EVALUATION OF TRANSPORTATION SAFETY AGAINST FLOODING IN DISADVANTAGED COMMUNITIES

May 2022 | Final Report

SAFE-D
SAFETY THROUGH DISRUPTION

Texas A&M Transportation Institute

San Diego State University
Leadership Starts Here

VT | TRANSPORTATION INSTITUTE
VIRGINIA TECH

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 05-101	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle EVALUATION OF TRANSPORTATION SAFETY AGAINST FLOODING IN DISADVANTAGED COMMUNITIES		5. Report Date 6. Performing Organization Code:
7. Author(s) Hassan Tavakol-Davani Vincent T. O'Hara-Rhi Sahar Ghanipoor Machiani		8. Performing Organization Report No. Report 05-101
9. Performing Organization Name and Address: Safe-D National UTC San Diego State University		10. Work Unit No. 11. Contract or Grant No. 69A3551747115/Project 05-101
12. Sponsoring Agency Name and Address Office of the Secretary of Transportation (OST) U.S. Department of Transportation (US DOT)		13. Type of Report and Period Final Research Report 14. Sponsoring Agency Code

15. Supplementary Notes

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.

16. Abstract

Flooding in urban areas, especially in low-income or disadvantaged communities, poses a serious problem to drivers. While techniques exist to map and predict flooding events, a knowledge gap exists in accurate mapping and prediction of urban flooding. It is important to have an understanding of how much flooding a region may experience given a certain weather event so that drivers may preemptively avoid flooded areas. This paper synthesizes several approaches to build an understanding of the spatial extent of urban flooding in the frequently flooded parts of San Diego, California. First, flooding reported during major storms was used as validation data for a Generalized Linear Regression model to create a map of flood risk. Then, a Support Vector Machine model was used to extract areas of possible flooding from a satellite image. Finally, model performance was compared. Each model provided robust and meaningful results, with the Generalized Linear Model indicating which areas of the city are most at risk for flooding and the image classification Support Vector Machine model successfully identifying water bodies during both dry and wet conditions.

17. Key Words Remote sensing, Flood control, Geographic information system, Support Vector Machines	18. Distribution Statement No restrictions. This document is available to the public through the Safe-D National UTC website, as well as the following repositories: VTechWorks, The National Transportation Library, The Transportation Library, Volpe National Transportation Systems Center, Federal Highway Administration Research Library, and the National Technical Reports Library.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 39	22. Price $0

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Abstract

Flooding in urban areas, especially in low-income or disadvantaged communities, poses a serious problem to drivers. While techniques exist to map and predict flooding events, a knowledge gap exists in accurate mapping and prediction of urban flooding. It is important to have an understanding of how much flooding a region may experience given a certain weather event so that drivers may preemptively avoid flooded areas. This paper synthesizes several approaches to build an understanding of the spatial extent of urban flooding in the frequently flooded parts of San Diego, California. First, flooding reported during major storms was used as validation data for a Generalized Linear Regression model to create a map of flood risk. Then, a Support Vector Machine model was used to extract areas of possible flooding from a satellite image. Finally, model performance was compared. Each model provided robust and meaningful results, with the Generalized Linear Model indicating which areas of the city are most at risk for flooding and the image classification Support Vector Machine model successfully identifying water bodies during both dry and wet conditions.

Acknowledgements

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program. The authors would also like to thank all of the subject matter experts who reviewed this work.

Table of Contents

List of Figures

List of Tables

Introduction

Flooding is a perennial problem exacerbated by urbanization and climate change. While catastrophic events such as hurricanes, dam failures, or hundred-year storms draw the most attention, less severe flooding also leads to property damage, infrastructure failure, and fatalities, especially in dense urban areas. Extreme events are well-documented and studied, and formal risk assessments such as the Federal Emergency Management Agency's National Flood Hazard Layer map are powerful tools for determining and managing flood risk. However, the complex urban topography and the small spatial scales at which urban flooding is relevant make modeling flood risk in dense cities challenging. There is no unequivocal definition for such minor flooding, but it is known by some and will be labeled here as Nuisance Flooding or NF, defined as minor flooding that can pose serious danger, especially to drivers [1].

Most climate models predict that precipitation in Southern California will become more seasonal. Rainfall totals are expected to remain at around current levels, but more extreme interannual variability and longer, drier dry seasons and shorter, wetter wet seasons will lead to both more severe drought and increased flooding, making water management more difficult [2], [3], [4]. Global sea levels are also rising, inundating coastal areas and overwhelming storm drainage infrastructure [5]. With these climactic changes and California's burgeoning population, it has become essential to understand the spatial extent of flooding—not only along floodplains such as that of the San Diego River, which is well-understood as being an area chronically at-risk for flooding—but also in built-up areas where flood risk, especially at NF levels, remains poorly understood. Mapping NF is a challenge for three main reasons: (1) the urban environment is highly complex, with waterways at submeter resolutions; (2) flooding is shallow and ephemeral; and (3) ponding means that the flood extent will be discontinuous, making analysis with a traditional hydrologic model difficult. While 2-D hydrologic models are effective at lower resolutions, the computational resources to properly model urban flooding are difficult to come by with current technology [6].

When considering NF in urban zones, an area as small as an intersection is relevant, and most techniques will have difficulty accurately mapping and predicting at such high resolution. Attempts have been made to map urban flooding and flood risk with more traditional methods, such as [7], which compared the efficacy of four hydrologic models ranging from a simple runoff-response model to more complex 2-D models; [8], which used a combination of hydraulic models and Random Forest (RF) and Multilayer Perceptron machine learning (ML) algorithms; and [9], which used building footprints to create topography for modeling flooding at a very small scale. These studies gleaned good results, but work remains to be done. The authors of [7] found success in their modeling but stated that the 2-D models required for high accuracy are not feasible for larger urban study areas, exemplifying the need for finding methods of mapping or predicting flooding that are less computationally intensive. Hydrologic models may be the best solution in the future, but better computational and more accurate inputs (such as those from

sensors placed in the correct locations), may need to be implemented before such models can be widely used for real-time flood mapping. Another tested method is to use a catchment's hydrologic properties and anthropomorphic variables as independent variables to create maps of flood hazard and vulnerability in urban areas. Many of these studies distinguish between flood risk—or the probability that an area will be exposed to flooding—and flood hazard—a combined metric combining risk and other vulnerability receptors such as income or infrastructure age. Some examples of this technique include [10], which used an ensemble of ML techniques and datasets to map flood hazard and risk in Iran; [11], which applies a similar method using the GARP and QUEST models; and [12], which gives an overview of flood hazard studies for cities in China.

Research has also been done in flood detection using remote sensing methods such as aerial photographs or satellite imagery. A summary of methods, such as Synthetic Aperture Radar (SAR) and LiDAR, is given in [13]. SAR is an especially promising technique due to the fact that, as an active sensor, radar can detect the Earth's surface no matter the local time of day or what cloud conditions prevail. Some notable studies in the detection of flooding with SAR include [14], which combined SAR imagery from COSMO-SkyMed and Landsat 8 Operational Land Imager data to map flooding along a river in northern China; [15], which employed RASARSAT-2 SAR images and flood stage data based on return period for the 2011 Richelieu River flood in Canada; and [16], which is the culmination of a series of studies using TerraSAR-X in tandem with very high-resolution aerial imagery to map flooding of the River Severn in England. A more in-depth overview of the techniques used in these studies and their relevance to this study will be discussed in the next section.

A combined approach has the double advantage of testing techniques presented in previous studies and running a rigorous analysis on the study area. This study consists of two analyses: first, the creation of a flood hazard map of the study area in the city of San Diego using a Generalized Linear Regression (GLR) model, hydrologic characteristics, and vulnerability variables, and second, flood detection using SAR imagery taken by COSMO-SkyMed and a Support Vector Machine (SVM) model. Lacking the robust validation data for flooding during a specific storm used in some other studies, such as high-resolution aerial imagery, the GLR model was then used to further test results gleaned from the SAR data. Finally, a flood hazard analysis was run for transportation networks in the City of San Diego to provide practical results that may be of use to improve flood preparedness. We also took into account flood hazard in disadvantaged communities and reflected this in flood vulnerability calculations. This study, outlined by a flowchart in Figure 1 below, seeks to build knowledge in the area of urban flood detection by testing methods in a new study area, that being a collection of seasonally wet urban drainage basins in San Diego, including the lowest reaches of the San Diego River watershed.

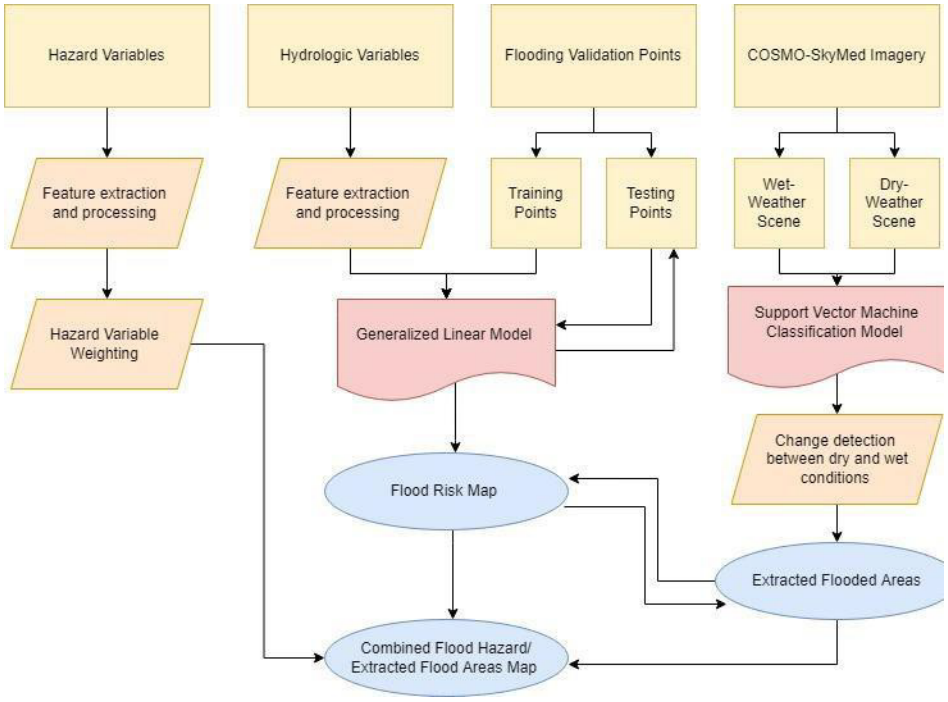

Figure 1. Project flowchart, showing each step of the project and relationships between inputs, outputs, and processes.

Background

Since its inception, remote sensing has proved quite valuable for scientists and engineers in many fields, allowing collection of data on a temporal and spatial scope that was once impossible. Though a powerful tool, remote sensing capability is limited by spatial and temporal resolution, conditions on the surface that might limit the sensor's functionality, and sensor availability. Flood mapping would be ideally carried out by high-resolution aerial photographs that could then be easily processed to extract water, but such flights are not always available and passive onboard sensors are limited by lighting and cloud conditions.

Flooding can occur over a short period of time and often underneath heavy cloud cover, making active sensors an ideal tool, especially for NF flooding, which may cover very small areas and quickly recede. While passive sensors require some input from the environment, whether it be visible light, ultraviolet radiation, or heat, active sensors such as SAR create their own inputs for the sensor to detect. SAR works by emitting microwave signals that bounce off the surface and are picked up by the sensor. This backscatter value σ, which is affected by the geometry of the surface that reflects the microwave, is then interpreted to glean information about the surface's nature. SAR is valuable because it can be used no matter what lighting or cloud cover conditions prevail [17]. SAR is applied to flood detection problems by seeking out a backscatter return that corresponds to smooth surfaces, such as floodwater. Water, along with other flat, smooth surfaces such as roadways or parking lots, returns a low backscatter value σ to a SAR sensor, appearing as dark regions on a radar image, as illustrated in Figure 2. In simple terms, studies

detecting flooding using SAR all sought out these regions of lower backscatter, either by training an algorithm to recognize the radar return of water through training samples or through change detection comparing images taken during flooded and unflooded conditions. This technique works best for regions in which a surface with a normally high radar return, such as cropland or low vegetation, is inundated and transformed into a smoother surface, which is then easily distinguished in a SAR image. Studies detecting floodwaters in rural areas, such as [14], [18], and [19], all reported good results, highlighting the efficacy of SAR in detecting surfaces whose reflectance characteristics have changed due to flooding.

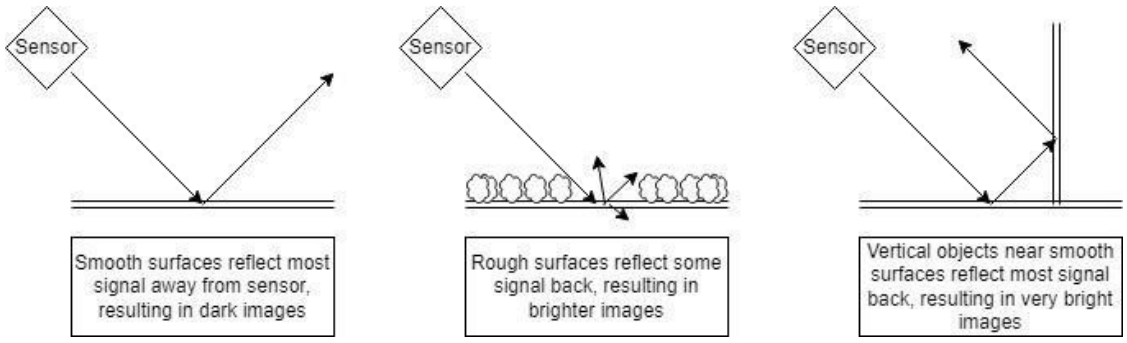

Figure 2. Diagram illustrating backscatter behavior of various surfaces [17].

SAR, however, also has limitations. Until recently, SAR data was considered insufficient for mapping flooding in urban zones due to low resolution and shadow and layover in the complex urban environment [20]. However, with a combination of more experience, higher-quality data, and innovative image processing techniques, there has been some success in using SAR for urban flood detection applications. An early foray into mapping urban flooding with SAR, [21] used TerraSAR-X data, a SAR simulator to remove shadow from buildings, and a region growing algorithm. [20] employed a gamma distribution to recognize the backscatter values of open water and another region-growing approach using open-water seeds for two case studies, while [16] and [22] improved on these aforementioned algorithms by including change detection processing and double-scattering recognition, respectively. Some studies, such as [16], extracted roadways and parking lots from flood mapping to avoid incorrectly classifying pixels. Other algorithms have also been used, such as [14], which employed a counter without edges (C-V) machine learning model to extract flooded areas or [23] and [24], which used interferometry, consisting of pairs of pre and post-flood images.

Another feature of flooding in urban areas is double-scattering, which has been explored by [22], [25], and [26]. Double-scattering is a phenomenon in which floodwaters on a surface will actually create a higher backscatter return. Double-scattering occurs in places where floodwaters adjoin a wall or other vertical surface, such as trees or vegetation. Radar signals will bounce first off the ground and then against the wall, being detected by the sensor in the same range band as signals deflected directly off the vertical surface. This creates a line of bright signal return at the edges of buildings, an effect that, in experiments conducted by [27] and [28], is enhanced by smoothing of the surface at the base of a building. A detailed review of techniques used and

results gathered by previous studies in the detection of urban flooding with SAR is included as Table 1.

Table 1. SAR image Classification Literature Review

Approach	Examples	Required Inputs	Results
Backscatter threshold, gamma distribution, region-growing algorithms	[21], [20] [29], [30]	SAR imagery and aerial/satellite imagery for validation.	Good results, with 76% of urban water pixels in [19] correctly classified. This method, however, excluded roadways and parking lots from consideration for having a radar return too similar to that of open water. In [29], rural flooding was used to extrapolate urban flooding during river flooding.
C-V model	[14]	SAR imagery and aerial/satellite imagery for validation.	93.70% total accuracy, but the C-V model is best applied to continuous flooded surfaces, such as overflowing rivers and other such large-scale flooding.
Double-scattering detection	[22], [25]	SAR imagery and aerial/satellite imagery for validation.	Good results, but lots of room for introduction of error. Care must be taken to ensure that other reflectors are not fudging results.
Change detection	[16]	SAR imagery and aerial/satellite imagery for validation.	Improvement on region-growing technique, some of the same restrictions still apply.
Interferometric pairing	[23], [24]	Interferometric SAR pairs and aerial imagery for validation.	89% accuracy for interferometry radar return data, 87% for just radar return. Interferometric pairs, however, require high revisit times and are much more costly to acquire.

