Charles River Watershed Water Resources

Assessing Flooding Vulnerability to Assist High Water Intervention and Urban Planning Programs in the Charles River Watershed

 **Technical Report**

Final Draft – November 19th, 2020

Trista Brophy (Project Lead)

Willow Coleman

Anna Garik

Will Peters

***Advisors***

Dr. Cedric Fichot, Boston University (Science Advisor)

Dr. Kenton Ross, Langley Research Center (Science Advisor)

**1. Abstract**

The Charles River watershed intersects 35 municipalities within the Boston Metropolitan Area and has a total population of 1.2 million, making it one of the most densely populated watersheds in New England. In recent years, the watershed has observed higher rates of flood inundation, mainly due to increased development, extreme precipitation events, and increased surface runoff. As the frequency of flood events increases and a changing climate poses an ongoing threat to local communities, governments and organizations in Massachusetts are in need of accurate flood risk assessments. This NASA DEVELOP project partnered with the Charles River Watershed Association, the Town of Natick’s Office of Sustainability, and the Massachusetts Audubon Society to assess the potential for watershed degradation, flood vulnerability, and flood susceptibility in the watershed. The team used Landsat 5 Thematic Mapper, Landsat 8 Operational Land Imager, Sentinel-1 C-Band Synthetic Aperture Radar, and Sentinel-2 MultiSpectral Instrument to assess the feasibility of identifying flood events using remote sensing. After identifying images that overlapped with the reported flood events, the team concluded that it was not feasible to use Earth observation data to detect localized flooding in this study. Instead, the Federal Emergency Management Agency (FEMA) 100-year floodplain was used as a proxy for areas where flooding may occur. The team used statistical analysis and supervised classification to develop a flood susceptibility map, incorporating flood conditioning factors like soil drainage, height above nearest drainage, and topographic wetness index. This was overlaid with demographic and socioeconomic data to create a flood vulnerability map. The flood susceptibility map captured over 2/3 of reported flood events in the watershed, an improvement over the 1/3 of events captured by the FEMA 100-year and 500-year floodplain maps.

**Key Terms**

remote sensing, flood vulnerability, flood susceptibility, urban flooding, Sentinel, Landsat, impervious surface

# 2. Introduction

***2.1 Background Information***

Flood risk is increasing around the globe primarily due to increases in urbanization and average global temperatures (Hirabayashi, 2013; Huong & Pathirana, 2013). Currently, flooding is estimated to cost $96 billion in economic activity globally; a number that is expected to increase to over $500 billion by the end of the decade (Ward et al., 2020). As urban centers expand, impervious surfaces increasingly replace natural vegetation that slows flooding by allowing precipitation to percolate into the ground, contributing to additional runoff and an increase in urban flooding. Increasing concentrations of greenhouse gases in the atmosphere further compound this phenomenon. As the atmosphere warms, precipitation patterns can be altered in the form of increased intensity, duration, and frequency of rainfall events (Tabari, 2020).

As temperatures rise in New England, a rise in the frequency and intensity of precipitation is predicted, leading to more flood events (Easterling, 2017). Coastal communities in particular, such as the Boston Metro Area, will be even more susceptible due to the combined effect of increased precipitation and sea level rise (Suarez, 2005). At the same time, population and economic growth are likely to increase land development in the region. The Boston Metro Area, which has 35 municipalities that intersect the Charles River watershed (Figure 1), is expected to grow 10% in population by 2030, compared to 2009 levels (Cheng et al., 2017). An increase in developed land increases not only the amount of infrastructure at risk, but also the number of impervious surfaces, and thus the runoff that impacts rivers, streams, and stormwater systems (Suarez, 2005).

Map

Description automatically generated

*Figure 1:* The project study area included the Charles River watershed and the Town of Natick, Massachusetts.

The Charles River begins in Hopkinton, Massachusetts and runs 182 km before emptying into the Boston Harbor. The total drainage area of the Charles River watershed is 798 km2 and consists of 50% urban landcover, 41% forest and wetland, and 3% water (Cheng et al., 2017). Although weather across the watershed is largely homogenous (Cheng et al., 2017), high streamflow events do occur, mainly due to seasonally high precipitation caused by coastal storms, seasonally high soil moisture, and antecedent streamflow due to snowmelt (Agel, 2019).

The municipalities that intersect the Charles River watershed are home to 1.2 million people (Cheng et al., 2017). Floods have different effects on communities in the Boston Metro Area, as a 2012 study found that the environmental justice communities of East Boston and Everett were left out of actively participating in planning processes due to insufficient knowledge of climate change or flood adaptations (Douglas et al., 2012). Equal engagement of local communities can increase their trust in the implementation of flood precautions, ensuring all municipalities are adequately prepared for increased flooding.

Flood control and prevention measures can reduce potential damage caused by floods (Billa et al. [2006](https://link.springer.com/article/10.1007/s13201-018-0710-1#ref-CR4); Huang et al. [2008](https://link.springer.com/article/10.1007/s13201-018-0710-1#ref-CR10)). Flood susceptibility is defined as the likelihood of a dangerous event occurring in an area on the basis of local terrain conditions (Santangelo et al., 2011) and can be used to identify areas with high flood potential to inform mitigation (Tehrany et al., 2015) and prevention strategies. Humans, infrastructure, and natural ecosystems may be susceptible to flood hazards (Samuels, 2009). Flood susceptibility can be analyzed using various types of models, including hydrological and data-driven models.

To create a flood susceptibility map, flood conditioning factors and their relationship to flooding need to be assessed (Liu & De Smedt, 2005). The selection of flood conditioning factors of an area varies based on the area’s individual characteristics (Tehrany et al., 2013). The team identified flood conditioning factors based on data availability and the Cheng et al. (2017) study on flood mitigation in the Charles River watershed. Flood vulnerability is a measure of the sensitivity of individuals, groups, or communities to their flood susceptibility (IPCC, 2007). The team created the flood vulnerability map by overlaying the flood susceptibility map with socioeconomic data. Flood risk combines assessments of the flood hazard, the susceptibility or exposure, the vulnerability of human and natural ecosystems, and often, the probability of occurrence (Randolph, 2012). The team used these methods to examine flood risk in the Charles River watershed beginning from January 2000 through September 2020.

