A geographically granular database of hurricane damage

Tatyana Deryugina[1]

Introduction

Natural catastrophe damage is increasing due to greater population exposure, growing wealth, and climate change. Yet damage estimates are notoriously imprecise. Margins of error for damage caused by some of the largest catastrophes are in the tens of billions of dollars, and geographically granular damage information is very rare. While insured damage is known to insurance companies, the sensitive nature of such data means that building-level claim data is seldom available to researchers. Moreover, widespread underinsurance means that observing insured damage is insufficient to understand the full effect of a catastrophe. With granular data on many individual-level outcomes becoming increasingly available, lack of geographically precise damage data prevents researchers from undertaking studies that would improve our knowledge of natural catastrophe impacts and how to mitigate them. For example, Medicare administrative data report the 9-digit ZIP code of beneficiary residence—a small geographic area often encompassing just a few houses—and tracks beneficiaries' health utilization and outcomes over time and space. However, without granular damage data, it is not possible to definitively identify individuals who were directly affected by a disaster.

We address this shortcoming by applying a computer vision model to disaster imagery data and creating a spatially detailed damage database for six major US hurricanes. The model uses Siamese Neural Networks—a deep learning technique—to detect buildings and classify their damage level. The network is trained using the XView2 dataset, which provides hand-coded building damage estimates for different disasters types—including floods and hurricanes—across 15 countries and the corresponding imagery. We applied a modified version of the trained model to disaster imagery from six recent US hurricanes to determine building-level damage in a subset of affected areas. The resulting database—comprising nearly 1.7 million buildings—can pave the way for several research projects, including assessments of mitigation strategies and individual-level responses to catastrophes.

Background

The xView2 AI Challenge was a competition organized by the Department of Defense's Defense Innovation Unit (DIU), in collaboration with the National Geospatial-Intelligence Agency (NGA),

[1] University of Illinois, Urbana-Champaign. deryugin@illinois.edu. This project was completed jointly with Rachel Young. Shubham Goel, Aaron Hoyle-Katz, Katie Kolodner, and Tejesh Bhaumik provided excellent research assistance. I thank Tejesh Bhaumik for helping to draft this report.

the Defense Advanced Research Projects Agency (DARPA), and the Office of Naval Research (ONR). The competition's goal was to accelerate research and development in object detection and image classification for overhead imagery using deep learning methods. Prior to the competition, the typical approach to surveying damage in the aftermath of a natural disaster was to manually annotate satellite and aerial images for building damage. This process was labor-intensive and typically took weeks. The goal of the challenge was to encourage the development of an algorithm that could automate this process, reducing the amount of time and resources required and ultimately expediting recovery responses. Challengers received access to the xBD dataset, which contains 850,736 hand-annotated buildings and spans 45,362 square kilometers of satellite imagery. Building damage was classified on a four-point scale, ranging from 0 (no damage) to 3 (completely destroyed). Submitted algorithms were assessed based on their performance in a hold-out sample.

The winning entry, whose code was subsequently made open-source, employed the U-Net Convolutional Neural Network (CNN) architecture, which is a popular algorithm for semantic image segmentation. Our project built on this algorithm using the open-source xBD database mentioned above and pre- and post-disaster imagery for six recent US hurricanes (Ida, Delta, Laura, Florence, Harvey, Michael) from Maxar's Open Data program. Pre-processing of these images was highly labor intensive. First, an analyst had to select images with minimal cloud coverage, visually determine areas of overlap between the pre- and the post-disaster imagery, and clip the images to the area(s) of overlap. Second, images had to be manually cropped to maximally exclude areas without buildings (e.g., farmland). Third, the resulting images were split to create 1024 by 1024 tiles to match the format of the xBD database. Finally, because the pre- and post-disaster images were sometimes captured at different times of the day, in different seasons, or simply under different lighting conditions, we performed histogram matching to ensure that such difference did not negatively impact model performance.

We then applied the algorithm to the resulting pre- and post-images. Unlike the xView2 testing set, which was used to determine the winning entry, these images were not part of the original xBD dataset and thus exhibited some systematic differences that could deteriorate model performance. The model thus required re-training to handle a wider variety of image types (e.g., different color balance), which was achieved by generating several systematically altered versions of the xBD dataset. After damage classification was completely, we used Microsoft's US Building Footprint database to identify buildings more accurately. In rare cases where the algorithm incorrectly labeled a single building as multiple buildings and assigned them different damage levels, we used their average as the damage level. Our final database thus contains some non-integer damage values.

