Tutorial: Flood Vulnerability Mapping

NASA DEVELOP Charles River Watershed Water Resources – Fall 2020

Trista Brophy, Willow Coleman, Anna Garik, Will Peters

Table of Contents

Preface:

The DEVELOP Charles River Watershed Water Resources team did not complete the NASA software release process on the Google Earth Engine code written that derived the data used on this project. As a result, in this document, we are unable to share any of the code the team created to comply with NASA software release regulations. The following tutorial documents the process for users to follow and write their own code. All screenshots, code links and descriptors in this document are from publicly available code and information provided by Google Earth Engine Tutorials and other open-source sites, and are cited accordingly. Please note there may be more than one way to correctly derive the same data through Google Earth Engine or other coding languages and software.

Part 1: Calculate Individual Flood Vulnerability Factors

The team developed a flood vulnerability index (FVI) to represent the amount of vulnerability in each census block group. The first step in this process is to calculate the fraction, or percent, of each of the three vulnerability factors. These factors include 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 fraction of residents living below the poverty line per census block group.

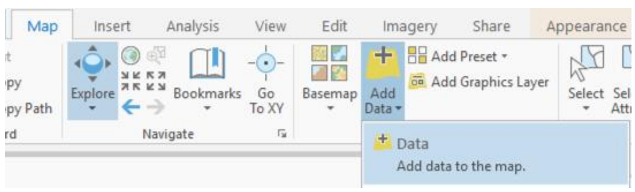
Step 1.1 – Download 2010 U.S. Census block group boundaries

Download the Census Blocks, Block Groups and Tracts shapefiles (~82 MB) from the following link:

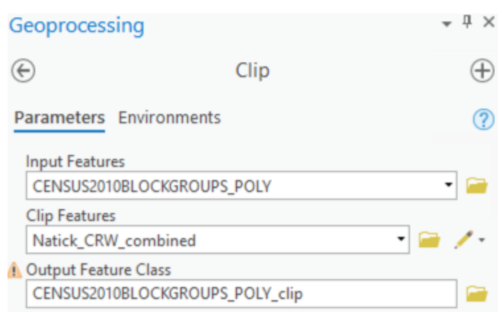
https://docs.digital.mass.gov/dataset/massgis-data-datalayers-2010-us-census

The folder CENSUS2010_BLK_BG_TRCT_SHP includes shapefiles of all census blocks, block groups, and tracts for Massachusetts.

Add the shapefile CENSUS2010BLOCKGROUP_POLYS.shp to the Map by using "Add Data" under the Map tab and selecting this .shp file in the CENSUS2010_BLK_BG_TRCT_SHP folder. It will be added to the map as CENSUS2010BLOCKGROUP_POLYS.



This map covers the entire state of Massachusetts, so it needs to be clipped to the study area (Charles River watershed + Natick, MA). Add a shapefile of the study area by using "Add Data" under the Map tab and selecting the desired .shp file. It will be added to the map as Natick_CRW_combined.



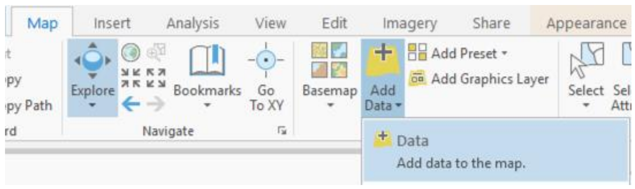
Step 1.2 – Download Census data

Download SF1 data dBase tables (~32 MB) from the following link:

https://docs.digital.mass.gov/dataset/massgis-data-datalayers-2010-us-census

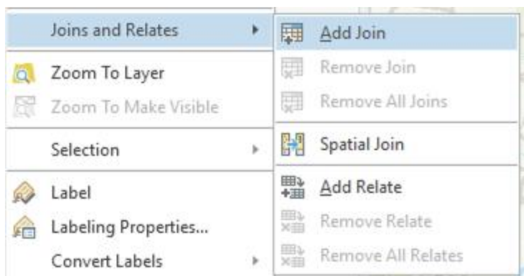
All data analysis will be performed at the census block group scale.

Add CEN2010_BG_SF1_POP_AGE_GENDER.dbf and CEN2010_BG_SF1_POP_RACE.dbf to the Contents panel as Standalone Tables by using Add Data under the Map tab and navigating the downloaded files from MassGIS.

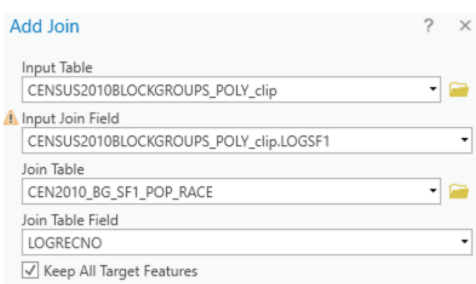


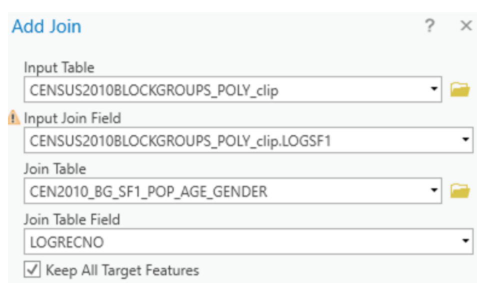
Step 1.3 – Join data tables

Right-click on CENSUS2010BLOCKGROUPS_POLY_clip in the Contents panel and navigate to Add Join under Joins and Relates.



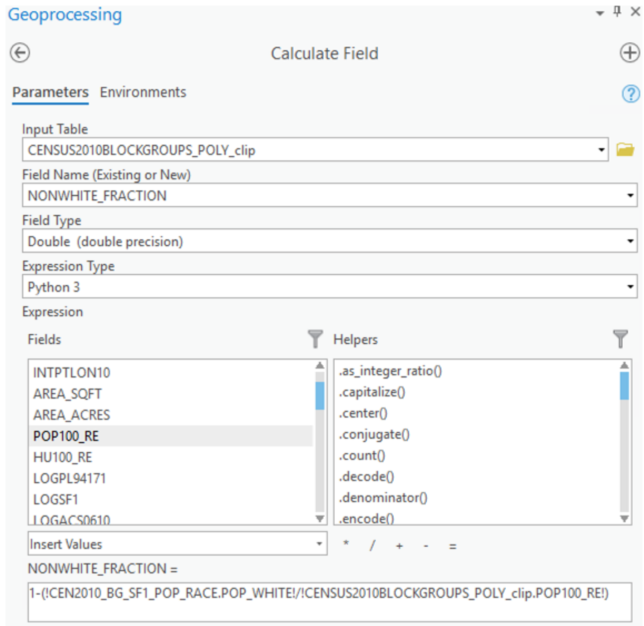
Perform two joins to join both SF1 database tables to the census block group data. For the first join, the input table is CENSUSBLOCKGROUPS_POLY_clip with join field LOGSF1, and this will be joined with CEN2010_BG_SF1_POP_RACE with join field LOGRECNO. For the second join, the input table is CENSUSBLOCKGROUPS_POLY_clip with join field LOGSF1, and this will be joined with CEN2010_BG_SF1_POP_AGE_GENDER with join field LOGRECNO.



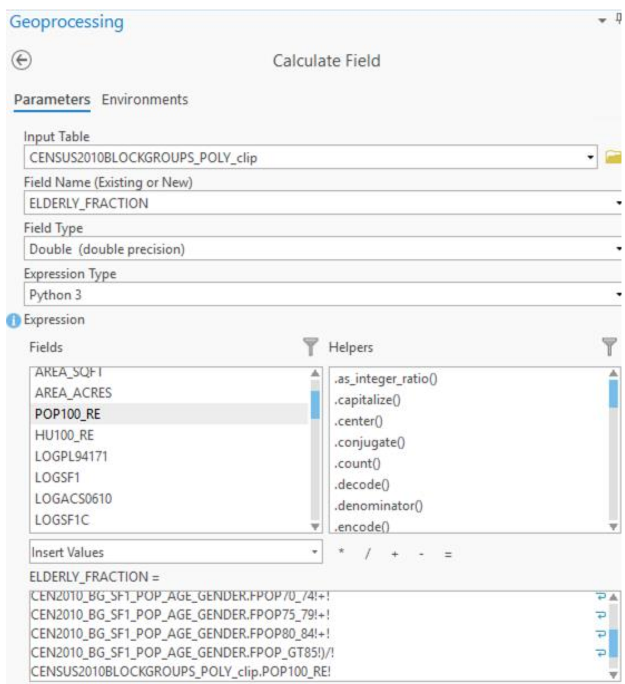


Step 1.4 – Calculate flood vulnerability factors from 2010 U.S. Census data

The "Calculate Field" function can be used to calculate the fraction of non-white residents per census block group and the fraction of elderly residents (age 65+) per census block group.



WHITE_FRACTION = 1-(!CEN2010_BG_SF1_POP_RACE.POP_WHITE!/!CENSUS2010BLOCKGROUPS_POLY_clip.POP100_RE!)



This material is based upon work supported by NASA through contract NNL16AA05C. Any mention of a commercial product, service, or activity in this material does not constitute NASA endorsement. 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 and partner organizations.

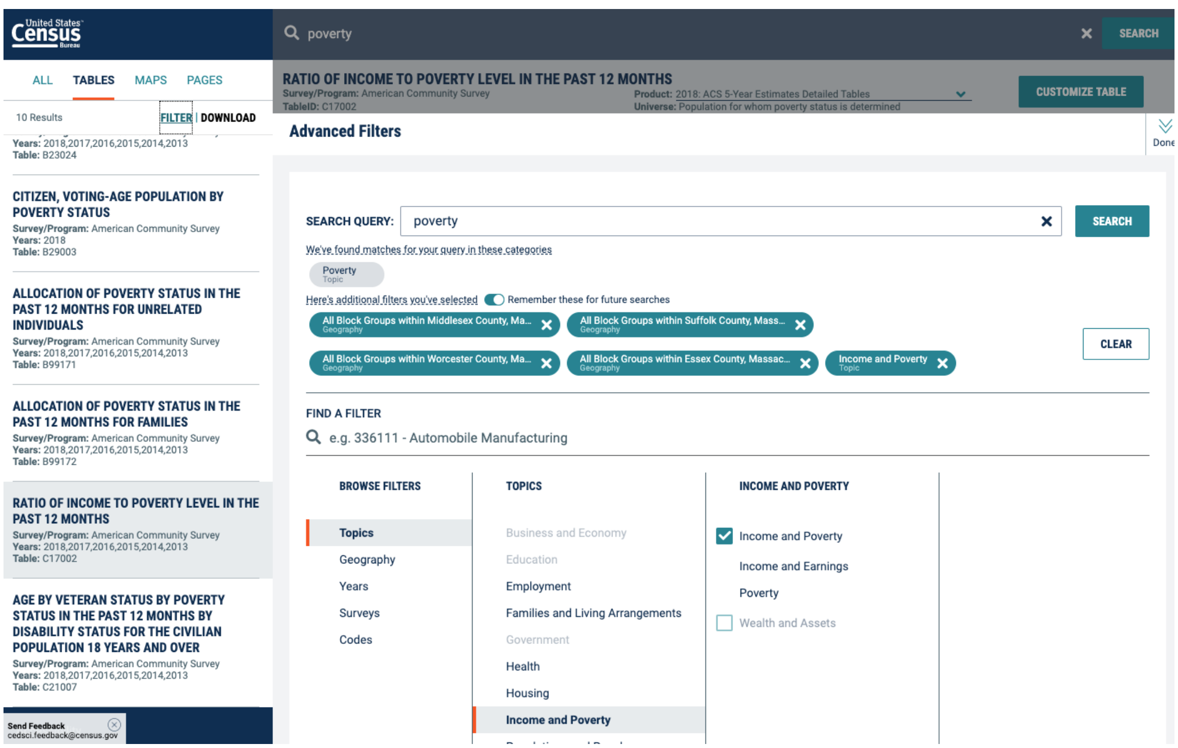
ELDERLY_FRARCTION = (!CEN2010_BG_SF1_POP_AGE_GENDER.MPOP65_66!+!CEN2010_BG_SF1_POP_AGE_GENDER.MPOP67_69!+!CEN2010_BG_SF1_POP_AGE_GENDER.MPOP70_74!+!CEN2010_BG_SF1_POP_AGE_GENDER.MPOP75_79!+!CEN2010_BG_SF1_POP_AGE_GENDER.MPOP80_84!+!CEN2010_BG_SF1_POP_AGE_GENDER.MPOP_GT85!+!CEN2010_BG_SF1_POP_AGE_GENDER.FPOP65_66!+!CEN2010_BG_SF1_POP_AGE_GENDER.FPOP67_69!+!CEN2010_BG_SF1_POP_AGE_GENDER.FPOP70_74!+!CEN2010_BG_SF1_POP_AGE_GENDER.FPOP75_79!+!CEN2010_BG_SF1_POP_AGE_GENDER.FPOP80_84!+!CEN2010_BG_SF1_POP_AGE_GENDER.FPOP_GT85!)/!CENSUS2010BLOCKGROUPS_POLY_clip.POP100_RE!

