A Framework to Analyze Flood Risk at a Continental Scale

Deb Smith
Penn State University
Advisor: Dr. Manzhu Yu
Abstract

Flooding is a consistently destructive issue, and as flooding increases due to climate change, more people will be affected by it. Flood maps show where floods might happen, but typically don’t show who and what is at risk. Flood risk assessments are needed that include hazard, exposure, and vulnerability data. The following is a proposal for a framework utilizing spatiotemporal analysis for flood risk assessments at a continental scale. Data used is SRTM elevation data and ECMWF runoff data, as well as various datasets showing exposure and vulnerability, including population density, poverty, and access to resources. This project utilizes spatiotemporal streamflow data to determine flood probability. By blending current methods in flood hazard prediction, with spatial data showing exposed and vulnerable populations, we can assess what areas are most at risk for flood events. The completed framework allows users to create a flood risk assessment of any region in the world because all the input data is globally available. This information can help leaders and organizations to make the proper preparations or do things to mitigate flood risk in the assessed areas of risk.

Background

In 2015, flooding displaced 150,000 people from their homes in South America (BBC, 2015). Due to El Nino, people from Paraguay, Argentina, Uruguay, and Brazil were forced to leave their homes. Floods are the most frequently occurring natural disaster, and from 1998 to 2017, floods affected over 2 billion people (Wallemacq and House, 2018). Flooding can have both a tremendous impact on humans and also cause extensive financial damage. Floods caused $656 billion in damages (UNISDR, 2017). The poor are disproportionately affected by flooding, making it difficult to recover (SAMHSA, 2017). According to the National Institute of Building Sciences, every dollar spent on preparing for disasters prevents $6 in spending on relief and recovery (NIBS, 2018).

According to the United Nations Global Assessment Report on Disaster Risk Reduction, risk is the combination of hazard, exposure, and vulnerability. Hazard is a dangerous phenomenon or natural disaster. Exposure is the structures, population, agriculture, business, or assets exposed to a hazard. Vulnerability factors are physical, social, economic, environmental, coping capacity, or adaptive capacity. All three elements are essential in understanding risk. Knowing what areas are most at risk would allow leaders, NGOs, etc. to preposition resources where they are needed most. For natural disasters, having more information and having enough time to act on it can mean life or death.

Flooding is a natural disaster caused by a variety of possible reasons. Excessive precipitation, snowmelt, dam break, tsunami or storm surge, etc. can cause flooding. Predicting where and when the next major flood will occur can be a difficult task. Flood maps display areas that could potentially be inundated in a flood event. Although it is helpful to know what areas would be affected, it isn’t always useful in understanding what areas are the most at risk. A site could be affected by flooding but not highly populated. Leaders need to know what areas are a priority if decisions need to be made regarding resources, aid, and prevention.
Floods are one of the most common natural disasters, and it's an issue that is getting worse. A study on the impacts of floods worldwide showed that the number of floods occurring worldwide is increasing. (Hu et al., 2018). Their findings indicated that more flood-related deaths arise not due to flood intensity but due to flood frequency. It is essential to understand what areas are impacted by frequent floodings. Since people are affected by frequency, understanding where that happens is necessary.

A spatiotemporal analysis is a standard method for analyzing flood risk due to studying when past flood events occurred and finding the frequency and probability of flood events for a particular area. Flood risk is often done on a specific flood plain or one specific area, not at the continental scale. One example of this is a spatial-temporal study done on flooding in Shanghai (Quan, 2014). The researchers used data from 251 to 2000, and it was event-based data. Another example is a study on the drainage basin of the Pinios River in Thessaly, Greece (Batherellos et al., 2018). This study also used historical flood events as the basis for their research. (Machado et al. 2015) also used historical data in their research to analyze flood frequency.

Another study used streamflow data with spatiotemporal aspects for their analysis (Adamowski et al., 2013). The purpose of their research was to use a spatiotemporal analysis of streamflow data to look at the effects of climate change on Canada's water. A method for determining the frequency of something is a Fourier transform. The technique used for spatiotemporal analysis was Continuous wavelet (CWT), and Cross-wavelet transforms (XWT). They claim that this kind of research is better than a Fourier transform because "wavelet coefficients are related to a specific period of time and frequency simultaneously...making it possible to trace the amplitude and phase of fluctuations of a specific wavelength through time." There are different methods for determining frequency.

