Natalie Memarsadeghi

Final Project Paper

GEOG788P

12/09/21

Creating a Time Lapse Visualization to Display Runoff Volume during a Rain Event in Warsaw, NY

Introduction

Flooding occurs naturally when there is excess water on land that is typically dry, and can put people, property, and livelihoods at risk. Flooding can be caused by heavy rain and storms, ocean tides, or when flood control infrastructure such as levees fail. Flooding is recognized as one of the costliest natural hazards as the annual cost of flood damages exceeds $2 billion, and causes the death of about 100 people a year in the United States (Ward et al., 2020, United States, n.d.). From 1980-2009 there were 539,811 deaths, 361,974 injuries, and 2,821,895,005 people affected by floods (Doocy et al., 2013). Additionally, flood impacts are only predicted to increase as we experience climate change. A study by the World Resources Institute modeled the economic and human impacts of flooding for three different climate scenarios, and found that the economic impacts of floods could as much as triple by 2050 (Ward et al., 2020).

Due to the large risk that flooding can cause, this study aims to visualize surface runoff volume for a rain event in upstate New York. Surface runoff is defined as all overland flow (Ramirez, 2000). Additionally, visualizing spatial data is a strong tool in understanding, analyzing, and making decisions based on data. Approximately 80% of all digital data includes geospatial referencing such as coordinates, addresses, etc, and there is an inherent complexity and interdisciplinary nature to natural and social sciences, creating a need for geovisualization (MacEachren and Kraak, 2001). In the context of this study, flood management is an interdisciplinary problem that involves local residents, businesses, government agencies, and environmentalists, and is location specific. Therefore, visualizing runoff volumes during a rainfall event, will allow for informed decision making and policy. Flood modeling is valuable in risk assessments, resilience planning and evacuation needs (Rosenzweig et al., 2021). Additionally, flood modeling allows us to understand the probability of severe flooding from rain events, allowing for more effective planning. Lastly, flood modeling also illustrates the people and property at risk, and influences property values (Rosenzweig et al., 2021). This study does not look at a major flooding event, but rather a small rainfall event in Warsaw, NY, in order to establish the basic methodology needed to create this type of geovisualization.

The study site of this project is Oatka Creek, at Warsaw, New York. Oatka Creek experienced a rainfall event from noon on October 3rd to midnight on October 5th, and total cumulative precipitation for this rain event was 1.4 inches, according to USGS Water Watch data (USGS, n.d.). This site was selected due to the availability of data, and the scope of this project. The goal of this project is to calculate and visualize runoff volume for this rain event, in a time lapse gif of hourly intervals, using python.

Literature Review

In Andrienko et al., 2010, an overview of visual analytics is provided. They state that the key features of visual analytics research are: emphasis on data analysis, problem solving and decision making, applying automated techniques for data processing, active involvement of a human in the analytical process, and support for the provenance and communication of analytical results (Andrienko et al., 2010). These are all important aspects to consider throughout this project. This visualization provides data analysis and analytical results through the visualization of runoff volume. The time lapse gif allows for easier understanding of the data, and to further understand the hydrological process at this site. Automatic techniques are used to create this result, specifically a for loop that produces each individual scene that corresponds with each hour interval. And lastly, I myself have been very involved in the analytical process, meaning this study meets these key parameters of visual analytics.

Additionally in Andrienko et al., 2010, it is illustrated that the need for visual analysis is very large and diverse (Andrienko et al., 2010). They provide a hypothetical scenario of a severe summer thunderstorm passing through a town, and several affected stakeholders. First is the insurance analyst, and as a result of the storm, they require data from weather services to quantify the damage from the storm for an initial damage assessment. Additionally, to assess agricultural impacts, satellite imagery is used. The second party described is a family, whose car was damaged, and would like to assess their exposure to risk. Third is the decision makers, who have to decide how to protect their citizens from future floods. And last is the community in general, who is increasing awareness of risk (Andrienko et al., 2010). All of these parties are linked by the need for spatio-temporal analysis. Therefore a geo-visualization would be beneficial to all, illustrating the large need for these tools.

