Final Project Report

Title: Flooding final project
Notice: Dr. Bryan Runck
Author: Maochuan Wang
Date: 2/27/2023

Project Repository: *https://github.com/MaochuanW/GIS5571/tree/main/Final%20Project*
Google Drive Link:
Time Spent: 20 days

Abstract:

The Red River Valley of Canada is vulnerable to frequent flooding due to its low-lying terrain, significant ice melt, high levels of precipitation, and multiple tributaries. In order to evaluate and model the flooding impacts and risks in the area, a project was conducted using multi-criteria decision analysis. The project used high-resolution digital elevation model (HRDEM) data, slope, flow direction, flow accumulation, and stream order, among other datasets, to build the flood model. The results showed that most of the river is classified as having a medium to high risk of flooding, with some parts having a low risk. The parts of the study that extend near the riverbanks and those further upstream are most vulnerable to flooding. The model identified areas highly vulnerable to flooding with a mean above 3.05 hazard zone.

Problem Statement

The Red River is in the north-central United States and central Canada. The river flows north through Manitoba and onto Hudson Bay from the Dakotas and Minnesota. Historically, the region has suffered catastrophic flooding in Spring every year, making it a priority to more fully understand the effects of flooding on the area. Several important factors that cause flooding in the region include poor drainage and heavy snowmelts. The Red River valley is extremely flat, and it flows through a trough formed by continental glaciers around 10,000 years ago. The river's elevation decrease from Fargo, North Dakota, to the Canadian border is approximately 130 feet (40 meters). The river meanders through the landscape due to the small slope; nevertheless, the areas surrounding the trough are more elevated. The poor drainage system causes flooding to occur at greater magnitudes in the central floodplain of the valley.
As previously mentioned, flooding occurs most years but a few examples of severe floods include the flooding in 1979, which caused damages of about $114 million in North Dakota; 1997, which resulted in $500 million of damage, in Manitoba alone; and the most recent 2022 flooding that resulted in several million dollars of damage.

The purpose of this project is to evaluate and model flooding hazards and impacts. I developed a model to assess flood hazards and then used those results to create tools that can inform policymaking and ensure that communities are aware of the hazards that they face. I believe that the models will be transferable and scalable, meaning that it will be easy to apply the same models and techniques to other areas of interest and streamline flooding hazard analyses across Canada and the United States.

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	High-Resolution Digital Elevation Model (HRDEM)	Elevation data for the study area	Yes		Natural Resources Canada	Mosaiced the many DEM tiles into a single elevation dataset for the project area
2	Land Cover	Land cover for the study area	Yes		Landsat 9 imagery and training data created manually	Google Earth Engine used for land cover classification with 79% accuracy
3	Precipitation	Precipitation data for the study area	Yes		Quarterly average precipitation data taken for nine cities/towns	Interpolation technique (IDW) used to estimate precipitation in the entire study extent
4	Demographic Information	Population and income data for the study area	Yes			Demographic measures were calculated and joined to the final hexagonal flood hazard dataset.

Input Data

#	Title	Purpose in Analysis	Link to Source
1	High-resolution Digital Elevation Model (HRDEM)	To create an elevation model for the study area and generate other datasets, including slope, flow direction, flow accumulation, and stream order	Natural Resources Canada
2	Landsat 9 Imagery	To classify the land cover of the study area	Google Earth Engine

3	Interpolated Precipitation Data	To estimate precipitation in the study extent	Interpolation Technique (IDW)

Methods

A. Land Cover Classification:

Land cover and land use are important factors in flood models since there are a variety of different variables, such as drainage and permeability, amongst others, that play a significant role in determining whether or not a given area is subject to flooding conditions. With this in mind, it was therefore imperative to create an updated land cover model of the study area to assess flooding hazards most accurately. To accomplish this, Landsat 9 imagery from September 2022 was used along with training data that was created, to predict land cover for the entire study extent. Google Earth Engine was the software used for this step, due to the ease of use, the ease of accessing data, and the scripting options available that make it easier to reproduce such an analysis. In Google Earth Engine, an analysis was first run to find the optimal imagery, in terms of a lack of cloud cover and covering both the desired spatial and temporal extents that were desired. After this, training data was manually drawn in so that the machine learning algorithm had an understanding of what different classes looked like, in terms of their spectral signatures. Overall, the model came out to about 79% accuracy, but some classes were more accurate than others. This was expected, though, due to the underlying similarities between the spectral signatures of a few of the classes. The land cover surface is easily sufficient for this study since it is only a small subset of the inputs for the actual flood model.

