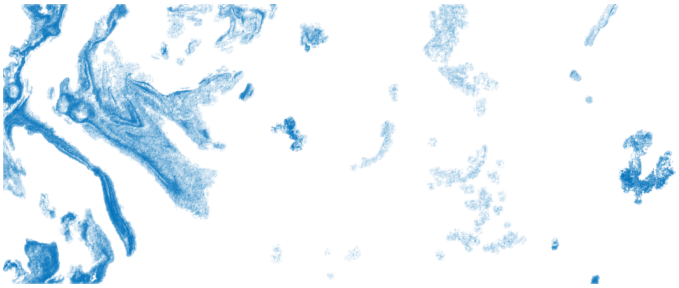
Narrated Presentation: https://youtu.be/mUIs2zT6vXo



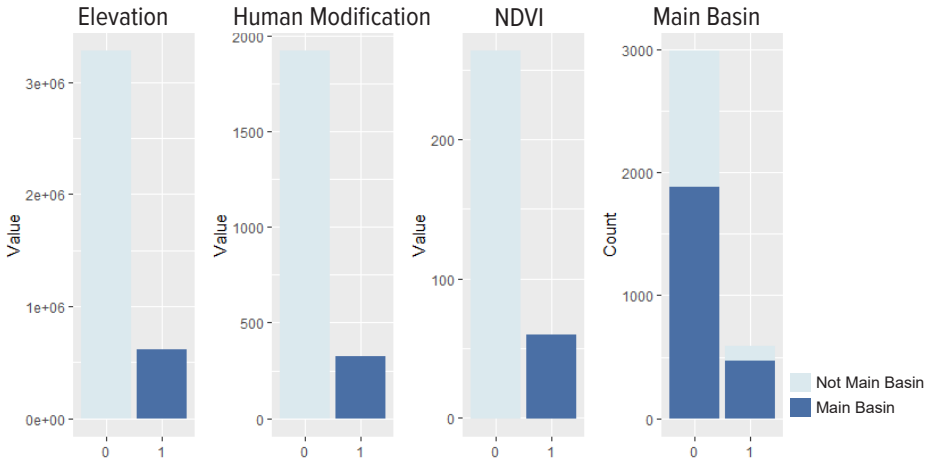
Predictive Modeling for Flood Resiliency

"Gillian" Xuezhu Zhao & Xinyi Qiu
Land Use and Environmental Modeling Course

Environment Features

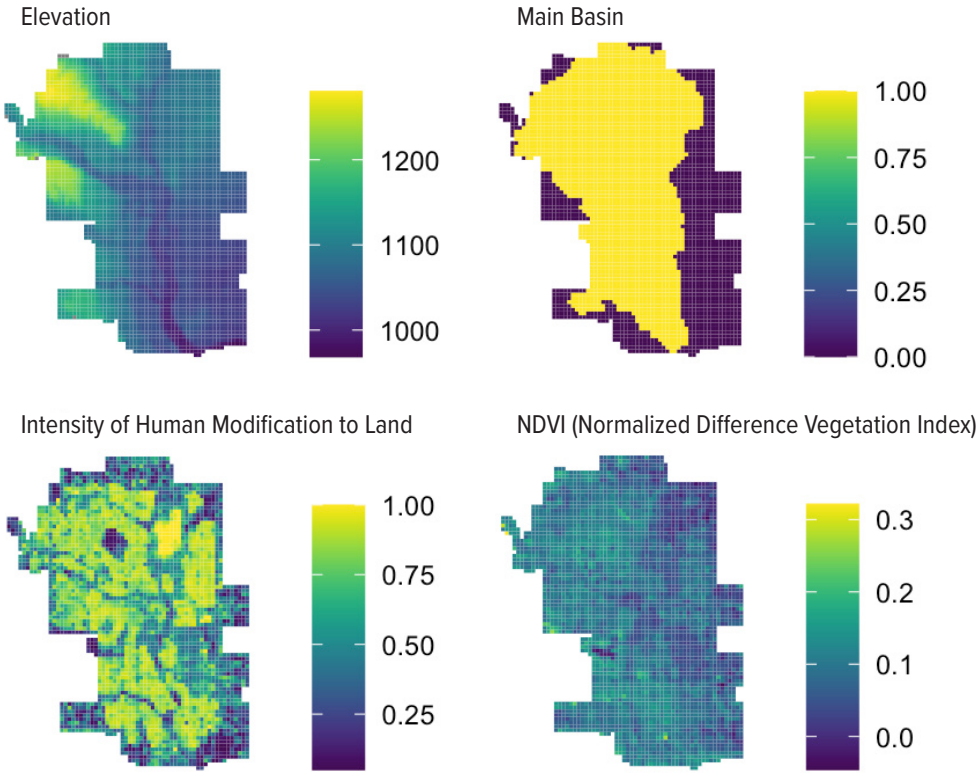
Exploratory Analysis of Features

Figure 1



Map of Four Important Features

Figure 2



The purpose of this project is to develop a predictive model that predicts the probability that an area will be inundated with flood water. The model has been trained and validated on Calgary, a Canadian City in Alberta, and then used to predict for a comparable city, Denver Colorado. This report will explain the planning motivation as well as illustrate the environment features, model performance, and prediction results.

