FORECASTING FLOOD INUNDATION IN CALGARY AND DENVER

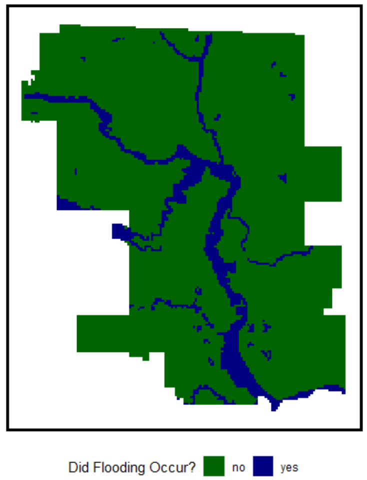
Kristin Chang & Jenna Epstein | *Land Use & Environmental Modeling (CPLN 675)*

MOTIVATION

According to the World Health Organization, floods are the most frequent natural disaster affecting countries across the globe. Floods have the potential to leave communities devastated not only because of the direct loss of life due to drowning and infrastructure damage, but also because of indirect impacts such as increased transmission of disease, higher risk of injury and hypothermia, disrupted or disabled infrastructure systems, and increased likelihood of causing other natural disasters. In the past ten years, climate change has exacerbated the effects of natural disasters like flooding, drought, sea level rise, and extreme precipitation and their frequency and intensity are expected to continue to rise if left unchecked.

Inundation in Calgary

Extent derived from raster from June 2013 flooding



This document outlines a machine learning algorithm that predicts which areas of a city are at the highest risk of flooding disasters. The model is created using past flood extent data from the City of Calgary in Canada to measure accuracy and then is applied to the City of Denver in the United States to measure generalizability.

SIGNIFICANT FEATURES

Mean Elevation

Elevation generally describes the topography of the land. Areas at lower levels of elevation are at higher risk for floods than areas at higher elevations. We explored calculating the lowest degree of elevation per fishnet grid cell in Calgary, but found that using the mean helped the performance of our model.

Distance to Nearest Stream

Understanding the structure of streams helps inform the intensity of potential flooding. If the water level of the stream exceeds the stream channel, then flooding occurs. Areas that are closer to streams are more vulnerable to floods than areas that are farther away. A stream network was generated based on the Digital Elevation Model for Calgary, and distance was calculated for each fishnet grid cell to the nearest stream.

Maximum Flow Accumulation

Flooding occurs when water levels exceed the stream channel. Based on the direction of the streams calculated above, using the maximum accumulation value in each grid cell allows us to evaluate at the highest risk level. In other words, the resulting model will present a "worst-case-scenario" picture enabling planners to be over-prepared rather than under-prepared.

Distance to Nearest Steep Slope

Steep slopes in the land can contribute to both flow direction and flow accumulation. Additionally, a steeper slope likely accelerates the speed of flow and can intensify in the case of a flood. Thus, it is important to be aware of the location of steep slopes in the topography and their relative distance to past floods.