* 1. ***Project Partners & Objectives***

For this project, the Fall 2020 Boston NASA DEVELOP team partnered with the Charles River Watershed Association (CRWA), one of the country’s oldest watershed protection organizations, the Town of Natick’s Office of Sustainability, and the Massachusetts Audubon Society (Mass Audubon), one of the largest providers of environmental education in Massachusetts. All three organizations are in the process of establishing climate preparedness plans for flooding due to weather extremes and are interested in incorporating cost-efficient, remotely-sensed data. The partners can use the maps to plan for future weather extremes and high-water events, specifically by identifying areas with the highest risk for future flood events. This will enable specific delegation of resources and aid to flood mitigation or recovery programs. The goal of this project was to conduct a rapid feasibility study on the use of remotely-sensed Earth observation (EO) data to assess flood risk in the Charles River watershed. The project objectives included: (1) mapping impervious surfaces and watershed degradation potential, (2) assessing flood susceptibility and vulnerability in the watershed, and (3) creating an ArcGIS StoryMap to communicate the results of the project with the general public and increase understanding of flood risks in the region.

# 3. Methodology

***3.1 Data Acquisition***

The team used Landsat 5, Landsat 8, Sentinel-1, and Sentinel-2 imagery from Google Earth Engine (GEE) in this study. All imagery was filtered by region, using a shapefile of the Charles River watershed, and a set of date ranges that overlapped with reported flood events (Table 1). Applying a cloud mask to Landsat 5, Landsat 8, and Sentinel-2 imagery minimized the effects of cloud cover on optical imagery. The resulting collection of satellite images was used to detect the extent of flooding in the Charles River watershed.

Table 1

*Earth observations used in this study*

|  |  |  |  |
| --- | --- | --- | --- |
| **Satellite** | **Description & Use** | **Dates** | **Source** |
| Landsat 5 Thematic Mapper (TM) Collection 1 Tier 1 Raw Scenes | Landsat 5 TM surface reflectance data were used to assess the feasibility of estimating the extent of recent flood events. | 2003 – 2010 | United States Geological Survey via GEE |
| Landsat 8 Operational Land Imager (OLI) Surface Reflectance Tier 1 | Landsat 8 OLI surface reflectance data were used to assess the feasibility of estimating the extent of recent flood events. | 2015 – 2020 | United States Geological Survey via GEE |
| Sentinel 1 C-band Synthetic Aperture Radar (C-SAR) Ground Range Detected, log scaling | Sentinel-1 C-SAR data were used to assess the feasibility of estimating the extent of recent flood events. | 2015 – 2019 | European Space Agency via GEE |
| Sentinel-2 MultiSpectral Instrument (MSI), Level-1C | Sentinel 2-MSI Level-1C data were used to assess the feasibility of estimating the extent of recent flood events. | 2016 – 2020 | European Space Agency via GEE |

The team acquired a 2016 1 m Land Cover/Land Use map of the state of Massachusetts as a geodatabase from MassGIS and then cropped it to the study area in ArcGIS (Table 2). The 2010 US Census data for Massachusetts were also downloaded from MassGIS and census block groups that intersected the study area were selected in ArcGIS for further analysis (Table 2). The team also downloaded critical infrastructure layers as point data sets from MassGIS and cropped them to the study area (Table 3). The American Community Survey (ACS) income/poverty data were downloaded in table form from the US Census Bureau data portal and then spatially joined to corresponding Massachusetts census block groups (Table 3).

Table 2

*Datasets used in watershed degradation maps*

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Description & Use** | **Dates** | **Source** |
| Land Cover/Land Use | Land cover/land use data were used to map impervious surface. | 2016 | MassGIS Bureau of Geographic Information |
| 2010 US Census | Census data aggregated to block groups were used to calculate population density per acre. | 2010 | US Census Bureau |

Table 3

*Ancillary data sets used in flood vulnerability and flood susceptibility maps*

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Description & Use** | **Dates** | **Source** |
| Critical Infrastructure | Acute Care Hospitals, Non-Acute Care Hospitals, Long Term Care Facilities, Police Stations, Fire Stations, and Colleges/Schools point data were overlaid with flood susceptibility maps. | 2015, 2018, 2019 | MassGIS Bureau of Geographic Information |
| ACS Income/Poverty Data | Ratio of Income to Poverty Level in the Past 12 Months (2018) dataset was used as a factor in the flood vulnerability index. | 2018 | US Census Bureau |
| USA Soils Drainage Class Map Layer | A map of soil drainage derived from the gNATSGO and gSSURGO databases was used as a flood conditioning factor. | 2019 | Esri, Natural Resources Conservation Service |
| Surficial Materials of Massachusetts | A 1:24,000-scale surficial materials geologic map database of Massachusetts that shows the distribution of nonlithified earth materials at the land surface was used as a flood conditioning factor. | 2019 | United States Geological Survey, Commonwealth of Massachusetts, Massachusetts Geological Survey, Executive Office for Administration and Finance |
| Land Cover/Land Use | Land cover/land use data were used as a flood conditioning factor. | 2016 | MassGIS Bureau of Geographic Information |
| National Hydrography Dataset (NHD) | The NHD flowlines were used to calculate height above nearest drainage (HAND), a flood conditioning factor. | 2020 | United States Geological Survey ScienceBase Catalog |
| Elevation/ Topographic Data | Topographic wetness index (TWI), slope, elevation, and HAND flood conditioning factors were derived from this digital elevation model (DEM). | 2005 | MassGIS Bureau of Geographic Information |
| Federal Emergency Management Agency (FEMA) National Flood Hazard Layer (NFHL) | The FEMA NFHL served as a proxy for flood extent. | 2017 | MassGIS Bureau of Geographic Information, Federal Emergency Management Agency |