Our main method of validating the results of our application of the algorithm was to test for the presence of outliers in the resulting dataset. Intuitively, hurricane damage should be spatially continuous, and cases where nearby buildings experience very different damage levels should be rare. To check for such outliers, we drew 1000 buildings at random for each hurricane. For every building, we created a rectangle around it with a diagonal length of 1 km. We then identified all other buildings in this area and calculated the interquartile range (IQR) of damage. We defined

an outlier as a building whose damage differed from the IQR by more than 1. Based on this definition, the outlier rate was negligible, indicating that the algorithm performed relatively well in this sample.

Results

Summary statistics

Table 1 shows the characteristics of the resulting damage dataset, which includes nearly 1.7 million buildings. Most of them (75.4%) are located in areas impacted by Hurricane Harvey, which made landfall in a large and densely populated metropolitan area (Houston, Texas). There is substantial heterogeneity in damage both within and across hurricanes. For example, very few buildings are completely untouched (damage level 0) or completely destroyed (damage level 3), except in the cases of Hurricanes Ida and Harvey, where 19 and 15 percent of buildings experienced no damage, respectively. For all but one hurricane, the “minor damage” (level 1) category is the most common, with about 70-80 percent of all buildings classified as experiencing minor damage. For Hurricane Laura, however, less than one-third of buildings fall into this category, whereas about two-thirds fall into the “major damage” (damage level 2) category. Low rates of zero damage are not surprising, as the Maxar disaster imagery deliberately selects areas with substantial damage levels. By contrast, low rates of complete building destruction indicates that the US building stock is relatively resilient to most hurricanes.

Table 1: basic characteristics of damage dataset

Hurricane name	Date of impact	No. of buildings	Distribution of damage (% of buildings)			
			0	1	2	3
Harvey	Aug-17	1,272,181	15.35	72.20	8.03	4.43
Florence	Sep-18	231,865	1.41	82.25	12.86	3.48
Michael	Oct-18	39,670	2.59	72.67	18.90	5.84
Laura	Aug-20	4,641	0.13	31.14	65.31	3.43
Delta	Oct-20	5,455	3.12	82.73	12.92	1.23
Ida	Sep-21	134,049	18.81	78.97	2.01	0.21

Note: The table reports damage summary statistics for six hurricanes that made landfall in the US. Cases where damage values are non-integers are counted in the lower damage category.

Next, we demonstrate the richness of the dataset with damage heatmaps for each of the six hurricanes available in the Maxar dataset. We rescale the damage categories so that the lowest value is 1 and the highest value is 4. Light blues are used to indicate the lowest damage levels, followed by darker blue and light orange. Bright red corresponds to the highest damage levels. Note that the exact appearance of the heatmaps depends on the zoom level, as the mapping algorithm smoothes individual points for legibility, and the amount of smoothing increases with the zoom level. The maps should thus not be viewed as exact representations of the underlying data but rather as illustrating the general damage patterns.

Hurricane Harvey

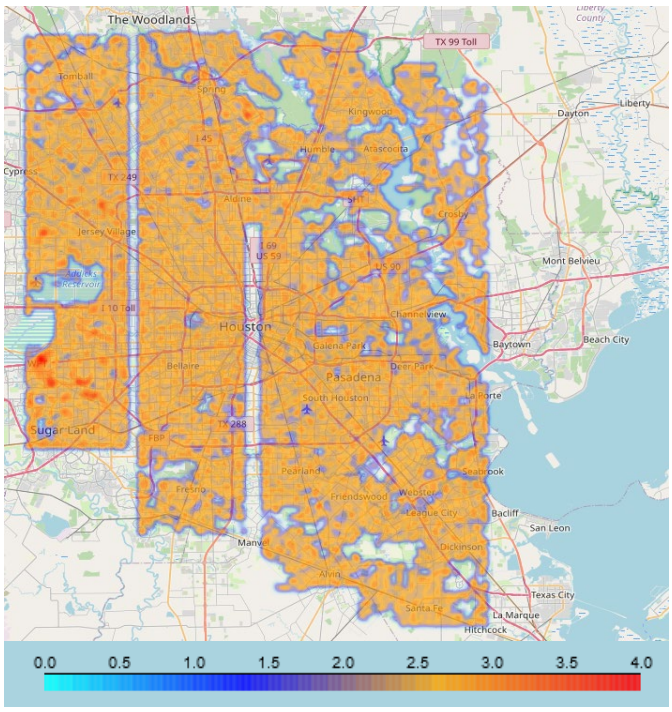

Figure 1. Damage map of Hurricane Harvey (August 2017)

