Creating a Model to Predict High-Risk Areas for Pluvial Flash Flooding in the Urban Areas of Houston, Texas

Aadya Jain, Dae San Kim, Gabriela Gomez Hernandez, Nishchay Jaiswal, Sanan Khairabadi

NASA SEES – Earth System Explorers

Abstract

One of the major flaws of modern flood maps is the lack of consideration towards precipitation flooding, which can turn into a type flash flood known as pluvial flooding. This type of urban flooding has not been explored in depth because of variations in urban infrastructure and a heavy focus on coastal flooding. However, urban flash flooding does happen, and residents are often unprepared and uninformed. The prominence of impervious surfaces – artificial structures that do not absorb water like asphalt and concrete – makes urban areas particularly vulnerable to pluvial flash flooding. Therefore, our project addresses the question: where will water accumulate in the event of a pluvial flash flood in Houston, Texas? This research uses previous data related to elevation, precipitation, and flash flooding in Houston, Texas. The data, collected from several sources including the GLOBE Observer land covers, OpenTopography elevation data, and QGIS analysis features, was developed into a Python program that predicts areas at high risk for severe flash flooding. The model analyzes digital elevation model rasters using a D8 flow algorithm and determines specific elevation points that collect large amounts of water from other points, indicating areas particularly vulnerable to water accumulation in a flash flood. After analyzing 6 areas of interest in the Houston area, we found that water is most likely to pool on impervious surfaces such as streets and roads. This knowledge can help urban planners and citizens to prepare for flash floods through informed drain placement and avoiding building houses in areas at high risk of flooding. Furthermore, our model can be used by other scientists studying flooding for similar urban environments, providing a valuable tool for predicting and mitigating the impacts of urban pluvial flash flooding.

TABLE OF CONTENTS

Acronym List

AOI: Area of Interest	**SIFT:** Scale-Invariant Feature Transform
DEM: Digital Elevation Model	**CSV:** Comma-Separated Values
IDW: Inverse Distance Weighted	**TIN:** Triangular Irregular Network
TIFF: Tag Image File Format	**HTML:** HyperText Markup Language
GLOBE: Global Learning and Observation to Benefit the Environment	

I. Introduction

Flooding is a common natural disaster that occurs very often in areas near water all around the world. Flooding can be caused by natural and man-made factors: earthquakes, hurricanes, rapidly rising sea or river levels, or dam failures. One of the more common flood types is coastal floods, which can be unpredictable and often cause severe damage and several casualties. Nevertheless, with the proper research and tools developed by environmental scientists and meteorologists, these coastal areas are warned with time to evacuate. However, several flooding tools and maps have a flaw: the lack of information and data on pluvial flooding. Pluvial flooding is any flooding related to precipitation, which is also associated with flash floods, due to its sudden nature. This type of flooding commonly happens in urban areas due to factors such as pavement and poor drainage. The pavement and prevalence of other impervious surfaces such as asphalt and concrete do not allow water to be absorbed into the ground, thus resulting in pooling water and flooding. In addition, many urban areas do not have the proper drainage systems for flash floods, and this oversight can be harmful to its residents, especially small children, the elderly, and animals, who are vulnerable to the strength of flash floods. Currently, there are limited tools to predict areas of vulnerability in the event of a flash flood. Nonetheless, several databases, such as GLOBE Observer, a program run by NASA that allows citizens to submit photo-based data of their surrounding environments, in order to contribute to comprehensive dataset that does not solely rely on satellites, are available to aid in the identification of high-risk areas for flash flooding, and serve in developing a way to warn residents of severe flooding coming their way.

This paper will address the combination of different databases to create a model that will predict high-risk areas for flash flooding in urban areas. The locations for this case study are several areas in Houston, Texas, which will be used for a flood simulation using previous elevation and fluid mechanics data. The anticipated end result of this model is to create a tool used to predict areas with a high risk of flash flooding by simulating a flash flood at any Area of Interest (AOI) point(s) supported by GLOBE Observer in the United States.

II. Literature Review

The following literature review was compiled using search engines such as Google and Safari. These articles were chosen for their focus on pluvial and flash floods in cities and other similar urban areas, as well as the effects on the people residing in said areas.

Pluvial Flood Risk and Opportunity for Resilience

One of the main reasons that makes flash flooding so difficult to predict, let alone allow a finding of high-risk areas, is the limited data collected in urban cities. More often than not, it is due to indifference or misleading assumptions about pluvial flooding. One incorrect assumption is that flooding is only a minor nuisance that causes minimal damage. However, there is evidence that the effect of several small floods is much more destructive than a few severe floods. Many times, city infrastructures neglect to maintain and expand their storm drainage in response to previous pluvial floods, further endangering the residents and properties.

