	1				
SeVI	-0.10	0.10	0.09	0.01	0.00	0.00	-0.06	0.02	-0.21	-0.49	-0.13	1			
DI1	-0.22	-0.01	0.01	0.13	0.17	0.17	0.03	0.06	0.21	0.16	0.23	-0.17	1		
DI2	-0.22	-0.01	0.01	0.13	0.17	0.17	0.02	0.06	0.21	0.16	0.23	-0.17	1.00	1	
DI3	-0.24	-0.08	-0.06	0.13	0.19	0.20	0.02	0.12	0.26	0.25	0.31	-0.24	1.00	1.00	1

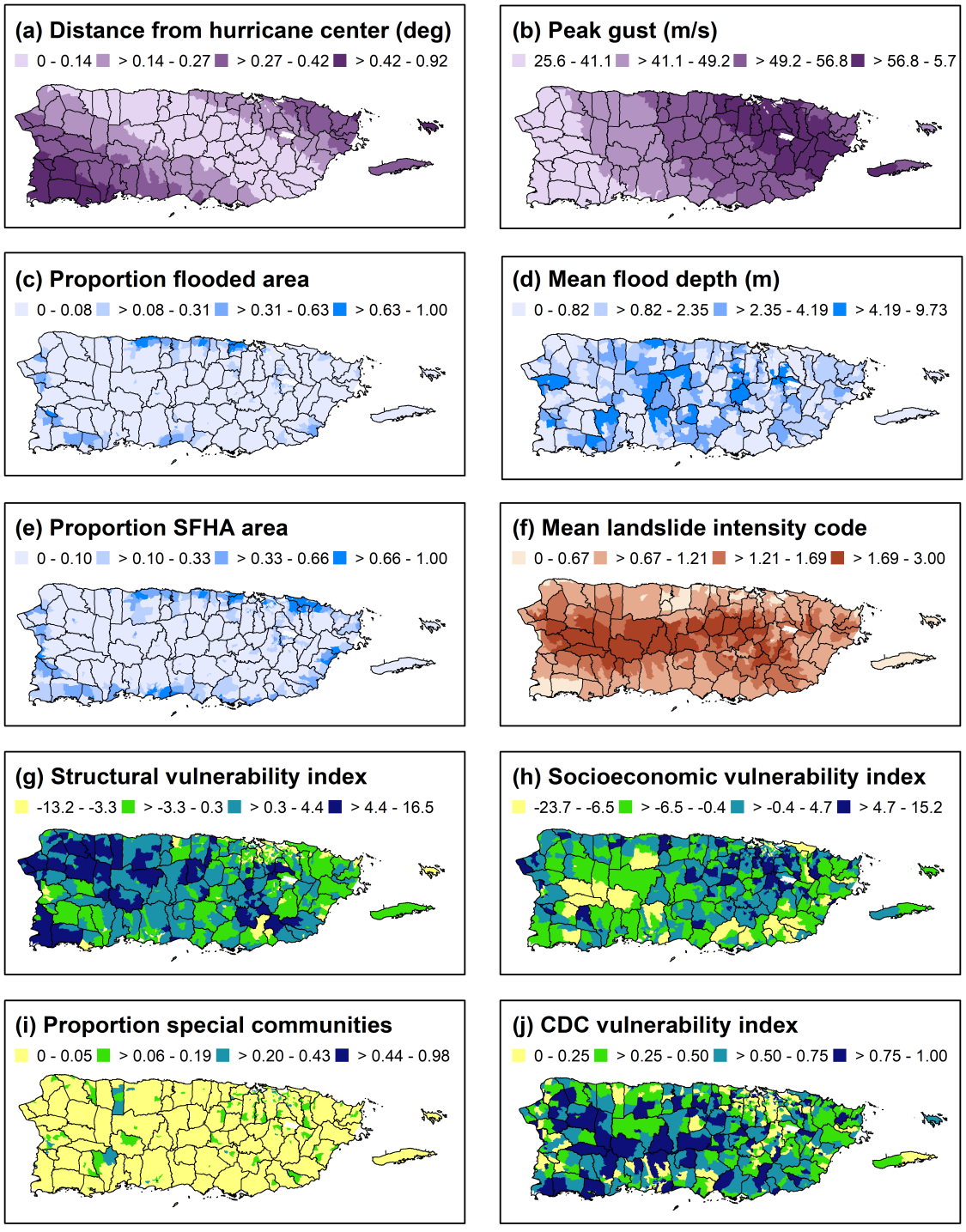

Figure 4.3: Predictive features selected for the machine learning analyses. Refer to Section 3.1 for data sources.

4.2 Model Performance

The automated tuning process searched for the optimal combination of 5 critical RF hyperparameters: maximum depth of each tree, maximum number of predictive features evaluated per tree, minimum samples required at a node, minimum samples required to split a node, and the overall number of estimator trees constructed. The best performing RF model, trained with DI2, limited each tree's depth to 21 levels to avoid overfitting. The maximum number of random features analyzed at each decision tree split was constrained to 8 features and a total of 2,300 predictive trees comprised the ensemble. It required the minimum number of samples to split a node (2 data samples) and the minimum number of samples required at each node (1 data sample).

Table 4.2: Performance comparison of all models run in this study.

		DI1	*DI2*	*DI3*	*Average*
RF	R^2 Evaluation	0.29	0.32	0.29	0.30
	R^2 Training	0.89	0.89	0.76	0.85
	ME	-0.0029	-0.0023	-0.00096	-0.0021
	MAE	0.019	0.019	0.019	0.019
SGBT	R^2 Evaluation	0.33	0.37	0.36	0.35
	R^2 Training	0.71	0.85	0.70	0.76
	ME	0.0026	0.0018	0.0020	0.0021
	MAE	0.018	0.018	0.018	0.018

The automated tuning process optimized an additional two hyperparameters in the SGBT models: random subsample considered at each tree and learning rate. Without evaluating the random subsample parameter, the SGBT algorithm would default to considering all data at each tree and would become simply the Gradient Boosting Trees algorithm, with no element of randomness. The learning rate and the total number of trees must be tuned in tandem to carefully maintain the balance between accuracy and computational expense. If the learning rate is lowered, then the number of trees should be increased because it controls

how drastically the next tree in the sequence corrects errors in the previous tree.