Hurricane Harvey was a Category 4 hurricane that made landfall in Texas in August 2017, causing widespread devastation and historic flooding in multiple areas along the Gulf Coast. However, Maxar only provides appropriate satellite imagery for the Houston Metropolitan Area (Figure 1). Although some hot spots are apparent (e.g., Sugar Land and its surroundings), damage is fairly event distributed throughout the area. Such spatially granular data could be useful for a variety of applications. For example, researchers could correlate damage with building-level characteristics within neighborhood, comparing damage to homes located in the same area but differing in some observable trait. A key application we will be pursuing going forward is linking these data to individual-level longitudinal data that contain address information (e.g., confidential tax records) or 9-digit zip code (e.g., confidential Medicare administrative data). Doing so will allow for novel insights, such as separating the effect of direct damage to one's home from being in or near a damaged neighborhood. Preliminary analyses linked to longitudinal data on individual locations (the Infutor dataset) have already yielded surprising results: not only did Hurricane Harvey not lead to widespread outmigration from the Houston Metropolitan area (a fact that is apparent in aggregate data and does not require detailed damage information), but individuals whose homes were destroyed were not more likely to relocate than individuals living in the same area whose homes were unharmed. This counterintuitive finding suggests that the context in which a disaster occurs—including the economic prosperity of the area, insurance coverage, and the government response—matter greatly for individual decisions in the aftermath of an extreme event. Future research will probe these hypotheses.

Hurricane Florence

Figures 2a-2d show damage heatmaps for Hurricane Florence, a Category 4 hurricane that made an impact in the southeastern United States in September 2018. The storm caused significant damage and disruption in North and South Carolina, particularly in low-lying areas and communities near rivers and streams. Thousands of people were displaced from their homes, seeking shelter in evacuation centers. Maxar provides Hurricane Florence satellite imagery for three coastal areas: the North Charleston area, the Myrtle Beach area, and the Cape Fear River Basin. Figure 2a shows the general locations of these areas.

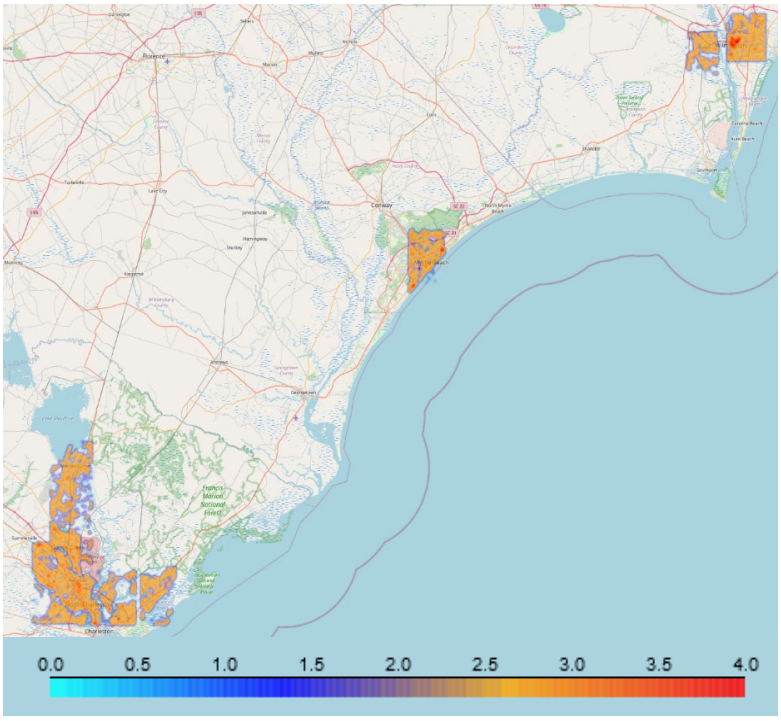

Figure 2a. Areas most impacted by Hurricane Florence (September 2018)

While North Charleston, South Carolina, did not experience the most severe impacts of Hurricane Florence in 2018, the city and the broader Charleston metropolitan area were still affected by the storm's outer bands and associated weather conditions (Figure 2b). The city did experience localized flooding in low-lying areas and some wind-related damage, including fallen trees and power outages. The impacts were relatively milder compared to areas closer to the storm's landfall, however. The majority of the homes experienced damage levels between two and three, but there are neighborhoods where damage levels average closer to four. However, the most intense impacts were concentrated farther north.