***3.2 Watershed Degradation Maps***

The National Oceanic and Atmospheric Administration (NOAA) Impervious Surface Analysis Tool (ISAT) is a tool for ArcGIS that calculates the percentage of impervious surface area within a given region and combines that with population density data to serve as a proxy for whether local water quality will be good, fair, or poor. NOAA ISAT uses the 30 m National Land Cover Database and other medium-resolution land cover databases to derive its impervious surface product, making it highly valuable in regions that lack a high-resolution land cover map. However, the MassGIS 2016 Land Cover/Land Use map is a 1 m high-resolution product from which impervious surface can easily be extracted. As such, the team chose to emulate the NOAA ISAT tool by combining calculations of impervious surface fraction per census block group derived from the 1 m product and population density (persons per acre) from 2018 ACS data for Massachusetts (Table 2). The team calculated population density based on ACS data at the census block group level using the ArcGIS Field Calculator and impervious surface fraction at the census block group level in GEE. Both datasets were reimported into ArcGIS and visualized using a bivariate symbology, which allowed the team to identify census block groups that had both high impervious surface fraction and high population density.

***3.3 Development of a Flood Inventory with Earth Observation Data***

In order to create a flood susceptibility map, the team had to first create a flood extent and inventory map. A flood inventory map provides information on locations where flood events occur most frequently in the study area as well as their geographic extent. The NOAA Storm Events Database provided a list of 57 flood events that occurred within the Charles River watershed between 2000 and 2020. The team used the NOAA Storm Events Database to populate a list of flood events that were classified as flood, flash flood, or heavy rain between 2000 and 2020. Remaining floods falling on the same date and in the same location with several reports were combined, resulting in 57 flood event dates that included 196 points (Figure 2). When exact coordinates were not available for the flood locations, approximate coordinates based on the town where the event occurred were used.

Map

Description automatically generated

*Figure 2:* A map of flood events in the Charles River Watershed within the study period, taken from NOAA’s Storm Events Database.

Each flood event included data on the cause of flooding, latitude and longitude of flood start and end locations, and other notes. The team found all Landsat 5, Landsat 8, Sentinel-1, and Sentinel-2 imagery that spatially and temporally overlapped with each recorded flood event and examined if flooding was visible in the imagery. If the team detected a flood in any of the satellite images that overlapped with a recorded flood event, a polygon was drawn around the flood extent in GEE and recorded. This product served as the flood inventory from which our bivariate, multivariate, and decision tree-based flood susceptibility models could be trained from.

To create the flood inventory, all image IDs that were captured within the four-day period before and the four-day period after each flood event were recorded. Optical images from Landsat 5, Landsat 8, and Sentinel-2 were then filtered by cloud coverage such that only images with less than 50% cloud cover were included in the final collection. For radar images (i.e., Sentinel-1 C-SAR), cloud filtering was not applicable because radar penetrates clouds. A median composite of the Landsat 5, Landsat 8, and Sentinel-2 ImageCollections was calculated in GEE and clipped to the study area to create three satellite base layers for comparison. The median image was assumed to represent normal, non-flood conditions because flood events do not persist for over a year. The base layer allowed the team to visually compare each potential flood image against normal conditions at the corresponding recorded flood location for each flood event.

Given that it is challenging to recognize water features in unprocessed Sentinel-1 imagery, the team performed a supervised image classification in GEE on Sentinel-1 imagery using a Random Forest (RF) classifier to distinguish between water and all other land cover types (non-water). Training polygons of water and non-water features were drawn using the Sentinel-2 median base layer as a reference and were used to train the classifier. The team applied the classifier to the Sentinel-1 median base layer to create a classified water and non-water map to serve as a reference under normal, non-flood conditions. The classifier was also applied to each overlapping Sentinel-1 flood image. The classified non-flood condition image and classified potential flood image were visually compared at the corresponding recorded flood location for each flood event.

***3.4. Development of a Flood Inventory with FEMA National Flood Hazard Layer***

Due to challenges with utilizing EO data for creating a flood inventory map (see section 4.2), the team instead utilized the FEMA NFHL. The team clipped the NFHL to the Charles River watershed, such that it represented the current effective flood risk for flood zones in the study area. The main flood risk classes in the NFHL are 1% annual chance flood event (100-year floodplain), 0.2% annual chance flood event (500-year floodplain), and areas of minimal flood risk. To select training points for the model, the FEMA map was filtered into two categories: 1) flood (within the 100-year flood zone) and 2) no flood (within the minimum flood risk zone). 6537 random training points were extracted from each category.

***3.5 Calculation of Flood Conditioning Factors***

Seven rasterized flood conditioning factors were selected based on previous literature and data availability: land cover, soil drainage, surficial materials, elevation, slope, TWI, and HAND. All flood conditioning factors were split into roughly 5 quantiles using Reclass by Range of Values in ArcGIS Pro. The 2005 MassGIS DEM was used to calculate elevation, slope, TWI, and HAND. All other flood conditioning factors were used as provided. Slope was calculated from the DEM by applying the Slope tool in ArcGIS Pro. To calculate TWI, a multidirectional flow algorithm (TWIMD8) was used in conjunction with the ArcGIS Pro Flow Direction and Flow Accumulation tools (Grabs et al., 2009). TWI is defined at a certain point in the catchment as the upslope area per unit contour length divided by the local gradient (Equation 1), where is the upslope area per unit contour length and is the local gradient (Grabs et al., 2009).

(1)

HAND was calculated by applying the D-Infinity method to the DEM and NHD flowline datasets using the Fill, Flow Direction, Flow Accumulation, and Flow Distance tools in ArcGIS Pro. HAND is a drainage-normalized version of a DEM that gives the vertical distance from a surface cell and its respective outlet-to-the-drainage cell, which can be used to find the difference in level between two cells that belong to a mutually existing flow path (Nobre et al., 2011).