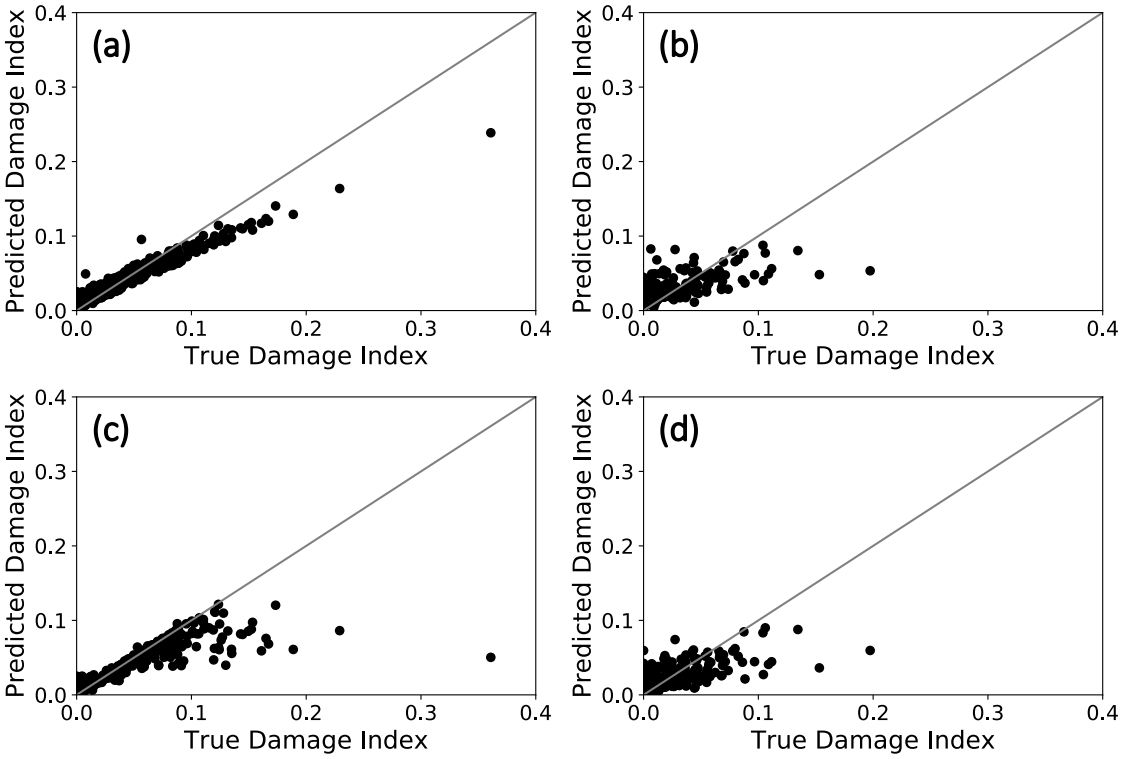

Figure 4.4: True and predicted DI1 values for (a) RF training data; (b) RF evaluation data; (c) SGBT training data; (d) SGBT evaluation data.

The best performing SGBT algorithm, also trained with DI2, permitted each tree built to learn from a random subset of 90% of the training data, learned at a rate of 0.09, constructed 500 trees total each with a maximum depth of 11 levels, evaluated a maximum of 3 predictive variables per tree, required a minimum number of 15 samples to split a node as well as 3 samples per node.

Table 4.2 summarizes model performances. On average, the RF algorithm accounted for 85% of the variance (R^2) in the training data and 30% of the variance in the evaluation data, with an average ME of -0.0021 and MAE 0.019 on the evaluation data across all damage indices. The SGBT algorithm, on average, accounted for 76% of the variance in the training data and 35% of the variance in the evaluation data, with an average ME of 0.0021 and

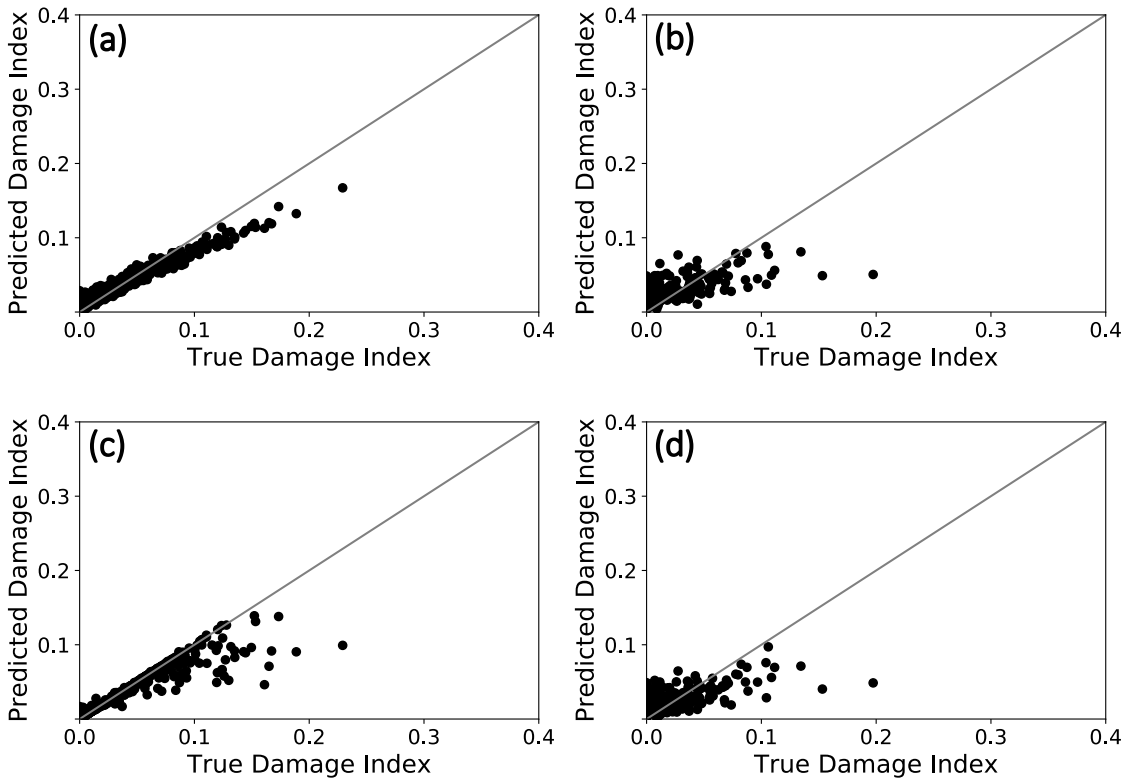

Figure 4.5: True and predicted DI2 values for (a) RF training data; (b) RF evaluation data; (c) SGBT training data; (d) SGBT evaluation data.

MAE 0.018 of on the evaluation data across all indices.

Both algorithms performed best when trained with DI2 (damage index excluding the highest outlier point). The leading RF model captured 89% of the training variance and generalized to represent 32% of the variance on the separate, unseen evaluation dataset. The MAE measured 0.019 while the model's ME measured 0.0023 for the evaluation dataset. The best SGBT model capture 85% of the training variance and generalized to represent 37% of the variance on the separate, unseen evaluation dataset. The MAE measured 0.018 while the model's ME measured -0.0018 for the evaluation dataset.

Averaged across the entire evaluation dataset, the RF models overpredicted damage, with negative ME scores, while the SGBT models underpredicted damage, with positive ME

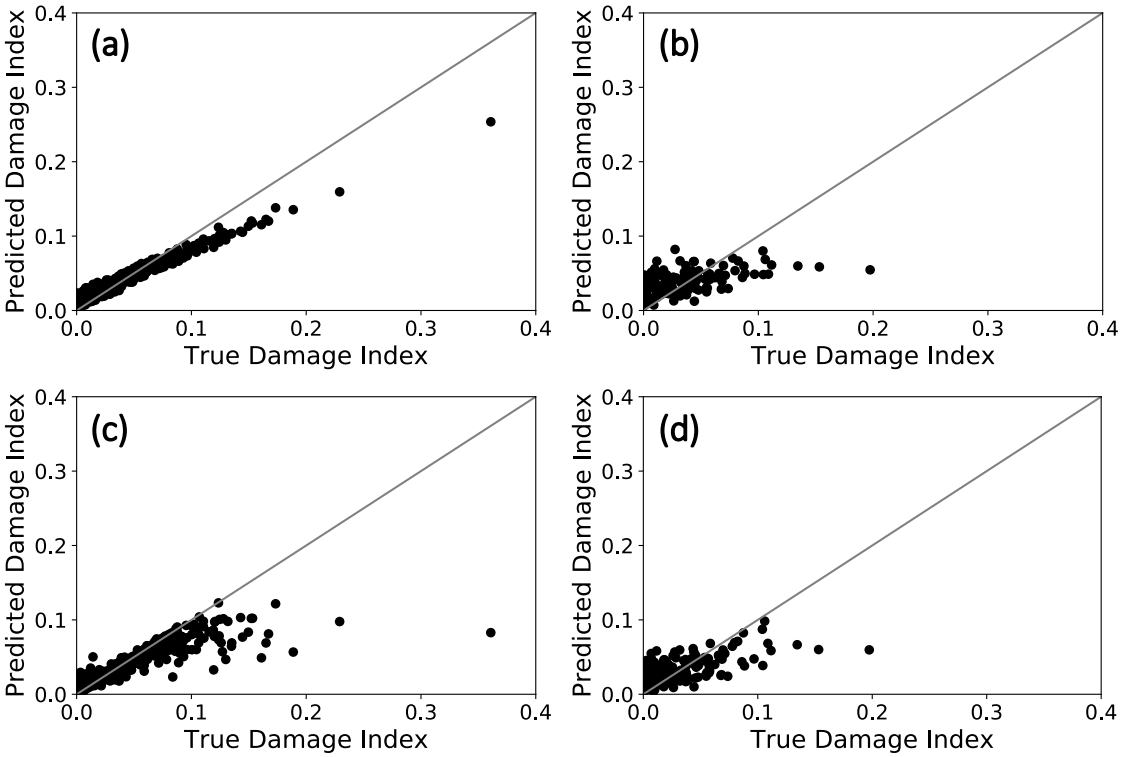

Figure 4.6: True and predicted DI3 values for (a) RF training data; (b) RF evaluation data; (c) SGBT training data; (d) SGBT evaluation data.