B. Elevation and Hydrological Modeling

Elevation and subsequent data products are the most essential components to flood modeling, and without adequate input data, any flood model is flawed. I took extra steps to verify and validate the accuracy only makes the end flood model even more useful. The first step was retrieving the most up-to-date high-resolution digital elevation model (HRDEM) datasets from Natural Resources Canada, the government agency responsible for most Canadian environmental and geophysical spatial datasets. LiDAR was recently collected in the Red River Valley in 2020, meaning that the input data is very relevant and useable. By mosaicing the many DEM tiles into a single elevation dataset for the project area, other datasets could then be generated through the plethora of tools available in ArcGIS Pro, including slope, flow direction, flow accumulation, and stream order, among others. With a stream network in hand, other datasets like drainage density and distance to streams were created, by using the line density and Euclidean distance tools, respectively. The bulk of the data used in the flood model all originated from the initial elevation model.

C. Flood modeling

There are a variety of different methods to model flood hazards, but the most robust is through multi-criteria decision analysis (MCDA) which can use binary, weighted overlay, or fuzzy logic to identify suitability, or a lack thereof, for some phenomena, given a set of inputs. MCDA can be used across many domains, ranging from ecology to retail site selection, and of course, flooding. For this model, a weighted overlay was used so that individual inputs could be weighted according to their importance for determining flood hazards. In the model, the weights used were as follows: elevation (30%), distance to streams (10%), drainage density (20%), slope (5%), and land cover (35%). An important intermediary step was to normalize the inputs into a standard scale so that the weights would have the correct intended impact on the outcome. Normalizing was completed by reclassifying the values of each input dataset onto a scale from 1 to 5, with 5 representing the highest risk values for a particular input and 1 representing the lowest risk values. After normalizing and calculating the weighted overlay, the results were then aggregated into hexagons, for a multitude of reasons, including better integrating with the hazard analysis, better visualizing the results, and smoothing out the results by calculating the average hazard score per each hexagon.

D. Hazard Analysis

With a flood hazard model in hand, the next step of the analysis was to incorporate various other datasets with the model, to identify the areas that are most prone to facing the consequences of flooding. Using Esri's Business Analyst, key demographic measures were calculated for the hexagonal flood hazard dataset. These measures were then joined to the dataset in ArcGIS Pro, before publishing the resulting final dataset to ArcGIS Online. These variables could be of use to policymakers, as the socioeconomic makeup of areas could be used as a factor in hazard planning projects.

The other part of the hazard analysis was running an optimized hot spot analysis on a finer-resolution hexagonal dataset, with which the mean hazard values were calculated. Using the optimized hotspot analysis, high (hotspots) and low (coldspots) clusters could be identified at different confidence intervals, to determine where concentrated at-risk areas are located. This provides a different lens to analyze hazards but is equally as useful.

Results

The current stage of the model has reached its final stage with high accuracy. I came up with a dashboard that helps the user to know more about this project. This Flood Hazard Analysis is a hexagon model with the hazard level on a scale of five, where five indicates a high hazard zone and one represents a very low hazard zone. A dashboard has income and population data to better understand the socioeconomic and demographic makeup of individual hexagons. This project presents an overview of machine learning methods used in flood prediction.

B. Hazard Analysis

The mapping indicated that the areas near the riverbanks and those further upstream are most at risk of flooding (figure 3). In addition, this river is most likely to experience flooding during heavy rain and snowmelt events, especially during spring. In the past 60 years, the floods in 1950, 1997, 2009, 2011, and 2020 have caused substantial damage, especially in the Winnipeg area.