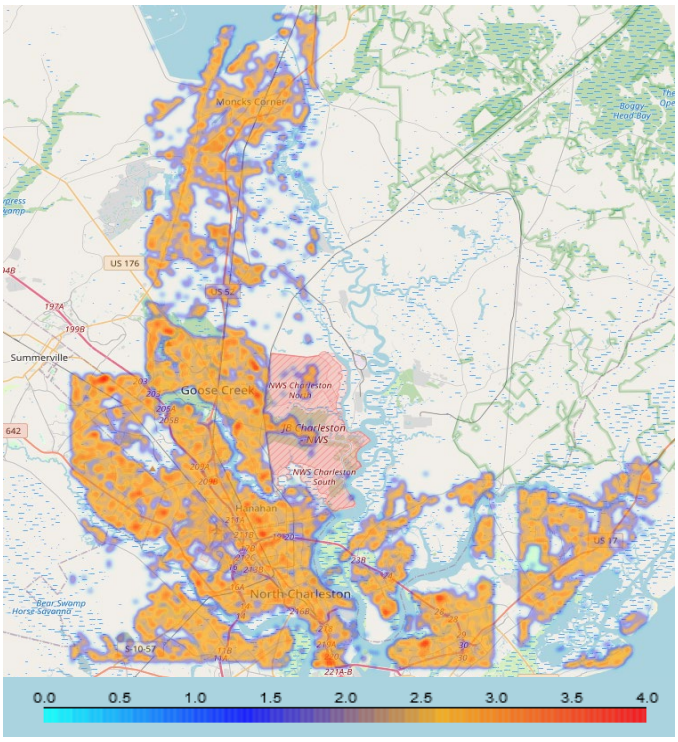

Figure 2b. Impact of Hurricane Florence in the North Charleston area

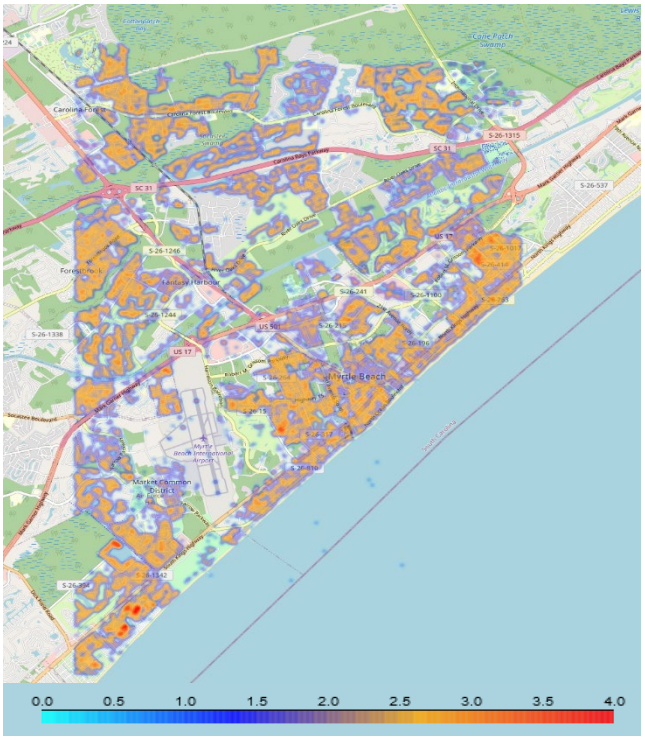

Figure 2c. Impact of Hurricane Florence in the Myrtle Beach area

Hurricane Florence had significant impacts on the Myrtle Beach area, located along the coast of South Carolina. While the city itself did not experience the most severe effects of the storm, it still faced a range of challenges due to the storm's heavy rainfall, strong winds, and potential for storm surge (Figure 2c). While the center of Hurricane Florence made landfall farther north in North Carolina, the storm's outer bands brought heavy rainfall to the Myrtle Beach area. The prolonged rainfall led to localized flooding in some parts of the city and surrounding communities. The rainfall also contributed to rising water levels in rivers and creeks. However, the general damage levels are noticeably lower than in Figure 2b, with most homes experiencing damage of less than three and only a few hotspots with high damage levels.