scores (Table 4.2). Figures 4.4, 4.5, and 4.6 plot the model predictions against the true data values for each algorithm and damage index. It can be seen from these plots that locally, models appear to overpredict at low damage index values and underpredict at high damage index values (Figures 4.4, 4.5, 4.6).

Predictive maps generated by the top performing model (SGBT-DI2) reveal the spatial patterns in model performance, seen in Figure 4.7. The model's overall spatial distribution of damage is representative (Figure 4.7a,b), areas with high error follow diagonally across the island (Figure 4.7c) and errors center around 0 (Figure 4.7d).

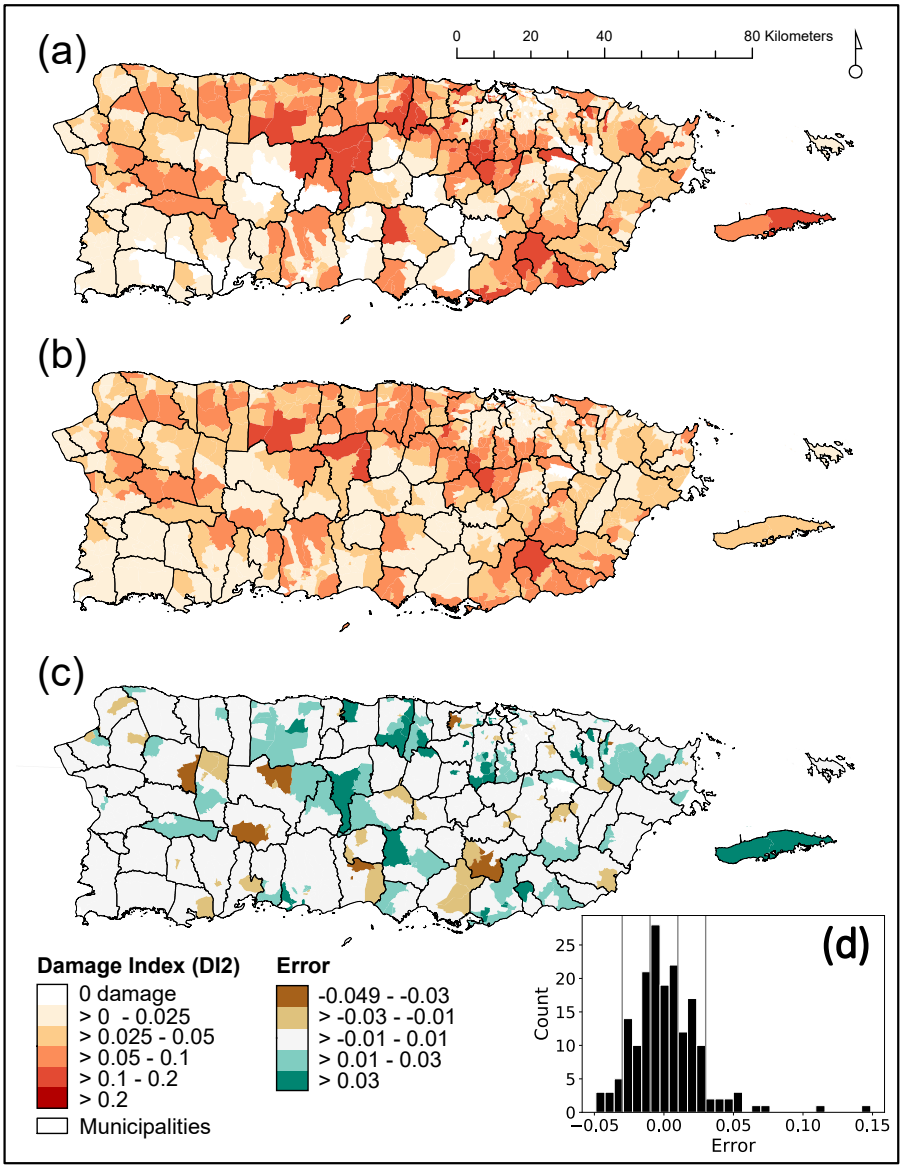

Figure 4.7: Spatial representation of (a) true DI2 values compared to (b) the predicted SGBT model output. Errors (true DI2 values - SGBT model predicted output) are displayed in (c) along with the error histogram (d).

4.3 Role of Vulnerability in Model Predictions

Each method to assess predictive feature importance revealed a unique interpretation of the model results (Figures 4.8 and 4.9). In the best performing model (SGBT-DI2) StrVI was the leading predictive feature (Figure 4.9b). All models consistently indicated that a vulnerability measure contributed the most predictive information. Randomly permuting the individual predictive variables revealed an interesting nuance. The CDCVuln variable became less important compared to the default measure of feature importance in all models while PropSC became more important in five of the six models. The four least informative features were often different orders of the three flood measures and AveLS. PropFA was the least predictive variable in all models.

When permuting predictive features categorically (randomly shuffling the four vulnerability variables, three flood variables, two wind variables, or the landslide variable) and measuring the subsequent drop in model variance, all models indicated that vulnerability was the leading predictive category, followed by wind, flood, and landslide (Figure 4.8III and 4.8III). Permuting the vulnerability variables in tandem resulted in approximately an 80% drop in model R^2, with wind reducing R^2 by approximately 50%, flood 25%, and landslide 10%.

The learned marginal impact of each predictive feature and the damage indices demonstrate a variety of relationships, Figure 4.10 displays the partial dependencies extracted from the best performing model (SGBT-DI2) as well as rug plots of data distributions. PropSFHA (4.10a), PropFA (4.10c), and AveLS (4.10d) exhibit relatively flat, horizontal relationships, indicating no pattern between how these variables relate to damage. For the majority of values, AveDepth (4.10b) appears to also be horizontal, but sharply increases at the variable's highest values. However, the distribution shows sparse data, therefore the sharp spike may be an over-interpolation due to lack of information. The predicted damage index de-