A study was done on South America (Vorosmarty et al., 2013) using data from 1960-2000 and combined high resolution geophysical and population datasets, using a risk-based approach. The study's limitations are that it can't predict specific events due to studying at the continental scale. Their goal was to identify continental-scale tendencies. The researchers aggregated the results to the 1st Administrative boundary, so they do not show a significant level of detail.

Wood et al. studied flood scenarios and incorporated spatial and temporal conditions into their analysis for disaster response planning (Wood et al., 2016). They assessed the chances of flood events over many different areas in the U.K. and what their return periods were. Using a simulation model, the researchers determined the events within the desired joint probability band, then assessed the probabilities using the mean of all the simulated river flows and the Weibull formula.

De Moel et al. (2015) analyzed flood risk at different scales. There are flood risk assessments at the continental and global scale. Still, they usually have a lower spatial
resolution (de Moel et al., 2015) Winsemius created a framework for a flood risk assessment at a global scale. It is a high resolution of ~1km². He incorporated hazard, exposure, and vulnerability, but only analyzed assets at risk by using GDP and land use datasets, not necessarily the people that are more at risk (Winsemius et al., 2013). Dotteri created a global flood hazard framework but didn't incorporate risk factors (Dotteri et al., 2016). Alfieri made a flood risk assessment at the continental scale for Europe (Alfieri et al. 2014) and created a flood risk based on climate change (Alfieri et al. 2015).

Existing flood analysis methods and flood risk analysis predominantly display areas potentially inundated by flooding but not places that would be most impacted by flooding. What is needed is to create a flood risk analysis that shows locations most affected by flooding due to population density, poverty, lack of access to resources, etc. Situational awareness of flood risk would enable leaders and decision-makers to allocate resources to minimize floods' impact on vulnerable areas. Current research is missing high-resolution risk mapping, a framework that brings different methods together to create a risk assessment, and a tool that is easy to use.

Goals and Objectives

The goal of this project was to create a comprehensive framework for flood risk analysis using spatiotemporal methods. This framework aims to show the areas most impacted by flooding due to population density, poverty, lack of access to resources, etc. The goal of this project was not to reinvent the wheel in terms of flood prediction or analysis. Many of the processes and models within the framework have been utilized many times before. The goal was to bring together current flood prediction methods and analysis with exposure and vulnerability data into one framework so that anyone, regardless of knowledge or experience with flooding, can look at an area and understand the flood risk to that area.

This project goes through the steps of the framework with South America as the study area to demonstrate that the framework can accommodate a continental scale. A secondary goal was to then turn the framework into a tool that can create a flood risk assessment for any interest area. The tool's value would also be that globally available is used so the tool could be used for any part of the world, without the need for any pre-made data.

Methodology

Data and materials

There are seven primary datasets for this project/framework, all of which are globally available. Three are used to show flood hazards, one for exposure, and three for vulnerability. Table 1 demonstrates the input data used in this study, and Figure 1 illustrates these data on the maps. The streamlines for this dataset were created using SRTM elevation data. SRTM stands for Shuttle Radar Topology Mission. It has a spatial resolution of 1 arc-second (30 meters) and has global coverage. SRTM was the best available global elevation dataset. TanDEMx is another globally available elevation dataset but is less accessible. Previously, the
Engineer Research and Development Center (ERDC) created flood hazard data for South America and other areas in the world but using 90m SRTM data. One of this project's goals was to re-create this data, but using 30m SRTM data to create a higher resolution end product.

Gridded runoff data was retrieved from ECMWF. The ECMWF is the European Centre for Medium-Range Weather Forecasts. It uses its forecast models and data assimilation systems to reanalyze archived observations, creating global data sets describing the recent history of the atmosphere, land surface, and oceans. (ECMWF, 2019) ECMWF offers historical data created through reanalysis. The dataset used for this project was the runoff data at a period of 6 hours from 1984-2019. Runoff is the water that drains away, either over the surface (surface runoff) or under the ground (subsurface runoff). The sum of surface runoff and subsurface runoff is referred to as 'runoff'. Runoff is a measure of water availability in the soil and can be an indicator of drought or flood. This dataset and the streamlines were used as the inputs for the RAPID model to simulate streamflow.

A global land cover dataset was used from Copernicus Global Land Service. Landcover is used as part of the AutoRoute process and is vital for getting Manning's roughness coefficient. Manning's roughness coefficient represents the resistance to flood flows in streams and floodplains. It shows how quickly water can spread or travel over a surface.