The similarity between the radar return of water and that of anthropomorphically-smooth surfaces is what makes the SAR approach to urban flooding so difficult. The flooding the authors of this report and most other authors wish to detect is in smooth, built-up areas, flooding which poses a hazard to drivers and infrastructure in impervious areas not well-suited to the flow of stormwater. Unlike many previous studies, however, our study area does not lie within the floodplains of a major river in a humid area, nor were the storms in question major, 100-year-storm events. Therefore, many of the techniques that were used to link flooding in urban floodplains and flooding on smooth urban surfaces will not be applicable to our study. In addition, the commonly-used region-growing technique, in which algorithms were used to grow "seeds" of known inundation will also not be applicable due to the discontinuous nature of urban flooding, where ponding, rather than river floodplains, is the primary cause of inundation.

Since this study's main purpose is to detect flooding along roadways, a different approach was necessary. We used simple, non-processing-intensive image classification and change detection techniques to extract regions of possible flooding from a COSMO-SkyMed image taken during a major storm on February 27, 2017, in San Diego. Our approach detects possible flooded areas by training an SVM algorithm to recognize the darker radar return of smooth surfaces. After these areas of possible flooding were identified, a change detection filter was then applied to remove false positives. These techniques will be explained in further detail in the methods section.

Methods

Study Area and Datasets

The study area within the San Diego region was chosen based on the following parameters: an area entirely within the extent of COSMO-SkyMed images taken during times of flooding, an area with the necessary hydrologic data to create a flood risk model, and an area containing a sufficient number of validation points for statistical significance. The final area that meets this criteria, shown in Appendix A (Figure A1), is entirely within the city of San Diego, covering about 155 mi^2 and includes the flood-vulnerable Mission Valley area as well as disadvantaged neighborhoods such as City Heights and Barrio Logan. The study area is the same for both the flood risk mapping and image classification portions of the study. The climate in the city is semi-arid Mediterranean, with long dry seasons and short wet seasons during the winter months. The climate, measured near downtown in the south-western corner of the study area, is generally warm and mild, with maximum average monthly temperatures at around 25°C in August and September and average monthly lows in December and January at about 14° C. Precipitation is highly seasonal on both the coast and in the more humid highlands. Precipitation in the city itself averages 9.79 inches a year, with around three-quarters of this value falling between December and March [31]. San Diego's climate is also subject to high degrees of inter-annual variability, due to the effects of the El Niño- Southern Oscillation [32].

Another element of the study involved the calculation of flood hazard, a metric combining flood risk, or the risk that flooding will occur at a location, and flood vulnerability, or the susceptibility of an area to flooding due to factors such as income or population density. This study will cover areas known as Disadvantaged Communities, mapped in Figure A2 by census tract [33]. SB-525 Disadvantaged Communities are areas designated by the State of California as meriting special investments from proceeds of pollution-cap programs. The goal is to improve not only quality of life and opportunities in these communities, but also to improve environmental conditions by reducing pollution. The criteria are based on communities in which at least 25% of the population is adversely affected by unemployment, poverty, or lack of access to healthcare or proper nutrition, as well as regions that are known to be polluted. These criteria are regularly updated by CalEnviroScreen, which scores census tracts based on the criteria previously mentioned.

Some examples of validation data used by previous studies are the aforementioned LANDSAT and aerial imagery, tweets, road closures, and river gauge data by [34]; social media data and topographic LiDAR data by [35]; and [36], which used photos taken during a 100-year flood in 1979 in the city of Jackson, MS to create a model for future flooding. We unfortunately lack robust validation datasets such as aerial photographs of high enough resolution to detect urban flooding, and social media sources such as tweets were found to be inadequate for ground-truthing. We used police and news reports of flooded intersections and roadways as ground-validation data. To ensure enough points for validation, road closure reports were gathered for four different storm events spanning three distinct water years: January 6, 2016 [37, 38];

February 28, 2017 [39, 40]; December 8, 2018 [41, 42]; and February 14, 2019 [43]. These events were chosen because they all caused significant flooding, are all included in reports of flooding by news outlets and police in sufficient numbers to be statistically significant, and all occurred recently enough that it is unlikely that the region's hydrological landscape significantly changed. Multiple storms were chosen to ensure enough validation points for a meaningful, robust GLR model. While it is true that flooding during specific storms may have been caused by factors such as blockages, any area flooded during heavy rainfall, even when caused by infrastructure failure, is an area vulnerable to flooding in the first place. These validation points do not represent areas that *will* flood during a given storm, but areas at a risk for flooding. The provenance by storm of each validation point is listed in Table A.1.

One storm was chosen for the SAR portion of the project. The temporal window for the SAR flood was much narrower than that for the GLR model, requiring heavy flooding during the precise moment the SAR image was taken. One difference between this investigation and most previous studies is the magnitude of the event under analysis. Our goal in this respect is to apply methods found to be effective for larger-scale flooding, such as the 2007 overflow of the River Severn, which saw the inundation of over 55,000 homes after the largest record rainfall since records began in 1766. After the 2017 hurricane in Houston, [44] or [24] conducted analysis on flooding to smaller-scale urban flooding in the San Diego, which has not seen catastrophic flooding on the level of these aforementioned events in many decades. The most catastrophic flooding in recent history was the 1916 storm, during which 11.5 inches of rain fell in under 80 minutes and the gage height in Santee, CA was over 35 feet [45]. Another such storm is likely to occur again, and one way to understand how it might affect the densely built-up city is to study the impact of smaller storms. To build an understanding of the variability of year-to-year water flow in San Diego, Figure A3 shows streamflow in the San Diego River over the past decade, highlighting low flows during a long drought in the first part of the decade and heavy flows during the latter half. The boundary conditions for a storm we could study were as follows: one that occurred recently enough to be imaged by satellites equipped with high-resolution SAR, and one large enough to create enough flooding to be noted by news and police agencies.

To assess the severity of a storm and the likely extent of flooding, we used rainfall and river stage data. While rainfall is far from homogenous across the San Diego area, we focused on the San Diego River at Fashion Valley, this being where most flooding was reported. Under the reasonable assumption that peak gauge height will correspond to peak flooding, we used flood gauge data from the United States Geological Survey to find peak storms over the past decade during which high-resolution SAR sensors were active.

One gauge height peak in February 2017 immediately stands out. Closer inspection reveals that the gauge height reached a peak of almost 10,000 cubic feet per second at around 1:00 on February 28, 2017. Rainfall records also reflect a short, intense burst of rainfall at around 23:00 on the 27th. It was likely this high-intensity period of rainfall following steady rain throughout the previous day that caused the high gauge height and flooding. Flooding for this day can be further verified with other qualitative data. A National Oceanic and Atmospheric Administration report

on heavy flooding through California in February of 2017 reported an "above-average" flood in Mission Valley on the 27th and 28th, near historic peaks in 2011 and 1995 [46]. A Times of San Diego article reported enduring flooding on the 28th and evacuations in the Lake Hodges and Mission Valley areas, as well as 702 collisions logged by the California Highway Patrol, nearly seven times the fair-weather average [39]. The precipitation and streamflow at Fashion Valley is shown in Figure A4. The temporal and spatial extent of the study was constrained by available datasets and processing time. Thanks to the low return period of COSMO SkyMed's satellite constellation, a scene of the San Diego area taken at 17:41 (local time) on February 27, 2017, was found. This was around the time of the most intense precipitation and just before the flood peak in the San Diego River. Additional datasets are presented in the Appendix: Figure A5–Figure A7 show the hydrologic variables, Figure A8 and Figure A9 show the flood vulnerability variables, and Figure A10 shows the SAR images used in the study.

Flood Risk and Hazard Mapping

The methodology of this research consists of two steps. First, a flood risk map was created using a GLR model and hydrologic variables, then regions of possible inundation were extracted from SAR imagery taken during a storm on February 27, 2017, in San Diego using image classification techniques. After these areas were extracted, the risk model was used to validate the SAR detection.

Hydrologic Variables

Any hydrologic model recreates, in some fashion, watershed characteristics. The simplest models might use nothing beyond the rational method, while the most complex might attempt to model a watershed down to individual culverts and pipes. Our approach does not include the traditional hydrologic model process of creating a "design storm" that "falls" onto the model and creates flow, but instead is a model of flood risk based on points of known flooding. We chose not to use a traditional hydrologic model because the area we wished to model was too large to efficiently compute with hydrologic models at the resolution we needed to model, and because the nature of urban flooding includes localized ponding that traditional flow models may have trouble detecting.

Our approach, therefore, was to use points of known flooding as training data for a GLR model using hydrologic variables as explanatory variables. We chose our variables based on availability, statistical significance, and model testing. Some variables that were used in other studies, such as [11], [47], or [48], included variables such as slope, elevation, groundwater depth, distance to drains, drainage density, and soil type. Some of these datasets were suitable for our study area while others were not, and we added new ones based on newly developed techniques and urban hydrologic factors. Soil type, groundwater depth, rainfall distribution, and elevation were all discarded as potential hydrologic variables for our analysis. Hydrologic soil maps of the area do not include underlying soil underneath heavily urbanized regions, labeling them simply "X," so an imperviousness layer was used instead. Similarly, groundwater depth maps for the San Diego region simply show the presence of aquifers, all of which are underneath valleys, therefore adding little insight into the spatial distribution of urban flooding.

Rainfall distribution over our small coastal study area varies little—there are only two weather stations with good data availability, and it is likely that storm-specific variation would be the most important rainfall variable that would need to be included. Land use was discarded as an explanatory variable because our validation points, nearly all being reported flooding along roadways, would have introduced bias if land use were included to explain their spatial distribution. Finally, absolute elevation was not considered as an explanatory variable, as it is unlikely, in our small study area, that elevation would have any impact on rainfall and including it would have likely introduced error. These variables were tested during the model training portion of the project and were found to either overfit the model or to lack enough variation to be significant. The variables chosen after training iterations were presence of bluespots, distance to nearest flowline, slope, local watershed imperviousness, and local watershed drainage density. Variables that required a Digital Elevation Model (DEM) were found using an already processed 0.76-meter resolution hydro-conditioned DEM provided by San Diego Association of Governments' GIS division.

Bluespot is a term for a landscape depression that is likely to fill up during storm events, taking into account stream networks and presence of landscape sinks. More details about this technique can be found in [50]. Before bluespots were found, building footprints were burnt onto the previously mentioned hydro-conditioned raster to account for the diversion of flow around buildings, as well as to prevent bluespots being detected on building footprints. The bluespot technique can be used as a high-resolution model of spillover given certain rainfall amounts, but our approach employed bluespots as a method to identify sinks in the urban environment susceptible to flooding. Further analysis could be conducted using bluespots to find areas of flood risk given specific storm rainfall totals and even rainfall spatial distribution, but this is outside the scope of this study and our study area was too large to conduct such as analysis efficiently.

Distance to flowlines is a way to model areas that will receive higher volumes of water. Flow lines were found using a flow accumulation raster, itself derived from a flow direction raster. These flow lines can represent actual channels and streams, or they could also be areas such as streets that see high volumes of flow during intense rainfall events. For this study, we defined flowlines as being pixels in the flow accumulation raster which received at least 1,700,000 pixels of flow, equivalent to about 247 acres contributing flow to a certain point. Euclidean distance to flow lines was then calculated, and the output raster used as an explanatory variable. Local imperviousness was calculated by finding the average imperviousness of sub-watersheds. This was found to be a better explanatory variable than imperviousness as a rough estimate of the likelihood of water in the sub-watershed to become runoff.

Flooding in urban areas generally occurs in areas with high imperviousness with little infiltration capacity, so the purpose of imperviousness data as an explanatory variable is to model the impact imperviousness has on runoff. Areas with low imperviousness have increased runoff, which is what this element of the model is simulating. Local drainage density was also calculated by sub-watershed. Features which accept flow, such as inlets, were added over each area, and features which add flow, being outlets and clean outs, were then subtracted from the total amount of

inlets. This metric, referred to as drain density, is presented in the datasets section. Surface channels in the study area were found to be sparse and with little spatial correlation to flooding. The slope of the hydro-conditioned raster was used as a simple filter for areas with higher slopes where flooding would be very unlikely. No threshold was chosen, but the model was run by training it using slope values at validation points.

Validation Points

As discussed in the previous section, we collected 47 validation points from news reports covering four storms in the San Diego area from 2016–2019. We split the validation points into 37 points for training, and 10 points were tested based on the 80:20 rule. Ten points were randomly sampled from across the study area as testing points and the rest were used to train the GLR model. The spatial extent of flooding for the training points was estimated from the reports, and the data was then converted to a raster where flooded points were assigned a value of 1 and unflooded points a value of 0.

Generalized Linear Regression Model

Flood risk was mapped using a GLR model through ArcGIS Pro. GLR is a commonly used technique with a range of applications, such as predicting the impact of sea level rise on NF [51], likelihood of flash flooding [52], and modeling of flood risk [10]. The GLR model was chosen because, as mentioned in the background, it was found in the comparative study [11] to yield decent results in comparison to other models, and this technique was resource and processing-efficient. A GLR model is a generalized form of the linear regression model, which allows response variables to be related via a link function and the variance of measurements to be functions of their predicted values. [53]. GLR models are known for being able to handle multiple independent variables, including mixtures of categorical and continuous variables.

GLR models consist of three components: a random component, referring to the probability of the response variable Y; a systematic component specifying the independent variables; and a link function, which specifies the relation between the response variable and linear predictors gleaned from explanatory variables [54]. The link functions serve as a way of linking non-linear relationships between explanatory and dependent variables to the linearity of the overall model. In this case, a GLR model was used to create a flood risk map by linking the previously mentioned datasets: $Y = \beta_0 + \beta_1 X_1 + \ ... \ \beta_n X_n$. The regression equation consists of a random component Y, systematic component X, and regression coefficients β. As the dependent variable of concern is the presence of flooding, a binary logistic GLR equation was chosen. The link function for binary logistic GLR models is as follows: $\mu = \log\left(\frac{\pi}{1-\pi}\right)$.

ArcGIS Pro's GLR GeoAnalytics tool [33] was used to run the analysis on the explanatory variables. The binary regression model was chosen, and the input variable, presence of flooding, and the five explanatory variables were then rasterized and converted to a 5 m resolution point grid. We chose 5 m as a resolution both low enough to be processed efficiently and high enough to capture the high variability of the urban landscape and small bluespots that, though only a few

meters wide, could still pose a hazard. The probability column of the output table was then converted to a raster, clipped, and normalized.

Flood Vulnerability Variables, Weighting, and Flood Hazard Mapping

Flood vulnerability represents how impactful flooding might be in a certain area. In this study, flood vulnerability is a metric that applies only to the city's transportation network, as this research is focused on the effect of flooding on transportation. Most of our vulnerability variables are mapped and processed by census tract, a commonly used unit for measuring demographic statistics in the United States. Population density is a simple metric to measure how much of an effect flooding will have in an area based on how many people will be impacted by inundation. This also allows us to filter out sparsely-populated areas that might suffer less human cost than densely- populated regions. This metric was classified based on natural breaks present in the data. Income per census tract allows us to measure and represent how wealthy certain neighborhoods are, and how resilient they might be to damages created by flooding. Thresholds were chosen based on California income brackets, with the highest vulnerability assigned to census tracts with average income below the California poverty line. Average drain infrastructure age is an estimate of the age of hydraulic infrastructure by census tract. Aging infrastructure is more likely to fail and puts regions at higher risk for flooding [55] so this was also included as a vulnerability metric, also classified by natural breaks. Street Overall Condition Index, or OCI, is a measurement used by the city of San Diego to determine which streets might soon be in need of repairs [56]. Streets are scored on features such as cracking, potholes, pavement collapse, or sub base failure. Streets are given a "good," "fair," or "poor" rating based on these characteristics. Street OCI was included in the study as a measure of vulnerability of transportation infrastructure to flooding.

Each variable was assigned an equal weight, and vulnerability was then calculated over roadways for which OCI data was available within the study area. The weights for each variable were calculated based on breaks in the data, discussed earlier for each variable. Weights were assigned by best judgement, with weight for population density and drain age (breaks for these variables having been determined by natural breaks in the data) in equal intervals to reflect the gradual, linear increase in flood vulnerability as infrastructure ages or population becomes denser. Income was weighed to reflect the disproportionate impact of income below the poverty line in California, while street OCI index was weighed similarly to represent the much higher flood vulnerability of streets rated to be in poor condition rather than fair or good condition. As our variable weighting is meant to compare vulnerability within the study area, the focus here is to ensure that different levels of each variable are distinct and will create unique results in the final flood hazard map. After these weights, or vulnerabilities, were calculated for each point, they were summed and multiplied by the risk factor, itself represented by a weight depending on classified risk index. The following calculation yields flood hazard: *Hazard* = *Risk* x *Vulnerability.* Calculations were performed at 5 m intervals. Table A.2 shows weight used by variable.

