GEO TUTORIAL

#QGIS
Dealing with Coastal Flooding series, part 6:
CALCULATING SPATIAL STATISTICS OF INUNDATED AREAS

Krzysztof Raczynski
Andrew Nagel
John Cartwright

Geosystems Research Institute
Mississippi State University

MAY 2025

This work was supported through funding by the National Oceanic and Atmospheric Administration Regional Geospatial Modeling Grant, Award # NA19NOS4730207.

Geosystems Research Institute, Mississippi State University, Box 9627, Mississippi State, MS 39762-9652

The Geospatial Education and Outreach Project (GEO Project) is a collaborative effort among the Geosystems Research Institute (GRI), the Northern Gulf Institute (a NOAA Cooperative Institute), and the Mississippi State University Extension Service. The purpose of the project is to serve as the primary source for geospatial education and technical information for Mississippi.

The GEO Project provides training and technical assistance in the use, application, and implementation of geographic information systems (GIS), remote sensing, and global positioning systems for the geospatial community of Mississippi. The purpose of the GEO Tutorial series is to support educational project activities and enhance geospatial workshops offered by the GEO Project. Each tutorial provides practical solutions and instructions to solve a particular GIS challenge.

CALCULATING SPATIAL STATISTICS OF INUNDATED AREAS

Krzysztof Raczynski [1, 2, 4, 5, 6, 8] chrisr@gri.msstate.edu
Andrew Nagel [3, 7] andrewn@gri.msstate.edu
John Cartwright [7, 9, 10, 11] johnc@gri.msstate.edu

Geosystems Research Institute
Mississippi State University

CRediT: 1: Conceptualization; 2: Methodology; 3: Verification; 4: Resources; 5: Data Curation; 6: Writing - Original Draft; 7: Writing - Review; 8: Visualization; 9: Supervision; 10: Project administration; 11: Funding acquisition

REQUIRED RESOURCES

- QGIS 3+

FEATURED DATA SOURCES

- Click here to access dataset used in this tutorial (7.835 MB).

OVERVIEW

Coastal areas across the United States face increasing challenges from changing water levels, which can lead to more frequent flooding and infrastructure strain. In communities like Bay St. Louis, Mississippi, rising water can make roads impassable, damage property, and disrupt daily life—posing serious concerns for homeowners and local economies.

As part of a planning team, your role is to assess how changing sea levels may impact the safety, infrastructure, and long-term growth of this Gulf Coast community. The focus is on protecting property, ensuring economic stability, and strengthening community resilience. This is the theme of the *Dealing with Coastal Flooding* tutorial series, which includes the following topics:

- Part 1: Creating Raster DEM from LiDAR Data
- Part 2: Spatial Predicates: Preparing Residential Data
- Part 3A: Using Unsupervised Machine Learning for Land Use Land Cover Classification
- Part 3B: Using Supervised Machine Learning for Land Use Land Cover Classification
- Part 4: Hydrologic Raster Preparation: Resampling and Burning Stream Network
- Part 5: Generating Flooding Extent with Raster Calculator
- **Part 6: Calculating Spatial Statistics of Inundated Areas**
- Part 7: Creating 3D Maps of Flooding Projections
- Part 8: 3D Map Animations
- Part 9: Creating and Animating Timeseries

In the previous parts of this series, we processed terrain data, prepared a building dataset, prepared land use land cover information (LULC), and forecasted water reach at the intermediate high sea level rise scenario for Bay St. Louis at several time horizons. In this tutorial, we will combine these results to calculate statistics for the inundated areas and related buildings and LULC. Make sure to check the remaining tutorials in the series to learn more about the entire analysis process.

DATA

For this tutorial we will use the data prepared in the previous parts: *buildings*, *LULC classification*, and sea level rise (*SLR*) *water reach*. If you don't have these data, you can use the **Featured Data Sources** link above to download the datasets.