Landscan was used for population data. Landscan comes from the Oak Ridge National Laboratory and is a global population distribution dataset with a 1km² spatial resolution. A global raster dataset was used to show increases in flooding from climate change, which was created from climate models (Hirabayashi et al., 2013). Access to resources was demonstrated with a global raster dataset showing the travel time to the nearest city (Nelson et al., 2019).

Poverty was shown by comparing areas with population density with nighttime lights (Ghosh et al., 2013). There were not any globally available datasets that show poverty. Still, a few studies show that poverty can be estimated by comparing areas with a population with nighttime lights, which was done for this project. For all these datasets, it was important for this project to use data that was not just specifically available for South America, but data that is globally available so that this framework can be used for any area in the world.
Figure 1. Input data: a) Population per 1km² b) Travel time to the nearest city. c) Flooding increases due to climate change. d) Poverty data. e) Watersheds. f) Land cover.

<table>
<thead>
<tr>
<th>Data source</th>
<th>Description</th>
<th>Spatial resolution</th>
<th>Temporal resolution/ Date of data creation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shuttle Radar Topography Mission (SRTM, NASA)</td>
<td>DEM</td>
<td>30m</td>
<td>2000</td>
</tr>
<tr>
<td>ECMWF</td>
<td>Runoff data</td>
<td>9km</td>
<td>6 hours</td>
</tr>
<tr>
<td>LandScan</td>
<td>Population Density</td>
<td>1km</td>
<td>2018</td>
</tr>
<tr>
<td>Nelson et al. (2019)</td>
<td>Travel time to nearest city</td>
<td>1km</td>
<td>2019</td>
</tr>
<tr>
<td>Hirabayashi et al. (2013)</td>
<td>Flooding increases due to climate change</td>
<td>18km</td>
<td>2013</td>
</tr>
<tr>
<td>NASA and Landscan</td>
<td>Poverty</td>
<td>1km</td>
<td>2018</td>
</tr>
</tbody>
</table>
Table 1. Input data description

Description of the study area

The study area for this project is South America. South America is an important study area for this project because "it is a continent with important emerging economies, expanding populations and urban centres, and a rising trend in flood-associated damage." (Vorosmarty, 2013) Disasters can also slow down or prevent growth and development. South America is also frequently affected by El Nino and la Nina, which are intricate weather patterns resulting from variations in ocean temperatures in the Equatorial Pacific. These weather patterns can sometimes cause flooding. South America is a continent with less available data than others. Other areas in the world have data such as rain gauges, and there's been a lot more research, so it is crucial to study and analyze the flood risk in South America.

Technology

The technology used for this project was ArcGIS, the ArcHydro Toolbox, and Python. These were used for analysis, creation of maps, and the tools for the framework. Python was used to decrease processing times. ArcHydro tools used to create the streamlines are fill, flow direction, flow accumulation, stream order, stream link, and stream to feature. More ArcHydro tools and RAPID toolbox tools were used for RAPID preprocessing. Catchments were created using catchment grid delineation and create catchment polygons.

RAPID is a river routing model, which stands for Routing Application for Parallel computation of Discharge. Cedric David created RAPID in 2007. "Given surface and groundwater inflow to rivers, this model can compute the flow and volume of water everywhere in river networks made out of many thousands of reaches." (David, 2011) RAPID can output streamflow data and can also calculate return periods. RAPID must be executed in a Linux environment. Ubuntu was used as a virtual Linux environment to run RAPID on a Windows computer.

AutoRoute was used to map out the flood inundations. AutoRoute is a flood inundation model. It uses a stream network, streamflow data, a DEM, and landcover to map out a flood inundation scenario (Follum, 2017). The specifics of the procedure can be set by the user. One can choose from a max flow event, or monthly averages, or a specified return period. AutoRoute was run in Jupyter Notebook.
Analysis and Methods

For faster processing of the data and continuity of the streamlines, the data was split up into 27 watersheds in South America. Watersheds can be determined using elevation data and some of the same ArcGIS Hydrology tools to create the streamlines. However, this project used a previously created South America watersheds dataset.

The first step in the analysis was to use 30m SRTM elevation data to create South America streamlines. The elevation was split up into each respective watershed using the split raster tool in ArcGIS. Next, any holes in the data were filled using the fill tool. Flow direction determines the angle/direction of the slope in elevation. This flow direction dataset was used to calculate the flow accumulation, using the flow accumulation, which determines the number of cells the flow into any particular cell. The next step can be adjusted depending on the project and the number of streams needed. For this project, a cell was determined as a stream if more than 1000 cells are flowing into it. This was done using the con tool. The following tools were stream order, stream link, and stream to feature. These tools are used to determine the stream network, how streams are ordered and linked, and then finally turn the stream raster dataset into a line feature dataset.