Extraction of Flooded Areas With SAR Imagery

A summary of various image classification approaches and reasoning for this study's use of an SVM classification model was given in the previous background section. A more in-depth description of the methods used in this analysis follows. Ours is a combined approach, applying three methods consecutively to detect flooding that may express itself in the SAR imagery in distinct ways. First, an SVM model classified SAR imagery, then a land use filter was applied to remove surfaces such as roadways where flooding would not be detectable with SAR, and finally a change detection filter was used comparing images taken during dry and wet weather. Using radar images to extract flooding in urban areas has been proven possible but is a challenging and still-incipient process. Many of these methods have yet to be proven with a wide range of study areas and validation data, and this study seeks to further test applications of SAR data to urban flood detection with more limited validation data. Without resources such as high-resolution aerial imagery taken during flooding, we draw upon previous experience, model assessment techniques, and results of our flood risk model to validate results of SAR image classification.

Classification of SAR Imagery Using an SVM Model

The goal of SAR image classification is to recognize backscatter values associated with a certain feature on the surface of the Earth. Our approach was to use an image classification algorithm, the SVM model, to learn which pixels belong to which land use type. The SVM model has already been used to classify SAR imagery, with some examples being [57], which used an SVM model to classify land use in Denmark, and [58], which used the same model to determine land type and tree canopy species on Tubuai Island, French Polynesia. We apply this same technique to partition SAR imagery into three land types: shrubland, corresponding to grass, chaparral, and some canopy; urban, consisting of areas on and around buildings and tall trees; and water, made up of both open water surfaces such as ponds and lakes and also anthropomorphic surfaces with a similarly dark radar return, such as freeways and parking lots.

Imagery was acquired from COSMO-SkyMed satellites 1 and 4, which took images including the entirety of the study area. Images were pre-processed by the data providers, with some speckle reduction and alignment to a local DEM (Table A.3). First, the prepared and processed SAR images were clipped to the study area and segmented. Image classification on a pixel-by-pixel basis would create too much noise, as SAR images are inherently noisy datasets [59]. Segmentation was performed to group together regions of similar pixels and smooth the image. In addition, an object-based approach was used in the image classification process. Therefore, regions with a certain backscatter distribution were classified rather than selecting a specific threshold beneath which a pixel might be considered to be water.

Segmentation was performed through ArcGIS Pro's Image Segmentation tool, which uses the Mean Shift technique to group pixels together based on average values. The Mean Shift approach employs a moving window that iteratively groups pixels together based on their characteristics [60]. Segmentation functions under three parameters: spectral detail, spatial detail, and minimum segment size. Spectral detail assigns a level of importance to spectral differences

in imagery, spatial detail determines the importance of proximity between features (with high levels more suitable to features which are clumped together), while minimum segments size specifies the smallest number of pixels a segmented region can contain. For segmenting SAR data, a relatively high spectral detail was chosen to be able to distinguish between relatively small variations in backscatter values, as well as a high spatial detail to pick out details in the urban landscape. To reduce speckle, a high minimum segment size of 2,000 pixels was chosen.

After the image was segmented, we then classified the segments using an SVM, which is a supervised machine learning model [61]. First, training polygons were defined using a land use raster as reference. To ensure statistical significance, at least 20 segments per class were chosen. The model was then run using the training polygons. The SVM model is a technique that searches for a linearly separable hyperplane, a decision boundary that separates members of one class from another. SVM models find this hyperplane using support vectors and margins, and are known for their ability to model complex nonlinear relationships and their high accuracy [54]. Like other ML models, the SVM seeks to find the characteristics that separate one set of data from another. If the SVM hyperplane is a plane of all dimensions equated to zero and θ a constant known as the bias, then:

$$\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 \ldots \theta_n x_n = 0.$$

An optimal hyperplane will be as far as possible from all observations, as it seeks to represent a division between classes. Therefore, the distance of each observation to the hyperplane must be computed, the smallest of these distances being known as the margin. The SVM algorithm then finds the hyperplane with the maximum possible distance to groups of observations, and classifies observations based on whether they lie "above" or "below" the hyperplane. Points closest to the hyperplane define the hyperplane's position; if these are shifted slightly, then the location of the hyperplane shifts too. These points can have their own planes drawn through them, which are known as *support vectors*. Points past these support vectors can shift, but the support vectors, and by extension the hyperplane that defines the division between classes, only shift if the maximal observations also change. These critical points define changes in classification. This method allows SVM to avoid overfitting, an important feature for a classification schema in which we want to avoid overprediction of flooded surfaces. The SVM algorithm can also account for nonlinear relationships through kernel functions. Note that open street surfaces, especially those in large parking lots or on freeways far from the effects of shadow or layover, will appear as water on a classified image. Therefore, the land use raster was used to filter any areas associated with transportation, parking, or other smooth anthropomorphic surfaces.

Land Use and Change Detection Filters

The similarity of the radar return of water and urban surfaces such as roadways means that applying the SVM to only the wet-weather scene would lead to an excess of false positive results, as freeways, parking lots, and some larger streets would be classified as water by the model. Therefore, a land use dataset was used to filter out any areas classified as water that were contiguous with roads and asphalt or concrete parking lots. Then, a change detection filter was

applied, removing any areas classified as water in the wet-weather scene that were also of the water land type in the dry-weather scene. This step allowed us to filter out any permanent water bodies, such as ponds, or any other areas with naturally low backscatter distributions before inundation.

Results

Flood Risk Mapping

Results of the GLR model for flood risk are presented in this section. These results consist of a flood risk map showing the spatial pattern of flood risk, a table with coefficients and t-statistics that give information about the model's performance and which variables were most relevant, and a graph of flood risk index to justify the classification used. Figure 3 shows classified flood risk for the study area, as calculated by the GLR model. Note the concentration of flood risk in coastal areas and along the floodplains of the San Diego River. The nature of urban flooding, however, means that areas with concentrated, spatially extensive flood risk are not necessarily those that require the most attention and analysis.

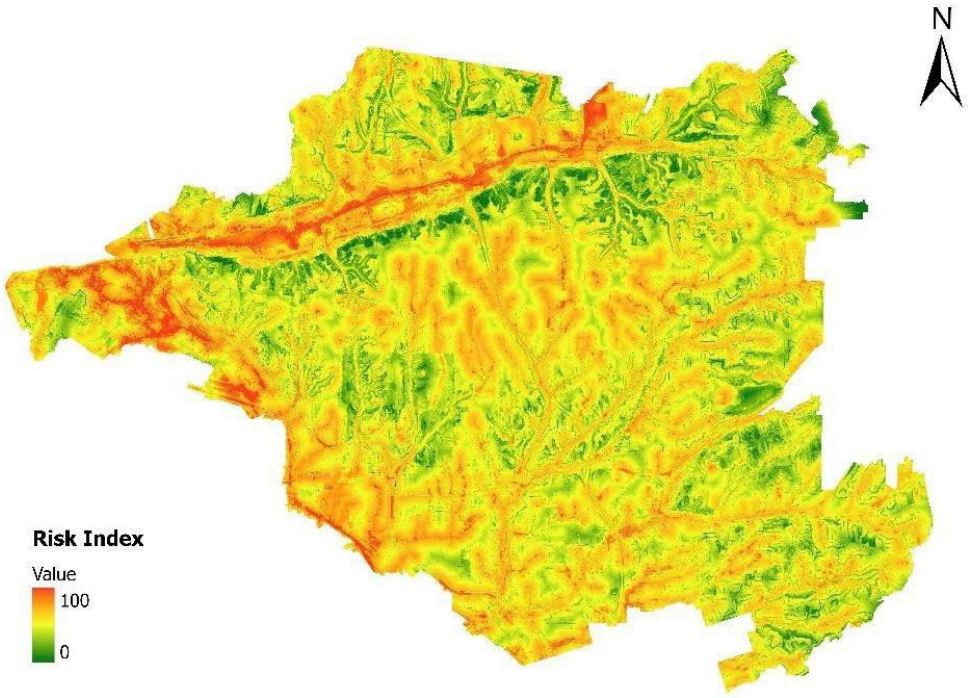

Figure 3. Flood risk map produced by GLR model.

GLR models are validated through a variety of methods, the goal of which are to test the efficacy with which the model can make predictions, and whether these predictions are more valuable than simply choosing random points. A typical receiver operating characteristic curve-area under curve, or ROC-AUC, was not generated for the GLR model because of a lack of true negative

validation data. To properly validate with this technique, a collection of points along transportation infrastructure known *not* to be flooded or to never experience flooding would need to be collected. Therefore, the model was tested using the aforementioned testing points, which were split from the validation dataset. Out of the 10 testing points, 8 showed a flood risk of at least 78, the cut-off between low and negligible flood risk. In addition, the performance of hydrologic variables in the model were evaluated through t- statistics.

The statistics tabulated in Table A.4 illustrate the weights assigned by the GLR model to each variable and their t-statistics, which measure statistical significance. The direction of each coefficient implies that higher values of average imperviousness and presence of bluespots (where 0 = no bluespot and 1 = presence of bluespots) and low drain density, low slope, and small distances lead to higher flood risk index values. These directions all make sense—presence of bluespots and high imperviousness can be expected to lead to a higher risk of flooding, just as proximity to streamflow, few drains, and low slope do. The t-statistics also give a sense of the importance of the GLR model assigned to each variable.

In Table A.4, the hydrologic variables are ranked by importance, in the order of bluespot presence, proximity of flow accumulation lines, average sub-watershed imperviousness, average drain infrastructure density in the sub-watershed, and slope in order of most to least importance. The t-statistics appear robust, showing that each variable used in the model contributed, though some more than others. We can be reasonably confident that the results are meaningful and map onto reality based on the coefficients and t- statistics gathered from the GLR model technique. Again, the value of detecting landscape sinks for urban flood risk analysis is highlighted. Mapping of these landscape sinks with high-resolution hydro-conditioned DEMs could be immensely useful in future studies.

The curve of flood risk index results (Figure A11) was used to classify values. Inspection of the curve reveals that the majority of pixels (~95%) lie on a flood index between 30 and 84. Instead of dropping back to near zero and behaving unimodally, however, the risk index curve rises to another local maxima, giving the data a bimodal distribution. This second maxima was interpreted to represent areas with flood risk. In addition, another group of values was also included to represent areas where flood risk is very low, but could still appear in extreme events or with infrastructure failure. In addition, this classification schema was found to have the best fit with testing data. With this classification schema, almost all areas assigned at least some flood risk are within bluespot extents, which follows the assumption that urban flooding will occur primarily in landscape depressions. Some flood risk also appears near flow accumulation lines, which could be interpreted as flooding occurring near streams in sub-watersheds with poor drain density and high imperviousness.

Extraction of Flooded Area

Explanation of image segmentation and classification parameters used and their justification are explained in the Methods section above. Results gleaned from image classification (Figure A12) show that the SVM model was able to properly classify large water bodies such as Mission Bay

and Lake Murray almost perfectly, with close to the same results under dry and wet conditions. The results can be visually understood through the classification map, with large continuous areas of low radar return, such as Mission Bay, Lake Murray, or stretches of the 5, 8, and 805 freeways easily visible on the map. Results can be better understood through the backscatter distribution of pixels associated with water, shrubland, and urban land use types, as shown in Figure A13.

To quantitatively compare backscatter characteristics, a gamma distribution was employed to compare the radar response of the three land type classes used by the SVM model. This approach allows us to present distributions without taking into account the number of pixels within each class, as the urban and shrubland classes cover a far greater amount of the study area than water and were trained using more pixels than were used for training the model to recognize water. Figure A14 presents the backscatter distribution of pixels used to train the SVM model during the wet condition, as laid out in the methods sections. There is some overlap between distributions, representing the effects of speckle and phenomena such as waves creating zones of brighter radar return in segments classified as water. However, these distributions illustrate clear shifts in spectral backscatter values for each land type class. Presented in Figure A13 are the gamma backscatter distributions comparing the backscatter response of pixels classified as shrubland or urban during dry conditions, yet water during wet conditions. These areas present a backscatter distribution almost identical to that of pixels used to train the model to recognize zones as urban during dry, unflooded conditions, yet with a distribution very close to pixels used to train zones as water during wet, flooded conditions.

Previous studies have found a distinct difference in backscatter distributions associated with various surfaces, and the results presented here reinforce these findings. Even a relatively simple model such as ArcGIS Pro's SVM image classification model was able to distinguish the backscatter distribution of different land cover types, and was able to reflect this same difference in distribution in areas classified as flooded before and after flooding occurred. In general, areas associated with water can be expected to have a short, sharp bimodal distribution concentrated at the lowest backscatter return values, while dry areas will have a distribution shifted towards higher backscatter return. Note that in both backscatter distribution graphs, digital number (DN) values above 750 were discarded.

The SVM model was tested by picking 60 random points in each classified image—30 on surfaces classified as water on the 2-m resolution land use raster, and 30 on surfaces classified as dry land (15 in shrubland-classified surfaces and 15 in urban). Regions populated with points for model testing were distinct from regions used to train the model. Results from these testing points are presented in Table A.5. SVM model parameters were chosen to keep over-detection to a minimum. Large segment sizes mean that small regions classified as water were merged with urban or shrubland segments in the model, best illustrated by the classification of narrower roadway surfaces (effectively, all surface streets) as shrubland rather than water, as flat anthropomorphic surfaces would be classified under ideal conditions. Therefore, the model tends towards under-detection rather than over-detection, with only 12 distinguishable, contiguous

regions, making it through the classification, land use filtering, and change-detection filtering steps of the technique. With this approach, we ensure that our model results are meaningful and that we avoid including a slew of questionable results that would require more time and resources to validate.

With the speckle-reducing effects of image segmentation and the discarding of areas in radar shadow, only a handful of contiguous regions were found that could be classified as flooded with a reasonable margin of error (Table 2). Note that they are all flat, vegetated surfaces that are extensive enough to be "seen" by the radar sensor over the tops of surrounding buildings. While the entire extent of the features listed did not have their classifications change from shrubland or urban to water, a significant amount changed in each, and a pool of standing water large enough to obscure vegetation on a field would be significant enough to cause NF.

Being all in a high- density urban area, every flooded region was found to be close enough to transportation infrastructure for flooding there to pose a possible risk, and most of either the regions themselves or some portion of adjacent transportation networks (83.3%) were classified as having at least some flood risk by the GLR model. Note that flood risk on adjacent transportation networks or flat blacktop surfaces (e.g. parking) were also considered, as this inundation would not be "visible," being either in shadow or not changing in backscatter distribution. Except for one region classified in the land use raster as shrubland, the regions classified as flooded were either marshlands by waterways or grass/dirt fields, either serving as undeveloped parking lots or recreational surfaces. The effect of waves on the backscatter distribution of water surfaces such as lakes or flooded fields is considered to be negligible, as the backscatter distribution of Mission Bay and Lake Murray during wet and dry conditions was found to be similar.

Table 2. List of Regions Classified as Flooded by SAR Imagery

Location	Description	Presence of Nearby Transportation Infrastructure	Flood Risk Presence
Parking lot near Lomita Park	Empty lot, dirt with some vegetation	Yes	Negligible
Morse High school Football field and track	Field, track, shrubland surrounding	Yes	Medium
Southern end of Keiller Neighborhood Park	Grass, some trees ringing park	Yes	Medium to High
Backyard of properties near Broadway and Klauber	Shrubland, dirt	Yes	Medium to High
Lot near Harriet Tubman Village Charter School	Empty lot, dirt with some vegetation	Yes	Negligible
McKinley Elementary Joint Use Park	Park, grass surrounded by blacktop	Yes	Low to Medium
San Diego River channel just south of Snapdragon Stadium	River channel, shrubs and trees	Yes	High
San Diego River channel just south of Hazard Center	River channel, shrubs and trees	Yes	High
Riverwalk Golf Course	Grass	Yes	High

Location	Description	Presence of Nearby Transportation Infrastructure	Flood Risk Presence
San Diego River just northwest of intersection of I-8 & I-5	Marshland, tall grass	Yes	High
Famosa Slough State Marine Conservation Area	Shrubland, marshland, tennis courts	Yes	High
Bill Cleator Community Park	Grass field	Yes	High

Validation of these regions would be immensely helpful in proving the efficacy of the SVM model in classifying flooding in high-density urban areas. However, any validation would need to take into account inter-storm variability. A region flooded during the heavy rainfall of February 27–28, 2017, might not necessarily be flooded during other, less severe storm events, or even storm events of equal or greater magnitude but with distinct spatial rainfall patterns. These points are useful, however, as a starting point for such analysis, and comparing what areas appear as flooded in SAR images during various storm events where spatial variation in precipitation is precisely known could give important insight to the effects of phenomena such as cloud bursts on flooding in the urban environment.

An interesting feature discovered by this analysis not discussed by other studies in the literature review is the notable darkening of rooftop surfaces under the wet condition. An example of this phenomenon and a comparison with similar results from a grass field is illustrated in Figure A15. Though the backscatter distribution on some of these rooftops was shifted significantly to the left, the SVM model did not classify them as water, likely since the segmentation parameters did not allow a small surface surrounded by very bright returns, such as a rooftop, to be classified as water. As the purpose of the study is to study flooding on the surface and along transportation networks, the capacity of the model to reduce speckle and limit false positives was not compromised to detect possible flooding on rooftops. Future research, however, could investigate the possibility of using SAR to detect standing water on rooftops during storms, and analyze which anthropomorphic surfaces see noticeable reductions in backscatter intensity during inundations, even if these inundations are very shallow.