***3.6 Flood Susceptibility Mapping***

The first data-based model used was the frequency ratio (FR) method, which is a form of bivariate statistical analysis. It is the ratio of the probability of a flood occurring to the probability of a flood not occurring (Equation 2, Tehrany et al., 2013). While FRs calculate the individual correlations between the flood inventory map and each conditioning factor, they neglect the potential relationships between the different flood conditioning factors (Lee et al., 2013; Tehrany et al., 2019). The value of the FR describes the relationship between the factor and flooding, where FRs larger than 1 indicate a strong relationship between the factor and flooding, while FRs less than 1 indicate a weak relationship (Tehrany et al., 2013).

(2)

The second data-based model was a multivariate logistic regression (MLR), a type of multivariate statistical analysis. It examines the multiple associations between the different flood factors and the flood inventory map simultaneously, taking into account the relationship between the flood conditioning factors, but ignores the influence of each flood conditioning factor quantile on the occurrence of flooding (Tehrany et al., 2013). The logistic regression equation (Equation 3) was used to determine the probability of flooding (P) given the values of flood conditioning factors in an area (Equation 4).

(3)

(4)

An ensemble method that utilizes both FR and MLR can overcome the disadvantages of using either FR or MLR individually (Tehnrany et al., 2013). The team used this ensembled FR/MLR method to develop a flood susceptibility map. To do so, the normalized FRs of each flood conditioning factor class were summed and used as weights in the logistic regression (Equation 5) and then used to find the weighted flood probabilities (Equation 3). The same training points were used from the MLR and FR methodology. The third data-driven model was RF, a type of decision tree (DT) algorithm that is commonly used to address the issues of multi-classification and prediction (de Santana et al., 2018; Quiroz et al., 2018; Tsagkrasoulis & Montana, 2018). In a study including three DT methods, Chen et al. (2020) concluded that the RF model outperformed the other DT models, so the team opted to utilize an RF model for predicting flood susceptibility.

(5)

***3.7 Flood Vulnerability Mapping***

A flood vulnerability map was created by incorporating socioeconomic and demographic factors, such as income and race, into the flood susceptibility map. Based on literature review and discussions with the partners, the most important population demographic factors for flood vulnerability were elderly, non-white, and impoverished residents. Elderly individuals tend to lack emergency preparedness plans, have limited mobility, and chronic physical or mental illness that can inhibit evacuation and recovery from flood events (Bukvic et al., 2018). Communities with large non-white populations tend to have disproportionately higher flood risk than majority white communities (Fielding, 2017). Lastly, flooding can exacerbate existing social inequalities in low-income neighborhoods because these areas typically greenspace to absorb flood water, and residents may lack the financial resources to evacuate and rebuild after the flood waters recede.

The team developed a flood vulnerability index (FVI) to represent the amount of vulnerability in each census block group. The FVI was calculated by summing the fraction of elderly residents (age 65 and above), the fraction of non-white residents (including Asian, Black, Hispanic, and Native/Pacific Island individuals), and the fraction of residents living below the poverty line per census block group. The team overlaid the FVI per census block group with the flood susceptibility map to understand where high-risk populations living in the Charles River watershed are most likely to be vulnerable to flooding. This was represented using a bivariate symbology in ArcGIS, where census block groups with a high FVI and a high fraction of area susceptible to flooding were considered the most flood vulnerable.

# 4. Results & Discussion

***4.1 Watershed Degradation Maps***

The impervious surface fraction per census block group was highest near downtown Boston, while the impervious surface fraction was lower in the rural and suburban middle and lower Charles River watershed (Figure 3, left). Also, census block groups with an impervious surface fraction of greater than 10% are likely to experience some form of watershed degradation or modification due to surface runoff. These results are important for the watershed degradation map because high impervious surface fraction increases the potential for surface runoff and flooding. In addition, the team noted that census block groups with both high impervious fraction and high population density exist in the Charles River watershed, notably near the mouth of the Charles River in downtown Boston. This suggests that there was a higher potential for watershed degradation in this area compared to the lower portion of the watershed (Figure 3, right). It is important to note that the potential for watershed degradation does not necessarily guarantee there will be degradation, due in part to mitigation efforts like green infrastructure and sustainability efforts taken on by municipalities.

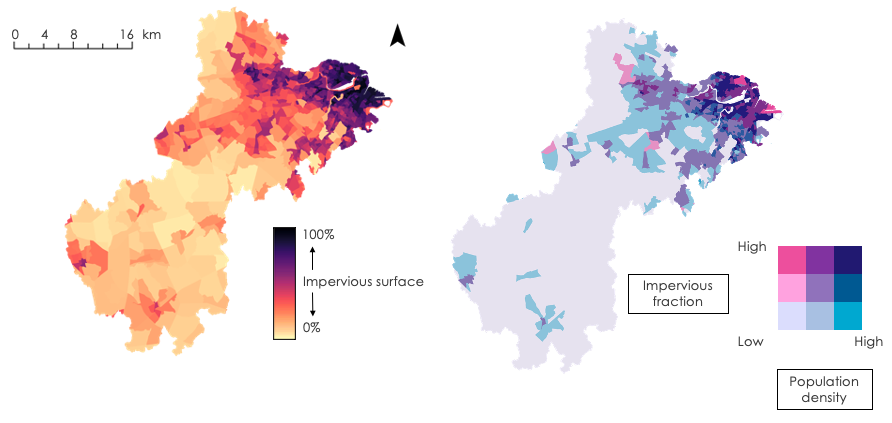


Figure 3: Impervious surface fraction per census block group (left) and estimated watershed degradation based on impervious fraction and population density (right).

***4.2 Feasibility of Earth Observation Data for Flood Detection in the Charles River Watershed***

The team aimed to develop an inventory of flood event extents in the Charles River watershed by visually comparing Landsat 5, Landsat 8, Sentinel-2, and classified Sentinel-1 images on known flood dates to non-flooded base imagery. However, the spatial and temporal resolution of Landsat 5 and Landsat 8 satellite imagery (30 m, 16-day repeat) was insufficient to capture the small-scale, localized flooding events recorded in the NOAA Storm Events database. While the spatial and temporal resolution of Sentinel-1 and Sentinel-2 imagery (10 m, 3-5-day repeat) was improved over the Landsat series, it was still insufficient to capture the reported flood events. This was likely due to the mixed pixel effect, which is common in remotely sensed imagery of urban and suburban areas due to their highly heterogeneous land cover at the 10 m or 30 m resolution. A mixed pixel means that a single pixel contains multiple land cover types, e.g., water, trees, and impervious surface. As such, one may not be able to visually identify a flood because the spectral signature was obscured by the other land cover types in the pixel.