Step 1.5 – Calculate flood vulnerability factors from 2018 American Community Survey data

Additional U.S. Census Bureau data can be downloaded from this link:

https://data.census.gov/cedsci/

Using the following filters, download a CSV of the table (TableID C17002) containing the Ratio of Income to Poverty Level in the Past 12 Months for Middlesex County, Suffolk County, Worcester County, and Essex County, MA at the Block Group scale.



Open the largest CSV file in Excel and insert a new column to the right of column GEO_ID that contains the last 12 characters of GEO_ID.

A	B	C	D
GEO_ID		NAME	C17002_001E
id		Geographic Area Name	Estimate!!Total
1500000US250173354002		Block Group 2, Census Tract 3354, Middlesex County, Massachusetts	2276
1500000US250173354004		Block Group 4, Census Tract 3354, Middlesex County, Massachusetts	696
1500000US250173354001		Block Group 1, Census Tract 3354, Middlesex County, Massachusetts	677
1500000US250173364012		Block Group 2, Census Tract 3364.01, Middlesex County, Massachusetts	2097
1500000US250173364013		Block Group 3, Census Tract 3364.01, Middlesex County, Massachusetts	2036
1500000US250173397002		Block Group 2, Census Tract 3397, Middlesex County, Massachusetts	1438
1500000US250173397003		Block Group 3, Census Tract 3397, Middlesex County, Massachusetts	984
1500000US250173523001		Block Group 1, Census Tract 3523, Middlesex County, Massachusetts	2700
1500000US250173532002		Block Group 2, Census Tract 3532, Middlesex County, Massachusetts	1141
1500000US250173532001		Block Group 1, Census Tract 3532, Middlesex County, Massachusetts	1639

GEO_ID	ClippedID	NAME
id		Geographic Area Name
1500000US250173354002	250173354002	Block Group 2, Census Tract 3354, Middlesex County, Massachusetts
1500000US250173354004	250173354004	Block Group 4, Census Tract 3354, Middlesex County, Massachusetts
1500000US250173354001	250173354001	Block Group 1, Census Tract 3354, Middlesex County, Massachusetts
1500000US250173364012	250173364012	Block Group 2, Census Tract 3364.01, Middlesex County, Massachusetts
1500000US250173364013	250173364013	Block Group 3, Census Tract 3364.01, Middlesex County, Massachusetts
1500000US250173397002	250173397002	Block Group 2, Census Tract 3397, Middlesex County, Massachusetts
1500000US250173397003	250173397003	Block Group 3, Census Tract 3397, Middlesex County, Massachusetts
1500000US250173523001	250173523001	Block Group 1, Census Tract 3523, Middlesex County, Massachusetts
1500000US250173532002	250173532002	Block Group 2, Census Tract 3532, Middlesex County, Massachusetts

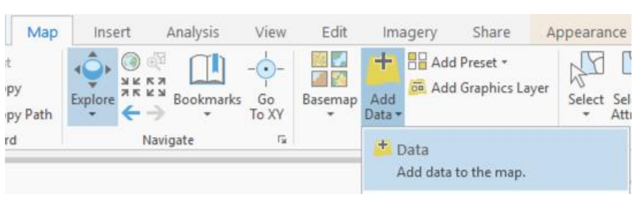
This table includes the ratio of income to poverty level, where ratios below 100% of the poverty level are considered impoverished.

Impoverished_Fraction = (Estimate!!Total!!Under .50 + Estimate!!Total!!.50 to .99)/ Estimate!!Total

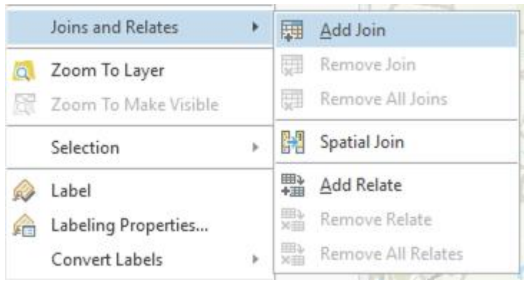
Delete all columns other than ClippedID and Impoverished_Fraction and export as acsPovertyData.csv.

A	B
ClippedID	Impoverished_Fraction
250173354002	0.025043937
250173354004	0.083333333
250173354001	0
250173364012	0.050548402
250173364013	0.06237721
250173397002	0.02433936
250173397003	0.024390244
250173523001	0.193703704
250173532002	0.247151621
250173532001	0.113483832
250173101001	0.286848073

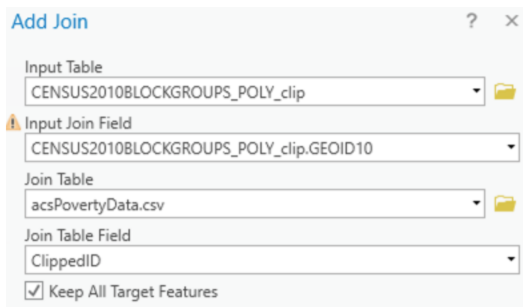
Add acsPovertyData.csv to the Contents panel as a Standalone Table by using Add Data under the Map tab and navigating the file you just edited in Excel.



Right-click on CENSUS2010BLOCKGROUPS_POLY_clip in the Contents panel and navigate to Add Join under Joins and Relates.



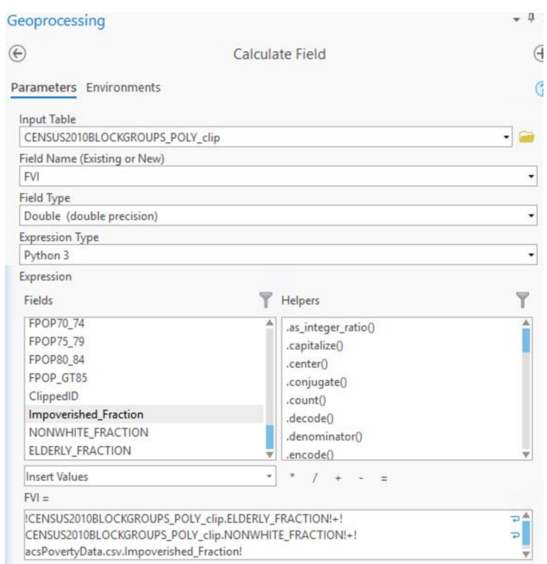
Perform a join to join the ACS Poverty data to the census block group data.



The next step was to sum the fractions of all three vulnerability factors to produce an aggregated FVI per census block group in order to understand where flood-vulnerable populations live in the Charles River watershed.

Step 1.6 – Adding a Flood Vulnerability Index field

The "Calculate Field" function can be used to calculate the FVI by summing Impoverished_Fraction, NONWHITE_FRACTION, and ELDERLY_FRACTION per census block group.



FVI =
!CENSUS2010BLOCKGROUPS_POLY_clip.ELDERLY_FRACTION!+!CENSUS2010BLOCKGROUPS_POLY_clip.NONWHITE_FRACTION!+!acsPovertyData.csv.Impoverished_Fraction!

Part 2: Flood Susceptibility Data Processing

Step 2.1 – Calculate flood susceptibility per census block group

The DEVELOP Charles River Watershed Water Resources team did not complete the NASA software release process on the Google Earth Engine code written that derived the data used on this project. As a result, in this document, we are unable to share any of the code the team created to comply with NASA software release regulations. The following tutorial documents the process for users to follow and write their own code. All screenshots, code links and descriptors in this document are from publicly available code and information provided by Google Earth Engine Tutorials and other open-source sites, and are cited accordingly. Please note there may be more than one way to correctly derive the same data through Google Earth Engine or other coding languages and software.

The calculation for determining flood susceptibility fraction per census block group in the study area is done in Google Earth Engine.

The following items need to be imported as Google Earth Engine assets:

1. floodMap.tif (raster of flood susceptibility, steps shown in Susceptibility Tutorial)
2. CENSUS2010BLOCKGROUPS_POLY_clip.shp (shapefile of census block groups)
3. Natick_CRW_combined.shp (shapefile of study area)

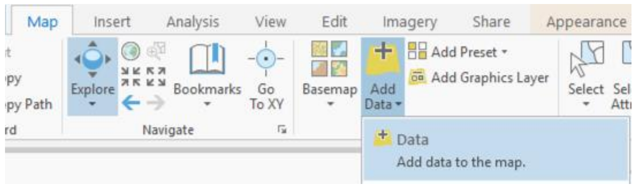
Using the reduceRegions() function with assets floodMap and CENSUS2010BLOCKGROUPS_POLY_clip, count the total number of pixels per census block group and the total number of flood pixels per census block group.

Flood susceptibility fraction is calculated and added as a feature to each census block group polygon by setting the variable Flood_Fraction to the number of pixels divided by the total number of pixels in a given block group polygon.

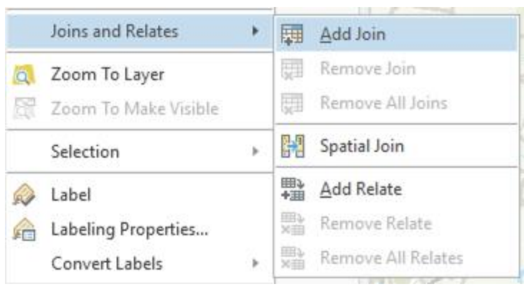
Select the variables 'Flood_Fraction', 'LOGSF1', and 'SHAPE_AREA' and export the Feature Collection to Google Drive as CENSUSBLOCKGROUP_FLOODFRACTION.csv using the function Export.table.toDrive().

Step 2.2. – Join flood susceptibility fraction map to flood vulnerability map

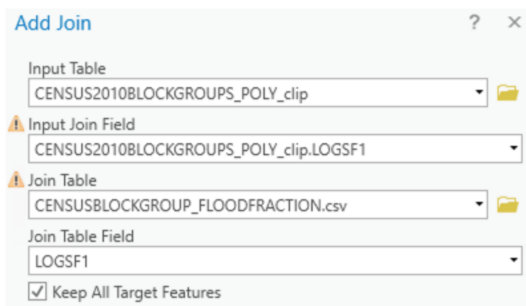
Add the table CENSUSBLOCKGROUP_FLOODFRACTION.csv to the map by using "Add Data" under the Map tab and selecting the file downloaded from Google Drive. It will be added to the map as a Standalone Table CENSUSBLOCKGROUP_FLOODFRACTION.csv.