Another phenomenon is the error created by changes in the urban environment, notably the presence of vehicles on transportation infrastructure. Individual cars were visible in the SAR images as small spots of bright radar return, and if grouped together, such as along a freeway or on a parking lot, the classification would change from water (smooth, flat surface) to urban (bright return from a high concentration of vertical surfaces such as walls). If not accounted for, the absence of cars on the same surfaces between the two images would cause a false positive reading in the SVM model. This error was accounted for by, as mentioned in the methods section, filtering out all blacktop surfaces from the final map of flooded areas during the storm of February 27–28.

Discussion

This study consists of three distinct elements brought together to fulfill a common purpose of mapping flood risk in an urban area. We seek to both produce a useful flood risk map that could be used to determine where monitoring stations should be placed and what parts of the city, especially disadvantaged neighborhoods, require better infrastructure, and to further knowledge in the field of remote sensing flood mapping. Each result conveys information, but combined, flood risk and flood vulnerability are mapped as flood hazard. Figure 4 shows a final presentation of the results discussed above, including both flood hazards along the transportation network and flooding detected by SAR.

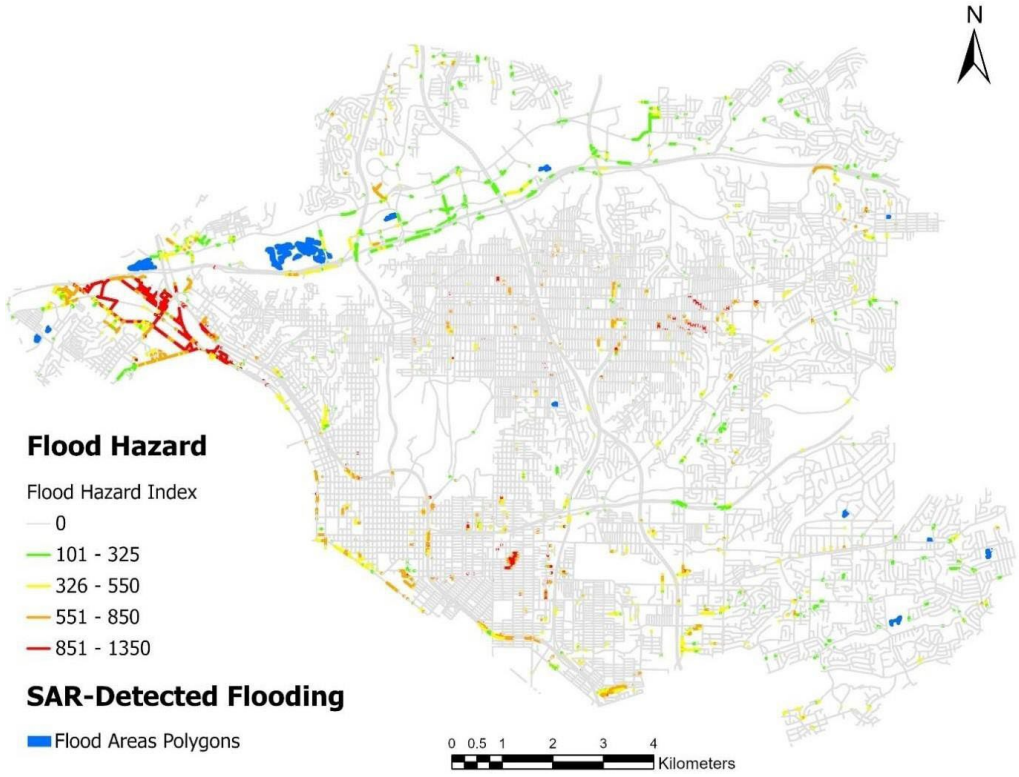

Figure 4. Final map of flood hazard index and areas detected as flooded by SAR imagery.

More insight can be gleaned from the flood hazard map by taking a closer look at specific zones. As expected, most of the study area has a flood hazard index of 0, reflecting the fact that in an urban area of generally topographically high relief, there will not be areas of extensive flooding, save in Mission Valley where the San Diego River occasionally bursts its banks and produces continuous flooding of the kind seen in more humid regions. A more in-depth review of the flood hazard map is shown in Figure A16.

Bluespot mapping proved a powerful tool for the modeling of urban pluvial flooding. Most previous studies of urban flooding were conducted in cities with large rivers running through them, facilitating approaches that mapped flooding associated with floodplains or using contour

models to recognize the borders of floodwaters. Other studies that did model urban pluvial flooding usually did so at very small scales, with techniques that would be time and processing-intensive to replicate over a large area. The bluespot technique allowed us to efficiently map flood risk at high resolutions over a large area. Of all the flood risk variables, bluespots had by far the largest impact. This disproportionate impact can be explained by the fact that in the final flood risk model, bluespots defined the extent of flooding. The other variables, except distance to flowlines, essentially determined which bluespots would "fill up" during a given storm. This follows the assumption that only landscape depressions would be likely to flood during storms, especially in the context of nuisance flooding. The extra flood risk near flowlines could represent, however, streams bursting their banks during extreme rainfall events.

The backscatter distributions of the three land cover classes defined in the study—water, shrubland, and urban—were distinct and generally followed the pattern found in other studies, with distributions for water shifted to the left of those of dry land cover classes. Segmentation was employed to reduce speckle and isolate regions of significant backscatter change in the image. After filtering out flat surfaces such as roadways, parking lots, rooftops, and regions in radar shadow, flooded regions corresponded with grassy fields or low shrubland. Finally, a significant change in backscatter intensity was also discovered on rooftops. The shift in distribution was less in magnitude than that for the areas deemed as flooded, but may still reflect some features and merits further investigation.

As would be expected, one of the main concentrations of flood risk was in Mission Valley, along the floodplains of the San Diego River. During very large flooding events such as a hundred-year storm, catastrophic flooding would be expected in the area, but our flood risk hazard method showed most of the Mission Valley area as having a low flood hazard index. This is due to a lower population density and higher income in this area compared to other neighborhoods in the study area. Another extensive area of continuous high flood risk was the Midway district, just south of the river's mouth. This area is notable for also having a high flood vulnerability, and therefore flood hazard, index. However, the ponded nature of urban flooding means that even discontinuous, geographically small flood risk zones are as important, as the San Diego River has been known to regularly burst its banks and inundate a nearby mall parking lot. These areas, especially when found in lower-income neighborhoods with aged infrastructure, are important to monitor.

The model was trained using flooding reported by police and news agencies. The exact depth and intensity of this flooding is unknown, but it is likely that much of it goes beyond the maximum depth associated with NF. This also applies to the flooding detected by SAR—flooding of a large enough extent and enough depth to be "seen" by the radar also likely goes beyond NF. Validation data specific to NF, however, would be very difficult to gather without extensive field work. Instead, areas of this study that correspond with low or moderate flood risk could be considered areas that could flood at NF levels during storms. The variables that affect flooding reported by news and police agencies would likely be the same variables that cause shallower NF, and the results of this model could also be used to detect and monitor NF. Future field work

could validate this by searching for points where flooding occurs at shallow depths during storms, and comparing these to flood risk model results.

Conclusions and Recommendations

The results gleaned from our models indicate that both the flood risk mapping and SAR classification elements of the study produced robust results. However, one issue was a lack of validation data on both the flood risk modeling and SAR detection sides of the study. Major storms occur infrequently in the San Diego region, and due to the interannual variability of rainfall in the region, years can pass without the major storm events that cause significant flooding in the region. While the validation points were sufficient to build a robust model, future investigation would be helpful to refine and improve the model. Monitoring intersections assigned a high flood hazard level and areas classified as flooded by the SAR image classifier during major storms could be an effective method for further validating both models. Another approach could be gathering data from stakeholders living in areas with high flood risk.

The obstacles of detecting flooding in urban areas, caused by the similarity of the radar return of water and smooth urban surfaces, and radar shadow and layover created by the complex urban topography, remain. One approach could be to detect for flooding in specific regions, such as a park or urban-adjacent vegetated area, known to be a signal for heavy flooding in urban areas nearby. One such example from our study is the Riverwalk Golf Club in the floodplains of the San Diego River, an open, vegetated area where our models detected flooding. Flooding in this golf course, and the extent of flooding, could be a signifier of flooding in urban areas adjacent. However, the nature of ponding in urban flooding suggests that this approach will be unable to catch all urban flooding occurring during a given storm. It is also possible that SAR images taken from an aircraft could be of better use than those from a satellite, but further research would need to be done in the area. Another direction research could be taken in this area is in mapping of flooding during specific storm events and comparing spatial distribution of rainfall to that of flooding. As both data availability and methodology improve, this could be carried out, giving better insight not only to what areas are at higher risk for flooding, but where flooding will occur during various storm events. Eventually, this method could lead to real-time flood extent mapping over large areas with only meteorological data as an input.

Additional Products

The Education and Workforce Development (EWD) and Technology Transfer (T2) products created as part of this project can be downloaded from the Safe-D website here.

Education and Workforce Development Products

This project provided support for a full-time graduate student in the Department of Civil, Construction and Environmental Engineering at SDSU toward his MS Degree. The graduate student, Vincent O'Hara-Rhi, has assisted with all project tasks outlined above, and successfully defended his MS thesis on January 12, 2022. This project was the one that generated data for Vincent's MS thesis. Throughout this project, Vincent has been provided with the opportunity to become familiar with the fundamental concepts around flooding impacts on transportation systems, flood control strategies, and Big Data Analytics. In addition, this project has provided Vincent with mentorship to excel his scientific communication skills. As a result, Vincent presented the findings of this study at the SDSU Student Research Symposium (SRS) on March 4, 2022, and has received the following feedback from the conference judge, as an example: "*Very interesting topic and good understanding of the issues with flooding.*" In addition, this project engaged an undergraduate student, Callum McRae, in the Department of Civil, Construction and Environmental Engineering at SDSU with GIS analyses of the flood-prone areas identified through this research. Callum has now been admitted to SDSU's Civil Engineering MS program, and is willing to continue his research in the PI's lab.

Furthermore, the PI has designed educational materials around flood control and remote sensing concepts for his graduate level course, "*CIVE 730, Advanced Topics in Water Engineering*" offered in Spring 2022. This is a graduate course at SDSU's Civil Engineering MS program. Specifically, the materials included instructional videos on advanced flood modeling and management using a numerical flood modeling software, followed by an assignment for the students to design a flood control measure for the SDSU campus to alleviate the flooding of the nearby transportation network; i.e. the culvert beneath the Freeway I-8. In addition to the above-mentioned videos, the PI has summarized the research methods and findings into a lecture for his CIVE 730 class introducing satellite imagery through Synthetic Aperture Radar (SAR) and its application in flood monitoring of the transportation network in San Diego.

Lastly, throughout the course of the project, the PI has repeatedly attempted to present the project's concepts at the SDSU STEM Exploration Day, an annual event conducted by the SDSU Pre- College Institute (PCI), which brings more than 300 K-12 students from disadvantaged communities around San Diego to inspire them with the state-of-the-art science and technology. However, per PCI's director, Ms. Nadia Rohlinger "*given the pandemic and concerns about student participation, our team decided to postpone STEM Exploration Day to Friday, November 4, 2022.*" Therefore, this workshop will take place on the above-mentioned scheduled date.

Technology Transfer Products

This project resulted into two different manuscripts for peer reviewed journals, one of which has been already published and the other which is still under-review:

- This manuscript summarizes Vincent's research and is currently under review by the *Journal of Water Resources Management*: **O'Hara-Rhi V., & Tavakol-Davani H. (*Under Review*) Evaluation of Urban Flood Hazard based on Remote Sensing and Machine Learning. Water Resources Management.**

- This manuscript summarizes Callum's research in collaboration with the iWERS Laboratory at the University of South Carolina, and has been published in the *Water Journal*: **Tanim, A. H., McRae, C. B., Tavakol-Davani, H., & Goharian, E. (2022). Flood Detection in Urban Areas Using Satellite Imagery and Machine Learning. Water, 14(7). https://doi.org/10.3390/w14071140**

In addition, with the assistance provided by the project champion Ms. Virginia Lingham, PE, who is a Senior Consultant working in WSP's Transportation Operations Strategies, this project was able to engage multiple stakeholders across the US through a webinar for transferring knowledge on our findings for informed future design of transportation network considering the flooding risks. This webinar took place on Feb 18, 2022, with 31 participants. The webinar flyer was distributed through multiple venues, including the Committee on Transportation System Operations (CTSO), National Operations Center of Excellence, and the TRB committee on Logistics of Disaster Response, Business Continuity, and Humanitarian Relief. Below is a link to the recording of this webinar, and the list of webinar attendees is presented in Table 3:

https://sdsu.zoom.us/rec/share/YnF6fZrtYZ-7b7kg5cWJpFadtg93JAzHi62xweLqgflEGonhkTCV-4X4jGVZxdZT.eGXAgzLKGnmVeE29?startTime=1645203423000

Table 3. List of Attendees to the Project Webinar on Feb 18, 2022

Name	Affiliation
Abdullah Al Mehedi	Villanova University
Alfred Gross	Colorado Department of Transportation
Amy Bailey	Tennessee Department of Transportation
Brad Darr	North Dakota Department of Transportation
Charles Tapp	Texas Department of Transportation
Christopher Poe	Director, Mixon Hill
Curtis Walker	National Center For Atmospheric Research
David Heller	South Jersey Transportation Planning Organization
Dennis Murray	Georgia Department of Transportation
Dominique Shannon	Kansas Department of Transportation
Jeff Williams	Utah Department of Transportation

Jim Cook	Ohio Department of Transportation
John Roberts	Arizona Department of Transportation
Josh Foster	Michigan Department of Transportation
Judith Peter	New York State Department of Transportation
Kate Homet	Villanova University
Katherine Jennings	Wisconsin Department of Transportation
Kayla Grubb	Volkert, Inc.
Khuzaima Mahdi	Tennessee Department of Transportation
Lee Mixon	CEO, Mixon Hill
Maria Rhi	Independent health practitioner
Matthew Miller	Texas A&M University
Peleg Kremer	Villanova University
Rob Marz	Michigan Department of Transportation
Sarah Gill	Michigan Department of Transportation
Thien Tran	Colorado Department of Transportation
Tony Coventry	Federal Highway Administration
Virginia Lingham	WSP USA
Virginia Smith	Villanova University
Xiaoxiao Zhang	University of Virginia

Data Products

The data products of this project are the maps of flood risk and hazard index presented in Figure 3 and Figure 4 above. This project did not collect any data through field measurements or social surveys.

References

1. Moftakhari, H.R., et al., *What Is Nuisance Flooding? Defining and Monitoring an Emerging Challenge.* Water Resources Research, 2018. **54**(7): p. 4218-4227.
2. Deitch, M.J., M.J. Sapundjieff, and S.T. Feirer, *Characterizing Precipitation Variability and Trends in the World's Mediterranean-Climate Areas.* Water, 2017. **9**(4).
3. Berg, N. and A. Hall, *Increased Interannual Precipitation Extremes over California under Climate Change.* Journal of Climate, 2015. **28**(16): p. 6324-6334.
4. Sivapalan, M., et al., *Linking flood frequency to long-term water balance: Incorporating effects of seasonality.* Water Resources Research, 2005. **41**(6).
5. Moftakhari, H.R., et al., *Increased nuisance flooding along the coasts of the United States due to sea level rise: Past and future.* Geophysical Research Letters, 2015. **42**(22): p. 9846-9852.
6. de Almeida, G.A.M., P. Bates, and H. Ozdemir, *Modelling urban floods at submetre resolution: challenges or opportunities for flood risk management?* Journal of Flood Risk Management, 2018. **11**(S2): p. S855-S865.
7. Henonin, J., et al., Real-time urban flood forecasting and modelling - a state of the art. *Journal of Hydroinformatics*, Vol. 15, No. 3, 2013. p. 717-736.
8. Hosseiny, H., et al., A Framework for Modeling Flood Depth Using a Hybrid of Hydraulics and Machine Learning. *Scientific Reports*, 2020. **10**(1): p. 8222.
9. Schubert, J.E. and B.F. Sanders, *Building treatments for urban flood inundation models and implications for predictive skill and modeling efficiency.* Advances in Water Resources, 2012. **41**: p. 49-64.
10. Darabi, H., et al., *Urban flood risk mapping using data-driven geospatial techniques for a flood-prone case area in Iran.* Hydrology Research, 2019. **51**(1): p. 127-142.
11. Darabi, H., et al., *Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques.* Journal of Hydrology, 2019. **569**: p. 142-154.
12. Yin, J., et al., *A review of advances in urban flood risk analysis over China.* Stochastic Environmental Research and Risk Assessment, 2015. **29**(3): p. 1063-1070.
13. Alsdorf, D.E., E. Rodríguez, and D.P. Lettenmaier, *Measuring surface water from space.* Reviews of Geophysics, 2007. **45**(2).
14. Tong, X., et al., *An approach for flood monitoring by the combined use of Landsat 8 optical imagery and COSMO-SkyMed radar imagery.* ISPRS Journal of Photogrammetry and Remote Sensing, 2018. **136**: p. 144-153.
15. Tanguy, M., et al., *River flood mapping in urban areas combining Radarsat-2 data and flood return period data.* Remote Sensing of Environment, 2017. **198**: p. 442-459.
16. Giustarini, L., et al., *A Change Detection Approach to Flood Mapping in Urban Areas Using TerraSAR-X.* IEEE Transactions on Geoscience and Remote Sensing, 2013. **51**(4): p.