FEMA Flood Maps Fail to Show Flood Risk of More Extreme Flooding Events

In recent years, there has been a surge of videos posted online of flash floods destroying cities and neighborhoods in areas that were not indicated by the Federal Emergency Management

Agency (FEMA) flood maps to be high risk for floods. This has caused thousands of people to pay out-of-pocket for flood repairs because there had been no indicators to buy flood insurance in past years. While FEMA maps are constantly being updated, they cannot keep up with the whole country, let alone the surrounding areas, putting the properties and lives of thousands up to a gamble. This model could aid in protecting the residents from being ill-informed and paying for unexpected flood damage.

III. Methods

i. Finding the Slope of Target AOI points

To generate accurate and clean data to enter into a flash flood model simulator, two different channels were utilized. The first channel used was from the AOI points collected from GLOBE Observer in the Houston, TX area. These points contain images of each AOI point at the North, West, East, and South. The second channel is from OpenTopography, which is used for its nearly worldwide satellite data. The benefit of using satellite data is its ability to get complete and widespread coverage. Since the AOI data points are scattered and do not give a complete view of Houston, satellite data must be used to patch up the holes left behind in order to make a complete DEM. Since satellite data is so comprehensive, it makes it easy for it to be used where civilian data is lacking. The fact that it is also government-backed, meaning it is more reliable than civilian data, ensures any data gleaned from the DEM created by SRTM can be trusted, rather than the potentially off proportions the civilian photos offer. Once the channels are established, the following series of steps is run to extract slope data from these images. First, the data is used to perform semantic segmentation using the U-Net architecture. This segmentation algorithm separates the ground from all other objects. Once the algorithm has identified the ground, a SIFT feature extraction algorithm is run, outputting a series of “key” points in the image. With these important points extracted, a monocular depth estimation algorithm, ZoeDepth, provides the 3D information on these ground features. Finally, the points are interpolated around the features, ultimately building a dataset of elevation nuances on the ground.

ii. Using the AOI Points to Create a Digital Elevation Model (DEM)

The DEMs were created using a combination of civilian data taken from the GLOBE Observer application and the SRTM GL1 elevation dataset accessed from OpenTopography. With the elevation data found from the AOI points and the SRTM GL1, QGIS features were used to create a Digital Elevation Model (DEM) to run the flood simulations on. Since the data is in XYZ format, or point form, the CSV file was added as a delimited text layer, allowing for the X mode and Y mode to be longitude and latitude, respectively. Because the AOI points are scattered, elevation data from OpenTopography is interpolated with the AOI data using IDW interpolation in order to fill in the blank spots to make as accurate a DEM as possible. Then, using the elevation as the interpolation attribute, TIN interpolation was applied to create a DEM, which is exported as a TIFF file (Figure 1), for better use in creating the simulation model.

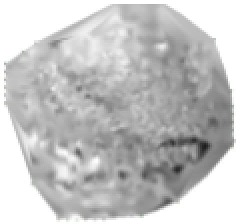


Figure 1: Sample TIFF file

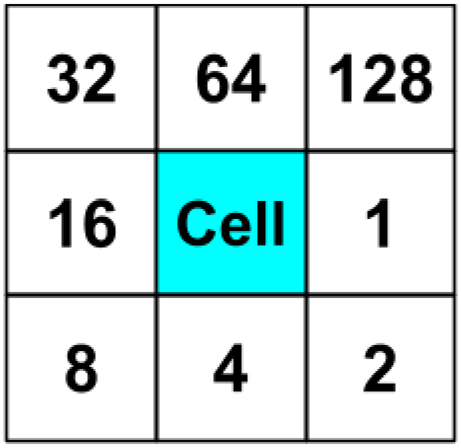
iii. Visualizing and simulating the floods using the DEM Data

Once the DEM data was developed, we designed a Python program to extract and visualize the elevation data from the DEM files, using Google Colab as a notebook. The first step was to remove any "sinks" in the elevation raster where water accumulates. Specifically, any cell not on the border that had an elevation lower than all eight of its neighbors had its elevation increased to its lowest neighbor. This was done because sinks are often an indication of faulty elevation data and can seriously hinder the calculation of water flow.

Then, the program performed a D8 flow analysis on the sinkless dataset, which is a type of raster analysis used to find the direction of the steepest flow. In the D8 method, flow is directed from each coordinate, or "cell", to its steepest downward neighbor based on elevation differences. This steepness or descent is calculated as:

$$Steepness = \frac{E_{cell} - E_{neighbor}}{distance}.$$

Furthermore, cells in a D8 analysis have a maximum of eight neighbors to potentially flow to. Based on the direction of a cell to its steepest neighbor, each cell is assigned a new direction "code" from 1-128 based on the following grid:



32	64	128
16	Cell	1
8	4	2

Figure 2: The D8 flow direction grid. If a cell's steepest descent is directly towards its right, the original cell will be assigned a value of 1, and so on.