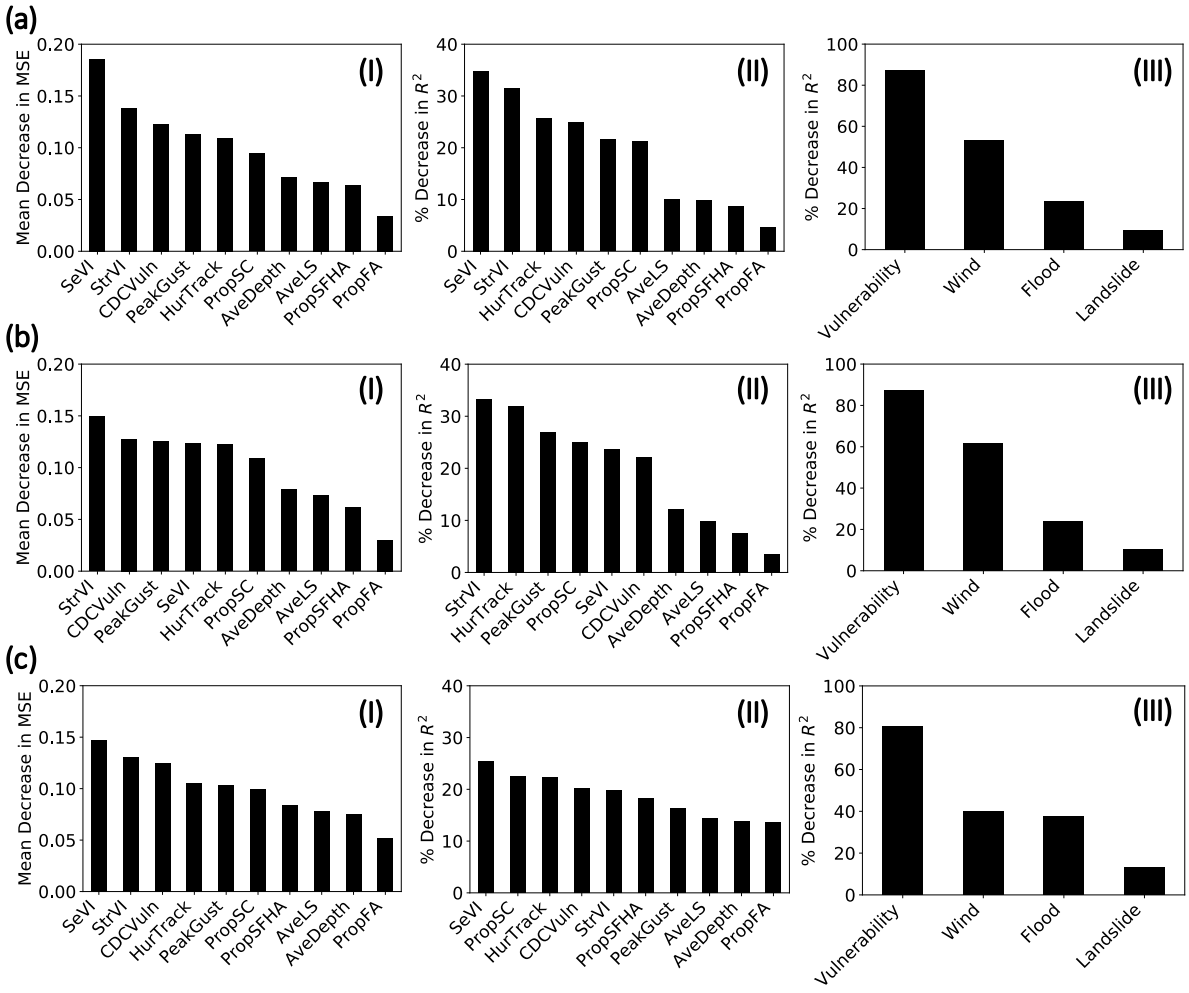

Figure 4.8: Feature importance analysis from the RF algorithm models for DI1 (a), DI2 (b), and DI3 (c). The default measure of mean decrease in MSE (I), the permutation measure of the percentage decrease in R^2 (II), and the group permutation measure of the percentage decrease in R^2 (III) provide unique indications of the important predictive features in each model.

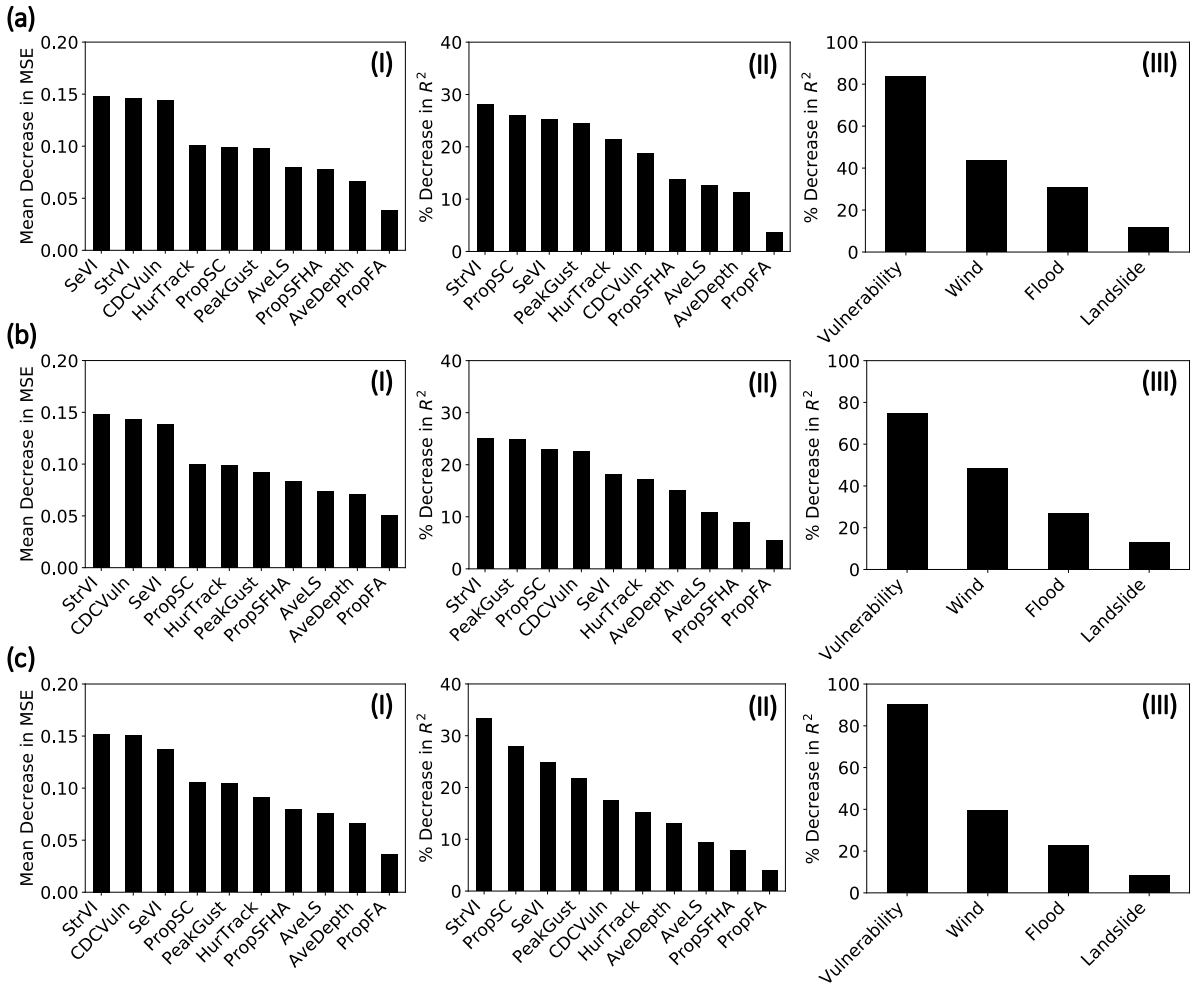

Figure 4.9: Feature importance analysis from the SGBT algorithm models for DI1 (a), DI2 (b), and DI3 (c). The default measure of mean decrease in MSE (I), the permutation measure of the percentage decrease in R^2 (II), and the group permutation measure of the percentage decrease in R^2 (III) provide unique indications of the important predictive features in each model.

creases as the distance away from the center of the storm (HurTrack) increases, although this relationship is more complex at very short distances away from the center of the storm (4.10e). Expected damage increases as PeakGust (4.10f), PropSC (4.10g), StrVI (4.10h), and CDCVuln(4.10j) increase. Damage appears to increase exponentially as StrVI increases (4.10h). Interestingly, SeVi (4.10i) exhibits a negative correlation with damage, also seen in the Spearman correlation values (Table 4.1).