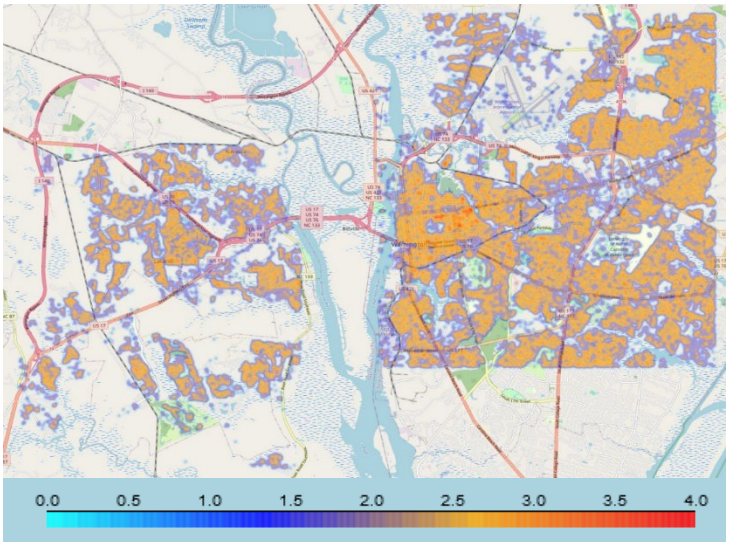

Figure 2d. Impact of Hurricane Florence in the Cape Fear River

The Cape Fear River Basin in North Carolina was one of the areas that experienced significant impacts from Hurricane Florence. The slow-moving nature of the storm, coupled with its heavy rainfall, led to catastrophic flooding in this region (Figure 2d). Rivers and tributaries within the basin, including the Cape Fear River itself, swelled well beyond their banks, inundating homes, businesses, roads, and agricultural fields. The slow-moving nature of the storm caught some residents off guard, and emergency responders had to conduct numerous water rescues as floodwaters rapidly rose. Roads, bridges, and transportation networks were washed out or compromised by the floodwaters. This hindered both rescue and recovery efforts, making it difficult for authorities to access and provide aid to affected areas.

Hurricane Michael

Hurricane Michael was a Category 5 hurricane which made landfall on the Florida Panhandle and then moved into Georgia in October 2018. Unfortunately, damage imagery from Maxar was only available for three Florida coastal areas impacted by the storm, shown in Figure 3a. These communities, including Mexico Beach, Panama City, and several other smaller communities, experienced the brunt of Hurricane Michael's impact. The storm brought destructive winds, storm surge, and heavy rainfall, resulting in widespread devastation. Many structures were destroyed, including homes, businesses, and public infrastructure. The communities in this region faced significant challenges in terms of power outages, water and sewage disruptions, and limited access to essential services.

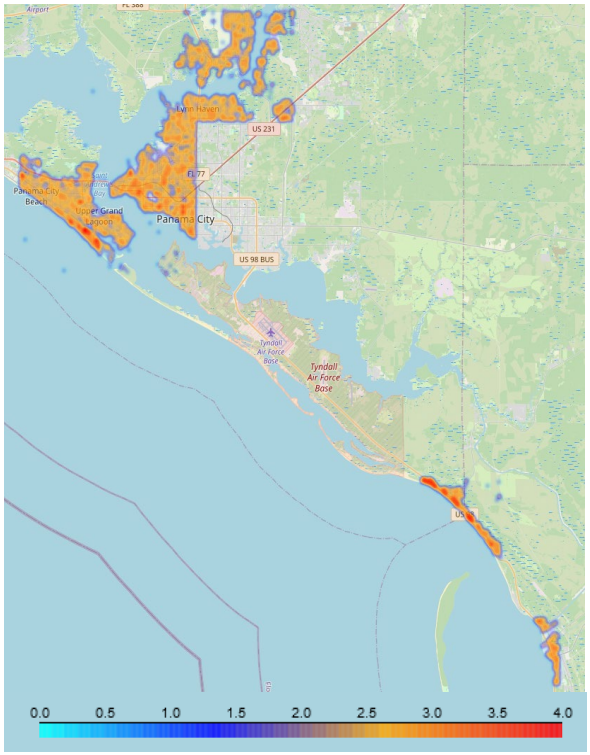

Figure 3a. Areas most impacted by Hurricane Michael (October 2018)

The Panama City area was one of the hardest-hit regions by the storm (Figure 3b). The storm brought sustained winds of up to 155 mph (250 km/h) at landfall, causing widespread destruction in its path. The Panama City area experienced extremely strong winds that tore apart buildings, uprooted trees, and caused extensive damage to infrastructure. Homes, businesses, and other buildings suffered varying degrees of damage, with many structures being severely damaged or destroyed. The storm's impact was particularly evident in beachfront and coastal areas.