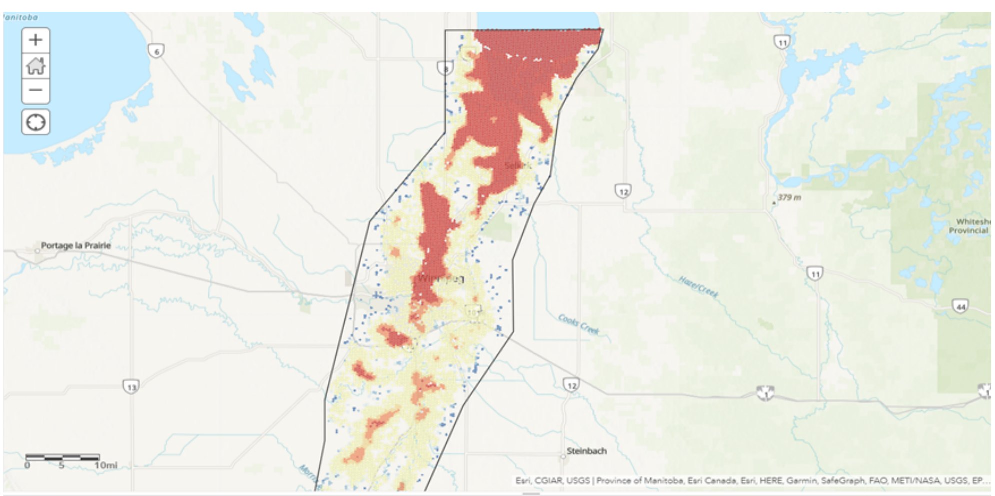

Figure 1. Concentrated levels of high vulnerability.

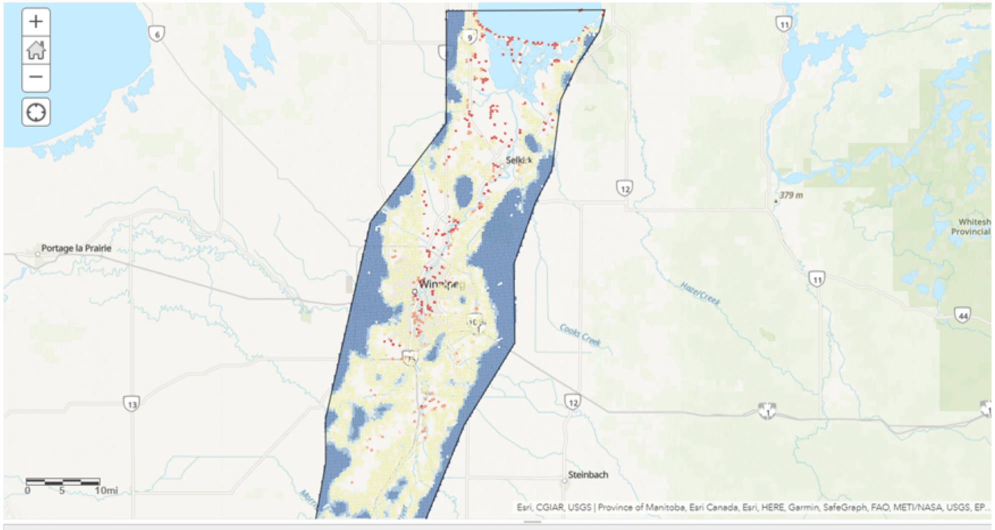

Figure 2. Concentrated levels of low vulnerability.

C. Flood Modeling

The flood risk zone modeling of Red Valley River showed that the area is particularly prone to flooding due to its low-lying terrain, enormous ice melt, high precipitation levels, and numerous tributaries. Areas of higher risk were identified along the main channel of the river, as well as in its tributaries and all around the huge water bodies. The results suggest that flood mitigation strategies should focus on protecting these areas. Additionally, the results highlight the need for improved infrastructure and land use planning in order to reduce flood risk. Furthermore, the modeling results can be used to inform decisions regarding flood insurance and emergency management. Overall, this project highlights the importance of understanding the potential risks associated with flooding and developing strategies to reduce the impacts of floods. The model identifies these are the areas highly vulnerable to flooding with a mean above 3.05 hazard zone (figure 3)

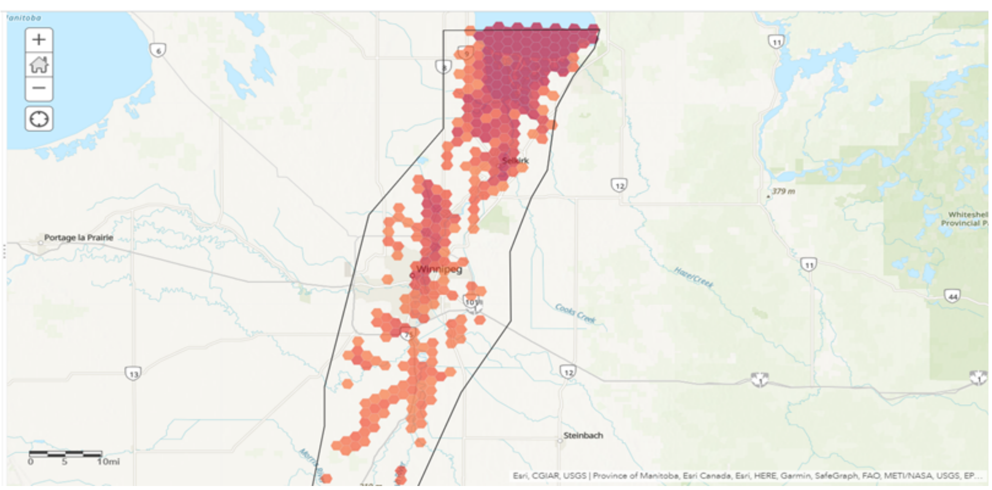

Figure 3. Areas of high vulnerability.