***4.3 Flood Inventory Map***

All Sentinel-1 pixels were classified as water or non-water using a binary supervised classification methodology in GEE. Although the supervised classifier had a high training accuracy (>95%), there were still misclassifications and errors. Common misclassifications included classifying golf courses and other irrigated grass as water or classifying high-albedo urban areas as water, when in fact they are both non-water features. Sentinel-1 imagery also underwent speckle filtering to reduce noise, which also coarsens the spatial resolution of the imagery and makes it challenging to identify small-scale flooding events. In addition, some Landsat 5, Landsat 8, and Sentinel-2 imagery had to be filtered and removed due to cloud cover, snow, and ice, all of which hinder the detection of flooding in optical imagery and introduces additional uncertainties.

***4.4 Flood Susceptibility Maps***

**Logistic Regression**

The output of the logistic regression gave the probability of flooding per pixel based on the values of the input flood conditioning factors (Figure 4; refer to Appendix A1). Applying Equation 6 to the entire study area yielded a map of flooding probability (refer to Appendix C2). These probabilities were divided into quantiles to create different levels of flood susceptibility: very low, low, moderate, high, and very high.

A picture containing timeline

Description automatically generated

Figure 4: Flood probability (left) and flood susceptibility (right) derived from MLR.

P = -0.467 + 0.363\*Soil Drainage – 0.184\*Slope + 0.108\*TWI – 1.351\*HAND

+ 0.947\*Surficial Materials – 0.561\*Elevation + 0.836\*Landcover (6)

**Ensemble method**

In general, the FRs for the flood conditioning factors aligned with the literature (refer to Appendix B1). Lower values of slope, elevation, and HAND had higher FRs while higher TWI values had higher FRs (refer to Appendix C1). Very poorly drained soils had the highest FR, as did early postglacial deposits and land cover classified as water. However, the low FRs for areas with impervious surfaces and grass/fields deviated from the literature. Tehrany et al. (2013) found that urban land use/land cover (mainly impervious surfaces) and grassland were among the factors with the highest FRs. This suggests the presence of a confounding factor in the data for impervious and grassland landcover, and for this reason, the team did not include the resulting map here.

**Random Forest**

Figure 5 shows the susceptibility map created by the team using the RF method and the previously calculated flood conditioning factors, with areas susceptible to flooding shown in blue. The team then overlaid the NOAA Storm Events Database with the susceptibility map. The team's map captured 68%, or roughly two-thirds, of the reported storm events, whereas the FEMA flood map only captured 33%, or one-third, of the reported flood events. It is important to note that this result was generated from a small selection of flood events from the NOAA Storm Events Database, which may have inaccuracies with regards to exact location and extent.



Figure 5: Flood susceptibility map derived from the RF model.

***4.5 Flood Vulnerability Maps***

In each of the individual vulnerable population maps, there was a greater proportion of these groups in the upper portion of the watershed (Figure 6). The non-white population was concentrated in the Boston metro area and the population in poverty was concentrated in pockets across southern Boston. There was a relatively small elderly population in the Boston metro area, except for a pocket in southwest Boston. The population in poverty was mainly concentrated in the Boston metro area. There was also a pocket of non-white and population in poverty in the Milford area of the watershed.