Right-click on CENSUS2010BLOCKGROUPS_POLY_clip in the Contents panel and navigate to Add Join under Joins and Relates.



The input table is CENSUSBLOCKGROUPS_POLY_clip with join field LOGSF1, and this will be joined with the flood fraction per census block group data in the table CENSUSBLOCKGROUP_FLOODFRACTION.csv with join field LOGSF1.

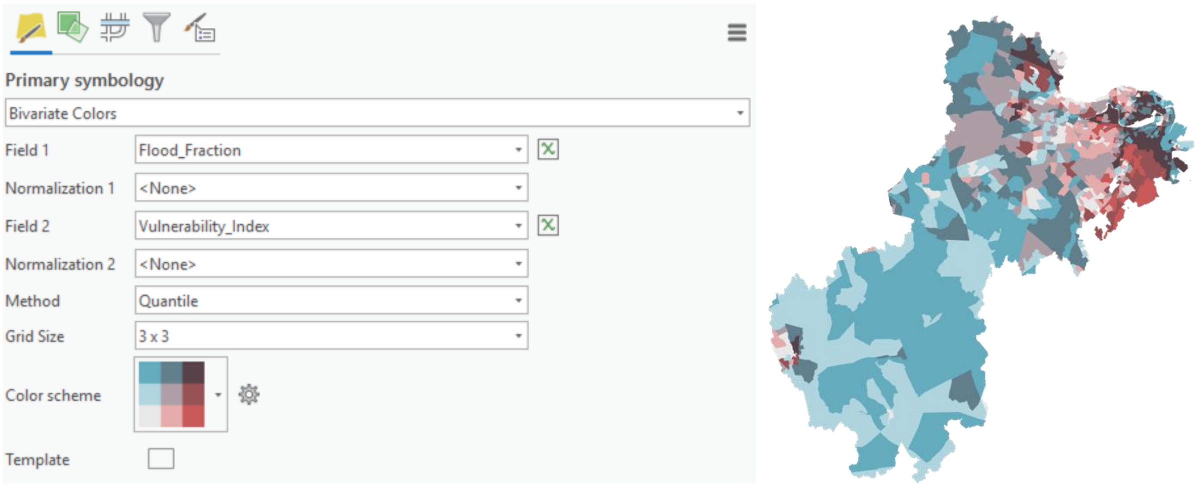


Now, there is a single attribute table in CENSUSBLOCKGROUPS_POLY_clip that contains information on the flood vulnerability factors, FVI, and flood susceptibility per census block group.

Part 3: Visualize flood vulnerability per census block group

Step 3.1. – Overlay FVI with flood susceptibility

The final step was to overlay the FVI per census block group with the flood susceptibility per census block group 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 (Vulnerability_Index) and a high fraction of area susceptible to flooding (Flood_Fraction) were considered the most flood vulnerable.



Flood vulnerability per census block group (map created by NASA DEVELOP)

Tutorial: Impervious Surface and Watershed Degradation Maps

NASA DEVELOP Charles River Watershed Water Resources – Fall 2020

Trista Brophy, Willow Coleman, Anna Garik, Will Peters

Table of Contents

Part 1: Impervious Surface Fraction

Step 1.1 – Download high-resolution land cover map

Download the Statewide File Geodatabase (~2.8 GB) from the MassGIS website from the following link:

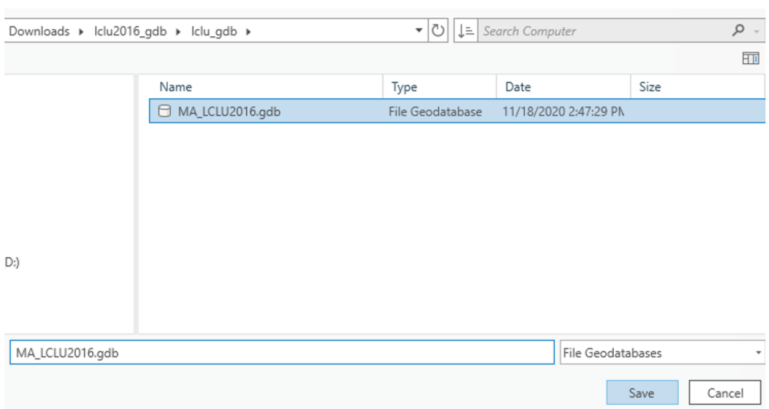
https://docs.digital.mass.gov/dataset/massgis-data-2016-land-coverland-use

This file is a vectorized 2016 land cover/land use map for the state of Massachusetts. First, you will need to unzip the downloaded file lclu2016_gdb.zip into lclu2016_gdb.

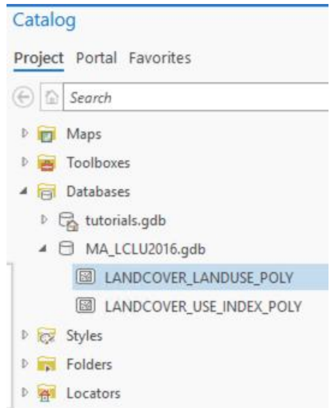
Step 1.2 – Process high-resolution land cover map

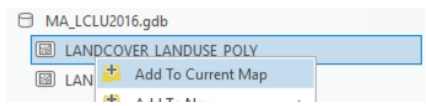
To add the land cover map in ArcGIS, add a New File Geodatabase by going to the Connections button in the Insert tab.

Navigate to the location of your geodatabase MA_LCLU2016.gdb located within the lclu2016_gdb folder, select it, and click Save.

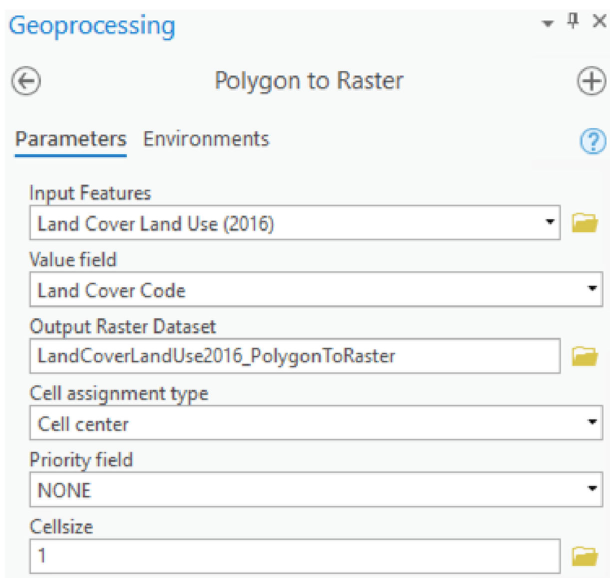


This process adds the geodatabase MA_LCLU2016.gdb to your catalog. Next, right click on LANDCOVER_LANDUSE_POLY in the catalog pane and select "Add to Current Map," which will add the statewide file to your ArcGIS map.



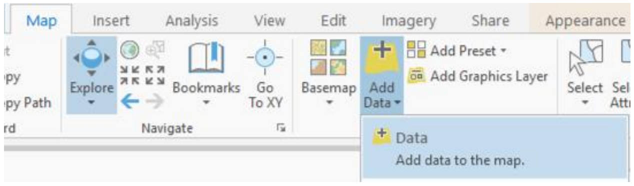


Next, convert the vectorized land cover map to a 1 m raster using the "Polygon to Raster" function. Note that this step may take some time because the geodatabase is large.

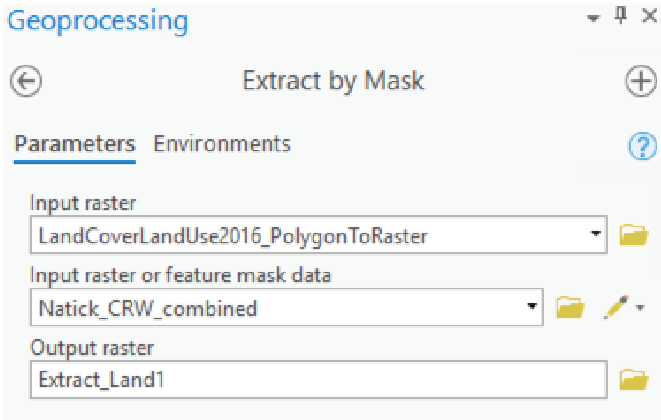


This map covers the entire state of Massachusetts, so it needs to be clipped to the study area (Charles River watershed + Natick, MA). Add a shapefile of the study area by using "Add Data" under the Map tab and selecting the desired .shp file. It will be added to the map as Natick_CRW_combined.

This material is based upon work supported by NASA through contract NNL16AA05C. Any mention of a commercial product, service, or activity in this material does not constitute NASA endorsement. 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 and partner organizations.



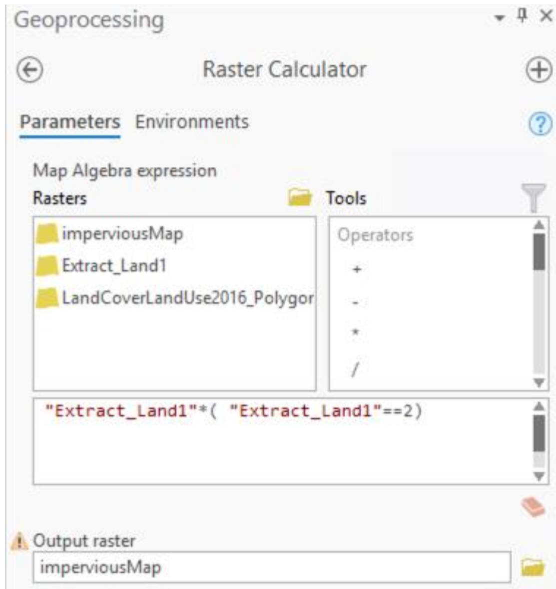
The last step of preprocessing the high-resolution land cover map is clipping it to the study area using the "Extract by Mask" function.



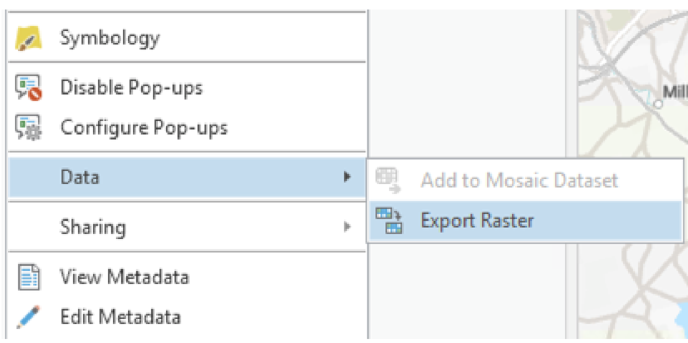
Step 1.3 – Create high-resolution impervious surface cover map

Next, use the "Raster Calculator" function to calculate a binary map that shows impervious surface fraction as Land Cover Code = 2 and all other land cover types as Land Cover Code = 0. The function is "Extract_Land1" * ("Extract_Land1" == 2), where 2 represents all pixels that classify as impervious surface and Extract_Land1 is the output from Step 1.2.