2417-2430.

17. Moreira, A., et al., *A tutorial on synthetic aperture radar.* IEEE Geoscience and Remote Sensing Magazine, 2013. **1**(1): p. 6-43.
18. Addabbo, A.D., et al., *A Bayesian Network for Flood Detection Combining SAR Imagery and Ancillary Data.* IEEE Transactions on Geoscience and Remote Sensing, 2016. **54**(6): p. 3612-3625.
19. Qiu, J., et al., *Flood Monitoring in Rural Areas of the Pearl River Basin (China) Using Sentinel-1 SAR.* Remote Sensing, 2021. **13**(7).
20. Matgen, P., et al., *Towards an automated SAR-based flood monitoring system: Lessons learned from two case studies.* Physics and Chemistry of the Earth, Parts A/B/C, 2011. **36**(7): p. 241-252.
21. Mason, D.C., et al., *Flood Detection in Urban Areas Using TerraSAR-X.* IEEE Transactions on Geoscience and Remote Sensing, 2010. **48**(2): p. 882-894.
22. Mason, D.C., et al., *Detection of flooded urban areas in high resolution Synthetic Aperture Radar images using double scattering.* International Journal of Applied Earth Observation and Geoinformation, 2014. **28**: p. 150-159.
23. Pulvirenti, L., et al., *Use of SAR Data for Detecting Floodwater in Urban and Agricultural Areas: The Role of the Interferometric Coherence.* IEEE Transactions on Geoscience and Remote Sensing, 2016. **54**(3): p. 1532-1544.
24. Li, Y., et al., *Urban Flood Mapping Using SAR Intensity and Interferometric Coherence via Bayesian Network Fusion.* Remote Sensing, 2019. **11**(19).
25. Lin, Y.N., et al., *Urban Flood Detection with Sentinel-1 Multi-Temporal Synthetic Aperture Radar (SAR) Observations in a Bayesian Framework: A Case Study for Hurricane Matthew.* Remote Sensing, 2019. **11**(15).
26. Pierdicca, N., et al. *Flood mapping by SAR: Possible approaches to mitigate errors due to ambiguous radar signatures*. in *2014 IEEE Geoscience and Remote Sensing Symposium*. 2014.
27. Ferro, A., et al., *On the Relationship Between Double Bounce and the Orientation of Buildings in VHR SAR Images.* IEEE Geoscience and Remote Sensing Letters, 2011. **8**(4): p. 612-616.
28. Watanabe, M., et al. *GB-SAR/PiSAR simultaneous experiment for a trial of flood area detection*. in *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*. 2008.
29. Mason, D.C., et al., *Improving Urban Flood Mapping by Merging Synthetic Aperture Radar-Derived Flood Footprints with Flood Hazard Maps.* Water, 2021. **13**(11).
30. Mason, D.C., et al., *Near Real-Time Flood Detection in Urban and Rural Areas Using High-Resolution Synthetic Aperture Radar Images.* IEEE Transactions on Geoscience and Remote Sensing, 2012. **50**(8): p. 3041-3052.
31. *Climate- San Diego* January 5, 2022]; Available from: https://www.weather.gov/wrh/Climate?wfo=sgx.

32. Pavía, E.G. and A. Badan, ENSO modulates rainfall in the Mediterranean Californias. Geophysical Research Letters, 1998. 25(20): p. 3855-3858.

33. *DAC Mapping Tool*. Tool Reference [cited 1/5/2022; Available from: https://gis.water.ca.gov/app/dacs/.

34. Schnebele, E., et al., *Real Time Estimation of the Calgary Floods Using Limited Remote Sensing Data.* Water, 2014. **6**(2).

35. Rosser, J.F., D. Leibovici, and M. Jackson, *Rapid flood inundation mapping using social media, remote sensing and topographic data.* Natural Hazards, 2017. **87**(1): p. 103-120.

36. Yerramilli, S., *A hybrid approach of integrating HEC-RAS and GIS towards the identification and assessment of flood risk vulnerability in the city of Jackson, MS.* American Journal of Geographic Information System, 2012. **1**(1): p. 7-16.

37. Adams, A., *Heavy Rain Causes Road Closures Across San Diego* 2016, 7 News San Diego p. https://www.nbcsandiego.com/news/local/flooding-causes-road-closures-across- san-diego/87591/.

38. *Major Flooding Causes Road Closures Around San Diego* 2016, KUSI San Diego. p. https://www.kusi.com/big-storm-in-store-for-county-tuesday/.

39. Jennewein, C., *Record Rain is Over, but Flooding Remains a Major Problem* 2017, Times of San Diego

40. Winkley, L., *Here Are All the Streets that are Flooded*. 2017, The San Diego Union Tribune. p. https://www.sandiegouniontribune.com/news/public-safety/sd-me-flooded-streets-20170227-story.html.

41. Saunders, M., *How San Diego's Flooding COmpares to FEMA's Historical Hazard Map*. 2018, 10 News San Diego p. https://www.10news.com/news/how-san-diegos-flooding-compares-to-historical-hazards.

42. Robbins, G., *Flash Flood Warning for San Diego Area Expires, but many Roadways Remain Inundated*. 2018, The San Diego Union Tribune.

43. Jennewein, C., *Updaterd List of Roads Closed by Flooding Across San Diego at 5 p.m.* 2019, Times of San Diego p. https://timesofsandiego.com/crime/2019/02/14/updated-list- of-roads-closed-by-flooding-across-san-diego-at-5-p-m/.

44. Harman, J., *Review of 2007 Summer Floods*, E. Agency, Editor. 2007. p. 4-5.

45. Ebert, H.D.M.a.F.C., *Southern California Floods of January, 1916*, U.S.G.S. Department of the Interior, Editor. 1916, Water-Supply Paper.

46. *Heavy Precipitation Events California and Northern Nevada January and February 2017*. [Webpage] 2017 [cited 2021 10/12/2021]; Available from: https://www.cnrfc.noaa.gov/storm_summaries/janfeb2017storms.php.

47. Eini, M., et al., *Hazard and vulnerability in urban flood risk mapping: Machine learning techniques and considering the role of urban districts.* International Journal of Disaster Risk Reduction, 2020. **50**: p. 101687.

48. Sowmya, K., C. John, and N. Shrivasthava, *Urban flood vulnerability zoning of Cochin City, southwest coast of India, using remote sensing and GIS.* Natural Hazards, 2015.

75(2): p. 1271-1286.

49. *Digital Elevation Models, Hydro-Flattening, and Hydro-Enforcement* January 31, 2017 [cited 1/12/2022; Available from: https://www.usgs.gov/media/videos/digital-elevation-models-hydro-flattening-and-hydro-enforcement#:~:text=Breaklines%20are%20used%20to%20create,entire%20water%20surface%20is%20constrained.
50. Balstrøm, T. and D. Crawford, *Arc-Malstrøm: A 1D hydrologic screening method for stormwater assessments based on geometric networks.* Computers & Geosciences, 2018. **116**: p. 64-73.
51. Vandenberg-Rodes, A., et al., *Projecting nuisance flooding in a warming climate using generalized linear models and Gaussian processes.* Journal of Geophysical Research: Oceans, 2016. **121**(11): p. 8008-8020.
52. Hosseini, F.S., et al., *Flash-flood hazard assessment using ensembles and Bayesian- based machine learning models: Application of the simulated annealing feature selection method.* Science of The Total Environment, 2020. **711**: p. 135161.
53. Zhao, Y., *Chapter 5 - Regression*, in *R and Data Mining*, Y. Zhao, Editor. 2013, Academic Press. p. 41-50.
54. Slavkovic, A. *STAT 504: Analysis of Discrete Data*. [Online Class] [cited 2021 December]; Available from: https://online.stat.psu.edu/statprogram/stat504.
55. Kessler, R., *Stormwater Strategies: Cities Prepare Aging Infrastructure for Climate Change.* Environmental Health Perspectives, 2011. **119**(12): p. a514-a519.
56. *Streets Overall Condition Index (OCI)*. 1/5/2022]; Available from: https://data.sandiego.gov/datasets/streets-overall-condition-index/.
57. Fukuda, S. and H. Hirosawa. *Support vector machine classification of land cover: application to polarimetric SAR data*. in *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*. 2001.
58. Lardeux, C., et al., *Support Vector Machine for Multifrequency SAR Polarimetric Data Classification.* IEEE Transactions on Geoscience and Remote Sensing, 2009. **47**(12): p. 4143-4152.
59. Hagg, W. and M. Sties. *Efficient speckle filtering of SAR images*. in *Proceedings of IGARSS '94 - 1994 IEEE International Geoscience and Remote Sensing Symposium*. 1994.
60. Comaniciu, D. and P. Meer, *Mean shift: a robust approach toward feature space analysis.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002. **24**(5): p. 603-619.
61. *Classify*. Tool Reference [1/6/2022]; Available from: https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/classify.htm.

Appendix. Supplemental Visualizations

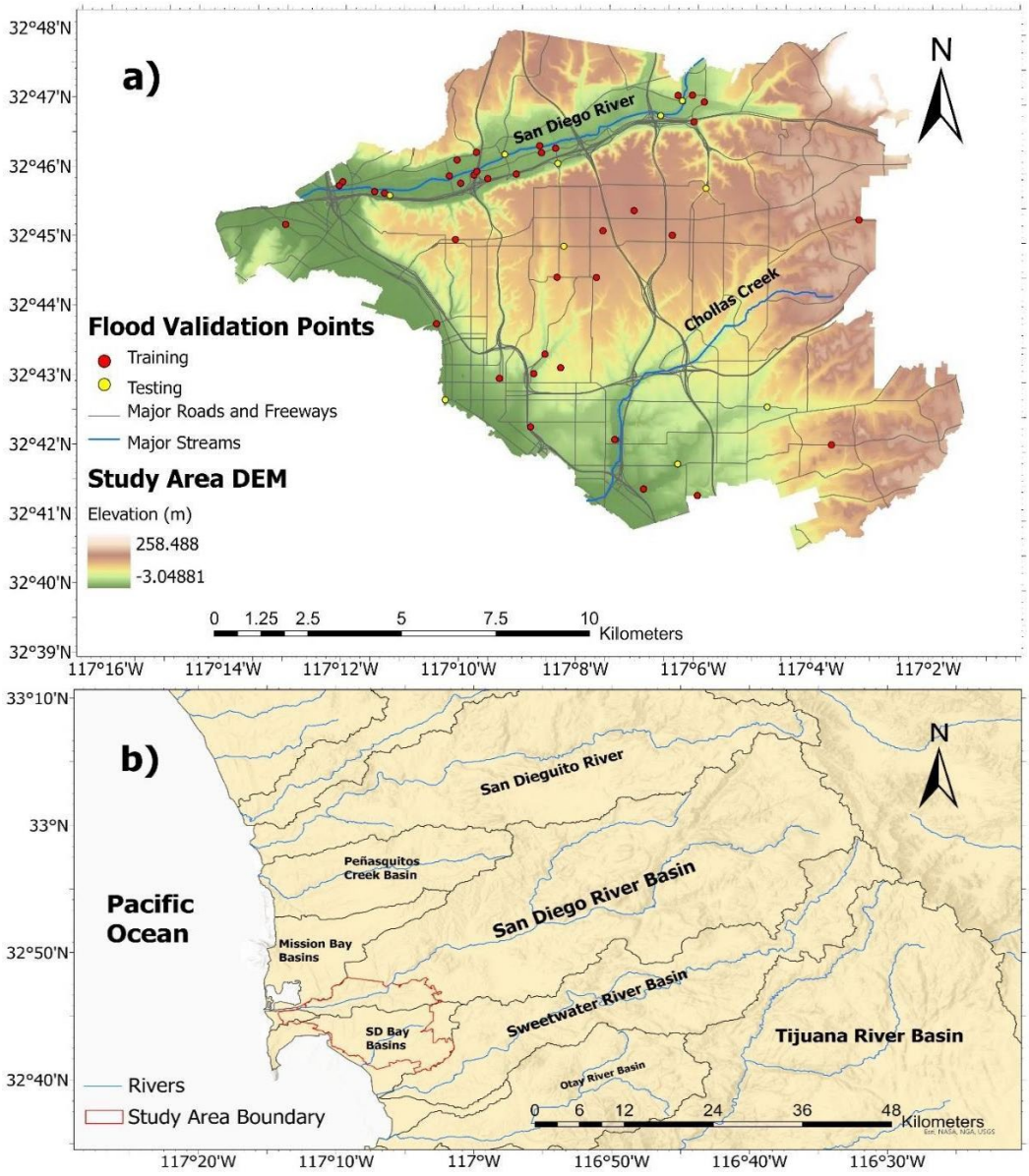

Figure A1. Location and geographic context of study area: a) shows study area DEM, network of major roads, major streams, and validation points; b) shows hydrologic and geographic context of region around the study area.

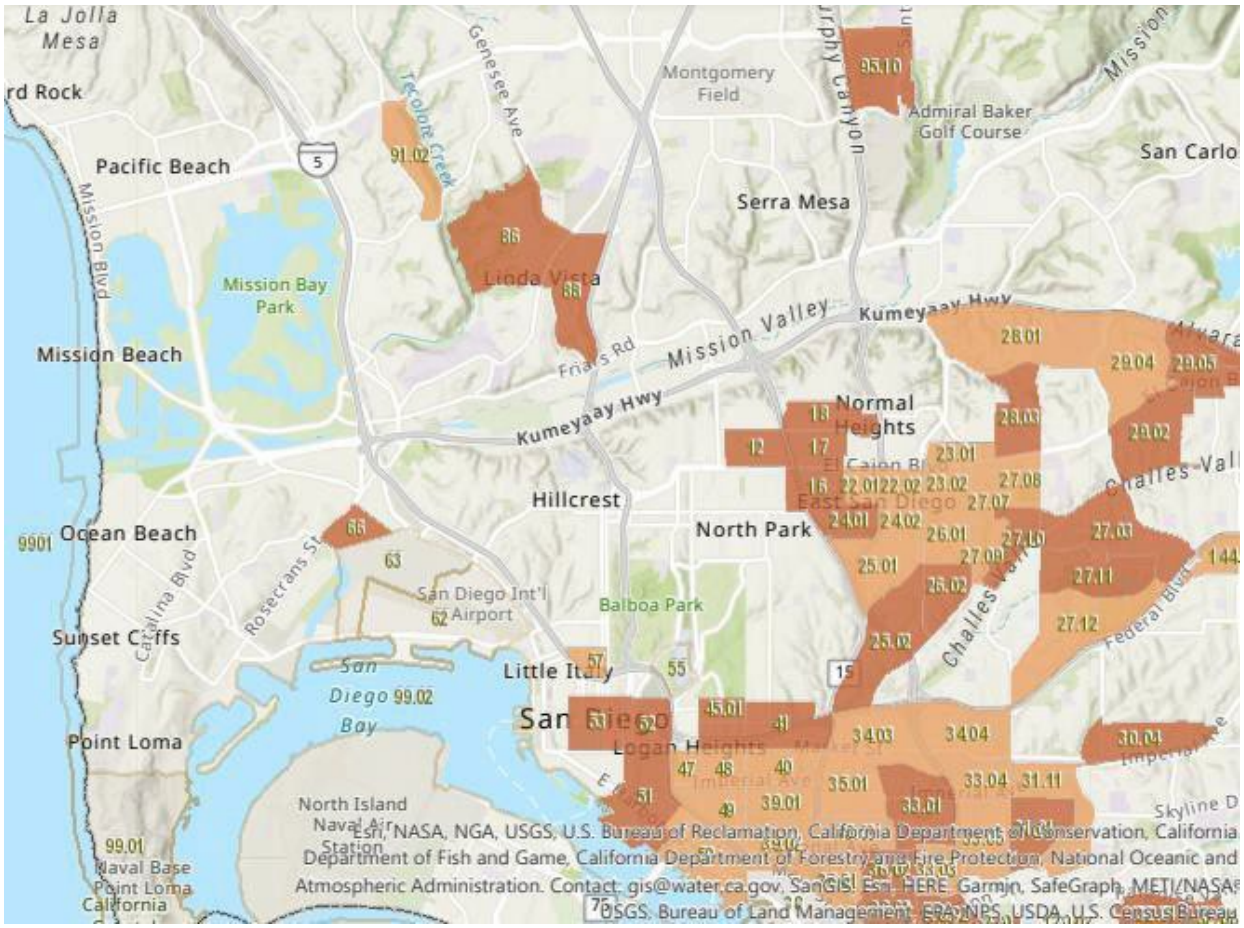

Figure A2. Disadvantaged communities by census tract in the study area as designated by CalEnviroScreen