Motivation

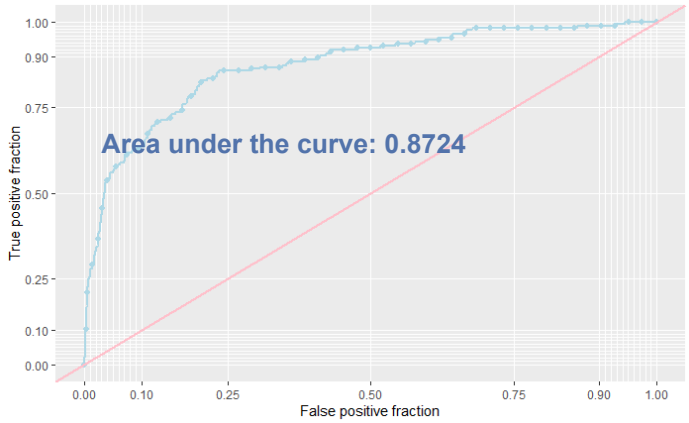
The motivation is to strengthen flood resiliency. Climate change is expected to affect water levels of not only coastal but also inland waters, as changing precipitation and outdated infrastructure have already led to alarming cases in places such as Calgary. While research and planning efforts have not been prevalent for riverine flooding, local governments suffer greatly from unexpected floods. Establishing a predictive model will be necessary and helpful to for governments to designate flood zones to allocate prevention resources.

Through exploratory analysis (Figure 1), we recognize that occurrences of inundation are related to environmental features. The distributions for four of them are mapped per 500 x 500 meters grid areas (Figure 2).

Logistic Regression Model

Model Performance

Receiver Operating Characteristic (ROC) Curve



With the environment features, we predicted them on logit regression. We used stepwise selection to choose features that are most significant to the flood probabilities. The features mentioned above along two slope characters are significantly related to flooding probabilities.

Figure 3

This ROC Curve indicates the goodness of fit of the model. With the area under the curve at 0.8724, our model could distinguish inundation and none inundation area quite well.

Confusion Matrix

Table 1

		Reference: No Flood	Reference: Flood
Prediction	No Flood	850	77
	Flood	41	98

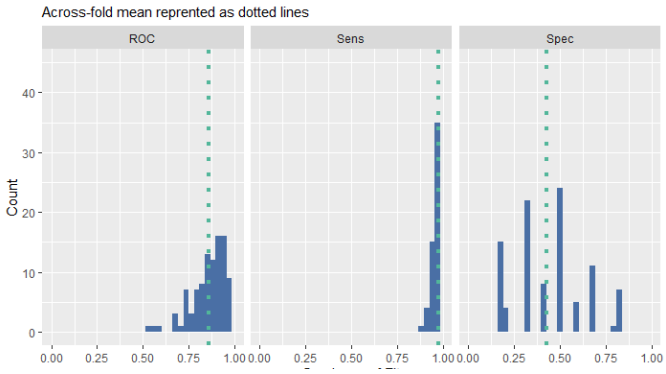
Accuracy:	0.8893
95% CI:	(0.8689), (0.9075)
No information rate:	0.8358
P-value [ACC>NIR]:	4.792e-07
Sensitivity:	**0.56**
Specificity:	**0.95**
Precision:	0.70

Confusion matrix describes prediction performances on flood and no flood. We hope to not miss any true flood occurrences but also make best use of prevention resources. Our model is able to predict 56% true flood occurrences correctly. Meanwhile, 70% of the predicted floods are correct.

Model Validation

Goodness of Fit Metrics from Cross-validation

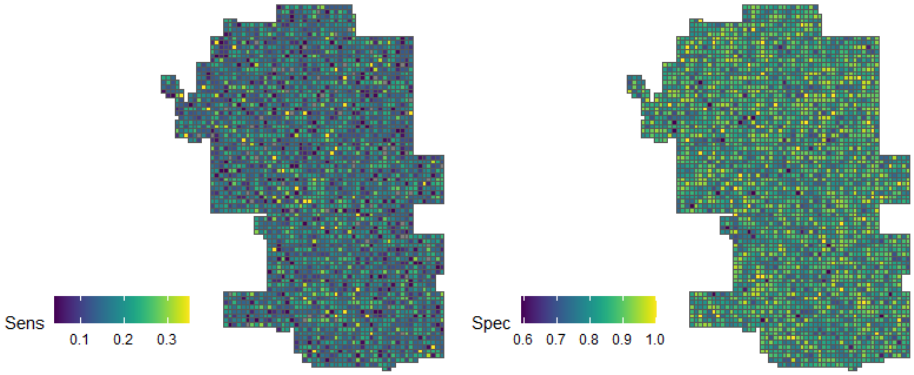
Figure 4



We cross validated it by iteratively testing on 100 random samples. The model predicts relatively well across samples.