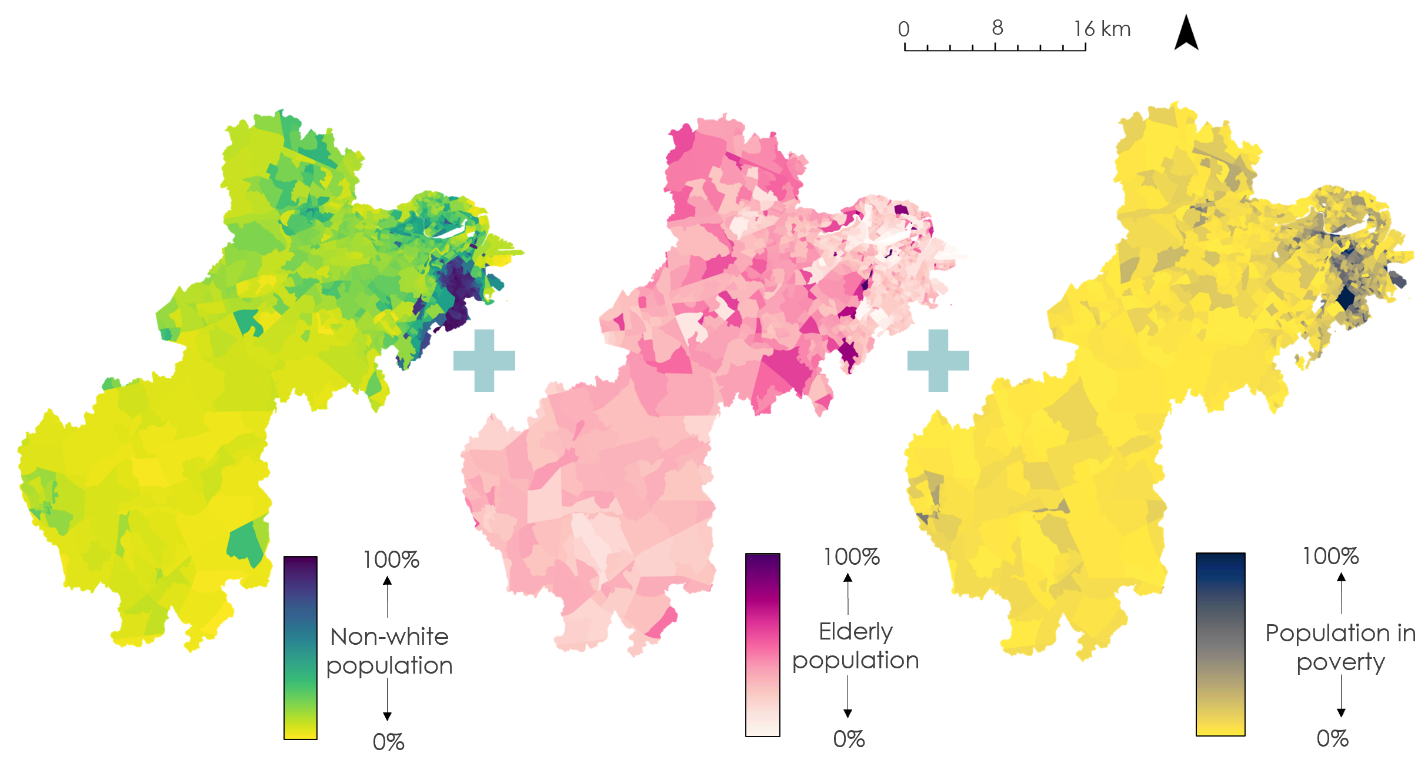


Figure 6: Flood vulnerability maps for the fraction of residents in each census block group that is non-white (left), elderly (center), or impoverished (right).

The FVI map shows the vulnerability index aggregated for the three vulnerable groups (Figure 7, left). As in the previous maps, there was a greater proportion of vulnerable populations in the upper watershed, and particularly across the Boston area, with higher vulnerability in South Boston. There was also a small pocket in the Milford area. The team developed a bivariate map using the susceptibility map and the vulnerability index (Figure 7, right). Areas with a high percent of pixels in the flood susceptibility zone and high vulnerability index were the most vulnerable to flood events. The areas with the highest flood vulnerability were the Allston and South End neighborhoods in Boston, as well as Milford, Cambridge, Dorchester, and Waltham.

Map

Description automatically generated

*Figure 7:* Flood vulnerability index map (left) and bivariate flood vulnerability map (right).

***4.6 Future Work Recommendations***

The authors recommend additional efforts to increase the strength of the findings. Including more socioeconomic and environmental data would further develop the vulnerability analysis. This could include the percentage of people living on the first floor in areas susceptible to flood, those without access to transportation, and the percent of people within close proximity to road networks for evacuation. The team recommends incorporating precipitation data to determine if observed flood events are more related to development patterns, changes in precipitation, or both, as well as incorporating long-term climate trends to understand if the future risk of extreme precipitation events in the study area. An ideal data source for this is the Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG), from NASA’s Global Precipitation Measurement Mission.

Mapping traditional stormwater conveyance, green infrastructure, and urban tree canopy could provide information regarding the role these are playing in flood susceptibility from a built environment perspective. This could be especially important in identifying why many of the reported floods were outside of the FEMA 100-year and 500-year flood plains. The scale could also be reduced to focus on identified areas identified with high vulnerability, including Milford, Cambridge, Dorchester, Waltham, and the Allston and South End neighborhoods in Boston. Finally, future work would be improved by incorporating higher resolution remote sensing imagery to understand the extent of flood events, as it becomes available. The NASA-ISRO SAR Mission (NISAR), scheduled to launch in 2022, would be important Earth-observing data to include in a future study. NISAR will have L-band and S-band SAR with a repeat cycle of 12 days with a spatial resolution of 3-10 meters, which is ideal for studying localized flooding and small water level changes.

**5. Conclusions**

Remote sensing is a very powerful tool that provides accessible, cost-efficient data. However, successful remote sensing results can require specific conditions and parameters. The use of NASA EOs for this particular project was found to not be feasible for detecting localized flooding in the Charles River watershed. This was due to the local scale of flooding, temporal and spatial resolution of each satellite, flood events used from NOAA’s Storm Event Database, and timing for this project. Using an alternative method, the team’s flood susceptibility map overlapped with 68% of the reported NOAA flood events, whereas the FEMA flood maps overlapped with 33%. In addition, the watershed degradation maps show that the highest potential for watershed degradation occurred near the mouth of the Charles River near downtown Boston. Combining these conclusions led the team to identify that the areas with the highest flood vulnerability were in Boston, Cambridge, Dorchester, Milford, and Waltham. These conclusions provided our partners with a better understanding of where and how impervious surface development can impact communities in the Charles River watershed, as well as which areas are the most susceptible and vulnerable to flooding. In addition, our findings were presented in an easily accessible and comprehensible format, via an ArcGIS Online StoryMap. As a result of this project, our partners, along with other governments and organizations in the watershed, will have enhanced data for decision making to better address community risk regarding these issues.

# 6. Acknowledgments

The authors would like to thank our municipal partners for their collaboration in this study. This includes Marianne Iarossi and Jillian Wilson-Martin from the Town of Natick, Danielle Perry from the Massachusetts Audubon Society, and Julie Wood from the Charles River Watershed Association. The team would also like to acknowledge Dr. Cedric Fichot of Boston University, Dr. Kenton Ross of NASA Langley Research Center, and Celeste Gambino of NASA DEVELOP for their guidance and mentorship throughout this project.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract NNL16AA05C.

This material contains modified Copernicus Sentinel data (2014-2020), processed by ESA.