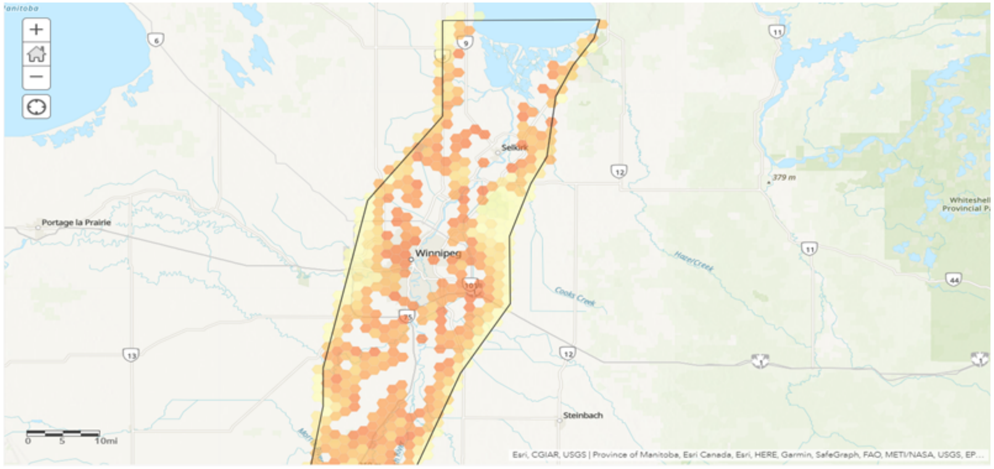

Figure 4. Areas of low vulnerability.

Discussion and Conclusion

The aim of this project is to evaluate and model flooding impacts and risks in the Red River Valley of Canada. The most important parameters for this flood study were chosen after considering the dynamics of the study extent and extensive literature review. It can be argued that existing weather conditions must be taken into account when determining the magnitude of floods, especially when the flood comes in spring when the snow melts. There are a number of flood models that don't consider precipitation data. However, I did use an interpolation technique (IDW) to estimate the precipitation in the entire study extent. Quarterly average precipitation data taken for nine cities/towns (4 inside the study extent and five outside of it) showed minimal variation. Since this project is about looking at general trends rather than the flooding risk after a specific precipitation event, the inputs deemed essential components of flood modeling were taken.

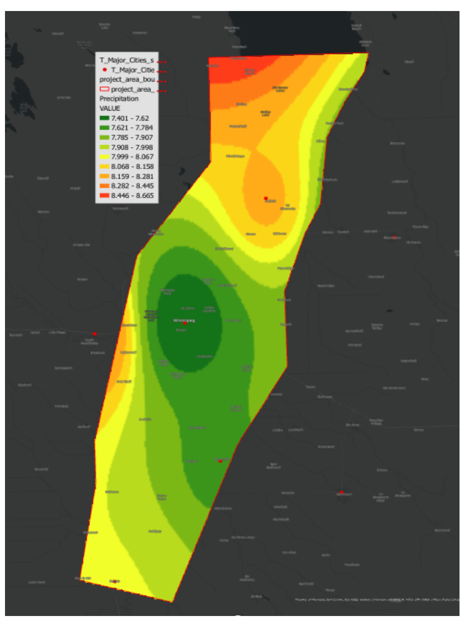

Figure 5. Interpolated Precipitation (inches) in the Study Extent

The flood risks were modeled using multi-criteria decision analysis after individual inputs were weighted according to their importance. I then decided to use hexagons for better visualization and smoothing of results. It was discovered that most of the river is classified as having a medium to high risk of flooding, with some parts having a low risk. Due to its low-lying terrain, significant ice melt, high levels of precipitation, and multiple tributaries, the risk zone modeling revealed that the region is vulnerable to floods. The parts of the study that extend near the riverbanks and those further upstream are most vulnerable to flooding, while the bordering areas are considered safe. The highest & lowest risk values were estimated as 4.61 & 1.4, respectively.