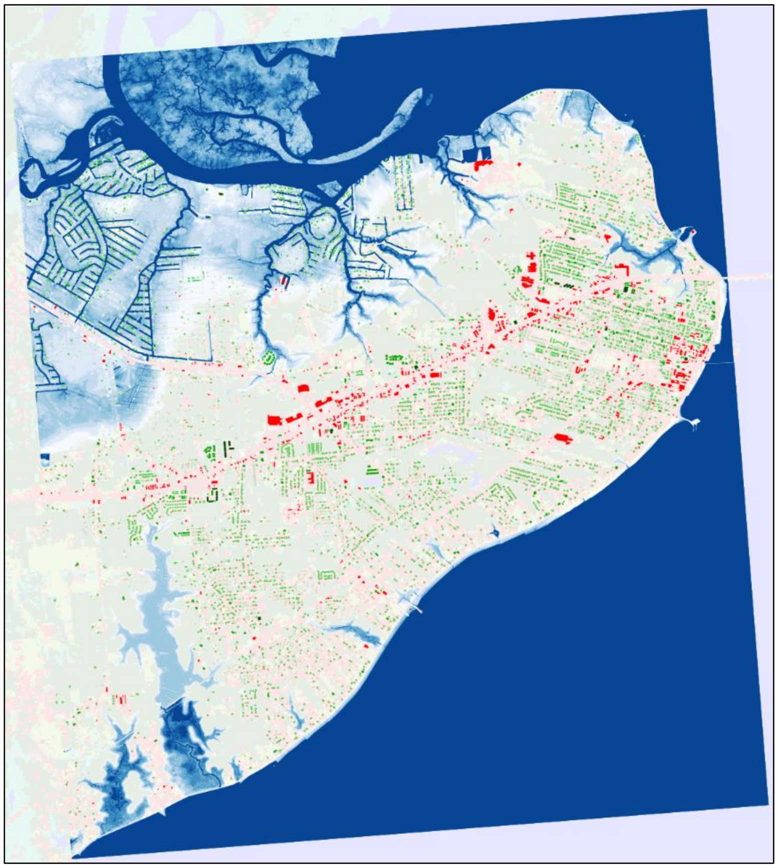

Fig. 1. *Classified buildings, LULC and SLR projections in the area of Bay St. Louis, Mississippi.*

AGGREGATING SLR LAYERS

The different time horizon SLR projections are in the separate layers. In the first step, we will merge all the outputs together and write the related inundation time into the attribute table. In the *Processing Toolbox*, expand *Vector General* and open the *Merge Vector Layers* tool. In the *Input layers*, click on the button to the right of the input field and select all vector layers with ***projected SLR*** (2030 to 2100), then click *OK*. Click *Run*. Once a new layer called *Merged* is created and added to your project, open its *attribute table*. Open *Field Calculator* and *Create a new field* with the *name* ***yearFlood***, of ***integer (whole number)*** *type* and ***4*** characters in *length*. Under the *expression*, write:

```
to_int(right( "layer", 4))
```

This expression is taking the *layer* attribute, which stores the projection scenario name (e.g., *IH_2050*), and then extracting only the 4 letters from the right, these indicate the year of the projection. Then the text representing the year will be transformed to an integer number. Click *OK*. You can now remove redundant columns: *DN*, *layer*, and *path* using *Delete fields*. This produced a layer where each feature represents a projected year of SLR. The *yearFlood* attribute contains information on the year projection.

DETERMINING A TIME HORIZON FOR HOUSES FLOODING

We can determine which buildings will be flooded for each projection by performing an *intersection* of our merged layer with the buildings layer. To do so, open the *Vector* menu, and in *Geoprocessing Tools*, select *Intersection*. Here set the *input layer* as the ***BayStLouis_buildings*** layer and the ***merged SLR*** layer as the *Overlay layer*. We can keep all the input fields, but at the *overlay fields* we only want to keep the ***yearFlood*** argument (Fig. 2). Make sure to save the output to a shapefile (.shp), as geopackages do not support non-unique feature IDs. The algorithm will generate all intersection cases (learn about spatial predicates in our Spatial Predicates: Preparing Residential Dataset tutorial) between the two layers.