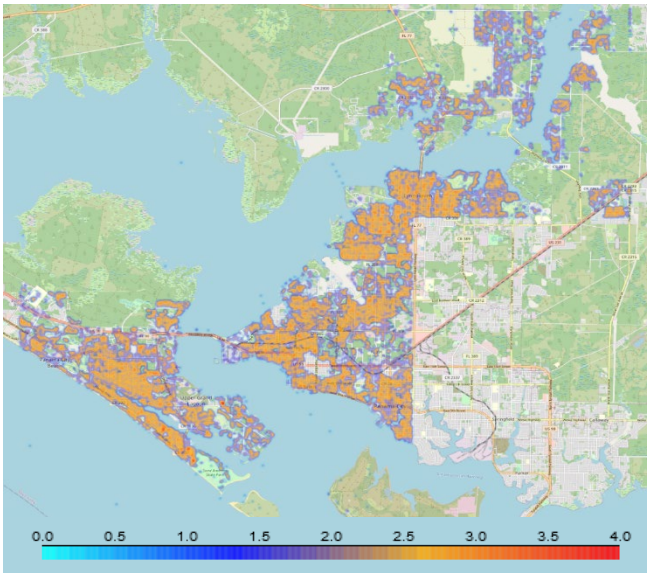

Figure 3b. Impact of Hurricane Michael in the Panama City area

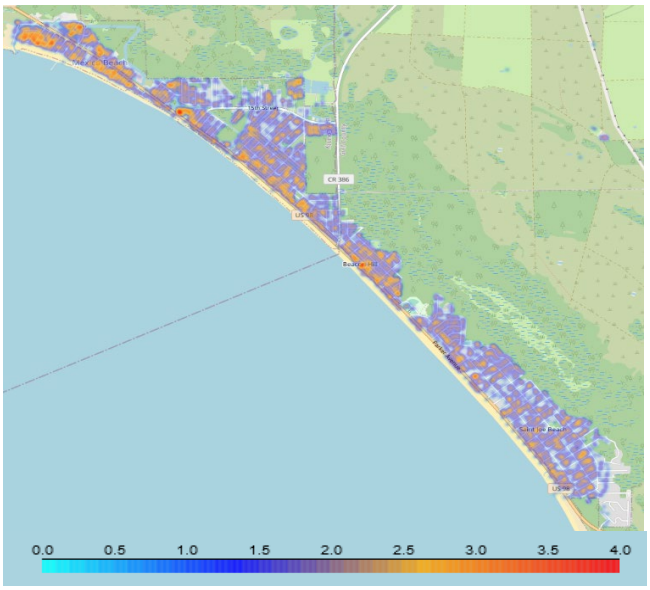

Figure 3c. Impact of Hurricane Michael in the Mexico Beach area

Likewise, Hurricane Michael had devastating effects on the Mexico Beach area, located along the Florida Panhandle. The storm's intense winds, storm surge, and destructive force left a lasting impact on this coastal community. Extreme winds demolished buildings, leveled homes, snapped trees, and left widespread debris in their wake. The storm surge, which reached over 15 feet in some areas, inundated coastal communities, causing flooding and washing away structures close to the shore. Debris was scattered across the landscape, making travel and recovery efforts challenging.

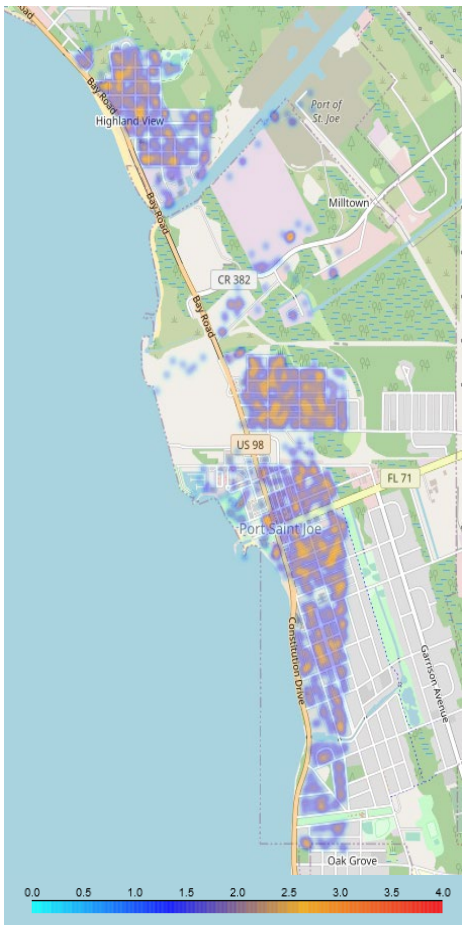

Figure 3d. Impact of Hurricane Michael in Port Saint Joe/Highland View

Hurricane Michael also had significant impacts on the Port Saint Joe and Highland View areas (Figure 3d). The strong winds associated with Hurricane Michael caused significant structural damage in the Port Saint Joe and Highland View areas. Homes, businesses, and other buildings were damaged or destroyed by the storm's impact. The storm surge from Hurricane Michael, combined with the astronomical high tide, led to coastal flooding in these areas. The surge, which reached several feet in some places, inundated coastal neighborhoods, causing flooding and contributing to the destruction of buildings near the shoreline. The destruction of businesses and infrastructure disrupted local economies, and the communities had to address the long-term effects on employment and commerce.