A Catchment is the area of land that flows into individual streams. For this project, catchments were created using the catchment grid delineation tool in the ArcHydro Toolbox, which uses flow direction and stream link datasets previously created to generate the stream
network. The catchment grid raster dataset was then turned into a feature dataset using the catchment polygons tool. Catchments are needed as a part of the RAPID preprocessing.

The next step was to download historical runoff data from ECMWF. The data was downloaded according to the extent of each watershed and was downloaded as a NetCDF file. The dates of the data are from January 1, 1980, to December 31, 2019. The temporal rate is every 6 hours.

The next step was to analyze streamflow data to get flood frequency/probability. Now that the stream network, catchment polygons, and runoff data were created, RAPID preprocessing started. RAPID preprocessing was done using RAPIDpy. RAPIDpy was used to create the various input files for RAPID. RAPID preprocessing steps include creating a connectivity file for the stream network, creating an inflow file from the runoff data, and creating a name list, which defines the RAPID process variables. RAPID was run in Ubuntu.

The AutoRoute model was used to map out the flood extent. Results of RAPID give us return years, the probability. Using AutoRoute, we can map out the extent of floods from a specified return year. This gives us the final flood hazard in a polygon dataset. AutoRoute was executed in Jupyter Notebook.

After the flood hazard was created, the other datasets were used to determine the exposure and vulnerability. Landscan population density data were used to determine the number of people potentially affected by floods, thus providing the exposure element of risk. Landscan is a raster dataset, but it was turned into a vector dataset to determine how many people are affected. A weighted overlay was used for the remaining datasets. A weighted overlay is a raster dataset that is utilized by assigning values on the same scale. In this case, the weight was applied towards factors more likely to cause flood risk. Areas more likely to be affected by floods due to climate change were given a higher value. Likewise, lower-income areas were assigned higher values due to difficulties in recovery. Use climate change, poverty, and access to resources data to determine vulnerability. The four datasets: landscan, access to resources, poverty, and climate change, were reclassified to a 1-5 scale and were averaged together.

Finally, combine hazard, exposure, and vulnerability to create a risk assessment. The flood hazard was intersected with the weighted overlay. With many features in the flood hazard polygons, intersecting the weighted overlay with them can take considerable processing time. The overlay was converted to a polygon feature class to save processing time. This was done by creating a fishnet from the overlay cells, converting the overlay to points using the rater to point tool, then executing a spatial join on the fishnet and the points. This gives the fishnet the original values of the weighted overlay, and analysis can be completed faster. Statistics were calculated.
Figure 4. Project Workflow
Results

This project's results are the completed framework, flood return periods, flood extent at a 20-year flood, a risk assessment map, and information on the number of people affected. Due to time constraints, only a 20-year flood was analyzed. Figure 5 shows the flood risk for a 20-year flood in South America. Figures 6 and 7 show close-ups of individual countries, and individual cities, respectively. Table 2 shows the number of people affected by a 20-year flood in each county. Table 3 shows the number of people affected in the areas with the most risk.

The results show large amounts of risk are in large cities because of significant populations. The results show areas at risk, and also that there is a concentration of regions in the continent at risk in lower Brazil, Uruguay, Argentina, and Colombia, Ecuador, and Peru. The
results could inform lawmakers, NGOs, humanitarian organizations, etc. about areas that they may not have known were vulnerable to flooding.

Discussion and Conclusion

Challenges and Limitations

There were several challenges throughout this project. Processing time is a huge factor. However, this project was quite large in scope, so processing time for a smaller total area would have fewer issues. Due to limited time for analysis, there are only products for a 20-year flood. One of this project's goals was to create results products for multiple return periods, so that goal was not met.

The framework cannot (at least at this moment in time) be turned into a python script or ArcGIS tool, which was another one of the goals. However, it may not be desirable to complete the entire process with one tool. For one reason, depending on the size of the area of interest, processing time would be an issue. It is essential to check the results of individual steps to ensure the results are correct, and there are no issues. Instead of running the entire process at once, it's essential to check the results at various steps along the way.

There were considerable difficulties installing and using RAPID, especially since it requires a Linux environment. There were also issues with the python environment for AutoRoute because it utilizes an older package version for GDAL. Since there were issues with the models used for this project, it limited the time for further analysis with the completed results. There may be other models better suited for this framework, but they may also be limited in different ways. Further research may be necessary to find which model is best for this framework.