If a cell has two or more directions with the same maximum descent slope, the direction codes are added. For example, if a cell had a maximum descent of 5m both directly east and south, it would have a code of $1 + 4 = 5$. Because the base codes are all powers of 2^n, all 255 unique combinations of directions correspond to exactly one code and vice versa. The advantage, therefore, of this labeling system is that all flow information of a particular cell can be condensed into an 8-bit integer. This analysis was performed on each DEM raster for the Houston AOIs selected. For each DEM, a new "directions_grid" NumPy array was created to store the direction codes for the AOIs. Then, topographical plots were generated using Python's Matplotlib library. The first was a color-based elevation topographical map using the raw DEM data. This plot visualized raw elevation data for each AOI and identified the areas of lower elevation, which could be at higher risk of flooding. The second depicted the same AOI but under the D8 flow analysis. Finally, and crucially, the D8 analysis raster underwent a "flow accumulation" function to determine what points or "cells" in the raster collected the most water from other cells. The flow accumulation function converts the newly created “directions_grid” array to a new array storing the amount of cells contributing flow at each cell. The cells with the highest accumulation scores are the most likely regions to accumulate water or flood. Taken together, these plots are help determine where water will pool in case of a flash flood incident.

The full Python program and other information are available on GitHub: https://github.com/dsk2025/SEES-2024-Flooding-Model. The first release is also available on Zenodo at doi.org/10.5281/zenodo.13147242.

iv. Making a Front-End Site for the Simulation

To make this tool accessible, a front-end website was developed, allowing users to select areas in Houston to simulate the flash flooding. We built the front end of the website using Python and HTML, specifically using Flask 3.0 to incorporate the back-end, and the Javascript library Leaflet to build the interactive map, where users could select and deselect AOI points (Figure 3). The center points for the six AOI areas we chose were manually put into the map via the backend in Python. Users can select an AOI center point, which prompts a pop-up of the latitude and longitude of the point. After finalizing their selection, users can press the “Confirm AOI” button to generate a DEM of the selected points on the website with an elevation graph, flow analysis graph, and two flow accumulation graphs (Figure 3). The website helps users understand how flash floods would occur at that AOI location and is a very useful tool to publicly implement the resource we created to aid with flash flood preparation for those living in that AOI area. This paper is focused on the Houston, TX area, but we hope to expand the website and implement it globally. In the future, we can expand the website by importing points into the website via a CSV file in SQLite. The Python and HTML files for the website can be found in the GitHub repository in the link above.

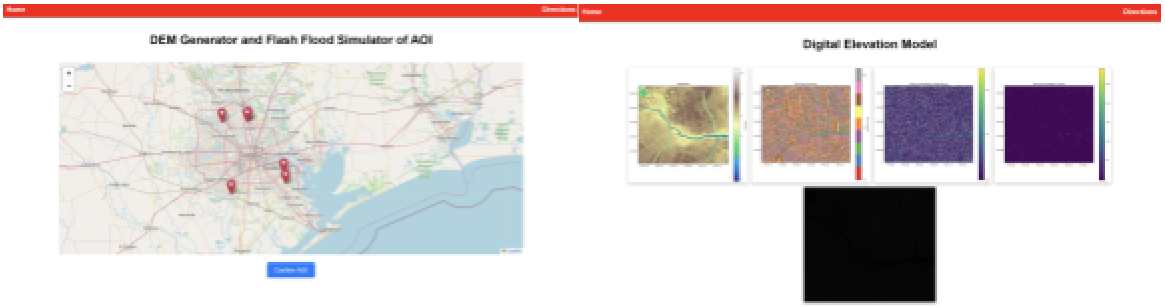


Figure 3: Website showing map with AOI center points and DEM of AOI #1.

IV. Results

The Shuttle Radiation Topography Mission (SRTM) is an elevation dataset created by a collaboration of NASA, NGA, and German/Italian space agencies in February 2000 using radar interferometry. With a raster resolution of 30 meters, OpenTopography is able to use this nearly global dataset to create a DEM of any highlighted area. When running the flooding simulation, instead of using all of Houston, the scattered AOI data was divided into groups of six to make a focused simulation. Since the end product is a website where users can see the effects of flooding in certain areas of Houston, using graphs and images that follow a smaller area makes for an easier experience and analysis.

GLOBE Observer has been taking pictures from large groups of civilians for several years in a variety of places, it allows for a widespread collection of data, which would be difficult to accomplish with just the resources available. Houston has a high AOI data concentration, allowing for lots of elevation data to be extracted. A benefit of relying on a civilian data resource is the ability to get a close-up of areas– something satellites are unable to achieve. Getting data from the ground instead of from a bird's-eye view also allows for easier height comparison, making elevation data easy to acquire. SEES interns are invited to take pictures of a 3 by 3-kilometer square in their area, also known as an AOI, and upload it to the GLOBE Observer database, the civilian data used for this model. Taking pictures of all cardinal directions, as well as directly up and down, allows for a cohesive model of the area, and therefore, a way to pull the height from these pictures. As such, the area chosen for this simulation was Houston, Texas due to the large number of pictures taken from citizen scientists

After running 6 Houston-area AOIs through the Python program, the following results were achieved:

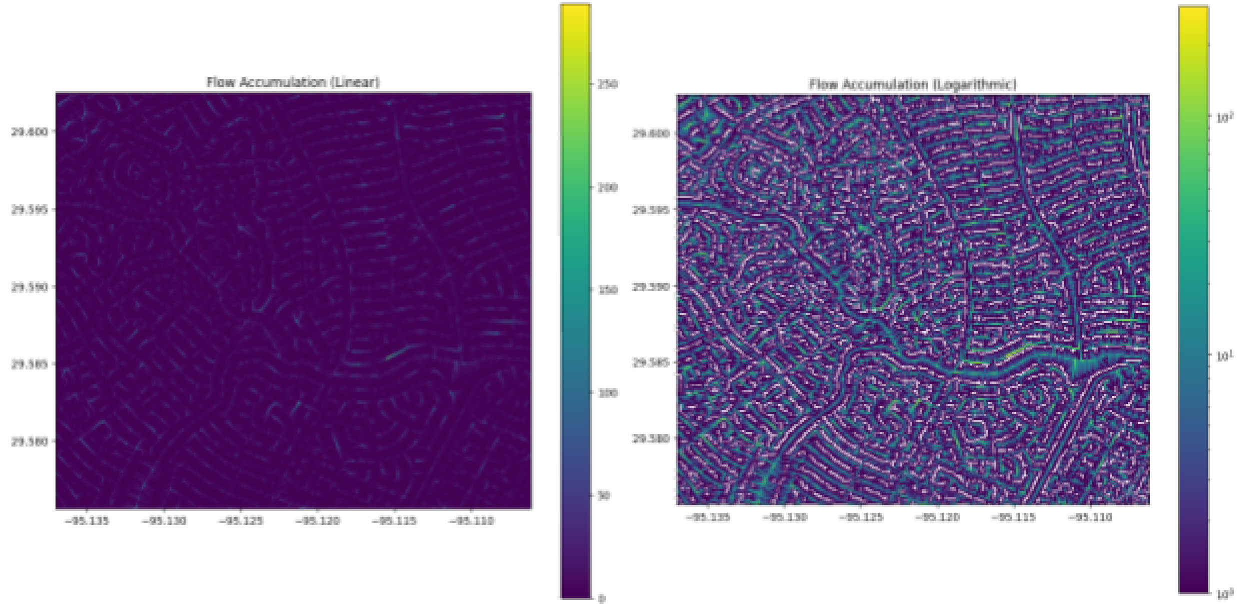


Figure 4: Test AOI #3

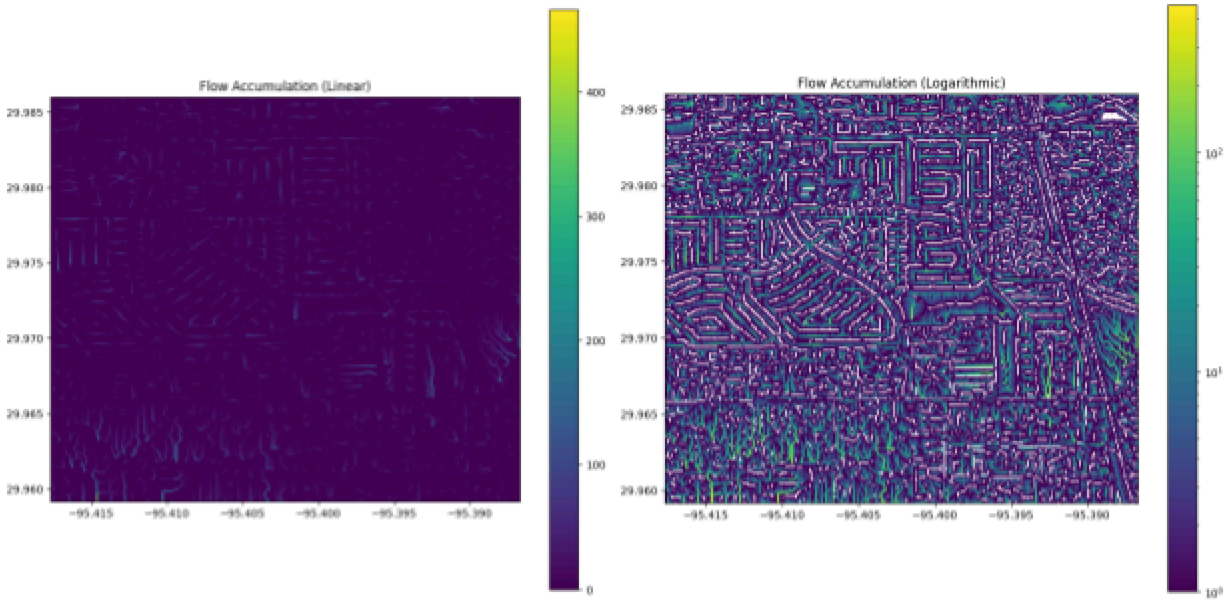


Figure 5: Sample AOI #8

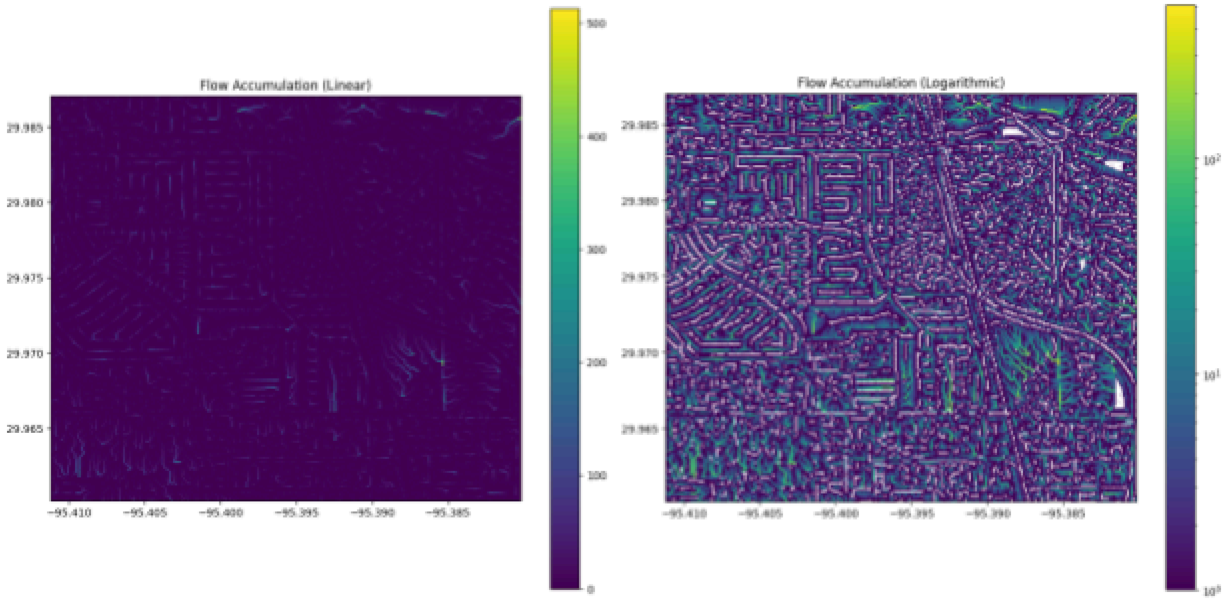


Figure 6: Sample AOI #13

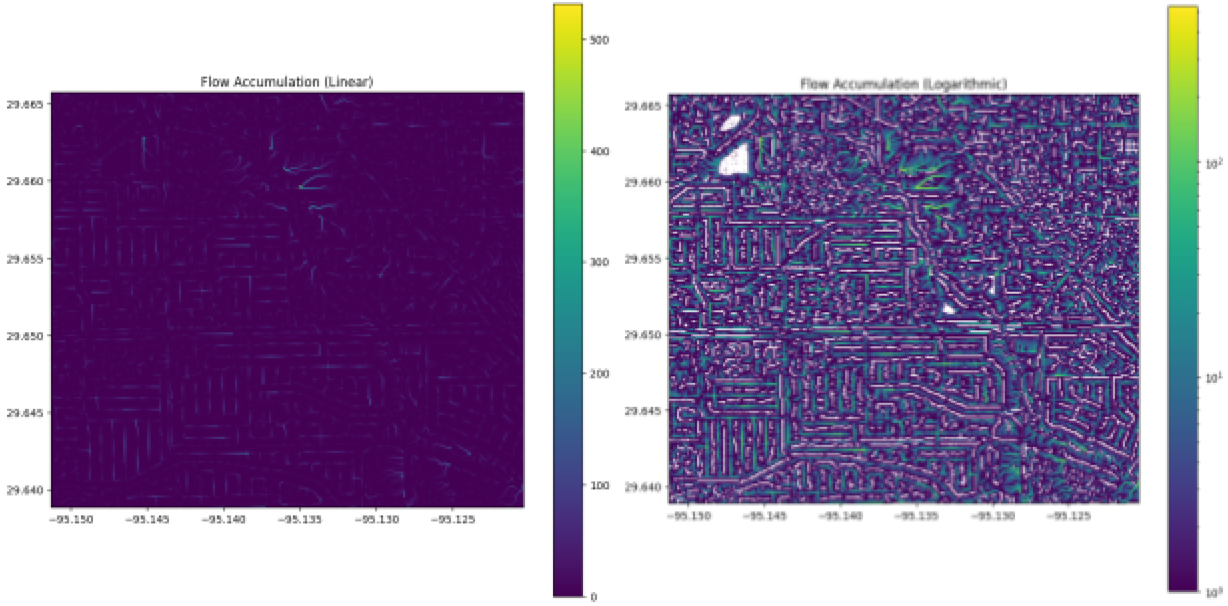


Figure 7: Sample AOI #24

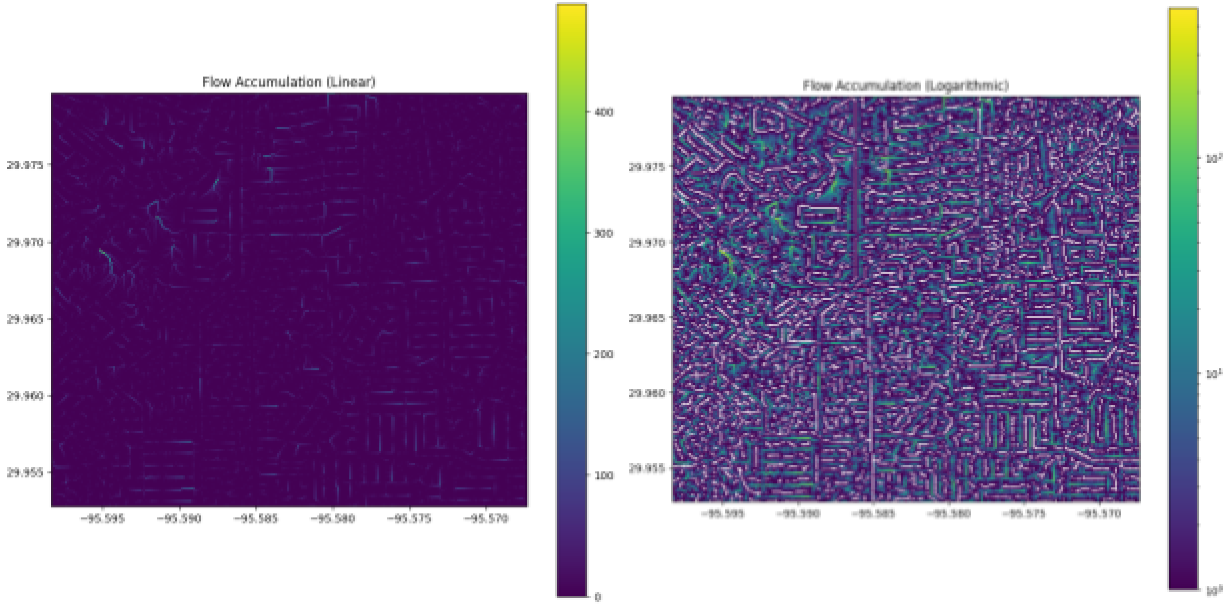


Figure 8: Sample AOI #35