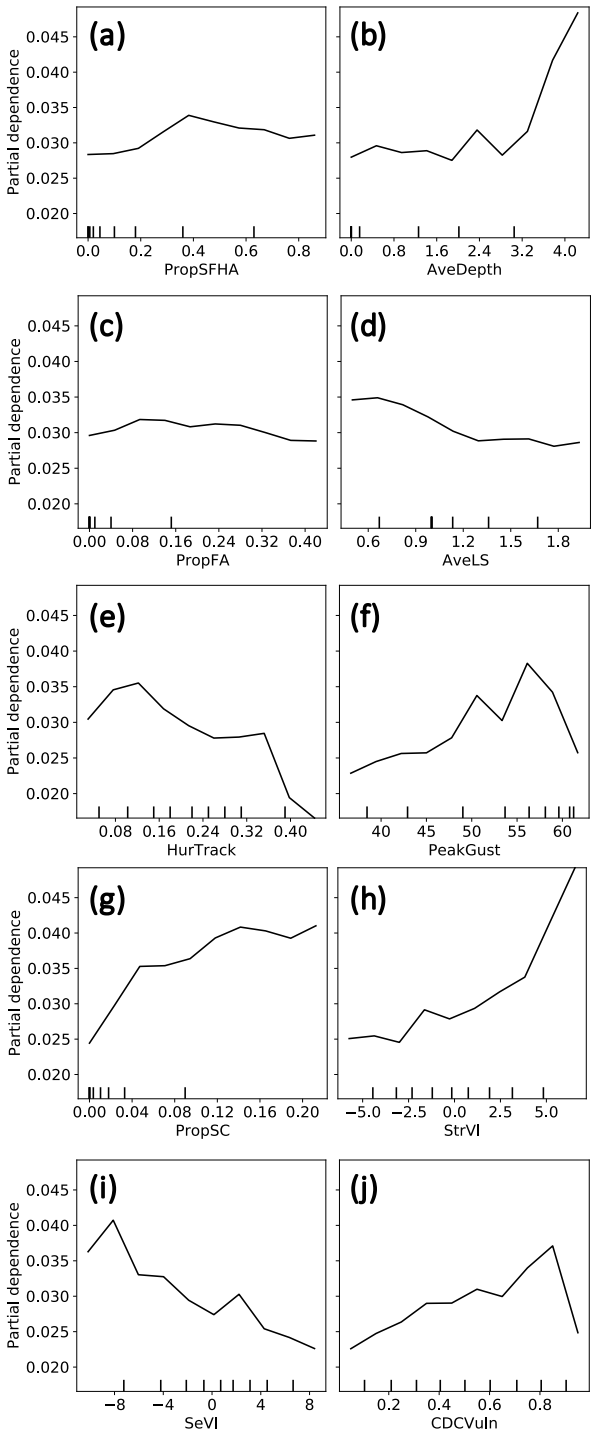

Figure 4.10: Partial dependence plots extracted from the SGBT Algorithm - DI2, where the y-axis is the change in the damage index prediction as function of different predictive features. A rug plot is provided for each feature to indicate data distribution.

Chapter 5

Discussion

5.1 Role of Vulnerability on the Damage Inflicted by Hurricane María

Vulnerability features correlated with Hurricane María's damage patterns and provided leading information to the machine learning models above all other wind, flood, and landslide variables. While PeakGust and HurTrack were considered important predictive variables, StrVI, PropSC, CDCVuln and SeVI were more often leading predictive variables. Categorically, vulnerability measures provided at least 30% more information to reduce model variance than wind, 55% than flood, and 70% than landslide measures. Overall, vulnerability measures played a critical role in the analysis of damage patterns.

By calculating importance with three different methods across two different algorithms, these results do not rely upon misinterpretations due to inherent biases of a single method. Upon individually permuting features, the CDCVuln variable became quantified as less important (Figure 4.8I,II and Figure 4.9I,II). This may be a relic of the bias from the default measure of

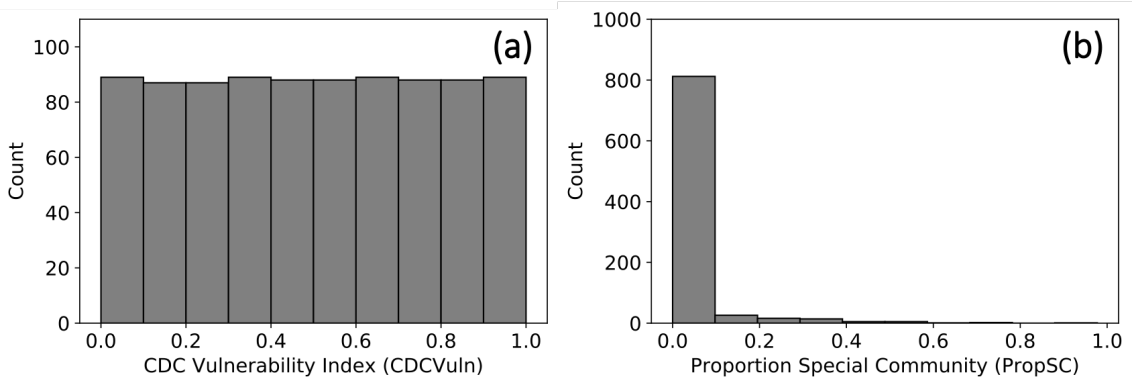

Figure 5.1: The CDCVuln feature (a) exhibits a wider and more evenly spaced distribution than the PropSC feature (b).

feature importance (mean decrease in MSE). Strobl et al. (2007) pointed out that the mean decrease in MSE method inflates the importance of variables with high cardinality. Upon permutation, the CDCVuln feature drops in importance and the PropSC feature increases in importance, possibly due to their respective distributions (Figure 5.1). CDCVuln exhibits higher cardinality than PropSC, potentially resulting in an exaggerated importance measure of CDCVuln when using the default method of mean decrease in MSE. Therefore, the findings of this study support previous assertions that the mean decrease in MSE measure can be biased and that using multiple strategies to calculate feature importance can avoid misinterpretations of model results.

The partial dependency analysis indicated that many predictive features had intuitive relationships with the damage index. Of note, increased StrVI led to an exponential increase in damage values (4.10h) and increasing PropSC resulted in a sharp increase in damage values (4.10g), levelling off as data became more sparse. One possible cause for the sharp drop at the end of the CDCVuln curve (4.10j) could be due to areas of high vulnerability that did not receive damage in geographically sheltered areas. Therefore, they do not conform to the general pattern of increased damage with increased vulnerability because of exposure to hazards of lower intensity. Similarly, the sharp drop at the high values of PeakGust (4.10f)

could possibly be due to areas that were exposed to intense hazards but were structurally very resilient.

Interpretation of the partial dependency of PropFA and AveDepth should be cautioned because their Spearman correlation value is 0.82 (Figure 4.1) and the analysis assumes independence between predictive features. Furthermore, the analysis captures the average effect of each predictive variable on the damage index and therefore does not capturing heterogeneities across individual data instances.