Discussion

The results show the areas where there is more flood risk. This is seen in lower Brazil, Uruguay, Argentina, as well as Colombia, Ecuador, and Peru. It seems this is due to the increase in flooding at those locations due to climate change. One thing of note was that the population density and access to resources datasets, for the most part, canceled each other out. More research would be necessary to determine if there are indeed populations without access to a larger city. It may be that the access to resources dataset is unnecessary and could be replaced by another dataset.

There are also other vulnerability factors not taken into account due to a lack of data availability. Another thing to note is that this risk assessment focuses on the people potentially affected by flooding, their ability to adapt and recover from a flood, and their environment. The missing vulnerability factor is the economic impact. It wasn't easy to find a global dataset displaying that information. The building material of infrastructure would help determine the amount of damage a flood would do to an area. While there is poverty data used for this project,
it is only an estimate, and a more accurate dataset of poverty would assist in identifying vulnerable areas. If that data becomes available, or a study has site-specific data, that can be added to this framework.

There are situations when this framework should and shouldn't be used. This framework should be used when a flood risk assessment is needed, but data are scarce. Since this framework uses globally available data, it can fill in any data gaps. The framework should also be used if there is no flood hazard data because, with this framework, it is made from scratch. It is helpful in analyzing a large area of interest. This framework should not be used if there is better data available or more area-specific data is available. This would add more value to a flood risk assessment. Also, it should not if the area of interest is small. If flood hazard data is already available, that section of the framework can be skipped.

**Conclusion**

The way forward from this project is to verify the risk assessment's accuracy and adjust the framework with better processes and data. A good idea would be to test different flood hazard models. Some might be better and more accurate; some might not be able to accommodate data on a large scale. It may be necessary to look at past recorded events to verify the flood hazard and risk assessment. Another way forward is to analyze different return periods. Finally, This framework is relatively simplistic, and it could go into a lot more detail and in-depth to add more value to the results. This framework takes the definition of risk and creates a risk assessment based on its elements. With a little bit of work and additional research, it could provide vital information to help leaders and decision makers reduce the amount of flood risk for any area in the world.

Tables and Figures
<table>
<thead>
<tr>
<th>Country Name</th>
<th>Total Population</th>
<th>Population Affected</th>
<th>Percent of Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>44,612,805</td>
<td>15,470,313</td>
<td>34.7</td>
</tr>
<tr>
<td>Country Name</td>
<td>Total Population</td>
<td>Population in High Risk</td>
<td>Percent of Population</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------</td>
<td>-------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Argentina</td>
<td>44,612,805</td>
<td>3,825,673</td>
<td>8.6</td>
</tr>
<tr>
<td>Bolivia</td>
<td>11,297,627</td>
<td>1,270,866</td>
<td>11.2</td>
</tr>
<tr>
<td>Brazil</td>
<td>206,853,482</td>
<td>11,943,770</td>
<td>5.8</td>
</tr>
<tr>
<td>Chile</td>
<td>17,694,593</td>
<td>747,250</td>
<td>4.2</td>
</tr>
<tr>
<td>Colombia</td>
<td>47,889,908</td>
<td>11,185,762</td>
<td>23.4</td>
</tr>
<tr>
<td>Ecuador</td>
<td>16,172,748</td>
<td>3,706,283</td>
<td>22.9</td>
</tr>
<tr>
<td>French Guiana</td>
<td>274,334</td>
<td>21,364</td>
<td>7.8</td>
</tr>
<tr>
<td>Country</td>
<td>Population</td>
<td>Number Affected</td>
<td>Flooding Rate</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>-----------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Guyana</td>
<td>702,088</td>
<td>213</td>
<td>0.03</td>
</tr>
<tr>
<td>Paraguay</td>
<td>7,018,605</td>
<td>279,129</td>
<td>4.0</td>
</tr>
<tr>
<td>Peru</td>
<td>31,198,237</td>
<td>2,868,840</td>
<td>9.2</td>
</tr>
<tr>
<td>Suriname</td>
<td>577,213</td>
<td>13,232</td>
<td>2.3</td>
</tr>
<tr>
<td>Uruguay</td>
<td>3,335,266</td>
<td>180,269</td>
<td>5.4</td>
</tr>
<tr>
<td>Venezuela</td>
<td>31,400,489</td>
<td>2,630,368</td>
<td>8.4</td>
</tr>
</tbody>
</table>

Table 2. Population affected by flooding in high risk areas. *Population numbers are based on Landscan estimates.*

References