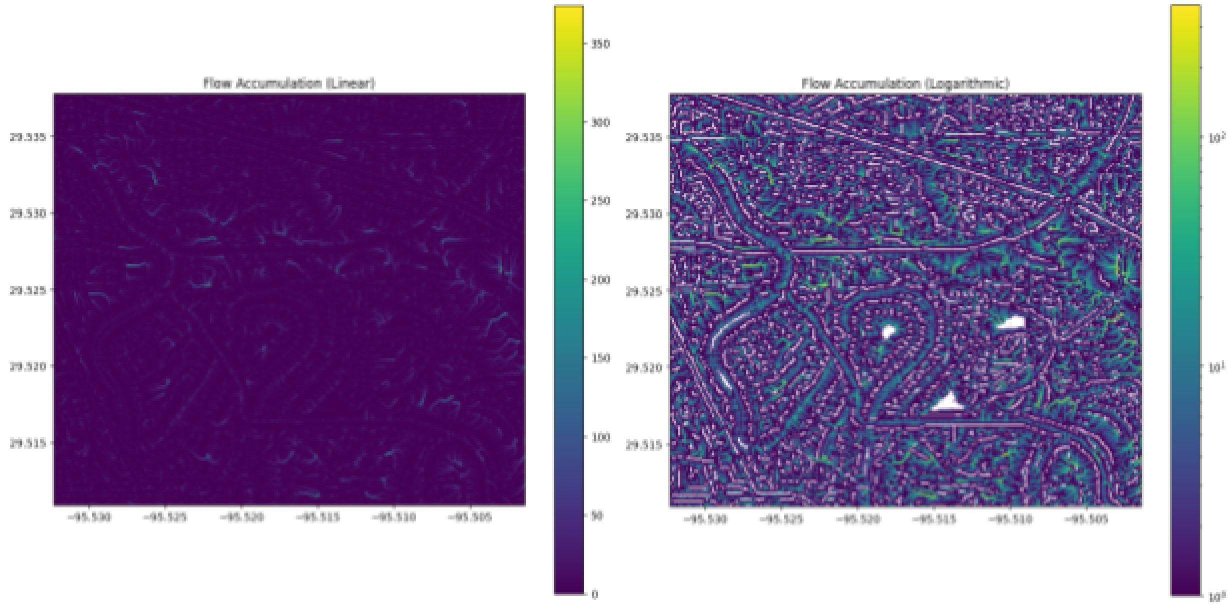


Figure 9: Sample AOI #37

These graphs depict the accumulation of water in the isolated AOI region in two formats, linear and logarithmic. The linear graphs use a linear color distribution to display, at each cell, the amount of other cells that contribute to the first cell. The logarithmic graphs demonstrate the same data but with a logarithmic color bar. This way, cells that accumulate an extremely large amount of water are more clearly depicted. Furthermore, white spots in the logarithmic plots depict cells that accumulate 0 other cells, meaning they are unlikely to collect much water in case of a flood.