SeVI exhibited an unexpected negative partial dependency relationship with damage and a Spearman correlation of -0.49 with the pre-existing CDC vulnerability index (CDCVuln). This could be due to to limitations of the Cutter et al. (2003) method. Upon developing the index, implementation of different minimum thresholds of variance and alternative methods of scaling resulted in an inconsistent variety of results. While SeVI was calculated with the inductive method of PCA, CDCVuln was calculated with a deductive method, utilizing a pre-identified consistent set of indicator variables to calculate the index. Furthermore, CDCVuln is vetted and applied in peer-reviewed research (Alméstica, 2018; Rudner, 2019), while SeVI is not as established. In order to remove model misinterpretation due to inaccuracies in the developed indices, the best performing model was trained without the SeVI and StrVI variables. Upon re-optimizing and running the best performing model (SGBT-DI2) without two variables, the remaining vulnerability measures persisted as leading predictive features, though with a lower magnitude due to the loss of the information (Figure 5.2).

Puerto Rico specifically may not be suitable for vulnerability quantification with traditional methods. Due to the prevalence of informal housing (Asociación de Constructores de Puerto Rico, 2018), census measurements may not capture information on the most vulnerable households, and therefore may lead to misrepresentations of true vulnerability, possibly explaining why the StrVI measure correlated with damage while SeVI did not. In order to

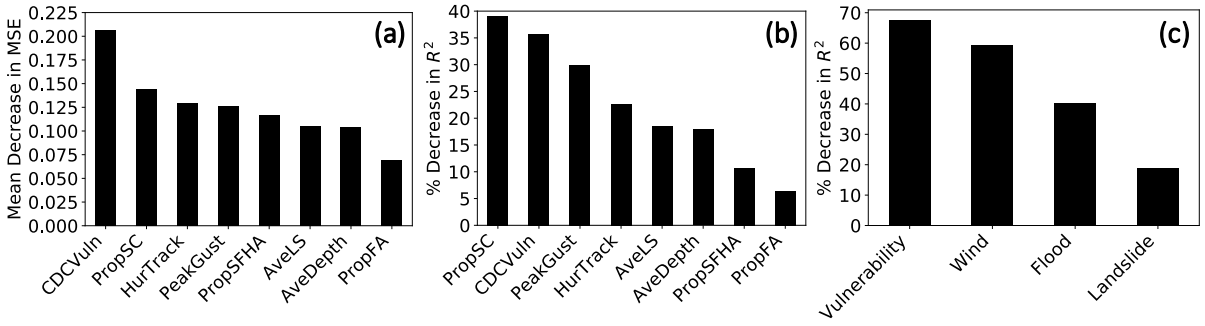

Figure 5.2: Results from SGBT-DI2 without the predictive features of SeVI and StrVI. Therefore the vulnerability category is comprised of CDCVuln and PropSC, wind is PeakGust and HurTrack, flood is PropFA and AveDepth, while landslide is AveLS. The default measure of mean decrease in MSE (a), the permutation measure of the percentage decrease in R^2 (b), and the group permutation measure of the percentage decrease in R^2 (c) provide unique indications of the important predictive features in each model.

improve the existing measures of Puerto Rican vulnerability, future work will be dedicated to developing a place-based methodology to create representative indices (see Section 5.4 Future Work).

5.2 Model Performance

SGBT models outperformed RF models. SGBT was able to account for more variance in the evaluation dataset, and was less prone to overfitting the training dataset. These models also had lower ME and MAE scores than RF models. Because SGBT incorporate boosting in addition to employing stochastic ensemble methods, it may be able to acquire better inductive rules than RF. SGBT has been found to outperform other algorithms, including RF, in previous studies as well (Shafizadeh-Moghadam et al., 2018; Rahmati & Pourghasemi, 2017; Youssef, Pourghasemi, Pourtaghi, & Al-Katheeri, 2016).

The reported model performance metrics indicate that the machine learning algorithms are struggling to generalize the data. This may be due to the imbalanced distribution in

target variable values and the small number of data samples (Figure 4.1). Generalizability is challenging on disproportionately distributed data with many outliers since machine learning algorithms optimize to reflect average conditions (He & Garcia, 2009). In this study, there may not enough instances of high data values to learn patterns without the tendency to overfit with unduly specific rules. This may be why DI2 performed best. Previous studies with similarly distributed targets have indicated that this is a common challenge (Sadler et al., 2018).

As seen in Table 4.2, model predictions based on the training data have high R^2 scores (0.70 - 0.89) while model predictions based on evaluation data account for less variance (0.29 - 0.37). Overall, the predictive performance of the RF and SGBT models is limited by the variance in the damage data. The dataset contains a high amount of variance due to a number of possible factors, including the scale of the analysis, passage of Hurricane Irma to the north of the island just weeks before Hurricane María, and a dearth of basic structural and building material data. Furthermore, the spatially diagonal pattern in the errors (Figure 4.7c), which appear to coincide with the center track of the storm (Figure 4.1a), could be indicative of the lack of topographic effects in the wind data (such as wind speed up through mountainous terrain). However, given these limitations, a predictive performance of 37% on an evaluation set is appropriate (Koch et al., 2019).

5.3 Implications

The findings of this study indicate that hazardous forces alone do not sufficiently explain damage patterns. Hurricane impact assessment models should include social factors as input variables to accurately depict areas of priority for decision makers, improve resource allocation, and, ultimately, ensure a more efficient and equitable response effort. Because this

strategy requires a large volume of data with diverse formats and sources, machine learning methods, including algorithms introduced in this study, can provide useful approaches to disaster management (Deparday, Gevaert, Molinario, Soden, & Balog-Way, 2019).

However, certain applications of machine learning methods should be cautioned. For example, inappropriately using machine learning approaches to inform insurance rates can result in unethical decisions. Machine learning models are incomplete approximations of reality that are prone to reflecting inherent biases that exist within society (Deparday et al., 2019). Vigilance is required to develop and interpret reputable models, especially in high-stakes disaster management decision making. Challenges, including social vulnerability, need to be overcome with the careful selection of a diverse suite of representative data, which can be challenging for data scarce regions, including Puerto Rico (Soden, Wagenaar, Luo, & Tjieesen, 2019).

While society does not control natural events, it indeed determines socioeconomic and infrastructure policies that precede them, either acting to increase a community's susceptibility to harm or, alternatively, promote resiliency and sustainability. The findings of this study contribute to the conclusions of existing literature which encourage deliberate pre-disaster mitigation investment in vulnerable communities (Fothergill & Peek, 2004). In order to prevent the exacerbation of social disparities in the wake of disasters, it is vital that the fundamental structural and institutional policies which lead to society-wide inequalities are reflected upon.

5.4 Future Work

The SeVI measure exhibited surprising correlations with both damage and CDCVuln. This study showed that different methods of quantifying vulnerability indicated different correla-

tions with damage and that traditional methods of quantifying socioeconomic vulnerability in particular can be misleading. It is especially pertinent in regions with fewer resources that existing methods of demographic data collection and vulnerability quantification are continuously refined to accurately represent the most vulnerable communities.

Collaborators in the Business Information Technology Department at Virginia Tech are currently investigating improvements to this widely accepted method of quantifying vulnerability. This will involve adapting the Cutter et al. (2003) into a place-based methodology. Because vulnerability factors will vary locally, broad methodologies that are not driven by community input will fall short of accurately representing critical phenomena. Therefore, if possible, this method will also incorporate community engagement strategies, directly sourcing stakeholder engagement in order to fine-tune vulnerability indices.

The data compiled during this project can also be applied to explore a variety of other relevant issues. Similar to vulnerability, post-disaster displacement is also challenging to quantify. Previous studies have relied upon school enrollment, housing market indicators, and FEMA assistance change of address requests as proxies for displacement (Crowley, 2006). Incorporating the aerial structural damage assessment data into a displacement analysis could provide a new method for displacement quantification.

By combining FEMA Individual Assistance data from Hurricane María with data used in this study, correlations between aid, damage, and vulnerability can be assessed. This would be another interesting application of the data already gathered for this study. Findings from that effort could determine if the needs of all communities have been fairly addressed and would provide recommendations for efficient post-disaster aid distribution.