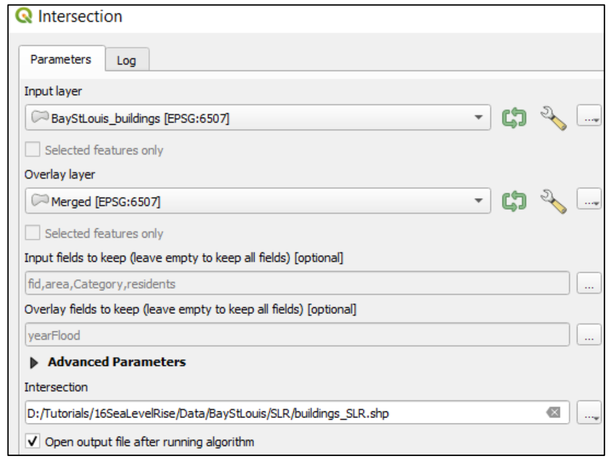

Fig. 2. Intersection can multiply features.

It is important to understand how the intersection works. For example, if we have one feature in a layer and 10 features in a second layer, the output will constitute a combination of each intersection separately. Therefore, if the feature from the first layer intersects all objects in the second layer, the output will have 10 copies of that feature, one for the intersection of each different feature from the second layer. Spatially, each copy of the building will be equal; however, their attributes will differ as they will have values added from the corresponding feature in the second layer (*yearFlood*). Note that only the buildings that are within the SLR projections are present in the layer.

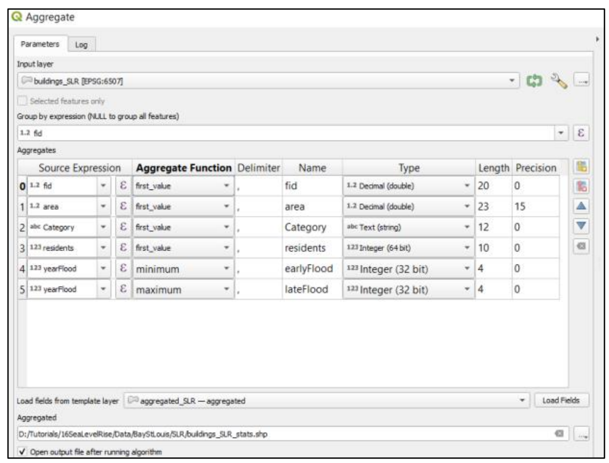

Fig. 3. Aggregating allows to grab multiple features into one geometry, following a group by expression.

Once we have generated all the *intersection* cases, we can *aggregate* the results to determine the specific situation for each house. Let's open the *Aggregate* tool (*Processing Toolbox*, *Vector Geometry* tab). We will produce two new attributes containing the date of the earliest house flooding projection and the latest. If the house won't be flooded, we will keep the argument empty. We can aggregate the buildings by their ID. Since each intersection of the building with different SLR projections references the same building, each copy will have the same ID. Set the ***fid*** in the *group by expression*. In the *aggregates window*, remove the *yearFlood* attribute, then add two new fields. *(earlyFlood, lateFlood)* For both, set ***yearFlood*** attribute as the *source expression* but change their *aggregate function* to ***minimum*** and ***maximum*** respectively. Set *types* to ***Integer*** and *length* to ***4***. This will produce two new columns, with the lowest value from each feature with the same fid, representing the earliest time SLR will reach the specific building, and the highest value, indicating the latest projection.