In AOI #35 (Figure 8), for example, we found that water appears to accumulate most heavily in streets, roads, and other areas high in impervious surfaces. The spacing of white cells indicates that water is unlikely to accumulate in homes and other high-elevation areas. This is because streets and roads are often lower than their surrounding buildings and thus accumulate more water. Thus, citizens and urban planners in AOI #35 should be aware that water is more likely to accumulate, in case of heavy rainfall, on their roads. They should therefore place drains in the roads as well as ensure that these drains stay unclogged to prevent flooding.

V. Discussion

The finished model served as a tool to find areas at high risk for flash flooding in Houston, Texas. Both linear and logarithmic models show a clear relationship between the elevation of each AOI point and the flow direction of precipitation. The relationship highlights how water will flow into the depressions on the map, resulting in a comprehensive map of where the floodwater is expected to pool. The models also show the most elevated locations where there is little chance of risk for flash flooding.

In comparison to data and previous models, the data types collected are very similar, and the models both take a very closely related approach to mapping areas. However, our model takes into account elevation and fluid mechanics as the primary data source, whereas this model (Figure 10) from Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) takes precipitation levels, folding depth, and previous damage as the primary data sources. Additionally, CMCC's model downscales AOI points at a 2 km scale for high resolution, whereas our raster resolution, which is at 30 meters, is not as precise as it could be. Theoretically, elevation should not undergo a drastic change, however, each 900 m^2 area is one pixel on the map, meaning the data taken from the DEM is not as pinpoint accurate as the civilian data can be, such as the height of individual buildings.

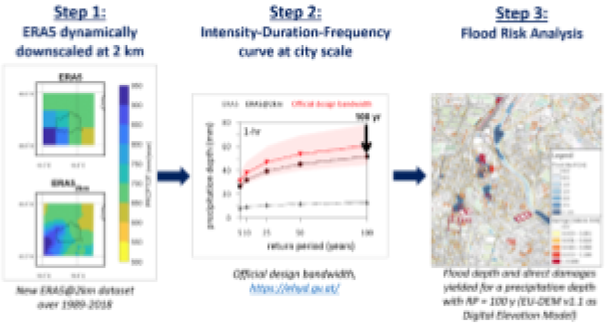


Figure 10 (Source: Copernicus)

This newly developed model suggests that precipitation flows into the streets, possibly because they are built slightly lower than the elevation level of urban infrastructure. These results could be influential to policymakers in improving the storm drainage system on roads and advocating for policies that keep the roads free of excess water. With these policies, more studies on pluvial flash flooding could be funded or approved by the government or private companies, as they become more relevant to the world of politics.

Due to the nature of how this research was conducted, this model has some potential for errors and oversight due to constraints and limitations in said research. As high school students, there was very limited monetary funding available to conduct our research. This affected the decisions made during the research process. For example, some of the software and datasets first outlined could not be used because they cost a significant amount of money, which was simply not available. Another limitation involved the nature of our communication. During the length of the research process, our communication was not as optimal as it could have been, had the team met in person to develop the model. Furthermore, the timeline for this research process was quite limited (about two weeks), which may have influenced the quality of our work and the help available to us. If there happened to be a longer time frame to further develop the model in the future, it could be possible to test larger areas, compare them to previous flooding, and improve the Python programming.

Additionally, since GLOBE Observer is composed of civilian-taken data, not each picture will be up to par. Some AOI points are missing all six pictures, forcing an incomplete evaluation of the area. We had to create a filtering algorithm that took into account the amount of terrain that was shown and also the quality of the images. Since we are applying various deep learning techniques to the images, the resolution is also very important to get accurate data. We have set a threshold for the resolution to make sure it isn't too low.

Lastly, the data on OpenTopography that was used was measured 24 years ago, meaning the urban infrastructure in Houston could have changed, making some pixels inaccurate, while the AOI data is more recent.

As a research team, we hope that our model is used and progressed on with more research from the scientific community, to eventually become a solution to the flaws in flood mapping. With time, factors such as precipitation levels, previous damage, and natural disasters could be added to the database, providing a fixed algorithm for predicting flash floods in the near future. The research conducted in the future could be crucial to advancing cities, preventing severe damage, and, most importantly, improving residents’ lives.