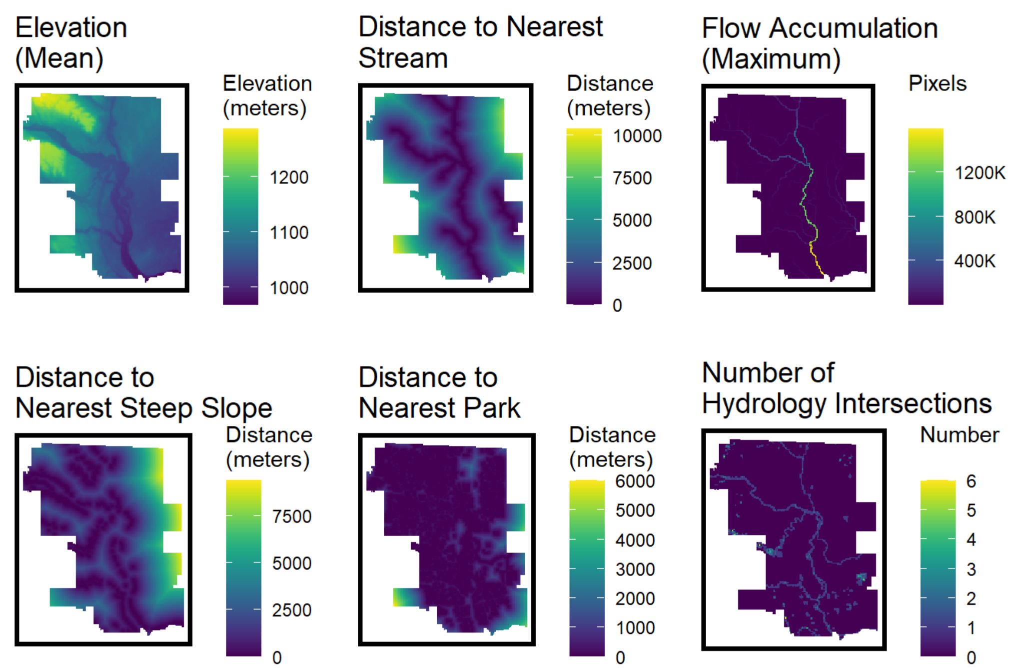
Distance to Nearest Park

Land cover or soil type can be indicative of areas prone to flooding. The more permeable the surface is, the more risky it is for floods. Initially, land cover data for Calgary (from the city's open data portal) was used to identify which fishnet grid cells were impervious based on their land cover type; however, this did not help the model perform well. Instead, distance to nearest park is incorporated into the model. Parks are generally covered in grass which is fairly permeable. Thus, the closer a fishnet grid cell is to a park, the more likely it is to become inundated.

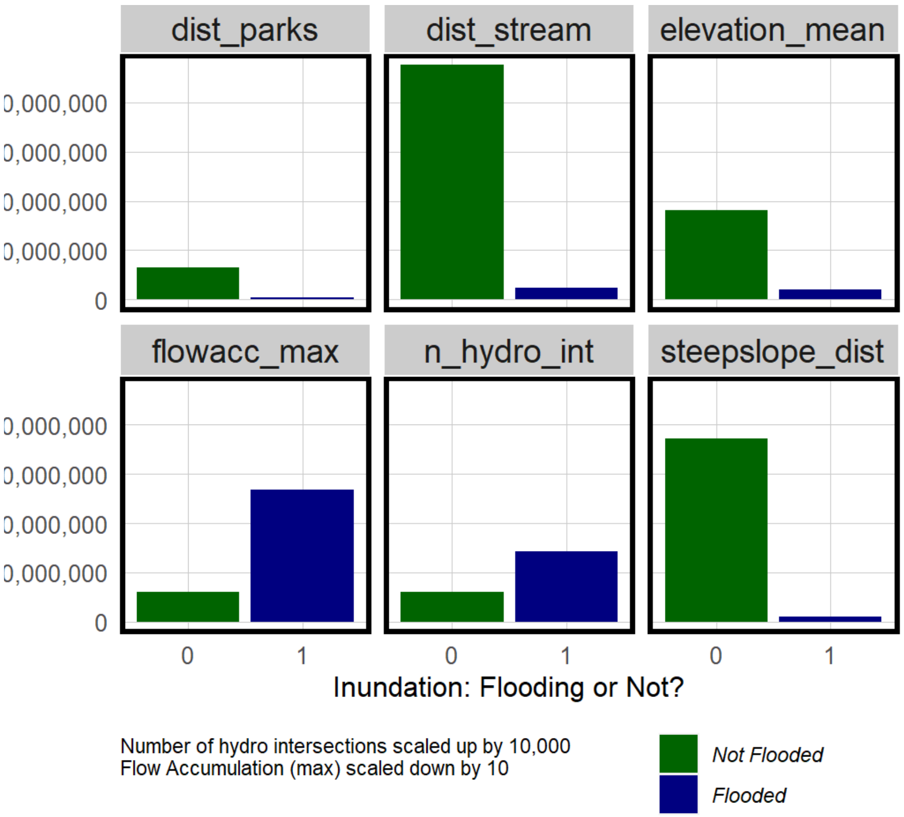
Number of Hydrological Feature Intersections

In addition to mere distance to stream lines, the frequency of intersections with hydrological features could indicate damper soils and areas at greater risk of inundation. Hydrology data was brought in from Calgary's open data portal, and the number of intersections per fishnet grid cell was calculated.

Mapping Significant Features



Plotting Significant Features per Inundation Outcome
These six selected significant features are shown below in the plots to show differences in across fishnet grid cells that flooded and those that did not, according to the classification from the 2013 inundation extent.