In the context of this study, there are several parties that might find these results useful. The local community benefits from understanding the hydrological processes of the area better, and policy makers will have more information for decision making. Insurance companies and property owners will use the information for determining rates and values. Lastly, hydrologists and environmentalists can use this information to understand the ecosystem and natural storm management.

However, there are several identified common problems of geovisualization that are necessary to consider in this project. For example, in Andrienko et al., 2007, the complex nature of the geographic space is introduced, with one of the main problems being the heterogeneity of physical spaces. In order to overcome this issue, it is necessary to distinguish the differences within physical space clearly (Andrienko et al., 2007). This proved to be a problem faced in this project. The original goal of this project included displaying runoff extent as well as volume. However, the data source used, USGS Water Watch, collected the data from a fixed location, a stationary rain gauge, meaning that the values collected are only precise to that exact location. Because the landscape and slope differs along the stream, it would be too large of an assumption to visualize runoff extent other than at that precise point using this data, and it was decided to only include volume in this visualization.

A second challenge presented in this project that has been discussed in literature is issues with outcomes and applications, specifically drawing meaning and reason from the visualization you are viewing. In Robinson et al., 2017, "developing visual analytical reasoning systems that help users add meaning to and organize what they discover from geospatial big data" was identified as a key challenge of geovisualizations (Robinson et al., 2017). Because this project has lost the parameter of runoff extent, it is a bit more challenging to draw conclusions when viewing the final results. The unit cubic feet is not commonly used in everyday life, and could be challenging for some to interpret. Similarly, in Çöltekin et al., 2017, through workshops with 72 experts in the field, the most persistent issues of geovisualization were determined, one being outcomes and applications, meaning there is a lack of connection between research and practice

(Çöltekin et al., 2017). This reinforces the challenge of drawing meaningful conclusions from visualizations.

Several other studies have been conducted with the similar goal of runoff modeling or the modeling of other hydrological properties. The U.S. Environmental Protection Agency has developed the Storm Water Management Model (SWMM), which has been established and used for decades. It is used to simulate single events and long term hydrologic events in urban settings. SWMM models hydrologic processes that generate runoff such as rainfall, snow melt, evaporation and infiltration, and models water routes through a hydraulic network including channels, pipes and pumps (Mcdonnell et al., 2020). There are however some limitations of this model, as it does not allow the user to interact with the SWMM model during simulation, and does not allow for access to all of the results. In Mcdonnell et al., 2020, the PySWMM library is introduced, and allows access to the SWMM model using python (Mcdonnell et al., 2020). It additionally enhances the model, and allows for user interaction and facilitates rapid prototyping. PySWMM provides the editing of network and hydrologic parameters, which allows users to streamline model optimization, controls and results post processing (Mcdonnell et al., 2020). PySWMM is actively used in industry, academia and government.

The Storm Water Management Model (SWMM) has been adapted and expanded for use in varying applications in several studies. In Rezazadeh Helmi et al., 2019, SWMM is merged with WetSpa-Python to create WetSpa-Urban, which is used for the modeling of continuous rainfall-runoff in urban areas (Rezazadeh Helmi et al., 2019). As stated by Rezazadeh Helmi, et al., WetSpa-Python is “an open-source, fully distributed, process-based model that accurately represents surface hydrological processes but does not simulate hydraulic structures” (Rezazadeh Helmi et al., 2019). SWMM however does simulate hydraulic structures through its network, so by combining these two tools in WetSpa-Urban, it allows for precise runoff estimations in urban areas. In this study, Wet-Spa Urban was applied to the Watermaelbeek catchment in Brussels, Belgium, an area that recently underwent rapid urbanization. The modeling was executed in three different stages, pre-processing, surface runoff calculations, and routing through hydraulic structures. Due to the high level of heterogeneity of the study area, runoff calculations were performed in a spatially distributed method, and the surface runoff was calculated at the sub-catchment level, rather than at the watershed as a whole (Rezazadeh Helmi et al., 2019). The results showed slight over estimation, and were validated using Nash–Sutcliffe efficiency and model bias.