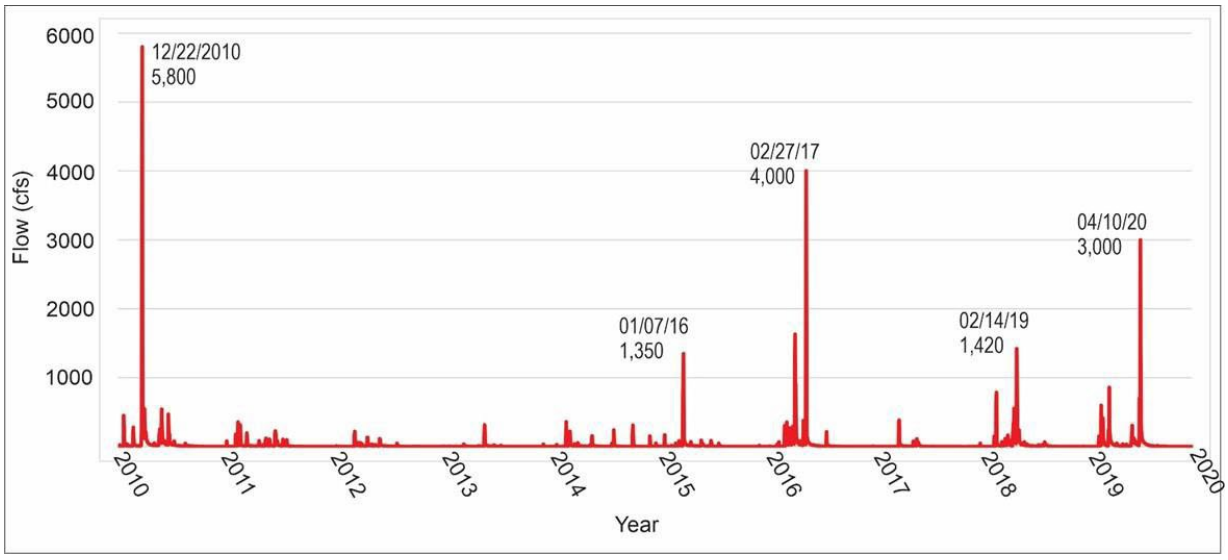

Figure A3. Streamflow in cubic feet per second in the San Diego River at Fashion Valley.

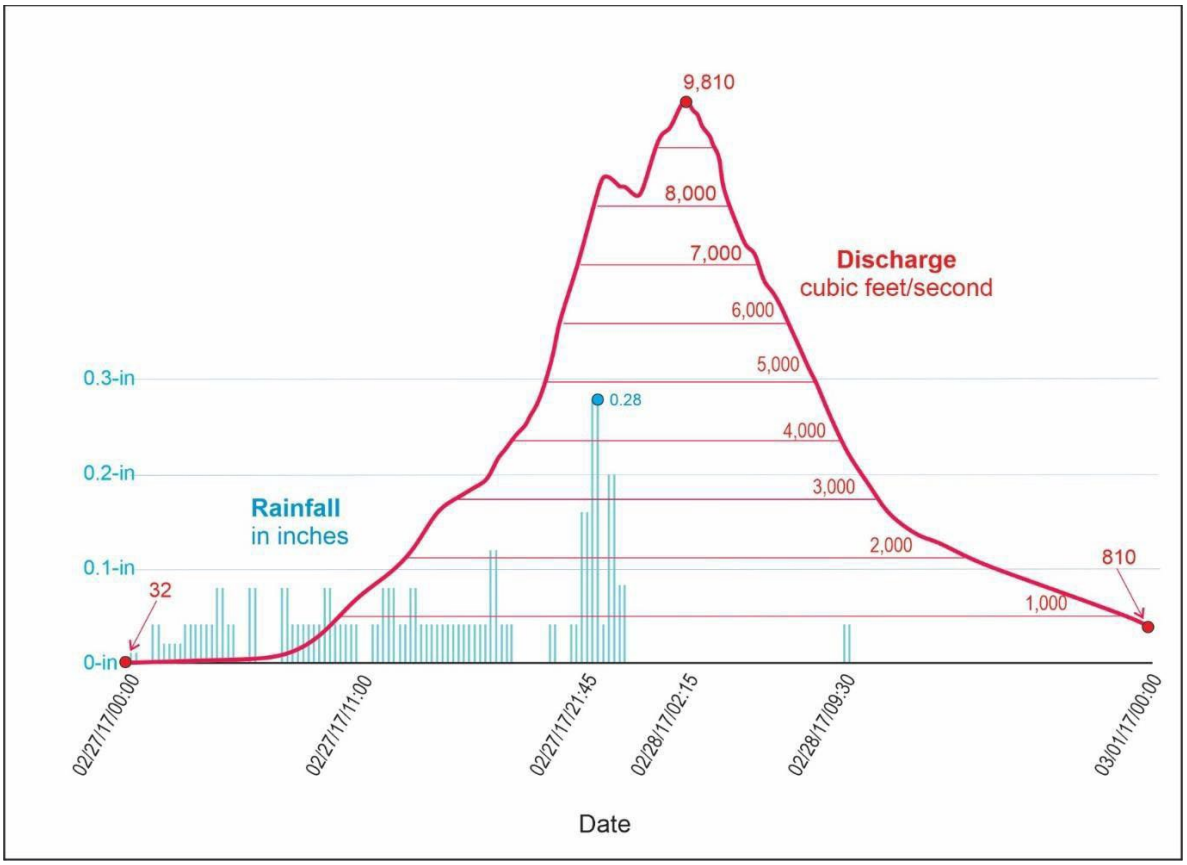

Figure A4. Average 15-minute streamflow in the San Diego river in cubic feet per second at Fashion Valley and rainfall at Fashion Valley, from 0:00 on February 27th, 2017 to 0:00 on March 1st, 2017.

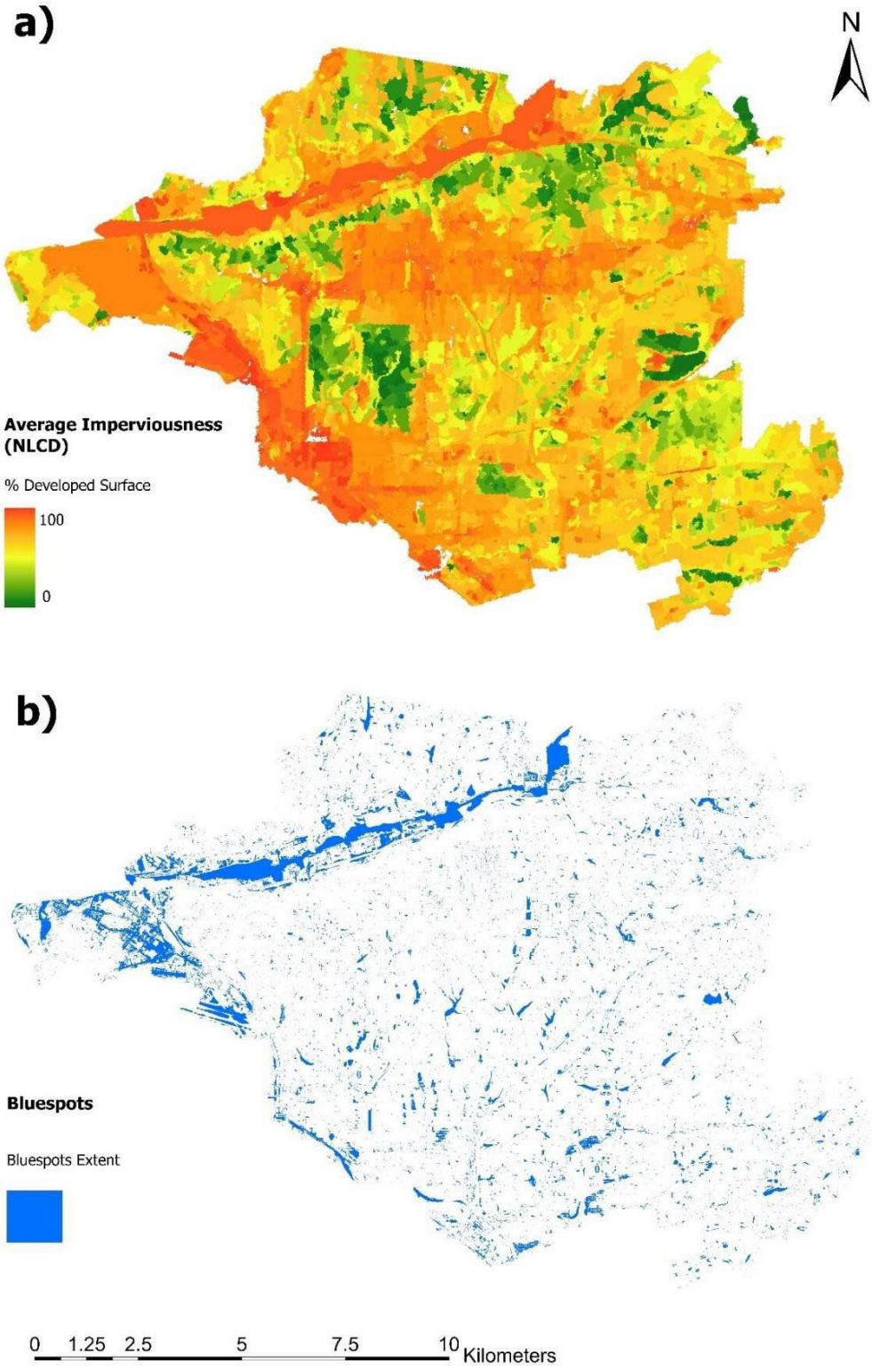

Figure A5. Imperviousness and bluespots: a) shows average imperviousness per sub-watershed; b) shows the extent of bluespots, or landscape depressions identified from a hydro-conditioned DEM.

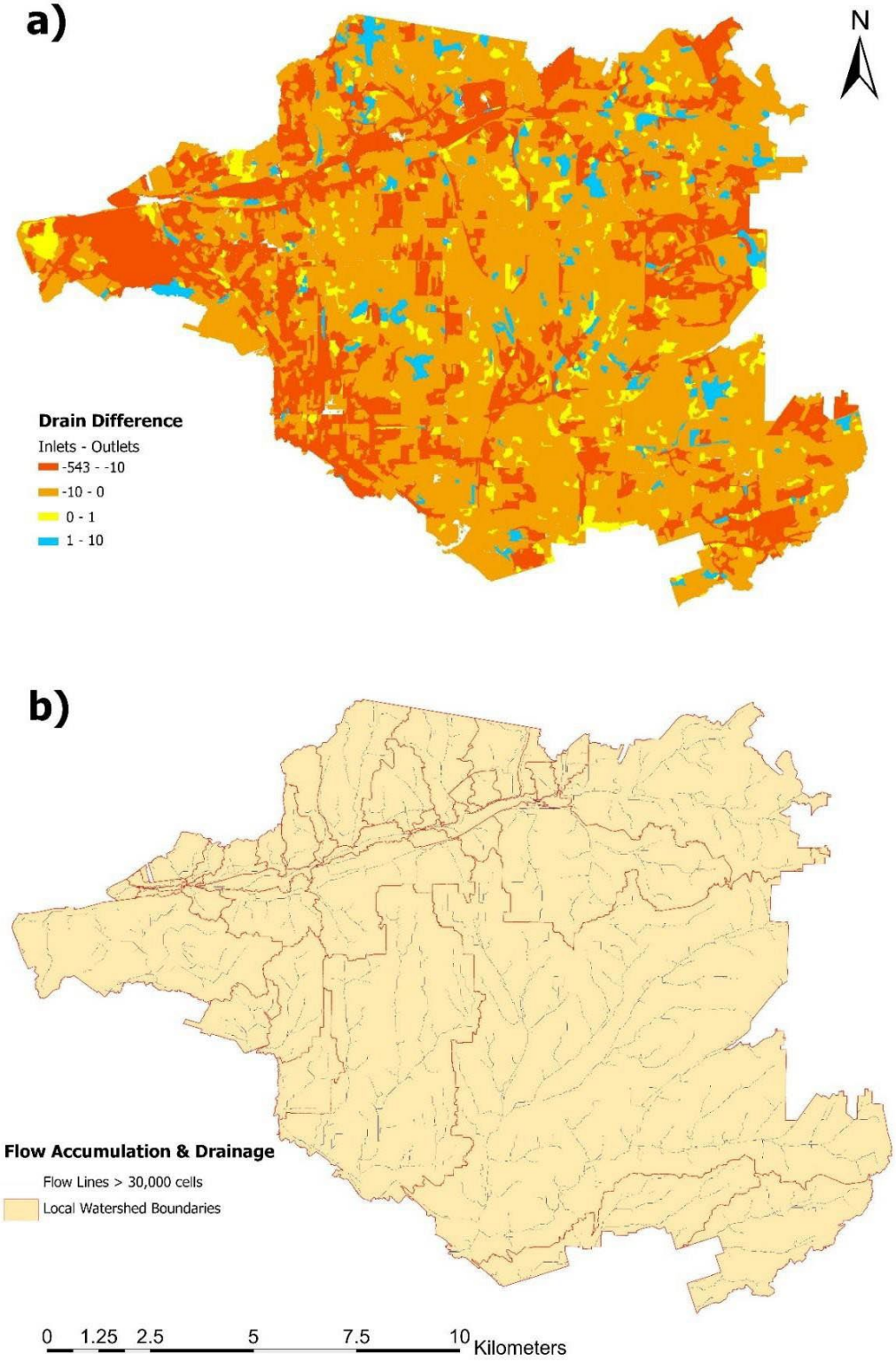

Figure A6. Drain density and flow lines/watershed boundaries: a) shows density of drains per sub-watershed; b) shows flow lines and local watershed boundaries. Note northern watersheds flow into the San Diego River, while those in the south flow directly into San Diego Bay.

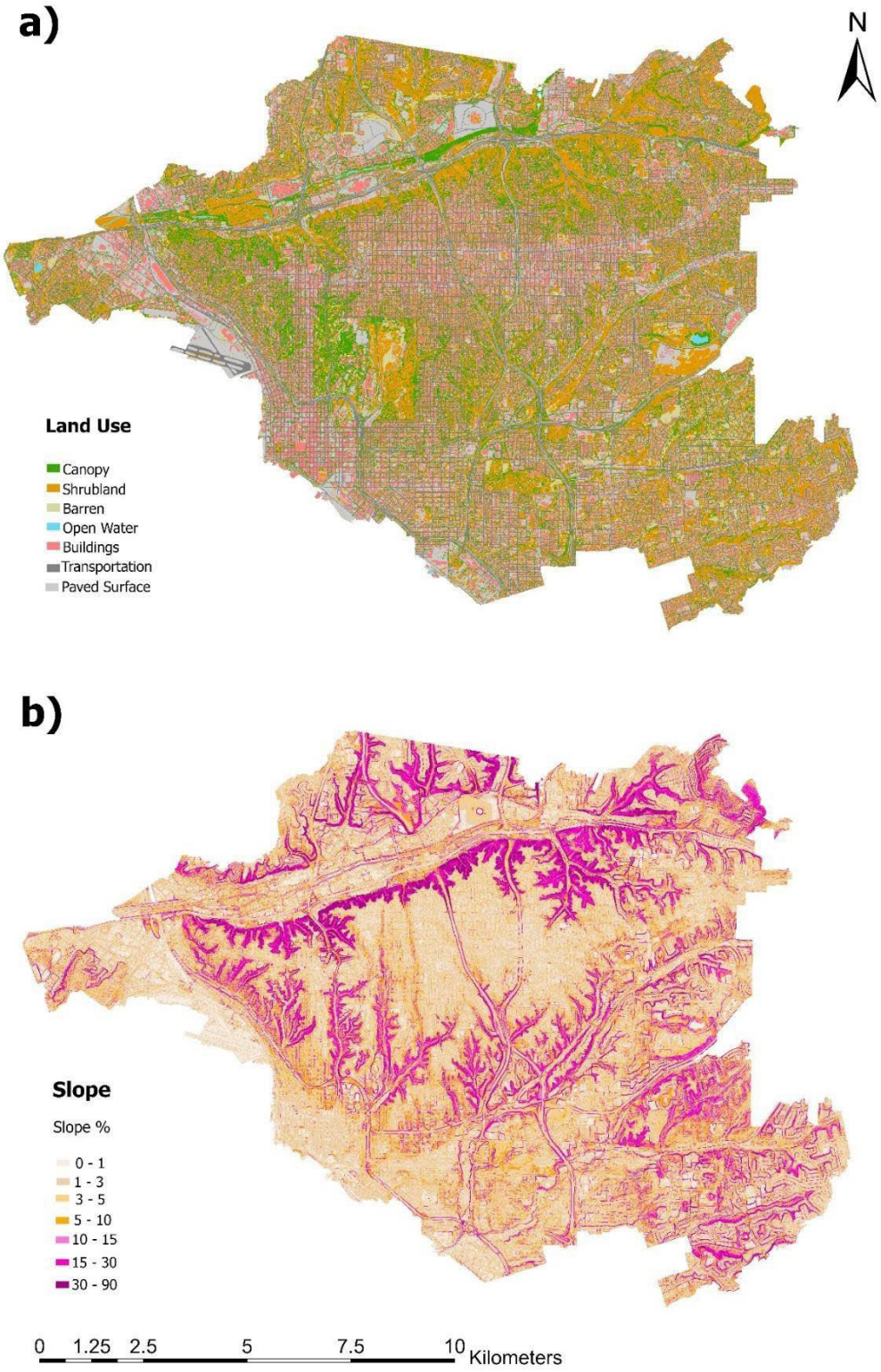

Figure A7. Land use and slope: a) shows land use in the study area; b) shows slope, found from a hydro-conditioned DEM.