Hurricane Laura

Figure 4 shows the damage heatmap of Hurricane Laura, a Category 4 hurricane that made landfall in southwest Louisiana and southeast Texas in August 2020. Unfortunately, Maxar damage imagery is only available for an area in Calcasieu Parish, Louisiana. This area was one of the areas severely impacted by the storm. Hurricane Laura brought destructive winds with sustained speeds of up to 150 mph (240 km/h) at landfall. The strong winds caused extensive damage to buildings, infrastructure, and trees throughout Calcasieu Parish. Many homes and businesses suffered varying degrees of damage, from roof and siding loss to structural collapse.

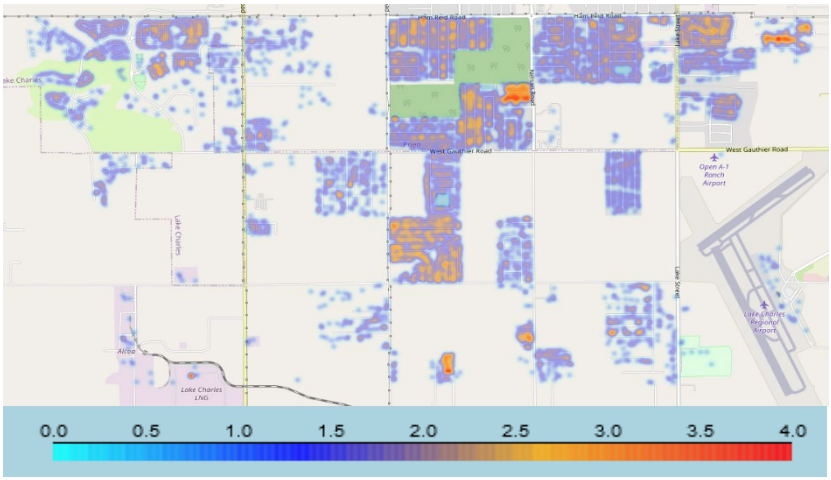

Figure 4. Damage map of Hurricane Laura (August 2020)

Hurricane Delta

Figure 5 shows the damage map for Hurricane Delta, a Category 2 hurricane that made landfall in southwestern Louisiana in October 2020. The storm brought strong winds, heavy rainfall, and storm surge, resulting in widespread damage to homes, businesses, and infrastructure. Many areas had already been affected by Hurricane Laura just a few weeks earlier, and Hurricane Delta compounded the damage. Structures that were weakened or damaged by Laura were further compromised by Delta's winds, leading to additional destruction. Maxar provides damage imagery for Erath and Delcambre areas of Louisiana. While Hurricane Delta was not as powerful as some other recent hurricanes, it still had significant impacts on these coastal communities, with the potential to cause damage to structures, uproot trees, and knock down power lines.

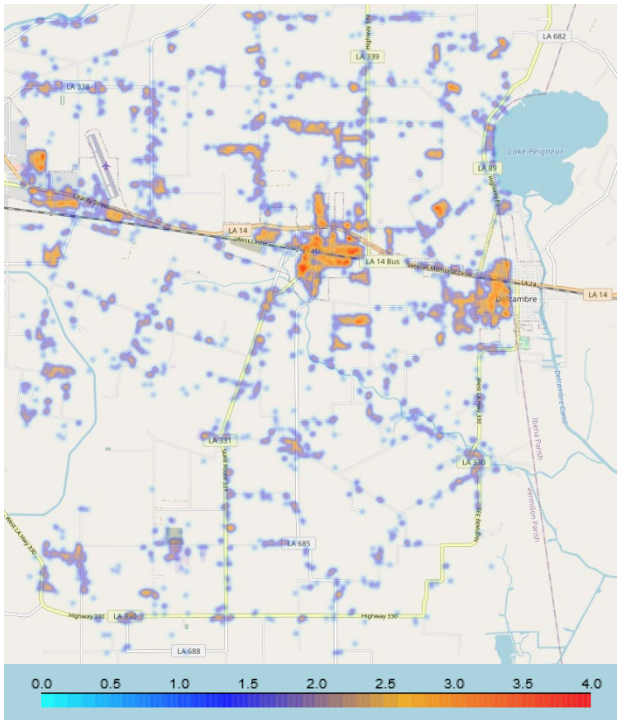

Figure 5. Damage map of Hurricane Delta (October 2020)