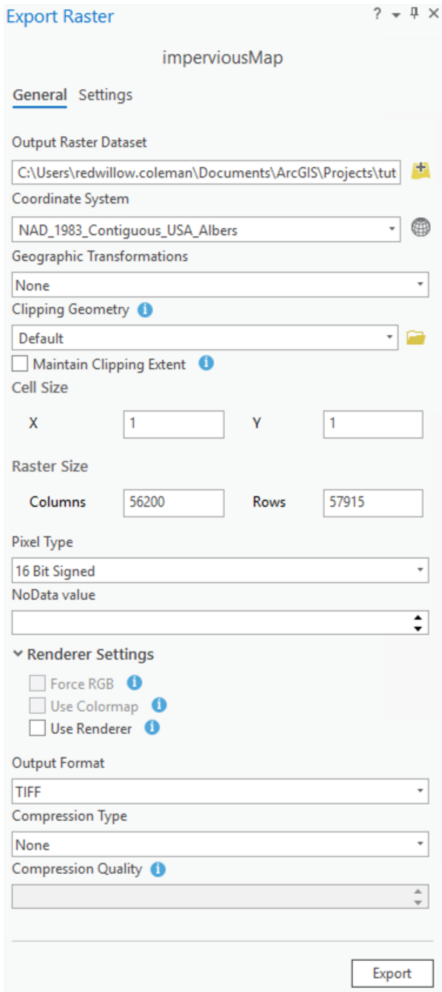
This material is based upon work supported by NASA through contract NNL16AA05C. Any mention of a commercial product, service, or activity in this material does not constitute NASA endorsement. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration and partner organizations.



If you want to complete Step 1.5 in Google Earth Engine, imperviousMap will need to be exported as a GeoTIFF file. Right-click on imperviousMap in the Contents panel in ArcGIS and navigate to Export Raster.



The following settings will export imperviousMap as a 1 m raster file.



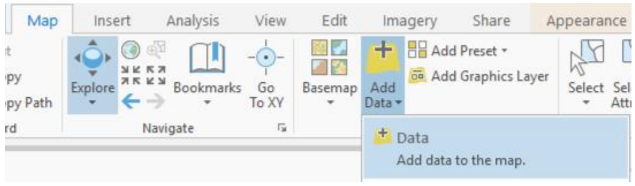
Step 1.4 – Download 2010 U.S. Census block group boundaries

Download the Census Blocks, Block Groups and Tracts shapefiles (~82 MB) from the following link:

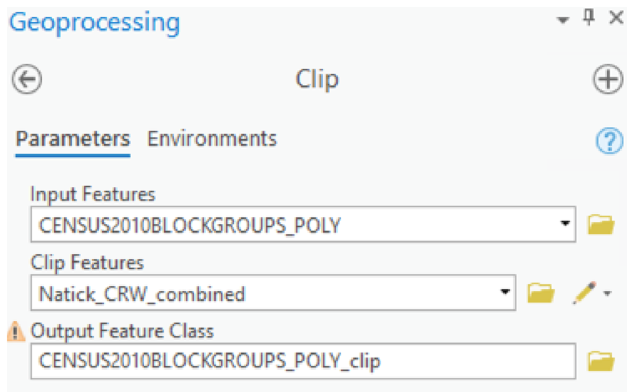
https://docs.digital.mass.gov/dataset/massgis-data-datalayers-2010-us-census

The folder CENSUS2010_BLK_BG_TRCT_SHP includes shapefiles of all census blocks, block groups, and tracts for Massachusetts.

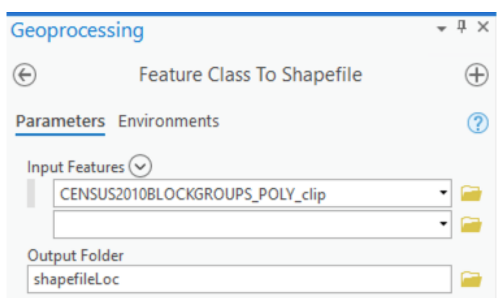
Add the shapefile CENSUS2010BLOCKGROUP_POLYS.shp to the Map by using “Add Data” under the Map tab and selecting this .shp file in the CENSUS2010_BLK_BG_TRCT_SHP folder. It will be added to the map as CENSUS2010BLOCKGROUP_POLYS.

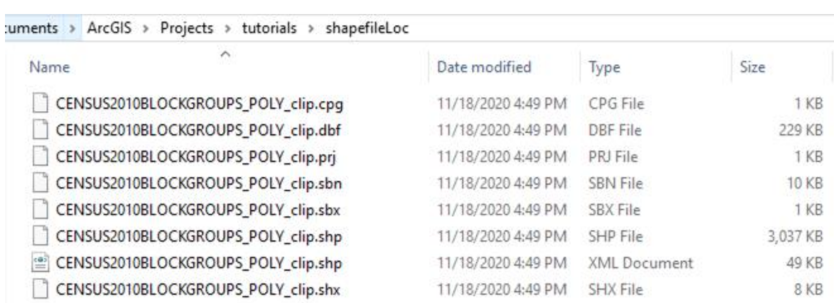


Again, this shapefile covers all of Massachusetts, so it will be necessary to clip it to the desired study area.



If you want to complete Step 1.5 in Google Earth Engine, CENSUS2010BLOCKGROUPS_POLY_clip will need to be exported as a shapefile using the function "Feature Class to Shapefile." All shapefile files will be located in the output folder after running the function.





Step 1.5 – Calculate impervious surface fraction per census block group

The DEVELOP Charles River Watershed Water Resources team did not complete the NASA software release process on the Google Earth Engine code written that derived the data used on this project. As a result, in this document, we are unable to share any of the code the team created to comply with NASA software release regulations. The following tutorial documents the process for users to follow and write their own code. All screenshots, code links and descriptors in this document are from publicly available code and information provided by Google Earth Engine Tutorials and other open-source sites, and are cited accordingly. Please note there may be more than one way to correctly derive the same data through Google Earth Engine or other coding languages and software.

The calculation for determining impervious surface fraction per census block group in the study area is done in Google Earth Engine.

The following items need to be imported as Google Earth Engine assets:

1. imperviousMap.tif (raster of binary impervious surface cover map)
2. CENSUS2010BLOCKGROUPS_POLY_clip.shp (shapefile of census block groups)
3. Natick_CRW_combined.shp (shapefile of study area)

Using the reduceRegions() function with assets imperviousMap and CENSUS2010BLOCKGROUPS_POLY_clip, count the total number of pixels per census block group and the total number of impervious pixels per census block group.

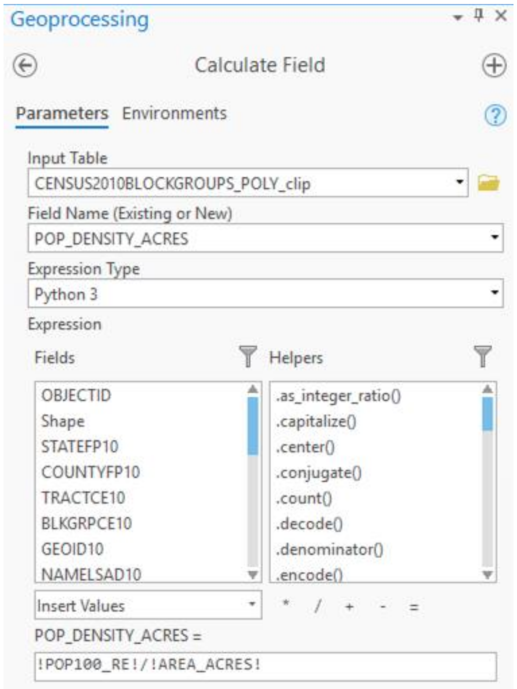
Impervious fraction is calculated and added as a feature to each census block group polygon by setting the variable Impervious_Fraction to the number of impervious pixels divided by the total number of pixels in a given block group polygon.

Select the variables 'Impervious_Fraction', 'LOGSF1', and 'SHAPE_AREA' and export the Feature Collection to Google Drive as CENSUSBLOCKGROUP_IMPERVIOUSFRACTION.csv using the function Export.table.toDrive().

Part 2: Watershed Degradation Potential

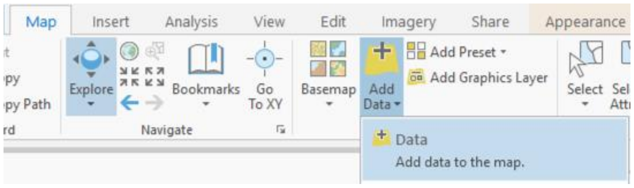
Step 2.1 – Calculate population density per census block group

The "Calculate Field" function can calculate the population density per census block group in acres by creating a new field POP_DENSITY ACRES = POP100_RE/AREA_ACRES in the table CENSUS2010BLOCKGROUPS_POLY_clip. POP100_RE is the number of individuals living in a census block group in 2010 and AREA_ACRES is the area of the census block group in acres.

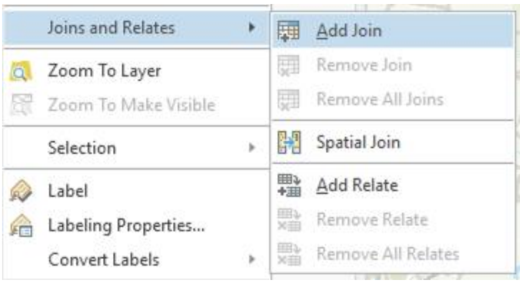


Step 2.2 – Join impervious surface fraction map to population density map

Add the table CENSUSBLOCKGROUP_IMPERVIOUSFRACTION.csv to the Map by using "Add Data" under the Map tab and selecting the file downloaded from Google Drive. It will be added to the map as a Standalone Table CENSUSBLOCKGROUP_IMPERVIOUSFRACTION.csv.

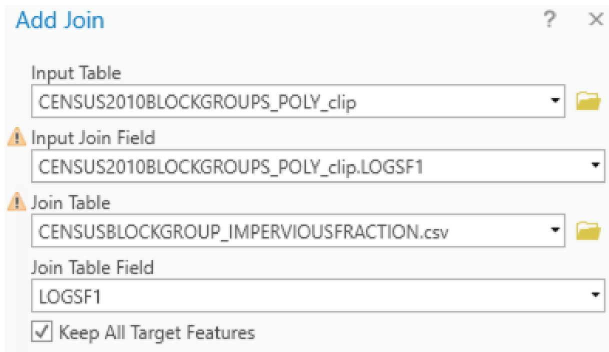


Right-click on CENSUS2010BLOCKGROUPS_POLY_clip in the Contents panel and navigate to Add Join under Joins and Relates.



This material is based upon work supported by NASA through contract NNL16AA05C. Any mention of a commercial product, service, or activity in this material does not constitute NASA endorsement. 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 and partner organizations.

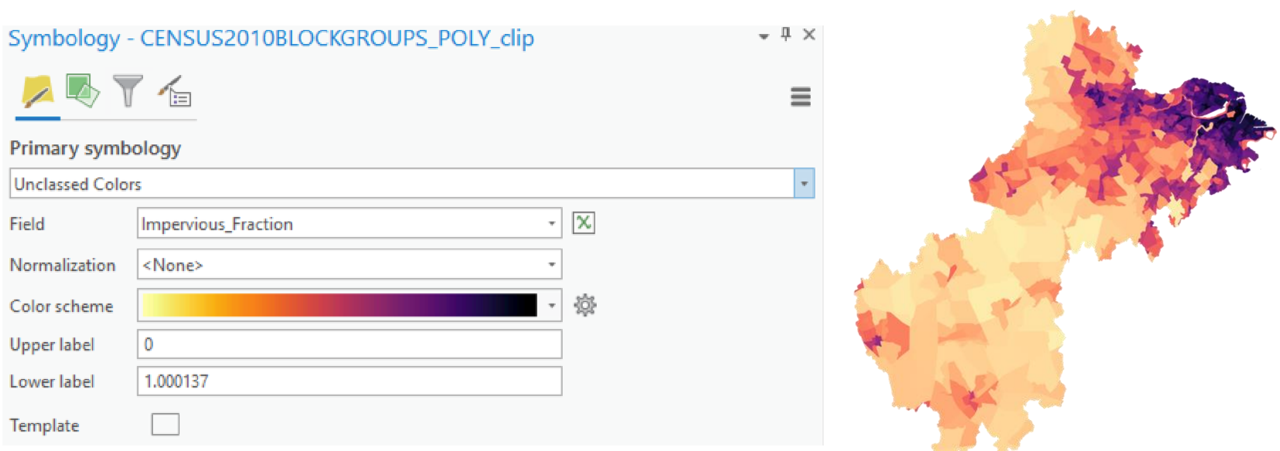
The input table is CENSUSBLOCKGROUPS_POLY_clip with join field LOGSF1, and this will be joined with the impervious surface fraction per census block group data in the table CENSUSBLOCKGROUP_IMPERVIOUSFRACTION.csv with join field LOGSF1.