While these models use input data that is collected through typical hydrologic monitoring, in Weeser et al., 2019 using crowd sourced water level data for hydrological modeling is explored (Weeser et al., 2019). Typically, complex and costly discharge measurements are required as inputs for hydrological modeling, but water level measurements are straightforward and can be collected using a crowdsourcing approach. In this study, six different calibration schemes based on monitored discharge values and crowdsourced water levels were used to assess the validity and value of using crowdsourced data for hydraulic models. The study area selected was the northwestern part of the Sondu‐Miriu‐River Basin in Western Kenya, using the temporal scale of January 1st- March 31st 2016 (Weeser et al., 2019). They use the Catchment Modelling Framework Version 1.4.1 for Python 3 introduced in Kraft et al., 2018. The results found that “A conceptual rainfall‐runoff model can be calibrated on crowdsourced water level data. The combination of
crowdsourced data and a rainfall‐runoff‐model might solve an often raised critical point when using crowdsourcing in hydrology, that is, data irregularity”(Weeser et al., 2019). Using

crowdsourced data could expand the field of hydrological modeling through the use of new data collection methods.

Data Sources and Limitations

There are three main data sources used for this project. The first is USGS Water Watch which provides the storm data that is necessary (USGS, n.d.). This includes the needed parameters of precipitation in inches and the average discharge in cubic feet per second, both in 15 minute intervals in a csv format. Secondly, NAIP imagery was obtained from USGS Earth Explorer to provide a basemap for this area, seen in Figure 1. Lastly, google earth was used to digitize and create a shapefile of the stream, seen in Figure 2.

Figure 1: NAIP Imagery used for Base Map

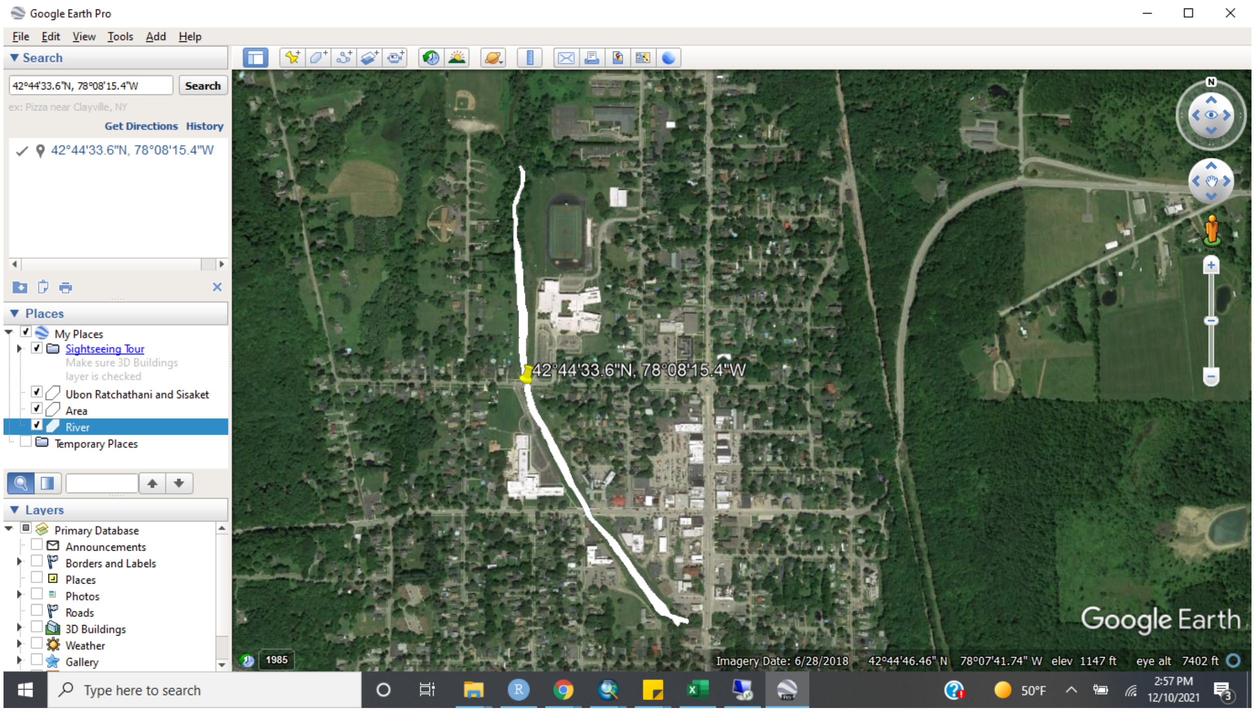

Figure 2: Digitized stream using Google Earth, point indicating location of the rain gauge