VI. Conclusion

In conclusion, this model has succeeded in mapping and identifying high-risk areas for flash flooding in the United States. While the algorithm for this tool can be complicated for the average person, the accessibility of this tool will make it possible for anyone to access a simulation for any set of points covered by GLOBE Observer. This research is the product of data provided by citizen scientists, which makes it more customized and familiar to the everyday citizens who will use it. By promoting the widespread use of this software, both in the scientific community and in the public domain, more awareness of flash flooding can be directed toward voters and policymakers for changes in flood prevention.

With some more work, this tool has the capacity to be on the cutting edge of weather and flooding maps, ultimately serving to protect the people living in high-risk areas and to prevent as much damage as possible. This research will empower residents by providing them with a public tool so they can make their own decisions about flood prevention, rather than depend on the statement of a single agency. Most importantly, the residents in the area will be able to regain control from the grip of unexpected floods, so they can enjoy what is important in their lives.

VII. References and Acknowledgements

Rosenzweig, B. R., McPhillips, L., Chang, H., & Welty, C. (2018, May 11). *Pluvial Flood Risk and Opportunities for Resilience*. ScholarWorks@GSU. Retrieved July 17, 2024, from https://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1036&context=urban_studies_institute

Mercogliano, P. (n.d.). *Pluvial Flood Risk Assessment in Urban Areas | Copernicus*. Copernicus Climate Change Service. Retrieved July 19, 2024, from https://climate.copernicus.eu/pluvial-flood-risk-assessment-urban-areas

Siqueira, V., Fleischmann, A., Jardim, P., Fan, F., & Collischonn, W. (2016). IPH-Hydro Tools: a GIS coupled tool for watershed topology acquisition in an open-source environment. In Revista Brasileira de Recursos Hídricos (Vol. 21, Issue 1, pp. 274–287). FapUNIFESP (SciELO). https://doi.org/10.21168/rbrh.v21n1.p274-287

Oakford, S., Muyskens, J., Cahlan, S., & Lee, J. S. (2022, December 6). *FEMA flood maps fail to show flood risk of more extreme flooding events*. Washington Post. Retrieved July 19, 2024, from https://www.washingtonpost.com/climate-environment/interactive/2022/fema-flood-risk-maps-failures/?itid=lk_inline_manual_5

NASA Shuttle Radar Topography Mission (SRTM). (2013). *OpenTopography - Shuttle Radar Topography Mission (SRTM GL1) Global 30m*. MyOpenTopo. Retrieved July 28, 2024, from https://portal.opentopography.org/raster?opentopoID=OTSRTM.082015.4326.1

U.S. Geological Survey. (n.d.). *LP DAAC - SRTMGL1*. LP DAAC. Retrieved July 28, 2024, from https://lpdaac.usgs.gov/products/srtmgl1v003/

Cheng J, Schloerke B, Karambelkar B, Xie Y (2024). leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.2.2.9000, https://github.com/rstudio/leaflet, https://rstudio.github.io/leaflet/.

A Special Thanks To Dr. Russanne Low, Ms. Cassie Soeffing, Mr. Peder Nelson, Dr. Erica Podest, Mr. Andrew Clark, Andrew Liu, Benjamin Hershman, and all others who provided guidance throughout the course of this research project.

Creating a Model to Predict High-Risk Areas for Pluvial Flash Flooding in the Urban Areas of Houston, Texas

Students: Dae San Kim, Sanan Khairabadi, Aadya Jain, Gabriela Gomez-Hernandez, and Nishchay Jaiswal

Mentors: Dr. Rusanne Low, Cassie Soeffing, Peder Nelson, Andrew Clark, Andrew Liu, and Benjamin Herschman

STEM Enhancement in Earth Science (SEES) Summer High School Intern Program

United States of America

9 August 2024

IVSS Badges

I am a Problem Solver: Our project deserves the problem solver badge because it uses Earth science to address the problem of flash floods, which can cause billions of dollars in damages and devastate entire communities. By using Earth science methods such as flow analysis and elevation data, we designed a model to predict which points in a given area are at highest risk of flooding. This model provides valuable insights to urban planners, policymakers, and citizens, helping them to plan for flash floods and mitigate their effects.

I am a Data Scientist: Our group took raw elevation data from select regions around Houston and determined which areas within those regions are at highest risk of flooding. To do this, we created a Python program to determine the direction of water flow at each point in the region. We used this new data to track water accumulation at each point, then identified the regions which had the most accumulation. From analyzing this data in the form of graphs, we found that surfaces like streets and roads were most likely to flood.

I am a STEM Storyteller: Our primary programmer wrote a summative blog in Mighty Networks outlining how publicly available elevation data can be used to determine high-risk areas for flooding. It is available here: https://earth-system-explorers.mn.co/posts/how-to-predict-water-flow-based-on-elevation.

Our website programmer wrote a blog outlining the preparation and learning done within the internship to prepare for the project, as well as the actual project summary. The blog can be found here: https://earth-system-explorers.mn.co/posts/63474015?utm_source=manual.

Sharing the process through blogs like this can enable citizen scientists to also use our tool and safeguard their homes and communities.