Now, there is a single attribute table in CENSUSBLOCKGROUPS_POLY_clip that contains information on the population density and impervious surface fraction per census block group in the study area.

Step 2.3 – Visualize impervious surface fraction per census block group

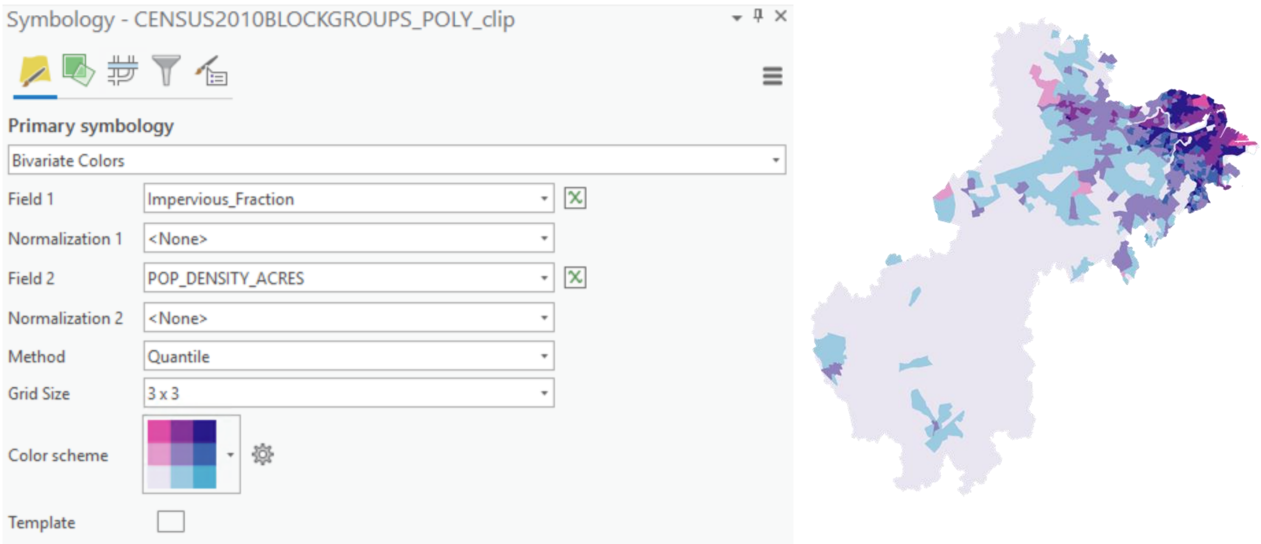
An unclassed colors symbology will show the distribution of impervious surface fraction across the study area.



Impervious surface fraction per census block group (map created by NASA DEVELOP)

Step 2.4 – Visualize potential for watershed degradation per census block group

A bivariate symbology combining POP_DENSITY_ACRES and Impervious_Fraction will show the potential for watershed degradation across the study area.



Estimated watershed degradation per census block group (map created by NASA DEVELOP)

Tutorial: Flood Susceptibility Mapping

NASA DEVELOP Charles River Watershed Water Resources – Fall 2020

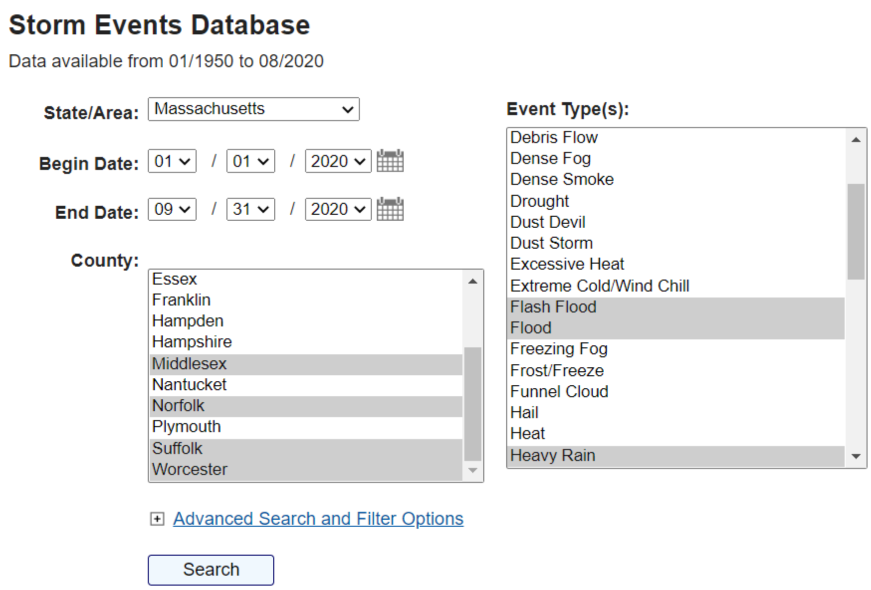
Trista Brophy, Willow Coleman, Anna Garik, Will Peters

Table of Contents

Part 1: Flood Event Data

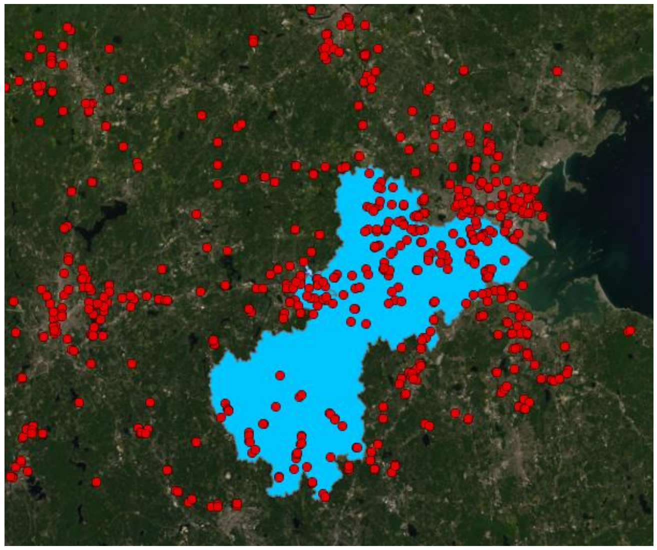
Step 1.1 – Selecting flood event data

Use [NOAA's Storm Events Database] to create a list of flood events in the counties containing your study area across your study period and choose your events of interest. We selected the Middlesex, Suffolk, Norfolk, and Worcester counties because the Charles River Watershed intersects them. We looked at floods, flash floods, and heavy rain between January 2000 and September 2020.



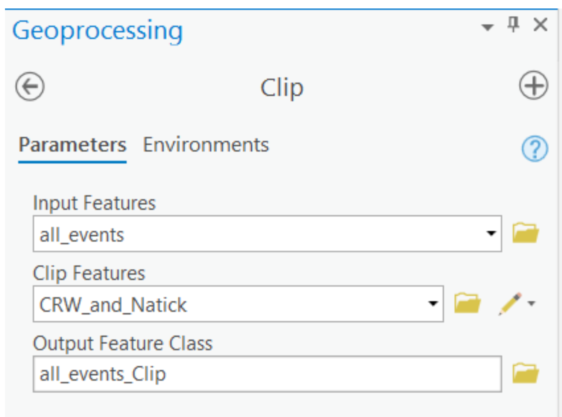
Step 1.2 – Download flood event data and map events in ArcGIS Pro

Download the flood events data using the "Data Export" CSV Download. You will now have a spreadsheet of flood event data including date, location (coordinates or name of the landmark/town), and a description of the event. For any events that do not contain coordinate data, the coordinates of the landmark or town where the event took place may be used as a proxy. Once all events have coordinate data, input the table into ArcGIS Pro and use the XY Table to Point tool to map all of the flood events. This will give you a map of the events across all of the selected counties.

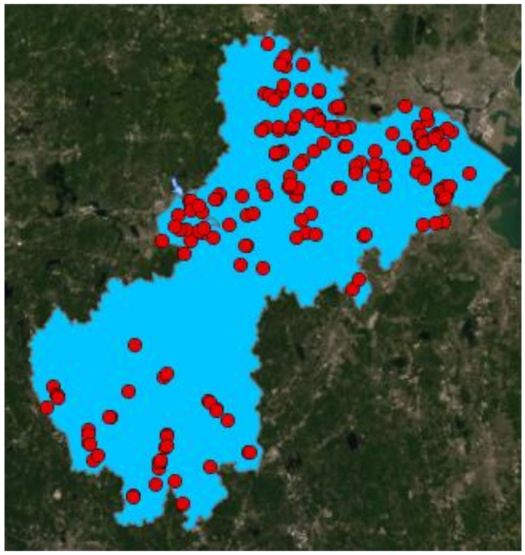


Step 1.3 – Clip flood events to study area

Clip the events to just your study area using the Clip function.



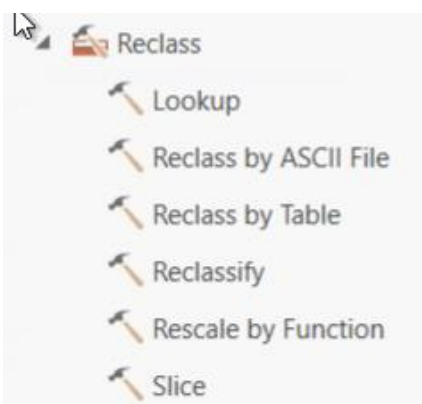
You now have a map of the flood events in your study area across your study period.

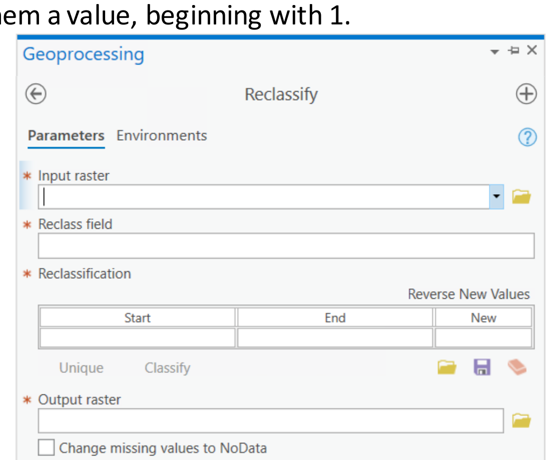


Part 2: Frequency Ratios

Step 2.1 – Download flood conditioning data

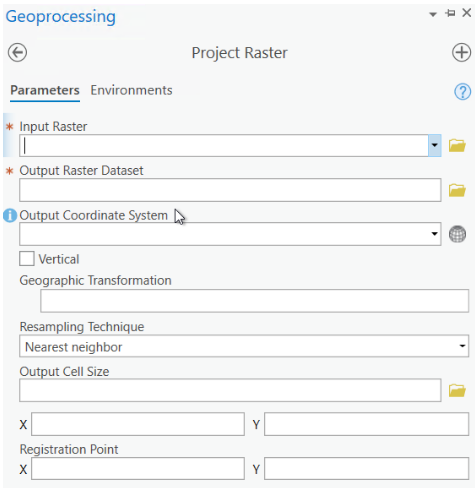
Download Digital Elevation Model (DEM), landcover, soil drainage, surficial materials, and FEMA's National Flood Hazard Layer (NFHL) raster datasets into ArcGIS Pro. Separate the numeric raster datasets into quantiles using the Reclassify tool. The team aimed to have around 5 quantiles for each flood conditioning factor, but this number fluctuated based on the datasets. Once all raster datasets are separated into categories or quantiles, assign them a value, beginning with 1.