The hydrologic data that was collected by USGS was collected at a stationary rain gauge located at 42°44'33.6" N, 78°08'15.4" W (USGS, n.d.). Data collection has occurred here since 1963, and the available parameters at this location were rain gauge height, precipitation and average discharge. At this rain gauge, the recorded maximum discharge occurred on July 8th, 1998 and was 4,110 cubic feet per second. The recorded minimum discharge occurred on September 28, 2007 and was just .87 cubic feet per second. The time period used for this study is October 3rd- 5th, 2021, and the maximum average discharge during this event was 926 cubic feet per second at 7:15 pm on October 4th. There is a spatial limitation with this data source due to the data being collected at one location. Because the data is collected at the location of the rain gauge, it is only precise to that exact location, and does not explicitly inform us of the differences in data along the stream. It was for this reason that runoff extent was not an included parameter of this project.

While the data collected from USGS was used in the runoff volume calculations, data for the visualization of the runoff volume was also needed. The National Agriculture Imagery Program acquires aerial imagery in the continental U.S. throughout the agricultural growing seasons (NAIP Imagery, n.d.). NAIP imagery from August 25th, 2019 was downloaded from USGS Earth Explorer to use in this project as a base map (Figure 1). Secondly needed for the visualization was a shapefile of the stream itself. To acquire this, Google Earth was used to vectorize the stream. Approximately 1 mile of the stream was digitized, half a mile north and half a mile south of the rain gauge (Figure 2). This presents a small limitation of data, because as the river was manually digitized, there lies a certain level of human error. However, because this shapefile is only used for visualization purposes, this small amount of error is acceptable.

Methods

After acquiring the data from the before mentioned sources, the first step of the project was to clean and format the data. For the storm data, this included cleaning the csv file to only include the necessary parameters by deleting unnecessary columns, and renaming the columns to correspond with the given parameter, seen in Figure 3. Some deleted columns include a column that specified that the data was collected in eastern time, the site ID, and the gauge height, as this was not needed in the project. The NAIP imagery was clipped and reprojected to correspond with the spatial scale and coordinate system of the river shapefile. There was no necessary pre-processing of the river shapefile.

	Site_Number	Date_Time	Discharge	Gage_Height	Precip
29	4230380	10/3/2021 0:00	12.5	3.21	0
30	4230380	10/3/2021 0:15	12.5	3.21	0
31	4230380	10/3/2021 0:30	12.1	3.2	0
32	4230380	10/3/2021 0:45	12.1	3.2	0
33	4230380	10/3/2021 1:00	12.1	3.2	0
...	...	...	...	...	...
312	4230380	10/5/2021 22:45	109	4.05	0
313	4230380	10/5/2021 23:00	107	4.04	0
314	4230380	10/5/2021 23:15	107	4.04	0
315	4230380	10/5/2021 23:30	105	4.03	0
316	4230380	10/5/2021 23:45	105	4.03	0

Figure 3: Data collected from USGS after cleaning

The next step of this project was calculating runoff volume. USGS provides discharge rates for each 15 minute interval in the unit of cubic feet per second. From these values, the total volume was able to be calculated. Runoff volume can be calculated by subtracting the baseflow, or the discharge that naturally occurs, from the current discharge, or the discharge that is occurring during the storm (Ramirez, 2000). This was performed in python by creating a new column that was defined by this equation. The baseflow of this site was 12.1 cubic feet per second, so each row within the discharge column subtracted this value and the results were added to their own column. This new column was then multiplied by 15 and 60 to determine the total volume of discharge within each 15 minute interval, by converting from cubic feet per second to cubic feet per 15 minutes. A new dataframe was then made that included cumulative time, or hours passed after the start of the rain event. Lastly, every four rows were summed in order to determine the total runoff volume per hour and added to the newly created dataframe. This resulted in a dataframe with two variables, hours that have passed from the start of the storm, and the corresponding runoff volume for each interval.