# 

# 7. Glossary

**Earth observations** – Satellites and sensors that collect information about the Earth’s physical, chemical, and biological systems over space and time

**Frequency ratio** – Bivariate statistical technique that focuses on the individual correlations between each conditioning factor and the flood inventory map; defined as the ratio of the probability of a flood event occurrence to the probability of a flood event non-occurrence for given flood conditioning factors and their classes

**Susceptibility** – The likelihood of a dangerous event occurring in an area on the basis of local terrain conditions

**Vulnerability** – A measure of sensitivity of individuals, groups, or communities to their flood susceptibility

# 8. References

American Community Survey. (2018). Ratio of Income to Poverty Level in the past 12 Months. United States Census Bureau. Table ID: C17002. Accessed on 10/08/2020. <https://data.census.gov/cedsci/table?t=Income%20and%20Poverty&g=0500000US25009.150000,25017.150000,25025.150000,25027.150000&tid=ACSDT5Y2018.C17002&hidePreview=false>

Billa, L., Shattri, M., Rodzi Mahmud, A., & Halim Ghazali, A. (2006). Comprehensive planning and the role of SDSS in flood disaster management in Malaysia. *Disaster Prevention and Management: An International Journal, 15*(2), 233–240. <https://doi.org/10.1108/09653560610659775>

Bukvic, A., Gohlke, J., Borate, A., &amp; Suggs, J. (2018). Aging in flood-prone coastal areas: Discerning the health and well-being risk for older residents. *International Journal of Environmental Research and Public Health, 15*(12), 2900. <https://doi.org/10.3390/ijerph15122900>

Cheng, C., Yang, Y. C. E., Ryan, R., Yu, Q., & Brabec, E. (2017). Assessing climate change-induced flooding mitigation for adaptation in Boston’s Charles River watershed, USA. *Landscape and Urban Planning*, *167*, 25–36. <https://doi.org/10.1016/j.landurbplan.2017.05.019>

Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., Wang, X., Bian, H., Zhang, S., Pradhan, B., & Ahmad, B. Bin. (2020). Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. *Science of the Total Environment*, *701*, 134979. <https://doi.org/10.1016/j.scitotenv.2019.134979>

de Santana, F. B., de Souza, A. M., & Poppi, R. J. (2018). Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *191*, 454–462. <https://doi.org/10.1016/j.saa.2017.10.052>

Douglas, E. M., Kirshen, P. H., Paolisso, M., Watson, C., Wiggin, J., Enrici, A., & Ruth, M. (2012). Coastal flooding, climate change and environmental justice: Identifying obstacles and incentives for adaptation in two metropolitan Boston Massachusetts communities. *Mitigation and Adaptation Strategies for Global Change*, *17*(5), 537–562. <https://doi.org/10.1007/s11027-011-9340-8>

Easterling, D.R., K.E. Kunkel, J.R. Arnold, T. Knutson, A.N. LeGrande, L.R. Leung, R.S. Vose, D.E. Waliser, and M.F. Wehner. (2017). Precipitation change in the United States. Climate Science Special Report: Fourth National Climate Assessment, Volume I. U.S. Global Change Research Program, Washington, DC, USA, 207-230. <http://dx.doi.org/10.7930/J0H993CC>

Federal Emergency Management Agency. (2005). National Flood Insurance Program Floodplain Management Requirements: A Study Guide and Desk Reference for Local Officials. FEMA 480. Accessed on 11/16/20. <https://www.fema.gov/media-library-data/1481032638839-48ec3cc10cf62a791ab44ecc0d49006e/FEMA_480_Complete_reduced_v7.pdf>

Fielding, J.L. (2017). Flood risk and inequalities between ethnic groups in the floodplains of England and Wales. *Disasters, 42*(1), 101-123. <https://doi.org/10.1111/disa.12230>

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., & Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change,* *3*(9), 816-821. <https://doi.org/10.1038/nclimate1911>

Huang, X., Tan, H., Zhou, J., Yang, T., Benjamin, A., Wen, S.W., Li, S., Liu, A., Li, X., Fen, S., & Li, X (2008). Flood hazard in Hunan province of China: An economic loss analysis. *Natural Hazards, 47*(1), 65–73. https://doi.org/10.1007/s11069-007-9197-z

Huong, H. T. L., & Pathirana, A. (2011) Urbanization and climate change impacts on future urban flooding in Can Tho City, Vietnam, *Hydrology and Earth System Sciences Discussions, 17*, 379–394. <https://doi.org/10.5194/hessd-8-10781-2011>

Hallegatte, S., Green, C., Nicholls, R.J., & Corfee-Morlot, J. (2013). Future flood losses in major coastal cities. *Nature Climate Change, 3*(9), 802–806. <http://dx.doi.org/10.1038/nclimate1979>

IPCC. (2007). Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 976pp.

Lee, S., Hwang, J., & Park, I. (2013). Application of data-driven evidential belief functions to landslide susceptibility mapping in Jinbu, Korea. *CATENA*, *100*, 15–30. <https://doi.org/10.1016/j.catena.2012.07.014>

Liu, Y. B., & De Smedt, F. (2005). Flood modeling for complex terrain using GIS and remote sensed information. *Water Resources Management*, *19*(5), 605–624. <https://doi.org/10.1007/s11269-005-6808-x>

MassGIS. (2019). *2016 Land Cover/Land Use* [data set]. Massachusetts Bureau of Geographic Information. Identifier: 9b5bc25a-9c85-4948-ac6d-3f4d23a0fecd. Accessed 10/08/2020. <https://docs.digital.mass.gov/dataset/massgis-data-2016-land-coverland-use>

MassGIS. (2020). *Data Layers from the 2010 US Census* [data set]. Massachusetts Bureau of Geographic Information. Identifier: f1df48e8-cf45-478f-956f-7792c8381715. Accessed on 10/08/2020. <https://docs.digital.mass.gov/dataset/massgis-data-datalayers-2010-us-census>

MassGIS. (2020). *MassGIS Data Layers* [data set]. Massachusetts Bureau of Geographic Information. Identifier: fc73d74e-d32d-42db-a401-e5032c851425. Accessed on 10/08/2020. <https://docs.digital.mass.gov/dataset/massgis-data-layers#inf>

Quiroz, J. C., Mariun, N., Mehrjou, M. R., Izadi, M., Misron, N., & Mohd Radzi, M. A. (2018). Fault detection of broken rotor bar in LS-PMSM using random forests. *Measurement: Journal of the International Measurement Confederation*, *116*, 273–280. <https://doi.org/10.1016/j.measurement.2017.11.004>

Randolph, J. (2012). Chapter 10 Landscape and Urban Ecology, Urban Forestry, and Wetlands. In *Environmental land use planning and management* (Second ed., p. 345). Washington: Island Press.