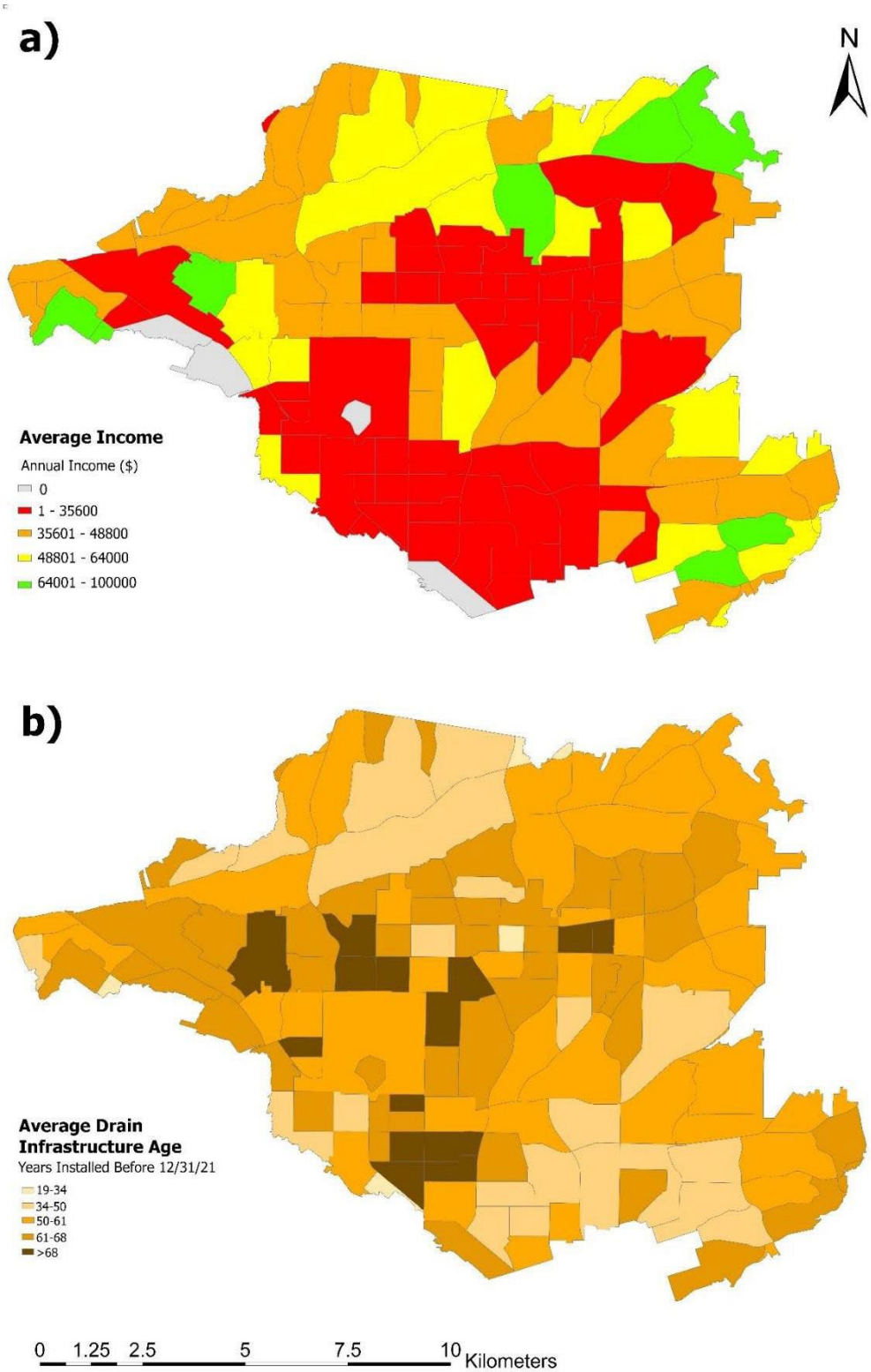

Figure A8. Income and drain infrastructure: a) shows average income; b) shows average drain infrastructure age, both by census tract.

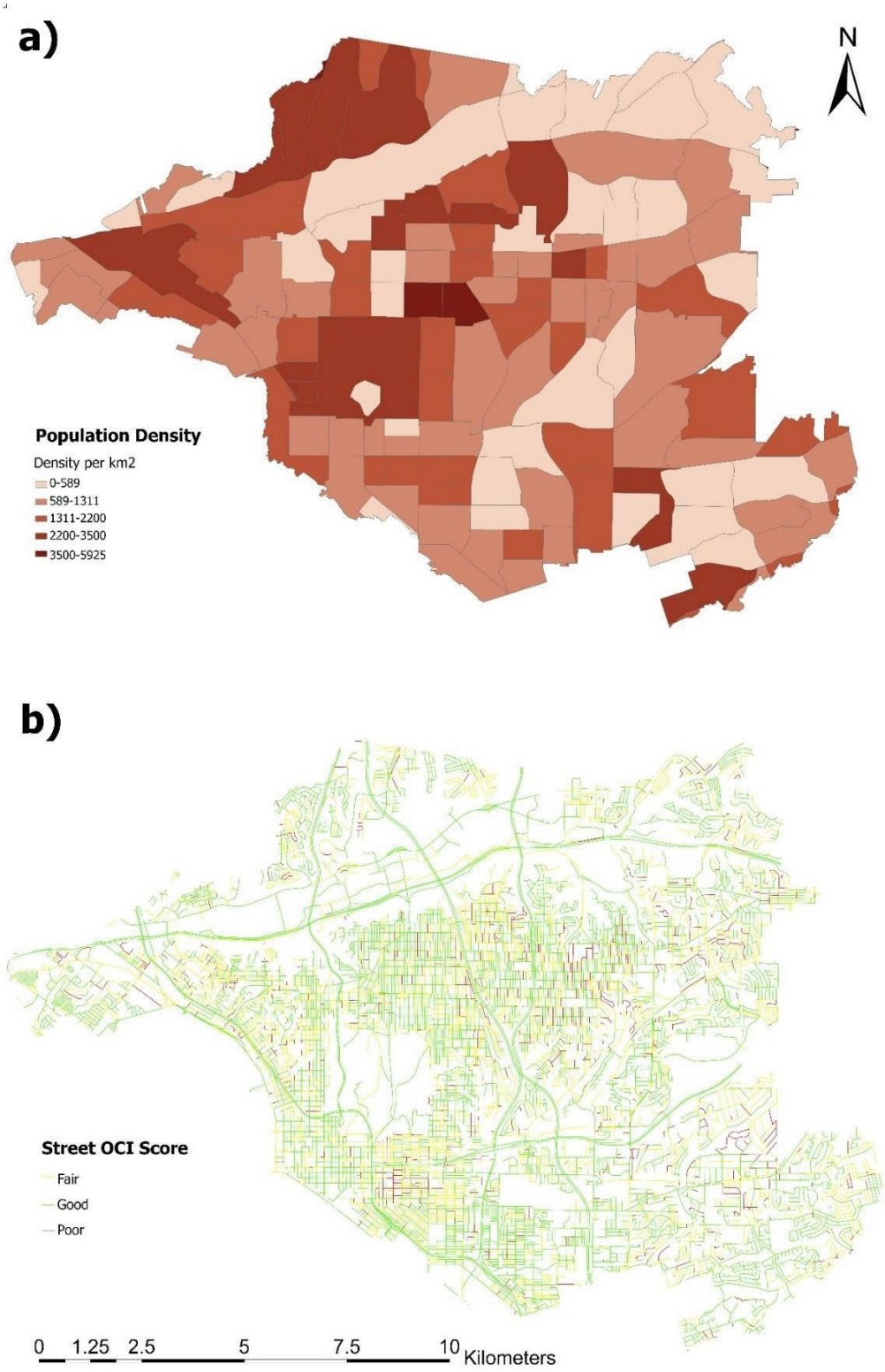

Figure A9. Density and OCI score: a) shows population density by census tract; b) shows street OCI score for all streets that were sampled.

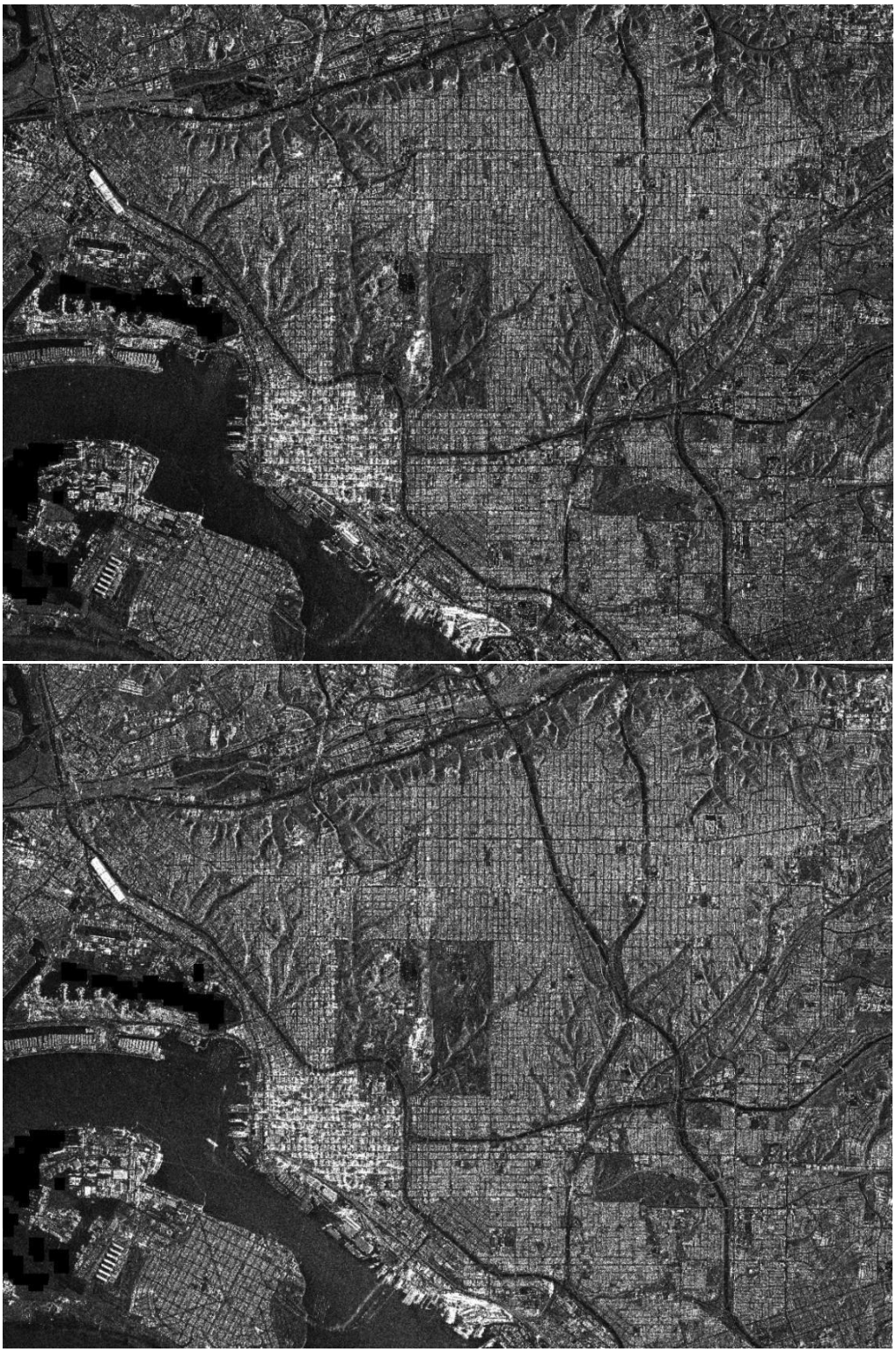

Figure A10. Images taken by COSMO-SkyMed satellite of the San Diego area on February 16, 2017 (top: dry condition) and February 27, 2017 (bottom: wet condition).

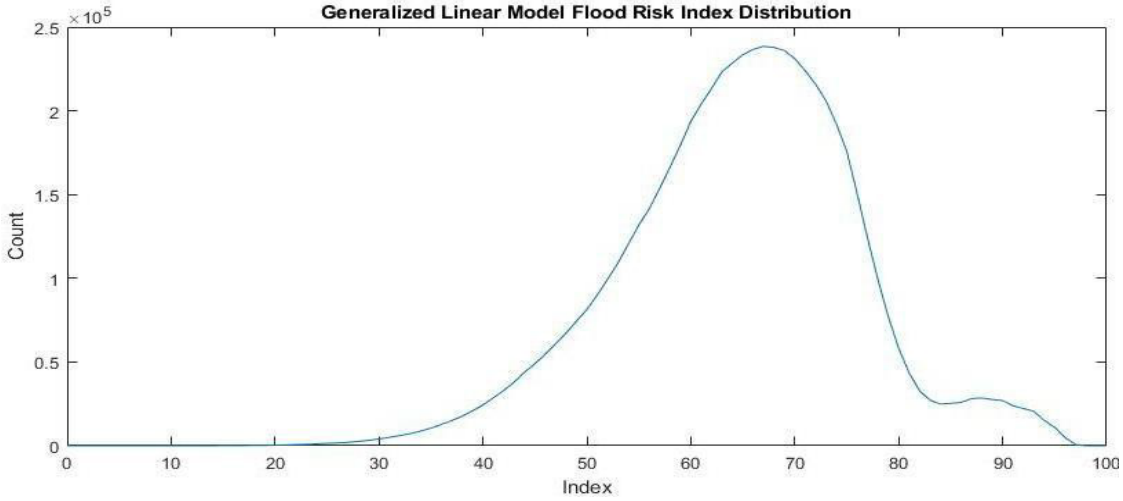

Figure A11. Distribution of pixels over normalized flood index values, with 0 being lowest flood risk and 100 highest. Note the bimodality of the distribution, and the concentration of values having a flood index from about 82 to 95.