We can now compute statistics for each projection in the area. In the *Processing Toolbox*, expand the *Vector Analysis* tab and run the *Statistics by Category* tool. Set the *input vector layer* as the ***aggregated*** buildings layer. Use the field called ***residents** to calculate statistics*. Under *fields with categories* set as ***earlyFlood***. The result will be a table presenting descriptive statistics. For our tutorial case, only *count* and *sum* fields are useful. You can compute cumulative sums to display the total numbers for each time horizon (Table 1).

Table 1. *Statistics on buildings and residents affected by Sea Level Rise in Bay St. Louis area for each projection year.*

SLR projection year	New Buildings affected by projection	Total buildings affected	New Residents affected by projection	Total residents affected
2030	262	262	461	461
2040	18	280	40	501
2050	41	321	83	584
2060	97	418	222	806
2070	438	856	907	1713
2080	550	1406	1137	2850
2090	324	1730	690	3540
2100	260	1990	539	4079

COMPUTING AREAS OF INUNDATED LAND USE LAND COVER CLASSES

To compute areas inundated by each SLR projection, we will use the *Raster Layer Zonal Statistics* tool. This tool, as the name implies, requires the input layer as well as the zones layer to be rasters. Previously, we transformed our SLR projections to vector; however, we still have the original outputs from the *Raster Calculator* (obtained from previous tutorial: Generating Flooding Extent with Raster Calculator). Recall that these presented raster layers with ***0*** and ***1*** values, where 1 indicated inundated areas and 0 dry areas. We can use these to perform this computation. We can build a *virtual raster* to store all the SLR projections in separate bands of the same file. To do this we will open the *Build Virtual Raster* tool available in the *Raster* menu in the *Miscellaneous* tab. For input parameters choose the ***2030 to 2100 raster projections***, then change the *resolution* to the ***highest***. Mark the setting to ***place each input into a separate band***. Under *NoData values for input bands*, set ***0***—this will set all non-affected cells to be excluded from the future analysis. *Run* the tool. A virtual raster with 8 bands will be added to the project.

We can now open the *Raster Layer Zonal Statistics* tool available in the *Processing Toolbox* under the *Raster Analysis* tab. Set the ***virtual raster*** as *input*. Note the *band number* setting. For now, leave it as ***Band 1***, which will result in a computation for the 2030 projection. Once completed, you can go back and choose the ***next band*** to handle next year's projection. For the *zones layer*, set the ***LULC classification raster***[1] (this raster will have only one band). Click *Run*. A new table, *Statistics*, will be added to your project showing information on the number of pixels falling under each class. *Rerun* the tool for each band in the virtual raster (alternatively, you can batch process the computation) to produce tables showing the results of the analysis. Compare your summaries to the table 2 results.

Table 2. *Area (in acres) of land inundated around Bay St. Louis for each projection year*

LULC class	Raster value	2030	2040	2050	2060	2070	2080	2090	2100
Forest	2	224	240	266	326	569	975	1401	1842
Fields	3	12	13	16	24	67	141	231	300
Wetland	4	399	583	891	1260	1738	2098	2224	2274
Urban	5	38	39	41	45	63	103	135	158
Residential	6	192	201	216	254	410	671	881	1046
Beach	7	23	24	25	32	46	61	73	91
Main Roads	8	75	76	78	83	99	119	135	146

[1] This raster was created in the third tutorial of this series: Using Supervised Machine Learning for Land Use Land Cover Classification, however, if you are using the one provided with this tutorial input dataset, then use *LULC_style.qml* file to load raster styling including labels for each category.

EXERCISE

Compute the information on areas of each building category affected by SLR (we have created this dataset in the Spatial Predicates: Preparing Residential Data tutorial) and total area of land covered by water in each projection. Create an additional vector layer that will include all buildings that are not affected by each of the SLR projections (use spatial predicates to extract buildings, then merge them into one layer, with an attribute indicating the projection year).

CONCLUSION

This concludes our GEO Tutorial, where you learned how to compute multiple statistics on spatially related data using different tools. If you are interested in expanding your knowledge and working on similar topics, please check out the remaining tutorials in this series.