MODEL

```
inundationModel <-
      glm(inundation ~ ., family="binomial"(link="logit"),
      data=(calgaryTrain) %>% as.data.frame()
      %>% dplyr::select(-ID_FISHNET, -geometry))
```

MODEL RESULTS

Observations	12977
Dependent variable	inundation
Type	Generalized linear model
Family	binomial
Link	logit

$\chi^2(6)$	4779.80
Pseudo-R² (Cragg-Uhler)	0.64
Pseudo-R² (McFadden)	0.56
AIC	3812.46
BIC	3864.75

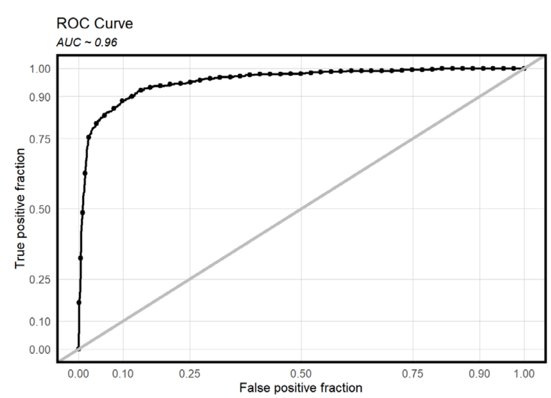
	Est.	S.E.	z val.	p
(Intercept)	37.84	1.89	20.06	0.00
steepslope_dist	-0.00	0.00	-20.60	0.00
dist_parks	-0.00	0.00	-2.95	0.00
flowacc_max	0.00	0.00	5.49	0.00
dist_stream	-0.00	0.00	-2.50	0.01
elevation_mean	-0.04	0.00	-20.64	0.00
n_hydro_int	3.14	0.10	32.37	0.00

Standard errors: MLE

The results of each scenario at the optimal threshold for accuracy (0.38) are displayed in the table below. These results support our above observation that the model predicts well for areas of no inundation than areas of inundation.

Type of Prediction	Description	Results
True Positive (TP)	Model predicts inundation and there is inundation	425
True Negative (TN)	Model predicts no inundation and there is no inundation	4873
False Positive (FP)	Model predicts inundation and there is no inundation	167
False Negative (FN)	Model predicts no inundation and there is inundation	96

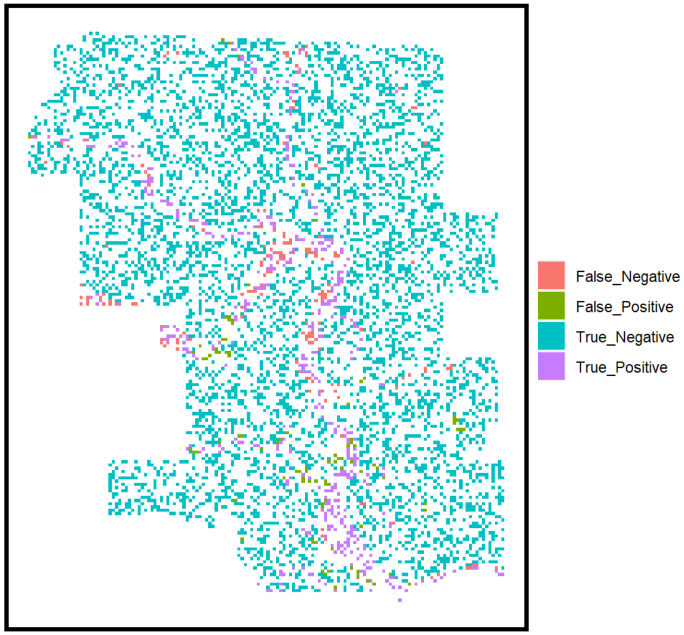
The Receiver Operating Characteristic (ROC) Curve for the model is a helpful goodness of fit indicator, while helping to visualize trade-offs between true positive and false positive metrics at each threshold. A line going "over" the curve indicates a useful fit. The area under the curve (AUC) here is about 0.96, indicating a useful fit. A reasonable AUC is between 0.5 and 1.