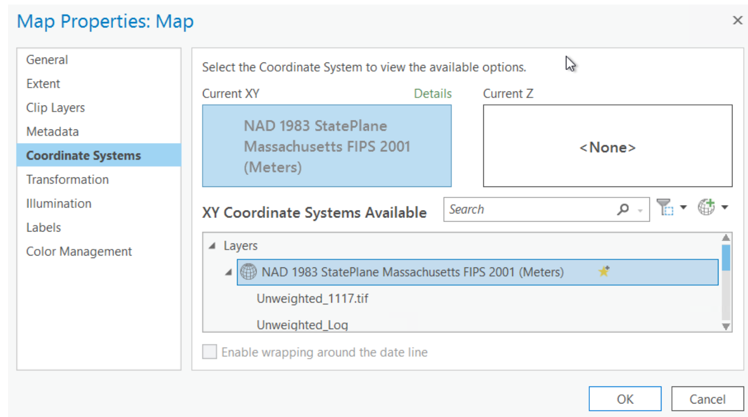




Step 2.2 – Standardize coordinate systems

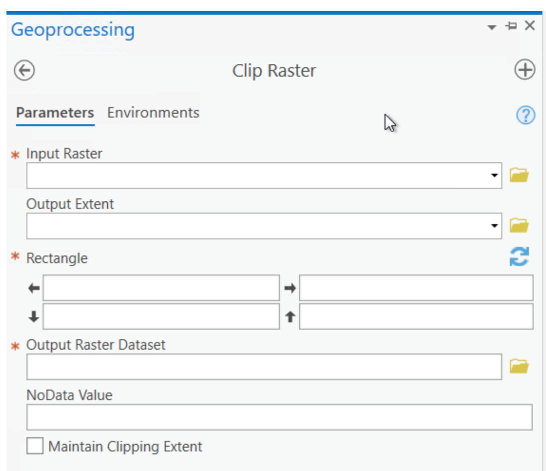
If the rasters and the overall map have differing coordinate systems, reproject them until all have the same coordinate system using the Project Raster tool or by changing the settings under Map Properties>Coordinate Systems.





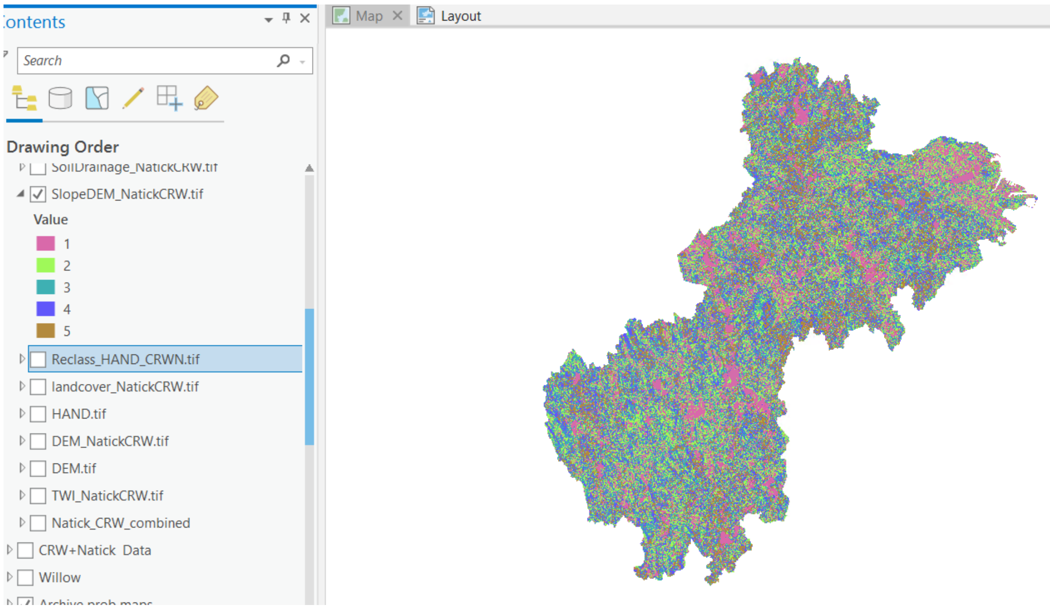
Step 2.3 – Clip flood conditioning factor rasters to the same size

If the rasters are not all the same size, clip them to the extent of the smallest raster using the Clip Raster tool.



Step 2.4 – Calculate slope

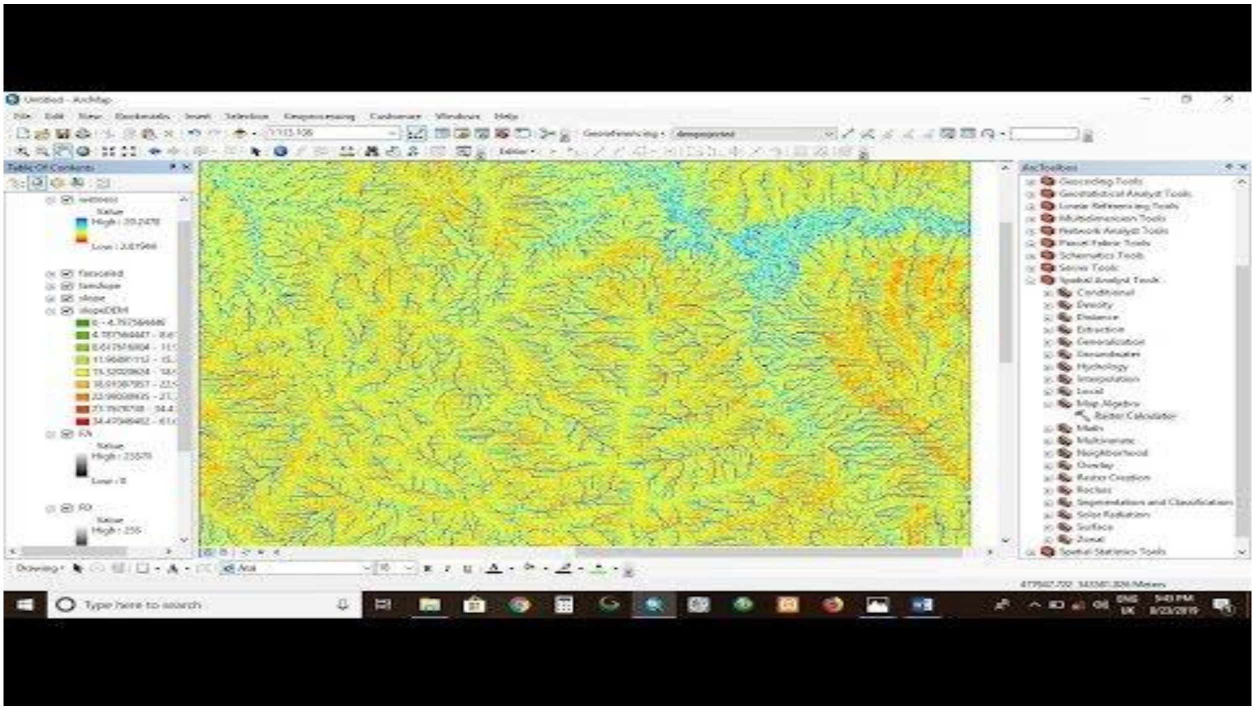
Calculate the slope raster by applying the Slope tool to the DEM raster. Once it is complete, separate it into quantiles and assign values to each according to the method described above in Step 2.1.



Step 2.5 – Calculate Topographic Wetness Index

Calculate the Topographic Wetness Index (TWI) raster with tools in the Spatial Analyst>Hydrology toolbox using the following tutorial. Once completed, separate it into quantiles and assign values to each according to the method described above in Step 2.1.

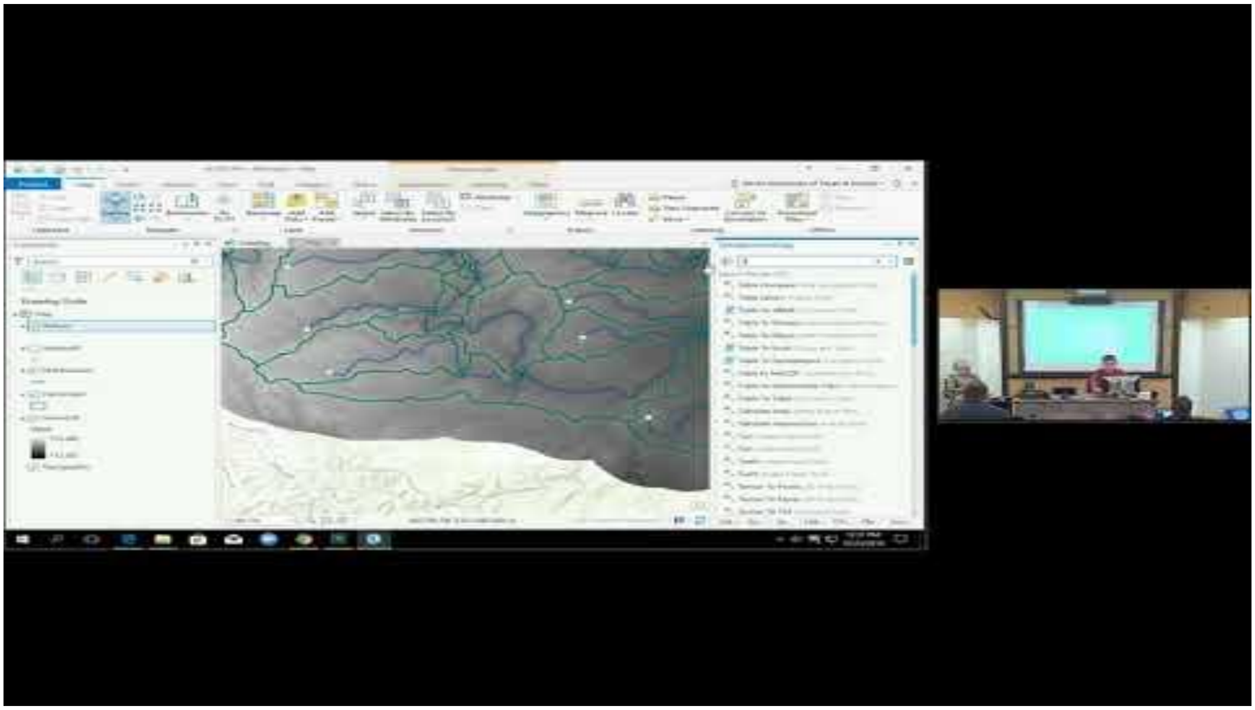
[How to calculate Topographic wetness index using ArcGIS]



Step 2.6 – Calculate Height Above Nearest Drainage

Calculate Height Above Nearest Drainage (HAND) with tools in the Spatial Analyst>Hydrology using the following tutorial. Once it is complete, separate it into quantiles and assign values to each according to the method described above.

[GISWR 16: Exercise 5: Height above nearest drainage (HAND) Flood Inundation mapping, 10/25/18]



Step 2.7 – Calculate frequency ratios

The DEVELOP Charles River Watershed Water Resources team did not complete the NASA software release process on the Google Earth Engine code written that derived the data used on this project. As a result, in this document, we are unable to share any of the code the team created to comply with NASA software release regulations. The following tutorial documents the process for users to follow and write their own code. All screenshots, code links and descriptors in this document are from publicly available code and information provided by Google Earth Engine Tutorials and other open-source sites, and are cited accordingly. Please note there may be more than one way to correctly derive the same data through Google Earth Engine or other coding languages and software.