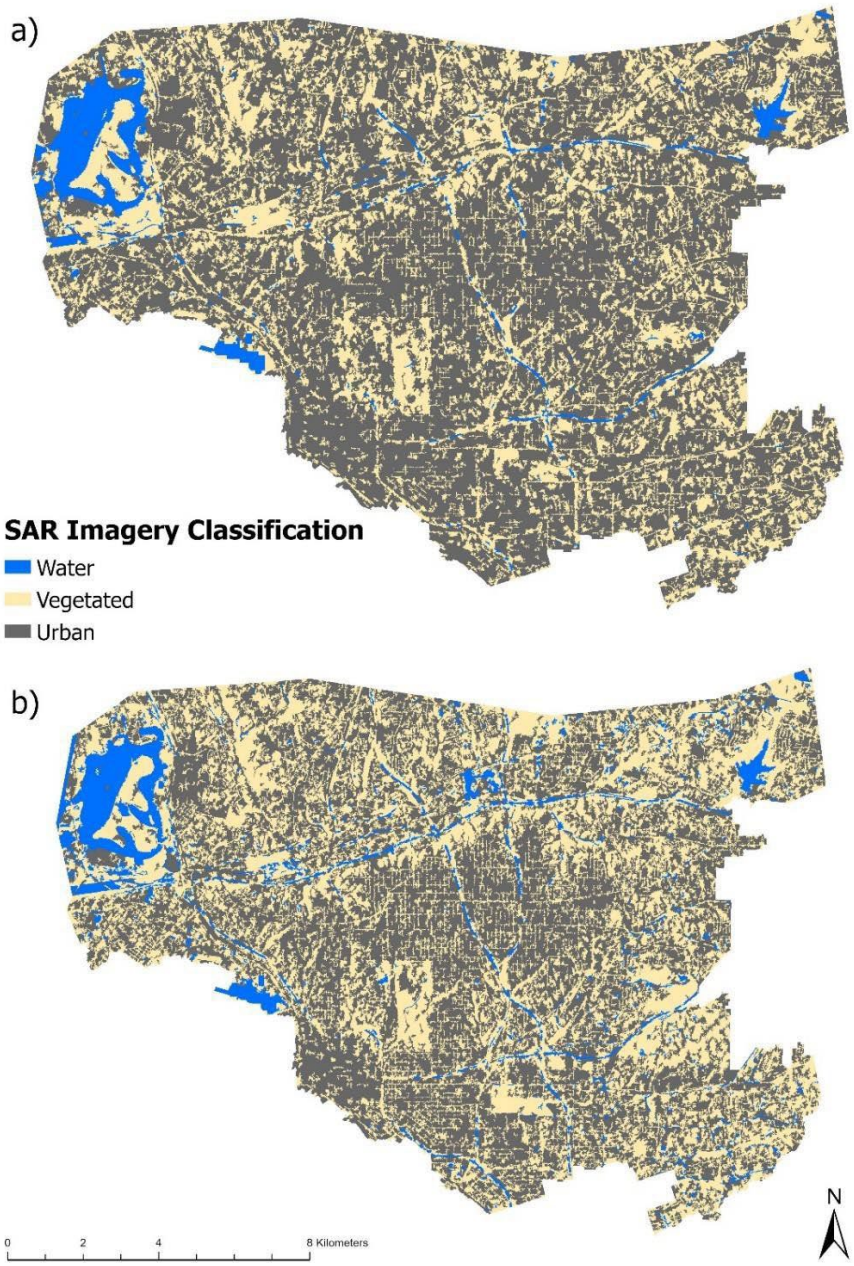

Figure A12. SAR Imagery classification for a) dry-weather condition and b) for wet-weather condition.

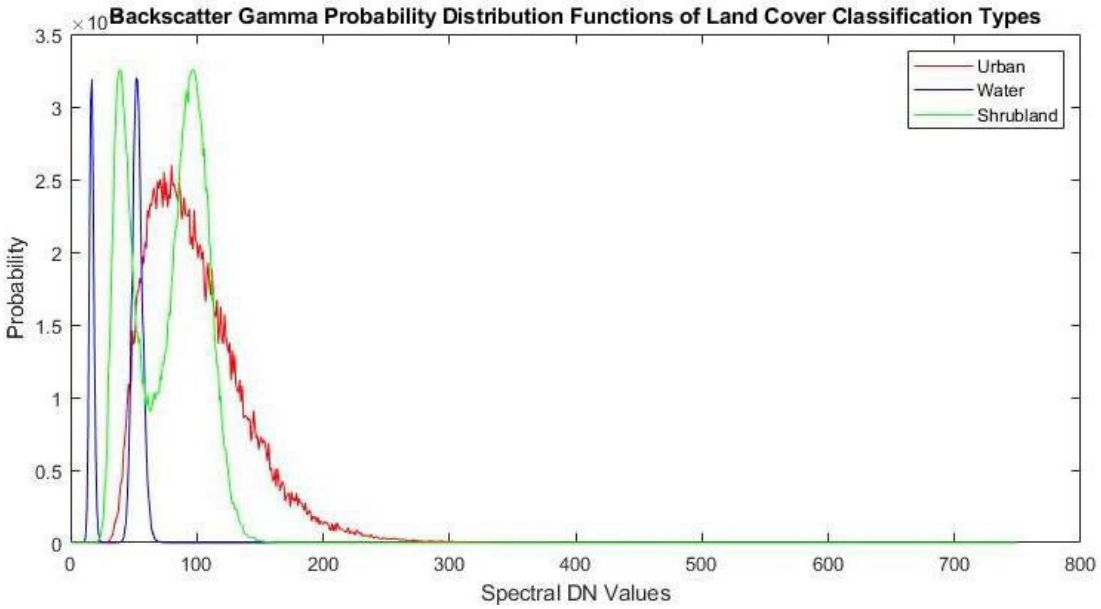

Figure A13. Gamma backscatter distribution of training areas for three land cover classes.

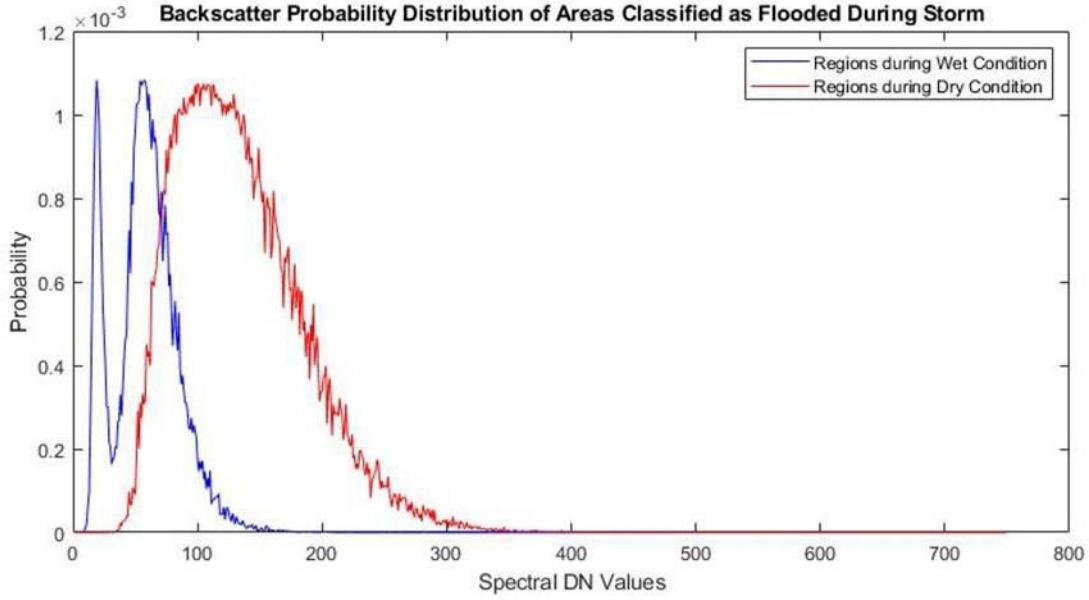

Figure A14. Gamma backscatter distribution for output regions classified as urban or shrubland during dry condition and water during wet condition.

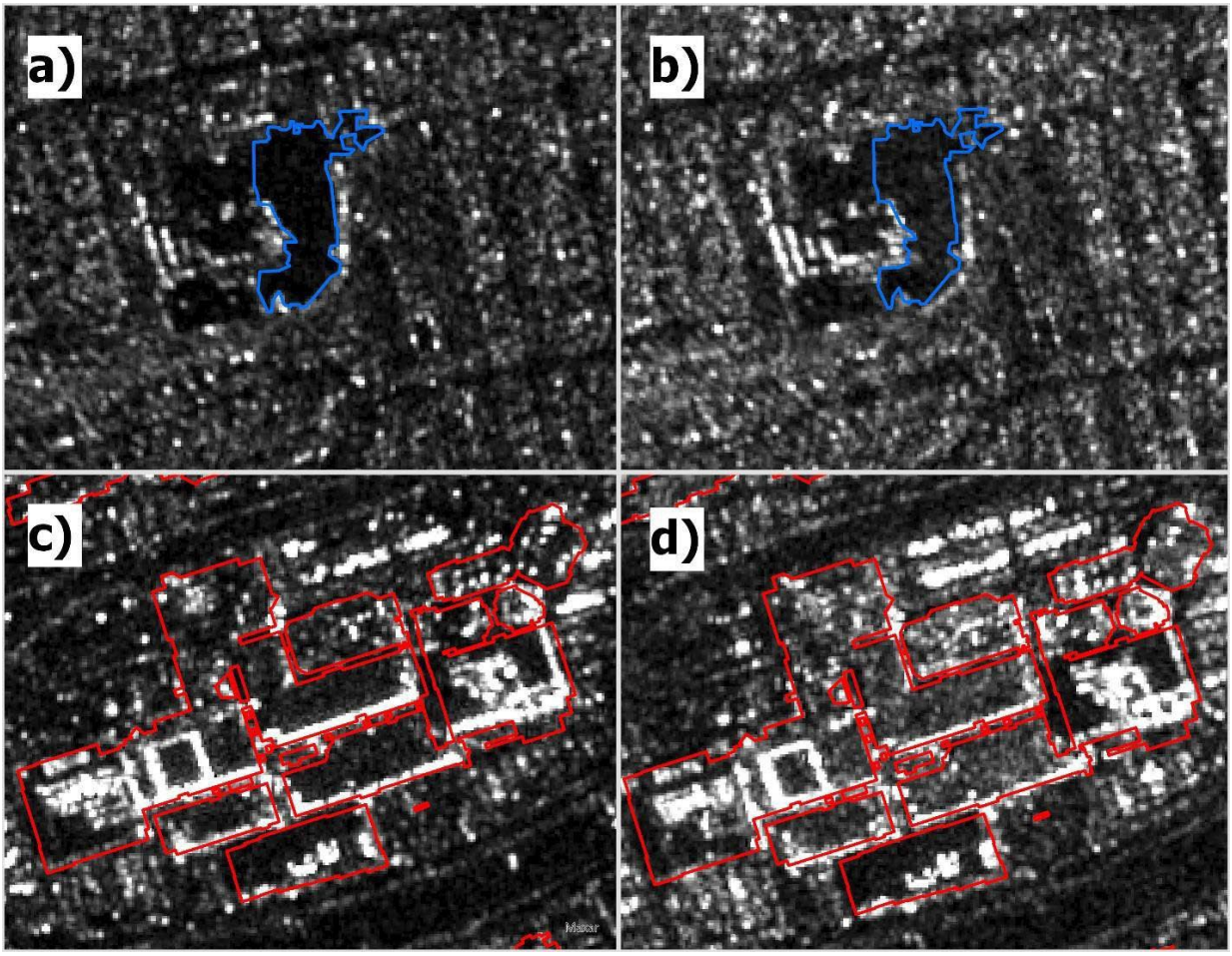

Figure A15. Illustration of changes seen in a grass and dirt classified as water during wet condition a) but shrubland during dry condition b), and example of the same phenomenon occurring c) with rooftops during storm in and d) during dry conditions.

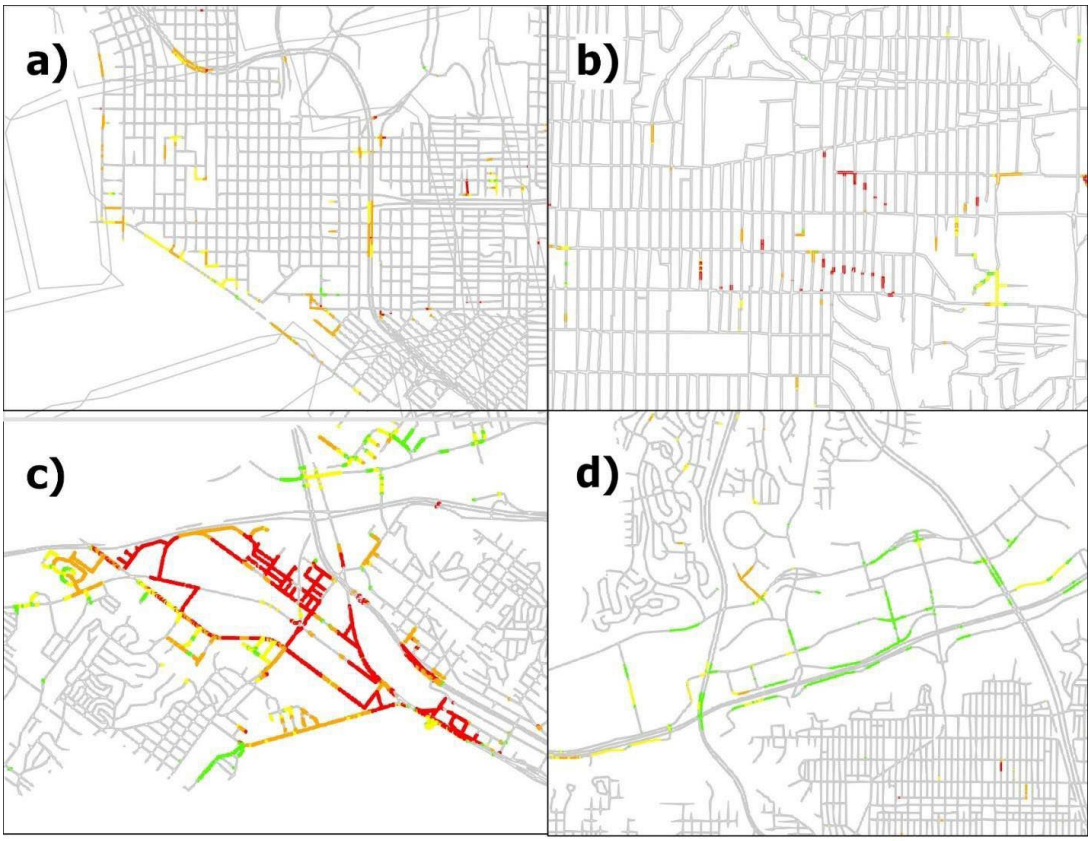

Figure A16. Four different areas of San Diego showing flood hazard levels. At the top, a) shows the downtown area of the city, b) a section of the Talmadge neighborhood, c) an area of elevated flood hazard , and d) the Mission Valley area around Fashion valley mall and the San Diego River. The area corresponding with c), the Midway District, is a mainly industrial and commercial area on a flat region between Mission Bay and San Diego Bay. Here the flood hazard index is high thanks to both high flood risk and high vulnerability indices, and it is a good example of high flood hazard in flat coastal areas. D) shows flood hazard along transportation networks in an area known for its flooding, Mission Valley.

Extensive, continuous flood hazard as would be expected in a floodplain area is present but note that the flood hazard is almost exclusively low in this area. This is not because flooding does not occur here- as the flood risk map shows, Mission Valley is a zone of concentrated flood risk- but because the mainly commercial, wealthier Mission Valley area scores lower in flood vulnerability indices. A) is the area around downtown, showing flood risk in the most built-up (though not necessarily most dense in terms of residents) neighborhoods of the city of San Diego. Once again there are extensive regions of high flood hazard along the coast. Also notable here are two concentrations of high flood hazard along the 5 freeway circumventing downtown. Finally, b) shows flood hazard in the relatively high-density neighborhood of Talmadge. This is an example of the sort of flood hazard typically seen in the city, with small stretches of roads or intersections having high flood hazard. This flood hazard, taking the form of NF, could be a nuisance or even a danger to drivers using roadways during storms.

Table A1. Validation Point Sources

Storm Date	Rainfall total over storm interval at Fashion Valley (inches)	Peak gauge height for San Diego River at Fashion Valley (feet)	Number of points reported flooded in study area	Flooded points shared with at least one other storm
January 6, 2016	3.41	11.6	7	4
February 17, 2017	2.78	14.2	23	9
December 8, 2018	2.19	8.2	25	8
February 14, 2019	1.78	12.1	20	5
Total: 47 unique flooded points				

Table A2. Flood Hazard Variable Weighting

Variable	Class	Weighting
Flood Risk		**Multiplier**
	High: 90–100	x3
	Med: 85–90	x2
	Low: 78–85	x1
	Negligible: 0–78	x0
Income		**+0.125**
	Below Poverty Line	0.125
	Middle Income	0.050
	High Income	0
Population Density		**+0.125**
	Very High	0.125
	High	0.100
	Middle	0.075
	Low	0.025
	Very Low	0
Drain Age		**+0.125**
	Oldest	0.125
	Old	0.075
	Recent	0.025
	Very Recent	0
Street OCI Index		**+0.125**
	Poor	0.125
	Fair	0.075
	Good	0
	Maximum Hazard Index = 1500	

Table A3. Cosmo-SkyMed Image Characteristics

Image	Sensor Mode	Resolution	Date	Direction	Polarization	Processing Level	Incidence Angle
Wet-weather scene	Stripmap	3m	02/27/2017, 1:22 UTC, 17:22 local time	Descending	HH	Level 1-D	26.56
Dry-weather scene	Stripmap	3m	02/16/2017, 1:44 UTC, 17:44 local time	Descending	HH	Level 1-D	26.57

Table A4. Generalized Linear Model Outputs

Explanatory Variable	Coefficient	Standard Deviation	T-statistic
Imperviousness	0.042571422	0.001504421	28.29753837
Bluespot Presence	2.218617379	0.039793941	55.75264266
Drain Difference	-0.001603797	0.0000829	-19.33972941
Flow Accumulation	-0.014684507	0.000382	-38.40768223
Slope	-0.090746418	0.00492696	-18.41833681

Table A5. Generalized Linear Model Additional Outputs

Dry	**True**	**False**
Predicted True	22	0
Predicted False	8	30
Wet	**True**	**False**
Predicted True	20	0
Predicted False	10	30