From there, the next step was creating the visualization. The first step for this was transposing the dataframe that was previously created, so that each column name is the hour, and the runoff values were all in one row rather than one column. This is necessary for the loop that was used to create the maps for each hour. Next, the river shapefile and the NAIP imagery basemap were loaded into python. The transposed dataframe was then joined with the river shapefile on the attribute of “SymbolID”. The symbol ID was a pre-existing parameter of the river shapefile, so an identical column was added to the dataframe to perform the join. Lastly before starting the loop, several parameters were defined. This includes the output path for each individual map, the minimum and maximum data values in order to keep a consistent scale in each individual map, and defining the column names as one variable, or a list of each hour of the storm, in order to read through in the loop.

Once these steps were completed, a for loop was run to create 73 individual maps in .png format. First the NAIP imagery was plotted as the basemap, and secondly the river shapefile was plotted. Through each iteration of the for loop, a unique map was created using the data of each of the predefined hours. The data scale is consistent through each image because each iteration of the loop uses the same predefined scale. An annotation was added in the bottom left corner that depicted how many hours after the start of the storm had passed in each image. Each image was saved to the same folder.

After each image was produced, the last step of the visualization was compiling them in a gif. To do this, first using the command “glob.glob”, all the png files that were created were read into a list. Next, using the function "natsorted", each image was sorted by numerical order so that the images would play in sequential order. Each image was then added in this order to an empty frame using a for loop, and then the completed fram was saved as a .gif file.

Due to the large range of data values in the runoff volume, specifically 0 to 3,500,000 cubic feet, during the first 30 hours of this rain event when runoff is low, it is difficult to distinguish the difference in these values. In order to account for this large range, a second gif was produced using the log base ten values of runoff volume through the same process. After taking the logarithmic values for runoff volume, the scale was reduced to 0-7, allowing the viewer to more distinctly visualize the rainfall event.

Results

In Figure 4, the cumulative precipitation throughout the rain event is displayed. The total cumulative precipitation for this rain event was 1.4 inches, and occurred 41 hours after the start of the rain event. There is a steady increase in precipitation until approximately hours 16-19, where there is a sharp increase in precipitation. Precipitation again remains steady until hour 39, where there is another sharp increase in precipitation, until the maximum is reached after 41 hours.

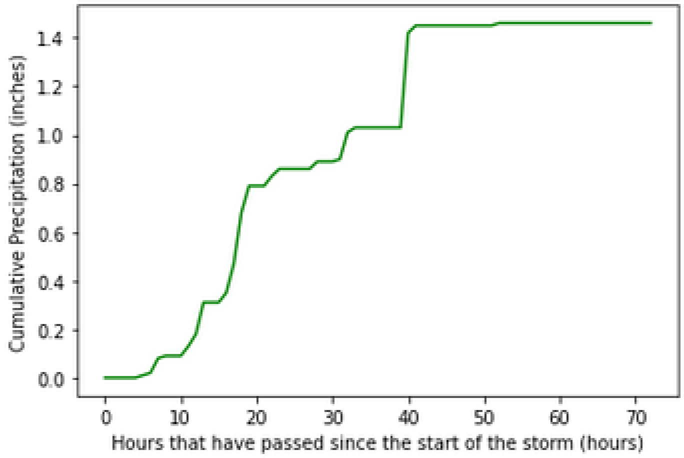

Figure 4: Cumulative precipitation

In Figures 5 and 6, runoff volume is displayed. Figure 5 displays the actual runoff volume, and Figure 6 displays the log values of the data. The peak runoff volume is reached at 43 hours after the start of the storm, with two time periods of sharp increase. The first is from hours 32-34, and the second is from hours 40-43. These two peaks follow 10 hours and 3 hours after the two peak precipitations respectively, indicating the lag time between maximum precipitation and peak runoff.