Frequency ratio calculations are completed in Google Earth Engine and Excel.

The following items need to be imported as Google Earth Engine assets:

1. soilDrainage.tif (raster of soil drainage)
2. surficial.tif (raster of surficial materials)
3. landcover.tif (raster of land cover map)
4. slope.tif (raster of DEM-derived slope)
5. twi.tif (raster of DEM-derived TWI)
6. elevation.tif (raster of DEM-derived elevation)
7. HAND.tif (raster of DEM-derived HAND)

8. fema.tif (raster of FEMA National Flood Hazard Layer)
9. boundingBox.shp (shapefile of study area, Charles River watershed + Natick, MA)

Clip all flood conditioning factors (soilDrainage, surficial, landcover, slope, TWI, elevation, and HAND) to the study area using the clip() function.

Flood pixels are any pixels that fall within the FEMA 100-year floodplain. Iterating through all flood conditioning factors and each of their quantiles, count the total number of pixels in a selected quantile (Total Pixels) and the total number of flood pixels in a selected quantile (Flood Pixels) using the reduceRegion() function over the entire study area.

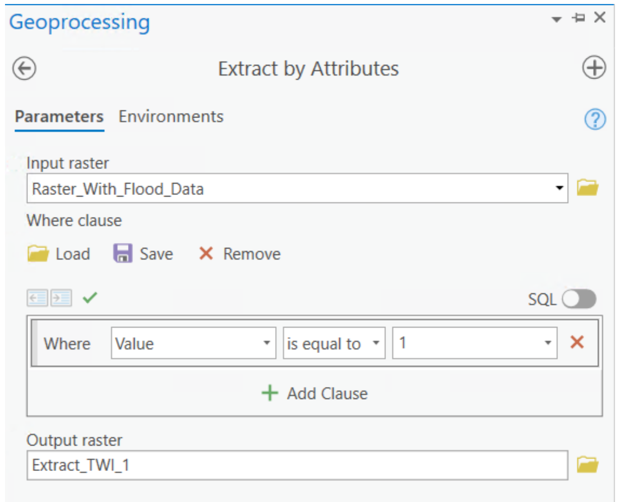
Factors	Class Name	Class Value	Total Pixels	% of Total Pixels in Quantile	Flood Pixels	% of Flood Pixels in Quantile	**Frequency Ratio**
Soil Drainage	Excessively drained	1	1015017	0.111282545	63189	0.072923146	0.655297251
	Somewhat excessively drained	2	1247172	0.136735123	34711	0.040058164	0.292961772
	Well drained	3	4119809	0.451679955	59156	0.06826887	0.151144343
	Moderately well drained	4	1056931	0.115877835	75759	0.087429531	0.754497452
	Poorly drained	5	373247	0.04092136	67705	0.078134827	
	Very poorly drained	6	1308904	0.143503182	565995	0.653185461	=G7/E7
		sum	9121080	N/A	866515	N/A	N/A
Landcover	Impervious	1	1795687	0.158285783	60638	0.047909121	0.302674822
	Grass/field	2	2415432	0.212914916	62945	0.049731845	0.233576146
	Forest/scrub	3	5002082	0.440922315	222770	0.176007041	0.399179255
	Wetland	4	1573593	0.138708695	624187	0.493160242	3.555366451
	Bare land	5	108681	0.009579987	28784	0.022741782	2.373884506
	Water	6	449113	0.039588304	266364	0.210449969	5.31596322
		sum	11344588	N/A	1265688	N/A	N/A

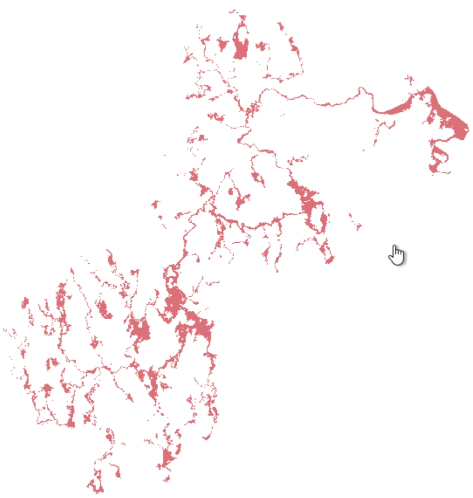
Frequency ratios are calculated with this equation: $FR = \frac{\%\ of\ flood\ pixels\ in\ quantile\ A}{\%\ of\ total\ pixels\ in\ quantile\ A}$, shown above in Excel.

Part 3: Logistic Regression

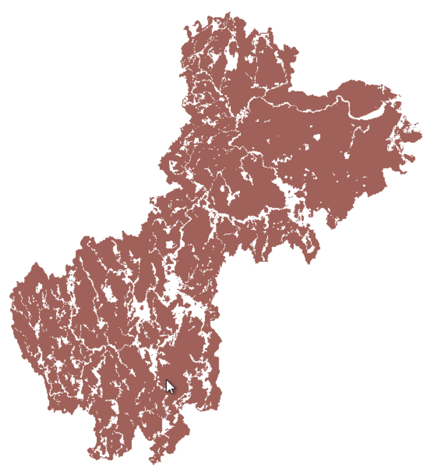
Step 3.1 – Select flooded and not flooded areas

Select training points to use for the regression. For this project, the team classified the 100-year flood zone on the FEMA NFHL as "Flooded", which was assigned a value of 1. Areas of minimal flood risk were classified as "Not flooded" and were assigned a value of 8. Use the Extract by Attributes tool on the FEMA raster twice to extract the areas that are classified as "Flooded" and "Not flooded".





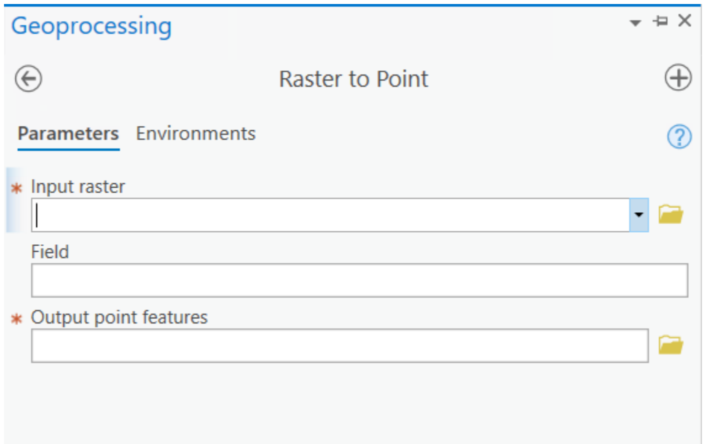
1. Extracted areas classified as "Flooded"



2. Extracted area classified as "Not flooded"

Step 3.2 – Convert flooded and not flooded rasters to points

Convert the Flooded and Not flooded rasters to points by using the Raster to Point tool. This will convert the entire raster image into a set of points.

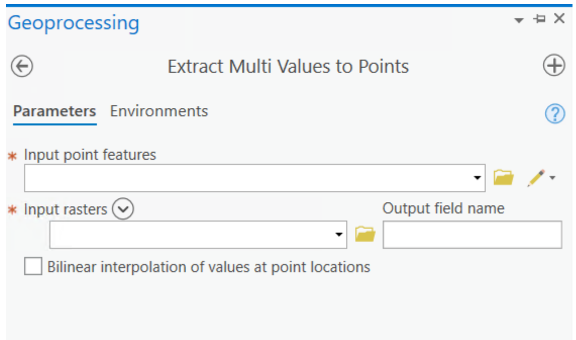


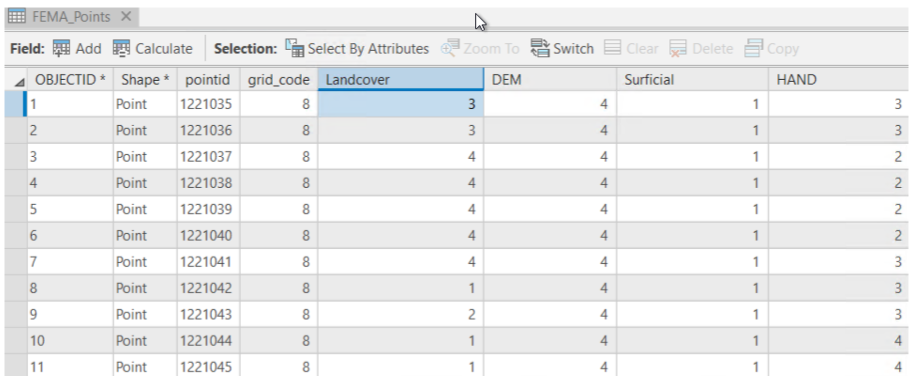
3. Areas designated as "Flooded" converted to points from a raster.

Step 3.3 – Create table with flood conditioning factor information for each training point

Join the points to all flood conditioning rasters to create an attribute table that has the flood conditioning data at each point by using the Extract Multi Values to Points tool.

This material is based upon work supported by NASA through contract NNL16AA05C. Any mention of a commercial product, service, or activity in this material does not constitute NASA endorsement. 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 and partner organizations.

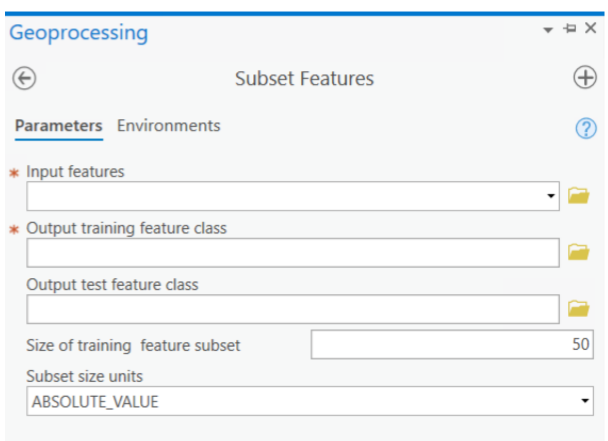




OBJECTID *	Shape *	pointid	grid_code	Landcover	DEM	Surficial	HAND
1	Point	1221035	8	3	4	1	3
2	Point	1221036	8	3	4	1	3
3	Point	1221037	8	4	4	1	2
4	Point	1221038	8	4	4	1	2
5	Point	1221039	8	4	4	1	2
6	Point	1221040	8	4	4	1	2
7	Point	1221041	8	4	4	1	3
8	Point	1221042	8	1	4	1	3
9	Point	1221043	8	2	4	1	3
10	Point	1221044	8	1	4	1	4
11	Point	1221045	8	1	4	1	4

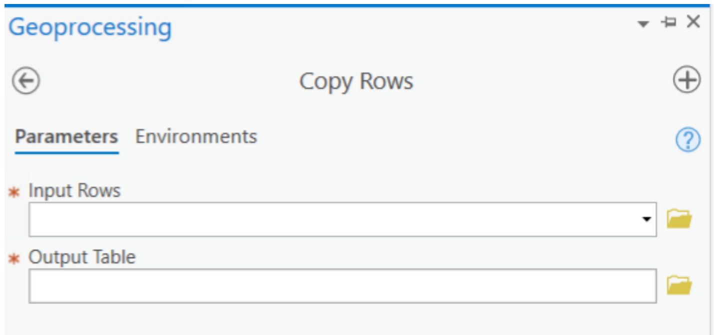
Step 3.4 – Filter to desired number of training points

Decide on the number of points you'd like to train your model with. There should be an equal number for the Flooded and Not flooded areas. Our team decided to use 10,000 points from each. Use the Subset Features tool on both point layers to randomly extract the number of points you decide on by using the "ABSOLUTE_VALUE" option. You can also choose the number of points as a percentage of the total number of points available using the "PERCENTAGE_OF_INPUT" option.