5.5 Summary

This study provides evidences that, (a) at the census tract level, vulnerable communities in Puerto Rico suffered higher levels of damage due Hurricane María, (b) vulnerability measures were more predictive of damage than wind, flood, and landslide hazards, (c) SGBT outperformed RF algorithms, (d) reliance on one interpretation of feature importance can incur bias, and (e) all methods of quantifying vulnerability are not of equal quality. Various disaster impacts may be reduced if prevailing vulnerabilities are addressed proactively before an event brings them to the fore. To support this effort, existing methods of vulnerability quantification need to be continuously refined and information regarding appropriate applications must be developed. Fortunately, data availability provides an opportunity for researchers to implement statistical learning approaches and studies that seek to provide situational awareness throughout the life-cycle of a disaster may find these approaches helpful. This study demonstrated that these emerging methods can analyze diverse datasets representing multiple drivers of disaster impacts, including social factors, and provide valuable, holistic estimates of damage patterns and intensities as well as quantify the influence of vulnerability variables.

References

Alméstica, E. V. S. (2018). María y la vulnerabilidad en Puerto Rico. *Revista de Administración Pública*, *49*, 13–38.

Applied Research Associates. (2017). *Hurricane María wind data.* `https://disasters.geoplatform.gov/publicdata/NationalDisasters/2017/HurricaneMaria/Data/Wind/ARA/`.

Asociación de Constructores de Puerto Rico. (2018). Situación de la industria de la vivienda en puerto rico: Recomendaciones de política pública. *Informe Final*, 1–63.

Bessette-Kirton, E. K., Cerovski-Darriau, C., Schulz, W. H., Coe, J. A., Kean, J. W., Godt, J. W., ... Hughes, K. S. (2019). Landslides triggered by Hurricane María: Assessment of an extreme event in Puerto Rico. *GSA Today*, *29*(6).

Bolin, R. (1985). Disasters and long-term recovery policy: a focus on housing and families. *Review of Policy Research*, *4*(4), 709–715.

Bolin, R., & Stanford, L. (1991). Shelter, housing and recovery: a comparison of US disasters. *Disasters*, *15*(1), 24–34.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123–140.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and regression trees. *Wadsworth Int. Group*, *37*(15), 237–251.

Burton, C. G. (2010). Social vulnerability and hurricane impact modeling. *Natural Hazards Review*, *11*(2), 58–68.

Center for Disease Control. (2017). *CDC social vulnerability index Puerto Rico.* `https://respond-irma-geoplatform.opendata.arcgis.com/datasets/39490368e512402ba5d7635458b18f30`.

Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H., Bui, D. T., Pham, B. T., & Khosravi, K. (2017). A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental modelling & software*, *95*, 229–245.

Chhotray, V., & Few, R. (2012). Post-disaster recovery and ongoing vulnerability: Ten years after the super-cyclone of 1999 in Orissa, India. *Global environmental change*, *22*(3), 695–702.

Crowley, S. (2006). Where is home? Housing for low-income people after the 2005 hurricanes. *There is no such thing as a natural disaster: Race, class, and Hurricane Katrina*, 121–166.

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. *Social science quarterly*, *84*(2), 242–261.

Dahl, K., Cleetus, R., Spanger-Siegfried, E., Udvardy, S., Caldas, A., & Worth, P. (2018). Underwater: Rising seas, chronic floods, and the implications for US coastal real estate. *Cambridge, MA: Union of Concerned Scientists. Online at www. ucsusa. org/sites/default/files/attach/2018/06/underwater-analysis-full-report. pdf*.

Deparday, V., Gevaert, C., Molinario, G., Soden, R., & Balog-Way, S. (2019). *Machine learning for disaster risk management.* World Bank.

Eaton, K. J. (1980). Low-income housing and hurricanes. In *Wind engineering* (pp. 7–21). Elsevier.

Eroglu, D. I., Pamukçu, D., Szczyrba, L., & Zhang, Y. (2020). Analyzing and contextualizing social vulnerability to natural disasters in Puerto Rico. *Proceedings of the*

16th International Conference on Information Systems for Crisis Response and Management ISCRAM 2020 (Eds. Amanda Lee Hughes, Fiona McNeill and Christopher Zobel). (in press)

Federal Emergency Management Agency. (2017a). *Modeled preliminary observations.* https://disasters.geoplatform.gov/publicdata/NationalDisasters/2017/HurricaneMaria/Data/DepthGrid/.

Federal Emergency Management Agency. (2017b). *Preliminary depth grid methodology – Hurricane María Puerto Rico and St. Croix.*

Federal Emergency Management Agency. (2018a). *FEMA 213: Answers to questions about substantially improved/substantially damaged buildings.*

Federal Emergency Management Agency. (2018b). *Fema p-2020 mitigation assessment team report: Hurricanes Irma and Maria in Puerto Rico.*

Federal Emergency Management Agency. (2018c). *Historical damage assessment database, public release.* https://communities.geoplatform.gov/disasters/historical-damage-assessment-database/.

Federal Emergency Management Agency. (2018d). *Puerto Rico, Commonwealth of: Effective products.* https://msc.fema.gov/portal/advanceSearch#searchresultsanchor.

Federal Emergency Management Agency. (2019a). *Flood insurance rate map (FIRM) database technical reference preparing flood insurance rate map databases.*

Federal Emergency Management Agency. (2019b). *FP 104-009-03: Individual assistance program and policy guide (IAPPG).*

Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., & Lewis, B. (2011). A social vulnerability index for disaster management. *Journal of homeland security and emergency management*, *8*(1).

Fothergill, A., Maestas, E. G., & Darlington, J. D. (1999). Race, ethnicity and disasters in the United States: A review of the literature. *Disasters*, *23*(2), 156–173.

Fothergill, A., & Peek, L. A. (2004). Poverty and disasters in the United States: A review of recent sociological findings. *Natural hazards*, *32*(1), 89–110.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, *38*(4), 367–378.

Ganguly, K. K., Nahar, N., & Hossain, B. M. (2019). A machine learning-based prediction and analysis of flood affected households: A case study of floods in Bangladesh. *International journal of disaster risk reduction*, *34*, 283–294.

Grömping, U. (2009). Variable importance assessment in regression: linear regression versus random forest. *The American Statistician*, *63*(4), 308–319.

Hartman, C. W., & Squires, G. D. (2006). *There is no such thing as a natural disaster: Race, class, and Hurricane Katrina*. Taylor & Francis.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, *21*(9), 1263–1284.

Highfield, W. E., Peacock, W. G., & Van Zandt, S. (2014). Mitigation planning: Why hazard exposure, structural vulnerability, and social vulnerability matter. *Journal of Planning Education and Research*, *34*(3), 287–300.

Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278–282).

Holand, I. S., Lujala, P., & Rød, J. K. (2011). Social vulnerability assessment for Norway: A quantitative approach. *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography*, *65*(1), 1–17.

Hong, H., Pourghasemi, H. R., & Pourtaghi, Z. S. (2016). Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology*,

259, 105–118.

Karagiannopoulos, M., Anyfantis, D., Kotsiantis, S., & Pintelas, P. (2007). Feature selection for regression problems. *Proceedings of the 8th Hellenic European Research on Computer Mathematics & its Applications, Athens, Greece*, *2022*.

Kitzbichler, S. (2011). Built back better? Housing reconstruction after the tsunami disaster of 2004 in Aceh. *Asian Journal of Social Science*, *39*(4), 534–552.

Klein, N. (2018). *The battle for paradise: Puerto rico takes on the disaster capitalists*. Haymarket Books.

Koch, J., Stisen, S., Refsgaard, J. C., Ernstsen, V., Jakobsen, P. R., & Højberg, A. L. (2019). Modeling depth of the redox interface at high resolution at national scale using random forest and residual gaussian simulation. *Water Resources Research*, *55*(2), 1451–1469.