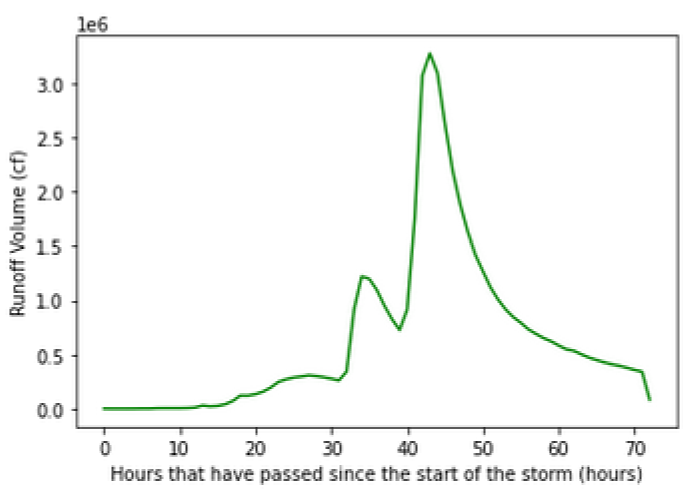

Figure 5: Runoff volume

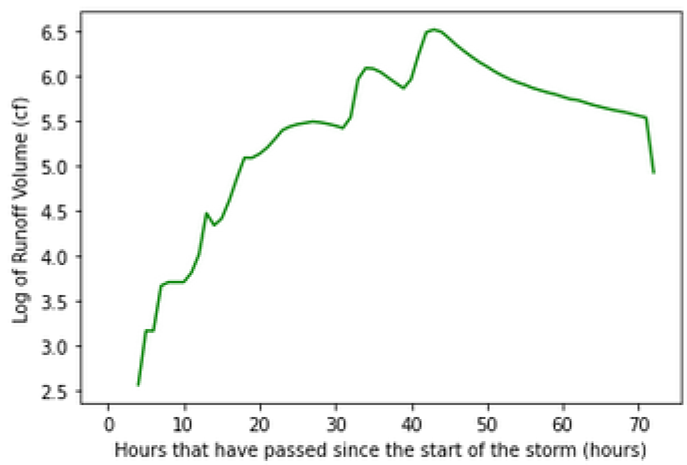

Figure 6: Log values of Runoff Volume

In total, 146 unique maps were generated for this project. The rain event lasted for a duration of 73 hours, so 73 individual maps were created for each hour of the storm using both the actual runoff volumes, and the log based values. These maps were successfully compiled into two gifs, which can be found in the github repository for this project, as well as the individual maps. Visually from the gifs, the peak runoff is discernible.

Discussion

Flood mapping proves to be an important tool in today's environment, and geovisualizations are essential to effectively communicate important information. While there are several established methods for flood modeling, the methodology provided here is a straightforward way to visualize past storms. This is necessary in order to effectively monitor and manage the ecosystem. Overall, this project was successful in calculating the runoff volume for the rain event that occurred at Oatka Creek in Warsaw, NY from October 3rd-5th, and in creating time lapse visualizations to display both the actual runoff values as well as the log based values. By creating both of these visualizations, we are able to see the progression of runoff volume throughout the rain event, and can more clearly interpret and visualize these values when using the log based values. These visualizations provide insight to the hydrological processes of the watershed and ecosystem, allowing us to draw meaningful conclusions. The data provided here is beneficial to storm management operations, the local community and local government.

There are several limitations of this project that have been briefly discussed. The primary limitation of this project is that the storm data used is not spatially distributed, but rather collected at a fixed location. This prevented the parameter of extent to be included in the study. If discharge were obtained in a spatially distributed format such as a raster layer, this would have been more feasible. However, this limitation does not influence the runoff volume that is displayed in the visualization. Secondly, because the stream bed was digitized manually, there is a certain level of human error we can expect. However, because the stream shapefile was used purely for visualization purposes to display the location of the stream, this small level error is not concerning and does not affect the results.

There are several ways in which this research could be continued and expanded on in future work. Other hydrological processes, such as the precipitation data could be visualized in the same manner. This would allow the viewer to watch both gifs simultaneously and gain a deeper understanding of how the hydrological processes of this area interact with one another. Additionally, this methodology could be replicated for different rain events at the same location or at different locations. This would allow for a comparative watershed analysis, by either comparing the runoff volume of different rain events at Oatka Creek, or through comparing the rain event at Oatka Creek with different watersheds. Lastly, social components such as population, housing infrastructure and economic parameters could be considered for a holistic analysis of the area. Through the results of this study, we are able to understand the hydrological process of runoff of Oatka Creek, and provide the framework for future work.