Sensitivity Rate by Fishnet Grid Area from Spatial Cross-validation

Figure 5



Our spatial validation method assumes the local spatial process from all other grid areas generalizes to the hold-out.

Sensitivity (True Positive rate)
measures the proportion of positives that are correctly identified.

Specificity (True Negative rate)
measures the proportion of negatives that are correctly identified.

Prediction Result

Performance Map

Confusion Matrix for Training Set

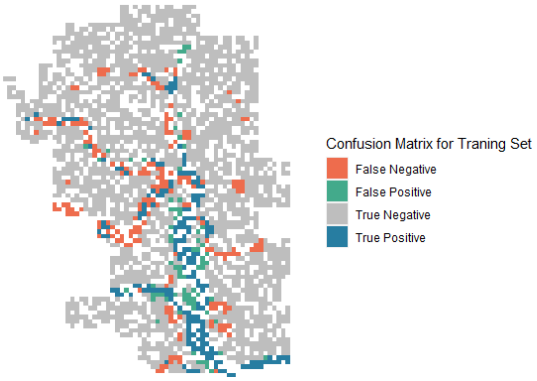


Figure 6

We advise policymakers to align local expertise with the probabilities map. It is especially important to be cautious of false negatives shown on the map in red. These are areas that were actually inundated but were predicted as not inundated.

Probability Threshold

Precision-Recall F Measure

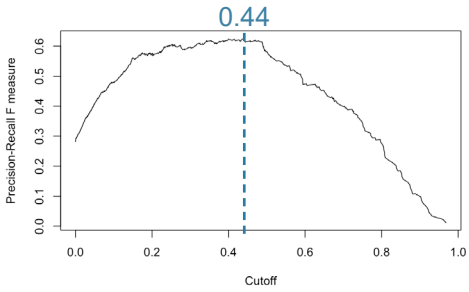
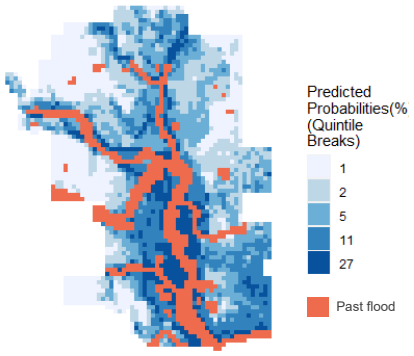


Figure 7

For flood zone designation, we hope that both the precision and sensitivity (also named recall) could be as high as possible. To balance them, we used F-1 measure to find the best probability threshold to be 0.44.

Prediction for Calgary, Alberta

Predicted Probabilitie



Predicted Flood

at 44% Probability Threshold

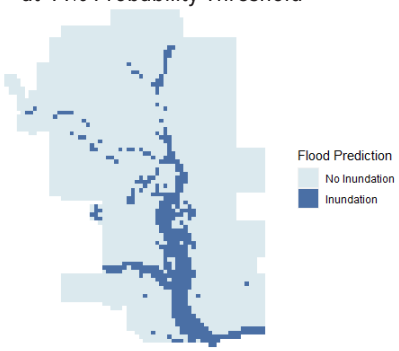


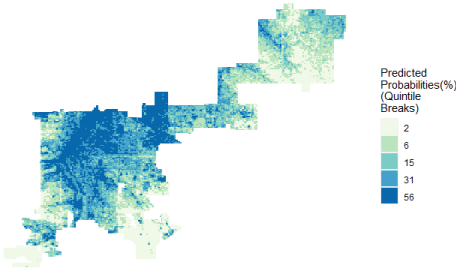
Figure 8

We can now see the probabilities of inundation by grid areas, overlaid with past flooding location as shown in red. The darker the blue, the more likely the inundation.

With our determined probability threshold at 44% - if the predicted probability for an area is above 44%, it is predicted to flood.

Prediction for Denver, Colorado

Predicted Probabilities



Predicted Flood

at 44% Probability Threshold

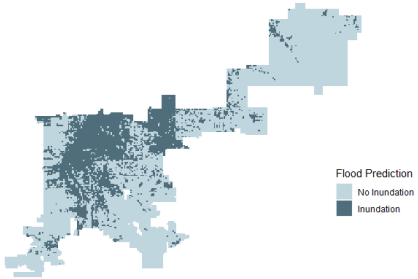


Figure 9

Finally we predict on a comparable city, Denver Colorado, to exemplify how the model can work for a client. Although at higher elevation and bigger in size, Denver is similar to Calgary in terms of hydrological features. The maps shows the predicted probabilities and the inundation prediction at 44% threshold.