The area of interest (AOI) has a total population of 875,953, so it is crucial to monitor and provide mitigation methods for this frequent flooding in and around the Winnipeg area. The created flood model is transferable and scalable, meaning that it can be applied to an entirely different AOI to streamline flooding hazard analysis. There is no ambiguity in the data flow, and every step is documented and justified.

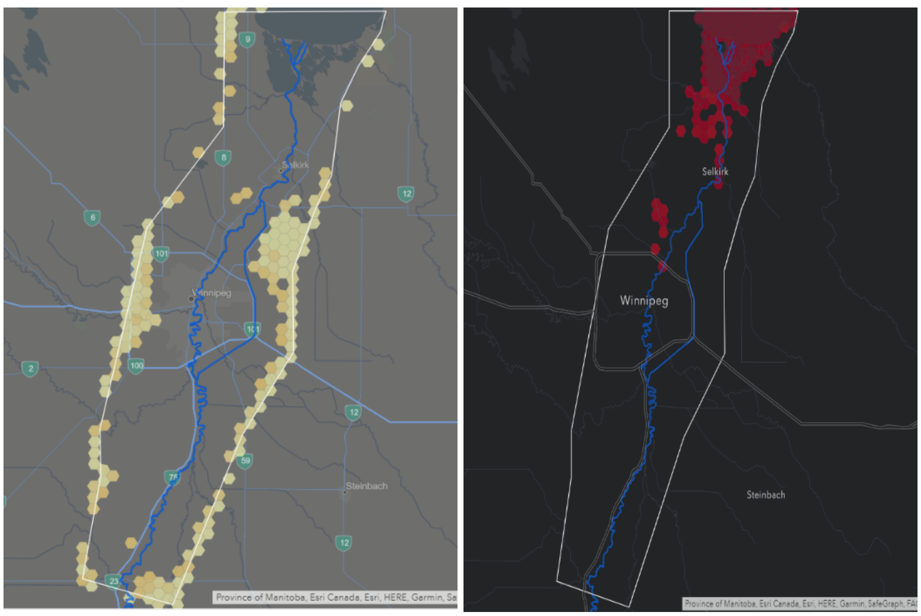

Figure 6. Concentrated at-risk *areas of* high (hotspots) and low (coldspots) *vulnerability.*

References

1. Afifi, Z., Chu, H. J., Kuo, Y. L., Hsu, Y. C., Wong, H. K., & Zeeshan Ali, M. (2019). Residential flood loss assessment and risk mapping from high-resolution simulation. Water, 11(4), 751.
2. Jeff Oppong: How to Use ArcGIS Pro to Map Flood Susceptibility. GISLOUNGE. Website. October 30, 2022. https://www.gislounge.com/how-to-use-arcgis-pro-to-map-flood-susceptibility/
3. Climate and Average Weather Year Round in Gimli. Weather Spark. Website. November 10, 2022. https://weatherspark.com/y/9098/Average-Weather-in-Gimli-Canada-Year-Round
4. Msabi, M. M., & Makonyo, M. (2021). Flood susceptibility mapping using GIS and multi-criteria decision analysis: A case of Dodoma region, central Tanzania. Remote Sensing Applications: Society and Environment, 21, 100445.
5. Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. Water, 10(11), 1536.
6. Michael Carlowicz: Red River Flooding is Worst in a Decade. NASA Earth Observatory. Website. November 05, 2022. https://earthobservatory.nasa.gov/images/149822/red-river-flooding-is-worst-in-a-decade

7. Paquette, J., & Lowry, J. (2012). Flood hazard modelling and risk assessment in the Nadi River Basin, Fiji, using GIS and MCDA. *The South Pacific Journal of Natural and Applied Sciences*, *30*(1), 33-43.
8. Rahmati, O., Zeinivand, H., & Besharat, M. (2016). Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis. *Geomatics, Natural Hazards and Risk*, *7*(3), 1000-1017.
9. Samanta, S., Koloa, C., Kumar Pal, D., & Palsamanta, B. (2016). Flood risk analysis in lower part of Markham river based on multi-criteria decision approach (MCDA). Hydrology, 3(3), 29.

Self-score

Category	Description	Points Possible	Score
Structural Elements	All elements of a lab report are included **(2 points each)**: Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score	28	**28**
Clarity of Content	Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**.	24	**24**
Reproducibility	Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified.	28	**28**
Verification	Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**.	20	**20**
		100	**100**