Bibliography

G. Andrienko , N. Andrienko , P. Jankowski , D. Keim , M.‐J. Kraak , A. MacEachren & S. Wrobel (2007) Geovisual analytics for spatial decision support: Setting the research agenda, International Journal of Geographical Information Science, 21:8, 839-857, DOI: 10.1080/13658810701349011

Gennady Andrienko , Natalia Andrienko , Urska Demsar , Doris Dransch , Jason Dykes , Sara Irina Fabrikant , Mikael Jern , Menno-Jan Kraak , Heidrun Schumann & Christian Tominski (2010) Space, time and visual analytics, International Journal of Geographical Information Science, 24:10, 1577-1600, DOI: 10.1080/13658816.2010.508043

Arzu Çöltekin, Susanne Bleisch, Gennady Andrienko & Jason Dykes (2017) Persistent challenges in geovisualization – a community perspective, International Journal of Cartography, 3:sup1, 115-139, DOI: 10.1080/23729333.2017.1302910

Doocy, S., Daniels, A., Murray, S., & Kirsch, T. D. (2013). The human impact of floods: a historical review of events 1980-2009 and systematic literature review. *PLoS currents*, https://doi.org/10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a

Alan M. MacEachren & Menno-Jan Kraak (2001) Research Challenges in Geovisualization, Cartography and Geographic Information Science, 28:1, 3-12, DOI: 10.1559/152304001782173970

McDonnell et al., (2020). PySWMM: The Python Interface to Stormwater Management Model (SWMM). Journal of Open Source Software, 5(52), 2292, https://doi.org/10.21105/joss.02292

NAIP imagery. USDA Farm Service Agency. (n.d.). Retrieved December 15, 2021, from https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/

Ramírez, Jorge A. "Prediction and modeling of flood hydrology and hydraulics." *Inland flood hazards: Human, riparian and aquatic communities* (2000): 498.

Rezazadeh Helmi, N., Verbeiren, B., Wirion, C., van Griensven, A., Weerasinghe, I., & Bauwens, W. (2019). WetSpa-Urban: An adapted version of WetSpa-Python, a suitable tool for detailed runoff calculation in urban areas. Water, 11(12), 2460

Rosenzweig, B. R., P. Herreros Cantis, Y. Kim, A. Cohn, K. Grove, J. Brock, J. Yesuf et al. "The value of urban flood modeling." *Earth's Future* 9, no. 1 (2021): e2020EF001739.

Anthony C. Robinson, Urška Demšar, Antoni B. Moore, Aileen Buckley, Bin Jiang, Kenneth Field, Menno-Jan Kraak, Silvana P. Camboim & Claudia R. Sluter (2017) Geospatial big data and cartography: research challenges and opportunities for making maps that matter, International Journal of Cartography, 3:sup1, 32-60, DOI: 10.1080/23729333.2016.1278151

United States, Department of the Interior, Geological Survey, "Floods and Floodplains," accessed December 11, 2021, http://pubs.usgs.gov/of/1993/ofr93-641/.

USGS current conditions for USGS 04230380 Oatka Creek at Warsaw NY. (n.d.). Retrieved December 15, 2021, from https://waterdata.usgs.gov/nwis/uv?cb_00045=on&cb_00060=on&cb_00065=on&format=gif_stats&site_no=04230380&period=&begin_date=2021-10-03&end_date=2021-10-05

Ward, P.J., H.C. Winsemius, S. Kuzma, M.F.P. Bierkens, A. Bouwman, H. de Moel, A. Díaz Loaiza, et al. 2020. "Aqueduct Floods Methodology." Technical Note. Washington, D.C.: World Resources Institute. Available online at: www.wri.org/publication/aqueduct-floods-methodology.

Weeser, B., Jacobs, S., Kraft, P., Rufino, M. C., & Breuer, L. (2019). Rainfall‐Runoff modeling using crowdsourced Water level data. Water Resources Research, 55, https://doi.org/10.1029/2019WR025248