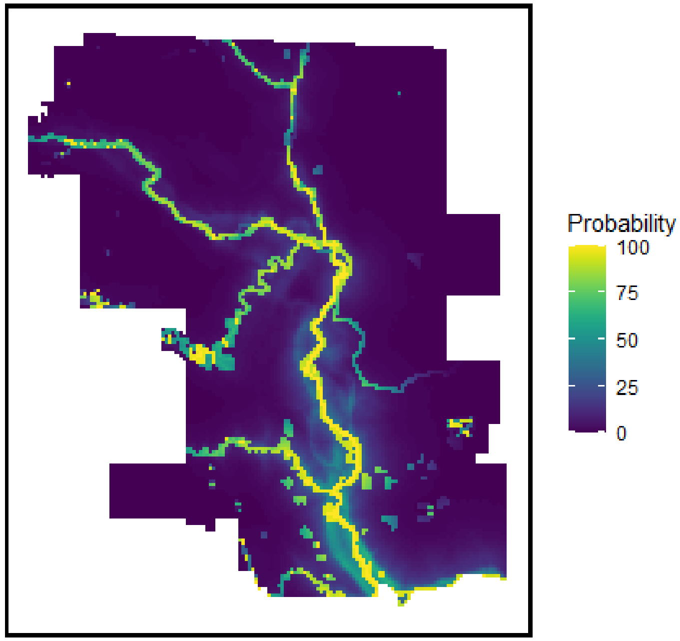
MAPPING CALGARY PREDICTIONS

The first map shows the confusion matrix results for the test set to better understand spatial arrangement of the outcomes.The second and third maps show predictions for inundation across Calgary are mapped for each fishnet grid cell. We elected to display the predictions on a probabilities scale to achieve a more continuous surface. This left-most map shows the same predicted probabilities, but with the 2013 flood extent overlay. This helps to compare the predicted inundation outcomes with the actual data used to train the model in the first place.

Prediction Classes on Calgary Test Set

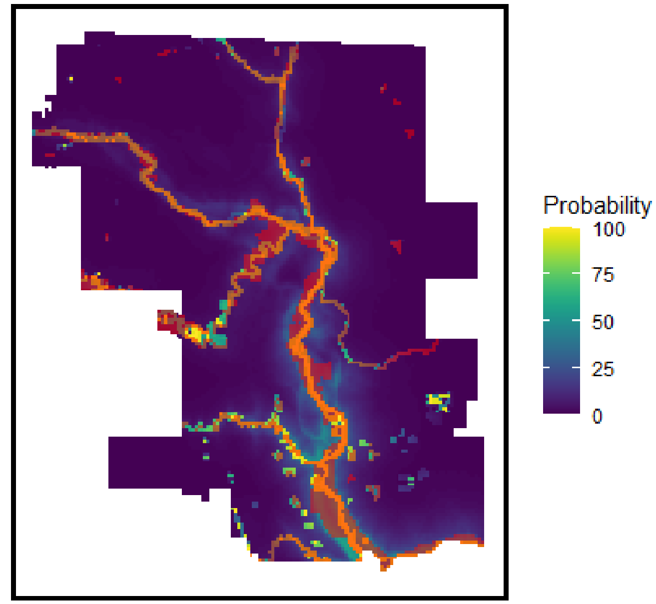


Predicted Probabilities for Inundation in Calgary



Predicted Probabilities for Inundation in Calgary

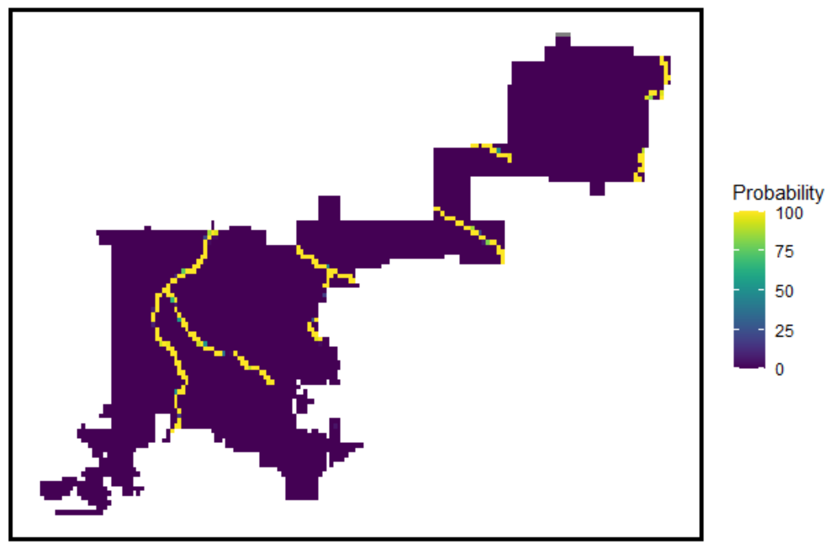
2013 flood extent in red overlay



MAPPING DENVER PREDICTIONS

Now that we have confirmed the model's high quality performance in Calgary, let's see if is generalizable to other cities. Denver, Colorado has a similar topography, infrastructure, and population size to Calgary. Below we perform the same feature engineering for the selected features of the model and run the regression for the City of Denver. A map of our predictions is displayed to the right.

Predicted Probabilities for Inundation in Denver



CONCLUSION

The results of this model are crucial to informing planners and allowing a city to be prepared for flood disasters. Successfully leveraging the knowledge and insight gained can save lives, cities, homes, and money. There are many implementation challenges when it comes to appropriately executing emergency preparedness programs such as funding, political and community support, and collaboration across departments. However, we are confident that the story being told through these data visualizations can be an effective tool in convincing relevant decision-makers.