Hurricane Ida

Hurricane Ida was a Category 4 hurricane that made impact along the Gulf Coast in late August 2021. Maxar provides damage imagery for the Greater New Orleans area. This region, including New Orleans and surrounding parishes, experienced significant impacts from the storm. The hurricane brought strong winds, heavy rainfall, and storm surge, resulting in widespread power outages, structural damage to buildings, and extensive flooding. Roofs were torn off structures, windows shattered, and signs were blown away. The storm surge from Hurricane Ida led to coastal flooding in parts of the New Orleans area, including low-lying neighborhoods and areas adjacent to bodies of water. The surge inundated roads, homes, and businesses, contributing to the destruction and making evacuation difficult in some areas. Many communities in this area faced challenges with infrastructure damage, including the loss of critical services such as electricity and water.

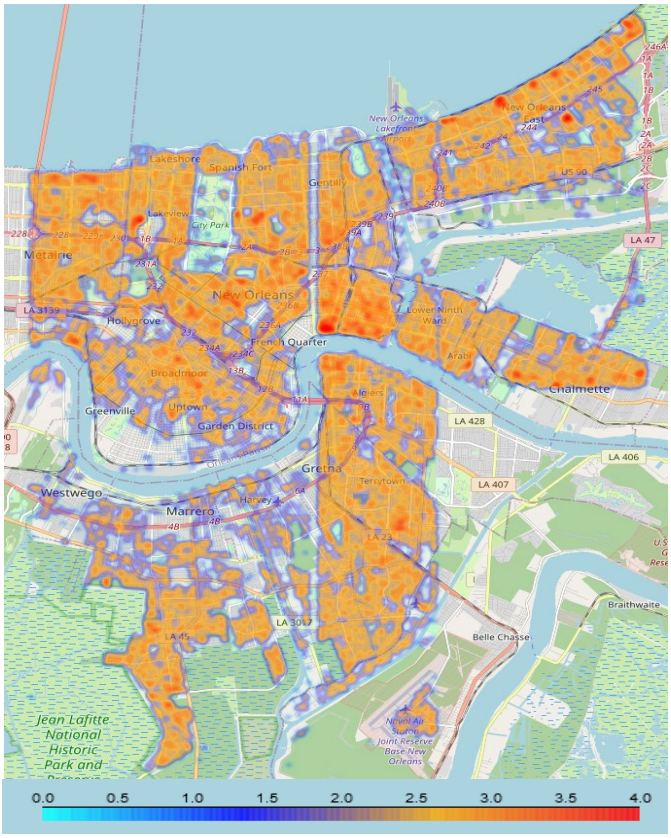

Figure 6. Damage map of Hurricane Ida (August 2021)

Limitations

Obtaining the required satellite imagery proved more challenging than expected, ultimately yielding a smaller sample of extreme events than desired. Locating appropriate pre- and post-disaster imagery is time-intensive, partly because narrative descriptions of impacted areas are often imprecise. It also requires substantial geoprocessing skills, and it was unfortunately difficult to recruit and retain appropriate personnel. Systematic collection and publication of satellite data have greatly expanded in recent years, however, promising greater data availability in the future but providing limited ability to obtain and process data for earlier extreme events.

Conclusion

The escalating impact of natural catastrophes resulting from factors such as increased population exposure, economic growth, and climate change highlights the urgency of understanding their effects and devising effective mitigation strategies. However, the imprecision of current damage estimates hinders comprehensive analyses and informed decision-making. This limitation is compounded by the scarcity of geographically granular damage data and the lack of access to

building-level claim information. The existing gap in knowledge regarding the individual-level impact of disasters remains a significant hurdle in devising targeted responses.

We present a promising solution to this problem that employs computer vision techniques to create a detailed spatial database of building-level damage for six major US hurricanes. Efforts such as these hold the potential to catalyze a new era of research aimed at understanding the intricate dynamics of natural catastrophe impacts. As natural disasters continue to pose significant challenges to communities and economies, the availability of such data-driven insights is invaluable in steering efforts toward more resilient and sustainable outcomes.