Samuels, P., Gouldby, B., Klihn F., Messner, F., Van Os, A., & Sayers, P. (2009). Integrated Flood Risk Analysis and Management Methodologies: Language of Flood Risk. Report Number T32-04-01. Second Edition. Accessed 10/05/2020. <http://www.floodsite.net/html/partner_area/project_docs/T32_04_01_FLOODsite_Language_of_Risk_D32_2_v5_2_P1.pdf>

Santangelo, N., Santo, A., Di Crescenzo, G., Foscari, G., Liuzza, V., Sciarrotta, S., & Scorpio, V. (2011). Flood susceptibility assessment in a highly urbanized alluvial fan: The case study of Sala Consilina (southern Italy). *Natural Hazards and Earth System Science*, *11*(10), 2765–2780. <https://doi.org/10.5194/nhess-11-2765-2011>

Suarez, P., Anderson, W., Mahal, V., & Lakshmanan, T. R. (2005). Impacts of flooding and climate change on urban transportation: A systemwide performance assessment of the Boston Metro Area. *Transportation Research Part D: Transport and Environment*, *10*(3), 231–244. <https://doi.org/10.1016/j.trd.2005.04.007>

Tabari, H. Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports, 10*, 13768 (2020). <https://doi.org/10.1038/s41598-020-70816-2>

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2015). Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. *Stochastic Environmental Research and Risk Assessment, 29*(4), 1149–1165. <https://doi.org/10.1007/s00477-015-1021-9>

Tehrany, M. S., Kumar, L., & Shabani, F. (2019). A novel GIS-based ensemble technique for flood susceptibility mapping using evidential belief function and support vector machine: Brisbane, Australia. *PeerJ*, *7*, e7653. <https://doi.org/10.7717/peerj.7653>

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2013). Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology*, *504*, 69–79. <https://doi.org/10.1016/j.jhydrol.2013.09.034>

Tsagkrasoulis, D., & Montana, G. (2018). Random forest regression for manifold-valued responses. *Pattern Recognition Letters*, *101*, 6–13. <https://doi.org/10.1016/j.patrec.2017.11.008>

Ward, P.J., H.C. Winsemius, S. Kuzma, M.F.P. Bierkens, A. Bouwman, H. de Moel, A. Díaz Loaiza, et al. 2020. “Aqueduct Floods Methodology.” Technical Note. Washington, D.C.: World Resources Institute. Available online at: [ww.wri.org/publication/aqueduct-floods-methodology](http://ttp://www.wri.org/publication/aqueduct-floods-methodology%22)

# 9. Appendices

***Appendix A: Flood Conditioning Factors***

Map

Description automatically generated*Figure A1:* Maps of all the flood conditioning factors.

*Flood conditioning factor maps with corresponding quantiles: (a) Land Cover; (b) Surficial Materials; (c) Soil Drainage; (d) Elevation; (e) Slope; (f) TWI; (g) HAND*

***Appendix B: Flood Conditioning Factor Frequency Ratio Table***

Table B1

*Flood conditioning factors and corresponding quantiles with calculated frequency ratios*

|  |  |  |
| --- | --- | --- |
| **Flood Conditioning Factor** | **Class** | **Frequency Ratio** |
| Soil Drainage | Excessively drained | 0.655 |
| Somewhat excessively drained | 0.293 |
| Well drained | 0.151 |
| Moderately well drained | 0.754 |
| Poorly drained | 1.909 |
| Very poorly drained | 4.552 |
| Landcover | Impervious | 0.303 |
| Grass/field | 0.234 |
| Forest/scrub | 0.399 |
| Wetland | 3.555 |
| Bare land | 2.374 |
| Water | 5.316 |
| Surficial | Glacial till and bedrock | 0.167 |
| Glacial stratified deposits | 0.845 |
| Postglacial deposits | 4.015 |
| Early postglacial deposits | 4.285 |
| Slope | ≤ 0.856 | 2.772 |
| ≤ 1.9988 | 0.99 |
| ≤ 3.7120 | 0.475 |
| ≤ 6.8530 | 0.334 |
| ≤ 72.8128 | 0.356 |
| TWI | ≤ 4.8738 | 0.375 |
| ≤ 5.8516 | 0.485 |
| ≤ 6.7315 | 0.686 |
| ≤ 8.0026 | 1.075 |
| ≤ 25.4061 | 2.373 |
| Elevation | ≤ 19.4520 | 1.412 |
| ≤ 47.6132 | 1.813 |
| ≤ 69.7399 | 0.762 |
| ≤ 98.9069 | 0.446 |
| ≤ 180.3733 | 0.085 |
| HAND | 0 | 4.832 |
| ≤ 4.2418 | 1.634 |
| ≤ 9.8976 | 0.138 |
| ≤ 19.0882 | 0.027 |
| ≤ 180.2778 | 0.007 |

***Appendix C: Weighted and Unweighted Logistic Regression Coefficients***

Table C1

*Flood conditioning factors and their corresponding highest frequency ratios*

|  |  |  |
| --- | --- | --- |
| **Flood conditioning factor** | **Class** | **Frequency ratio** |
| Soil drainage | Very poorly drained | 4.552 |
| Landcover | Water | 5.316 |
| Surficial | Early postglacial deposits | 4.285 |
| Slope | ≤ 0.8566 | 2.772 |
| TWI | ≥ 25.4061 | 2.373 |
| Elevation | ≤ 47.6132 | 1.813 |
| HAND | 0 | 4.832 |

Table C2

*Logistic regression coefficient and weighted logistic regression coefficients for each flood conditioning factor and the equation intercept*

|  |  |  |
| --- | --- | --- |
| **Flood conditioning factor** | **Logistic regression coefficient** | **Weighted logistic regression coefficient** |
| Intercept | -0.467 | -0.467 |
| Soil drainage | 0.363 | 0.044 |
| Slope | -0.184 | -0.037 |
| TWI | 0.108 | -0.022 |
| HAND | -1.351 | -0.203 |
| Surficial materials | 0.947 | 0.102 |
| Elevation | -0.561 | -0.124 |
| Landcover | 0.836 | 0.069 |