Step 3.5 – Export training point data as a CSV

Export the attribute tables of the point layers you will use as your training points by using the Copy Rows tool. Be sure to add .csv to the end of the name of your Output Table so it saves as one.



Step 3.6 – Input training point data to R and shuffle the rows

Input the data into R and combine the two training datasets into one dataframe and randomly shuffle the rows using the **set.seed** and **sample** functions.

Step 3.7 – Run logistic equation on the training points

Run the logistic regression with the FEMA column as the dependent variable and the flood conditioning factors as the independent variables. Use the **glm** function with family=binomial and link=logit. The output of this function will give you the intercept and logistic regression coefficients for each flood conditioning to use in the equation as shown below:

$$z = Flood\ conditioing\ factor_1 * data\,points + Flood\ conditioning\ factor_2 + Flood\ conditioning\ factor_3 * data\ points\ldots$$

This equation will be plugged into the equation below to calculate the probability of flooding at the pixel level based on the flood conditioning factors at that pixel:

$$P = \frac{1}{(1 + e^{-z})}$$

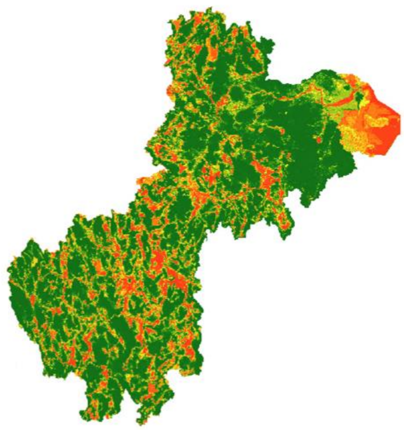
Step 3.8 – Calculate flood probability across the study area

Return to the map with the clipped flood conditioning factor rasters. Use the Raster Calculator and type in the equation shown above, plugging in the raster datasets for each flood conditioning factor for their respective logistic regression coefficient like so:

$$P = \frac{1}{(1 + e^{-FCF_1 * FCF\ Raster_1 + FCF_2 * + FCF\ Raster_2 * FCF\ Raster_3})}$$

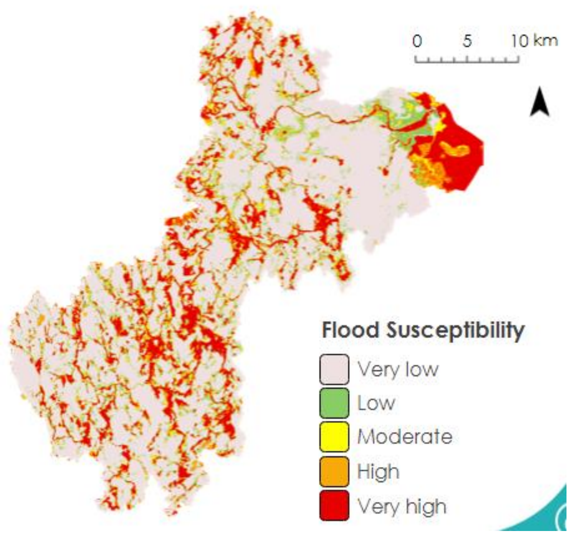
Where "FCF" stands for Flood Conditioning Factor

The resulting raster will be your flood susceptibility probability map! It will look something like this (though the initial map will be in black and white - this color scheme was chosen by the team):



Step 3.9 – Define different levels of susceptibility

If you'd like to have different levels of flood susceptibility, you can define quantiles and again use the Reclassify tool to define them. Our team created five levels, ranging from very low to very high. You now have your flood ranked flood susceptibility probability map!



Part 4: Random Forest Classification

The DEVELOP Charles River Watershed Water Resources team did not complete the NASA software release process on the Google Earth Engine code written that derived the data used on this project. As a result, in this document, we are unable to share any of the code the team created to comply with NASA software release regulations. The following tutorial documents the process for users to follow and write their own code. All screenshots, code links and descriptors in this document are from publicly available code and information provided by Google Earth Engine Tutorials and other open-source sites, and are cited accordingly. Please note there may be more than one way to correctly derive the same data through Google Earth Engine or other coding languages and software.

Step 4.1 – Import assets of the study area into GEE, as well as any other maps for context and comparison

In Google Earth Engine (GEE), you can import assets into your script. Assets can generally be imported as Images (GeoTiff or TFRecord) or Tables (shapefiles or .csv)

Upload a shapefile of your study area. This shapefile will allow you to run calculations within this area only, reducing processing time and providing results within your area of interest.

Depending on why you are creating a random forest classifier, you may need to use different maps or layers to visually compare the resulting data with. For instance, our team uploaded shapefiles of the FEMA Flood Hazard Layer to compare the results from the random forest classification with.

1. Upload shapefile of your study area; in this case, both Charles River watershed and the Charles River watershed + Natick (Files: Charles_River_Watershed.shp, Natick_CRW_combined.shp)
    a. Click on the "Assets" tab and choose the shapefile option.

    b. Once the asset is uploaded, hover over the asset and click on the arrow to the right. This will import the asset into your GEE script.

2. Upload any additional maps or layers needed for calculation (File: FEMA_clipped_CRW.tif)
    a. Repeat step 1 for any additional assets
        i. Depending on the type of asset, it may be a GeoTiff, PNG, Shapefile, or Table
3. Rename your assets to something convenient when performing calculations
    a. Go into your *Imports* section at the top of your GEE script, click on name of asset, change name

   b. Ex: study area name = 'boundingBox'
   c. Ex: FEMA Flood Hazard layer = 'FEMA'

Step 4.2 – Import each flood conditioning factor into GEE

In GEE, there are two output modes for the random forest model: 1) Classification and 2) Probability. For this project, the team used **7 Flood Conditioning Factors (FCF)** that were processed in the Random Forest Classifier. The results were a classification of whether or not each pixel is expected to flood, and the predictive power of the random forest model. You will need to upload any files that will be processed in your Random Forest Classifier that are relevant to your project.

4. Upload each FCF map as an asset into GEE (Files: DEM_NatickCRW.tif, landcover_NatickCRW.tif, Reclass_HAND_CRWN.tif, Slope_DEM_NatickCRW.tif, SoilDrainage_NatickCRW, SurficialMap_CRWNatick.tif, TWI_NatickCRW.tif).
   a. This process is the exact same as the explained in Step 2.1
      i. You should see a list of each FCF listed under the assets tab, as well as the Imports section of your script

Step 4.3 – Import assets of training data - flooded and non-flooded feature collections - into GEE

The Random Forest Classifier needs to know where to run its classification. In other words, you need to "train the classifier". This can be accomplished by uploading a file containing points within your study area. In our case, the team created a set of random points that are located in flood zones or non-flood zones (calculated in ArcGIS). Ideally, these points will be completely random so there is no bias as to why the classifier was run at certain locations.

1. Upload the flooded and non-flooded training datasets as assets into GEE (Files: FEMA1_pointsSubset_Reprojected, FEMA8_pointsSubset_Reprojected)
   a. This process is the exact same as explained in steps 2.1-2.2
      i. These will be uploaded as shape files, but will be considered "feature collections", since they will also import a data table with values associated to each training point, as well as a class property for each dataset.
   b. You can double check your training data is formatted correctly by clicking on the feature collection in your assets tab/imports section. Then, click on the "Features" tab
      i. Note that there is a class called "landcover" in both tables, while there are two different values for each table.

DESCRIPTION FEATURES PROPERTIES

Feature Index	CID (Long)	RASTERVALU (Integer)	landcover (Integer)	system:index (String)
0	0	1	1	
1	0	1	1	
2	0	1	1	
3	0	1	1	
4	0	1	1	
5	0	1	1	
6	0	1	1	
7	0	1	1	
8	0	1	1	
9	0	1	1	

DESCRIPTION FEATURES PROPERTIES

Feature Index	CID (Long)	RASTERVALU (Integer)	landcover (Integer)	system:index (String)
0	0	8	2	
1	0	8	2	
2	0	8	2	
3	0	8	2	
4	0	8	2	
5	0	8	2	
6	0	8	2	
7	0	8	2	
8	0	8	2	
9	0	8	2	

Step 4.4 – Merge the feature collections of the training points together

Merging the flooded and non-flooded feature collections will group the flooded and non-flooded training points together so they are located in one place. Each feature collection should have the same **class property,** but different **values.** This is imperative for the merging process, because it will allow the separate feature collections to merge together without interfering with their respective values.

1. Merge the feature collections using the ".merge()" function

Step 4.5 – Set parameters for the classifier

The bands for each flood conditioning factor need to be used when training the classifier. Thus, each band for each FCF must be renamed. Renaming these bands will make it easy to add them together, differentiate which band is which, and select when training the classifier.

4.5.1 - Rename, define, and assign the bands to of all the Flood Conditioning Factors

1. Apply the ".rename()" function to each FCF
2. Clip the results to the study area using the ".clip()" function

4.5.2 - Create a variable that holds each band from each flood conditioning factor

1. Apply the ".addBands()" function to each FCF

4.5.3 - Create a variable that lists each band for each flood conditioning factor

1. This is necessary because this list will define which bands will be used in the classifier
    i. Set a variable equal to a list of names for each band, like such:
        1. var *name* = ['band1', 'band2'…]

4.5.4 - Create a variable that holds necessary inputs for the classifier

Creating a variable that holds necessary inputs for training the classifier will make it easier to set parameters for the classifier. Inputs could be feature collection(s), scale, bands, and others that are relevant.

1. Use the ".select()" and ".sampleRegions()" functions to complete this step.
   i. Name the variable something relevant
   ii. Select the bands from each FCF
   iii. Use the same feature collection(s) that was previously defined
   iv. Use the property that is defined in the feature collections(s)
   v. Set the scale to 10, or whatever is relevant to your project

Step 4.6 – Train the classifier

Training the classifier will tell the classifier exactly what to do, given a specific set of inputs. This can be done by inputting where to run the classification (training points) and the parameters to use (bands from each FCF, the variable created in the previous step holding certain inputs/parameters, and any additional inputs).

1. Create a variable that is the name of the classifier. Using the ".smileRandomForest()", ".setOutputMode()", and ".train()" functions to complete this step. Define "Features" as the variable defined in the previous step that contains the list of inputs, "classProperty" as the 'landcover' class property, and "inputProperties" as the bands that were previously defined.

Step 4.7 – Run and display the classification

Use the ".classify()" function to run the classifier. Use ".select()" function to select the bands previously defined. The input for the classify function will be the name of the classifier that was defined in the previous step. Run the classifier.

Step 3.7.1 – Display the classification and other necessary layers

Use the "Map.addLayer()" function to add the classifier, as well as any other maps/layers needed for display.

Step 4.8 – Create a confusion matrix that shows the accuracy of the random forest classification

This confusion matrix will show how accurate the classifier is based on the training data that was used. Use the ".confusionMatrix()" function and the ".accuracy()" function on the classifier that was defined earlier and print the results. The result will give you a percentage score on how accurate the classifier is.

Step 4.9 – Export the resulting maps/images as a GeoTiff to Google Drive

Once your calculations are complete and the code has successfully run, you can export your results (a displayed map of the decision tree results) to Google Drive for easier access and shareability. Use the "Export.image.toDrive()" function to accomplish this.

1. Define the boundaries that this image will be exported to (either the Charles River watershed or Charles River watershed + Natick)
2. Select a scale that is appropriate for your map and define which image you will be exporting to Google Drive (the name of the layer in GEE)
3. Give this exported file a proper name for when it ends up in Google Drive.