Krieger, N. (2006). A century of census tracts: health & the body politic (1906–2006). *Journal of urban health*, *83*(3), 355–361.

Liaw, A., & Wiener, M. (2018). Package 'randomforest': Breiman and Cutler's random forests for classification and regression. *R package version*, *4*, 1–29.

Ma, C., & Smith, T. (2019). Vulnerability of renters and low-income households to storm damage: evidence from hurricane maria in puerto rico. *American journal of public health*(0), e1–e7.

McKay, G., & Harris, J. (2016). Comparison of the data-driven random forests model and a knowledge-driven method for mineral prospectivity mapping: A case study for gold deposits around the Huritz Group and Nueltin Suite, Nunavut, Canada. *Natural Resources Research*, *25*(2), 125–143.

Merz, B., Kreibich, H., & Lall, U. (2013). Multi-variate flood damage assessment: a tree-based data-mining approach. *Natural Hazards and Earth System Sciences*, *13*(1), 53.

Moran, P. (1948). Rank correlation and product-moment correlation. *Biometrika*, *35*(1/2), 203–206.

Mueller, E. J., Bell, H., Chang, B. B., & Henneberger, J. (2011). Looking for home after Katrina: postdisaster housing policy and low-income survivors. *Journal of Planning Education and Research*, *31*(3), 291–307.

Naghibi, S. A., Moghaddam, D. D., Kalantar, B., Pradhan, B., & Kisi, O. (2017). A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. *Journal of Hydrology*, *548*, 471–483.

National Hurricane Center. (2017). *NHC GIS Archive - Tropical Cyclone Best Track - Hurricane María.* `https://www.nhc.noaa.gov/gis/archive_besttrack.php?year=2017`.

Oficina del Coordinador General para el Financiamiento Social y la Autogestión. (2008). *Comunidades especiales.* `http://www.gis.pr.gov/descargaGeodatos/Delimitaciones/Pages/Comunidades.aspx`.

Oliveira, S., Zêzere, J. L., Queirós, M., & Pereira, J. M. (2017). Assessing the social context of wildfire-affected areas. the case of mainland Portugal. *Applied Geography*, *88*, 104–117.

OpenStreetMap. (2019). *HOTOSM Puerto Rico Buildings.* `https://data.humdata.org/dataset/hotosm_pri_buildings#`.

Pasch, R. J., Penny, A. B., & Berg, R. (2018). National hurricane center tropical cyclone report: Hurricane Maria. *TROPICAL CYCLONE REPORT AL152017, National Oceanic And Atmospheric Administration and the National Weather Service*, 1–48.

Peacock, W. G., Dash, N., & Zhang, Y. (2007). Sheltering and housing recovery following disaster. In *Handbook of disaster research* (pp. 258–274). Springer.

Peacock, W. G., Van Zandt, S., Zhang, Y., & Highfield, W. E. (2014). Inequities in long-term housing recovery after disasters. *Journal of the American Planning Association*, *80*(4), 356–371.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Dubourg, V. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, *12*(Oct), 2825–2830.

Prieto, C., Le Vine, N., Kavetski, D., García, E., & Medina, R. (2019). Flow prediction in ungauged catchments using probabilistic random forests regionalization and new statistical adequacy tests. *Water Resources Research*, *55*(5), 4364–4392.

Rahmati, O., & Pourghasemi, H. R. (2017). Identification of critical flood prone areas in data-scarce and ungauged regions: a comparison of three data mining models. *Water resources management*, *31*(5), 1473–1487.

Rahmati, O., Pourghasemi, H. R., & Melesse, A. M. (2016). Application of gis-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. *Catena*, *137*, 360–372.

Rodriguez-Galiano, V., Chica-Olmo, M., & Chica-Rivas, M. (2014). Predictive modelling of gold potential with the integration of multisource information based on random forest: a case study on the Rodalquilar area, Southern Spain. *International Journal of Geographical Information Science*, *28*(7), 1336–1354.

Ross, T. (2013). *A disaster in the making: Addressing the vulnerability of low-income communities to extreme weather*. Center for American Progress.

Rudner, N. (2019). Disaster care and socioeconomic vulnerability in Puerto Rico. *Journal of health care for the poor and underserved*, *30*(2), 495–501.

Sadler, J., Goodall, J., Morsy, M., & Spencer, K. (2018). Modeling urban coastal flood severity from crowd-sourced flood reports using poisson regression and random forest. *Journal of hydrology*, *559*, 43–55.

Santiago-Bartolomei, R. (2018). Notes for a planning and public policy framework for housing in Puerto Rico. *Center for a New Economy Housing and Land Initiative*, 1–6.

Santos-Burgoa, C., Sandberg, J., Suárez, E., Goldman-Hawes, A., Zeger, S., Garcia-Meza,

A., ... others (2018). Differential and persistent risk of excess mortality from Hurricane Maria in Puerto Rico: a time-series analysis. *The Lancet Planetary Health*, *2*(11), e478–e488.

SAS Institute Inc. (1989-2019). *Jmp pro, version 15.*

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768.

Shafizadeh-Moghadam, H., Valavi, R., Shahabi, H., Chapi, K., & Shirzadi, A. (2018). Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. *Journal of environmental management*, *217*, 1–11.

Soden, R., Wagenaar, D., Luo, D., & Tjieesen, A. (2019). Taking ethics, fairness, and bias seriously in machine learning for disaster risk management. *arXiv preprint arXiv:1912.05538*.

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, *8*(1), 25.

Suthaharan, S. (2016). Supervised learning algorithms. In *Machine learning models and algorithms for big data classification* (pp. 183–206). Springer.

Tesfamariam, S., & Liu, Z. (2010). Earthquake induced damage classification for reinforced concrete buildings. *Structural safety*, *32*(2), 154–164.

Trigila, A., Iadanza, C., Esposito, C., & Scarascia-Mugnozza, G. (2015). Comparison of logistic regression and random forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). *Geomorphology*, *249*, 119–136.

U. S. Department of Housing and Urban Development. (2018). *Housing damage assessment and recovery strategies report Puerto Rico.*

United States Geological Survey. (2019). *Map data from landslides triggered by Hurricane María in four study areas of Puerto Rico.* `https://www.sciencebase.gov/catalog/`

item/5ca3c65fe4b0b8a7f6334309.

U.S. Census Bureau. (2017). *2017 Census Bureau ACS 5-year Estimate.*

Van Zandt, S., Peacock, W. G., Henry, D. W., Grover, H., Highfield, W. E., & Brody, S. D. (2012). Mapping social vulnerability to enhance housing and neighborhood resilience. *Housing Policy Debate*, *22*(1), 29–55.

Van Zandt, S., & Rohe, W. M. (2011). The sustainability of low-income homeownership: the incidence of unexpected costs and needed repairs among low-income home buyers. *Housing Policy Debate*, *21*(2), 317–341.

Vickery, P. J., Skerlj, P., Steckley, A., & Twisdale, L. (2000). Hurricane wind field model for use in hurricane simulations. *Journal of Structural Engineering*, *126*(10), 1203–1221.

Walsh, K. J., McBride, J. L., Klotzbach, P. J., Balachandran, S., Camargo, S. J., Holland, G., ... Sobel, A. (2016). Tropical cyclones and climate change. *Wiley Interdisciplinary Reviews: Climate Change*, *7*(1), 65–89.

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, *527*, 1130–1141.

Ward, P. S., & Shively, G. E. (2017). Disaster risk, social vulnerability, and economic development. *Disasters*, *41*(2), 324–351.

Youssef, A. M., Pourghasemi, H. R., Pourtaghi, Z. S., & Al-Katheeri, M. M. (2016). Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides*, *13*(5), 839–856.