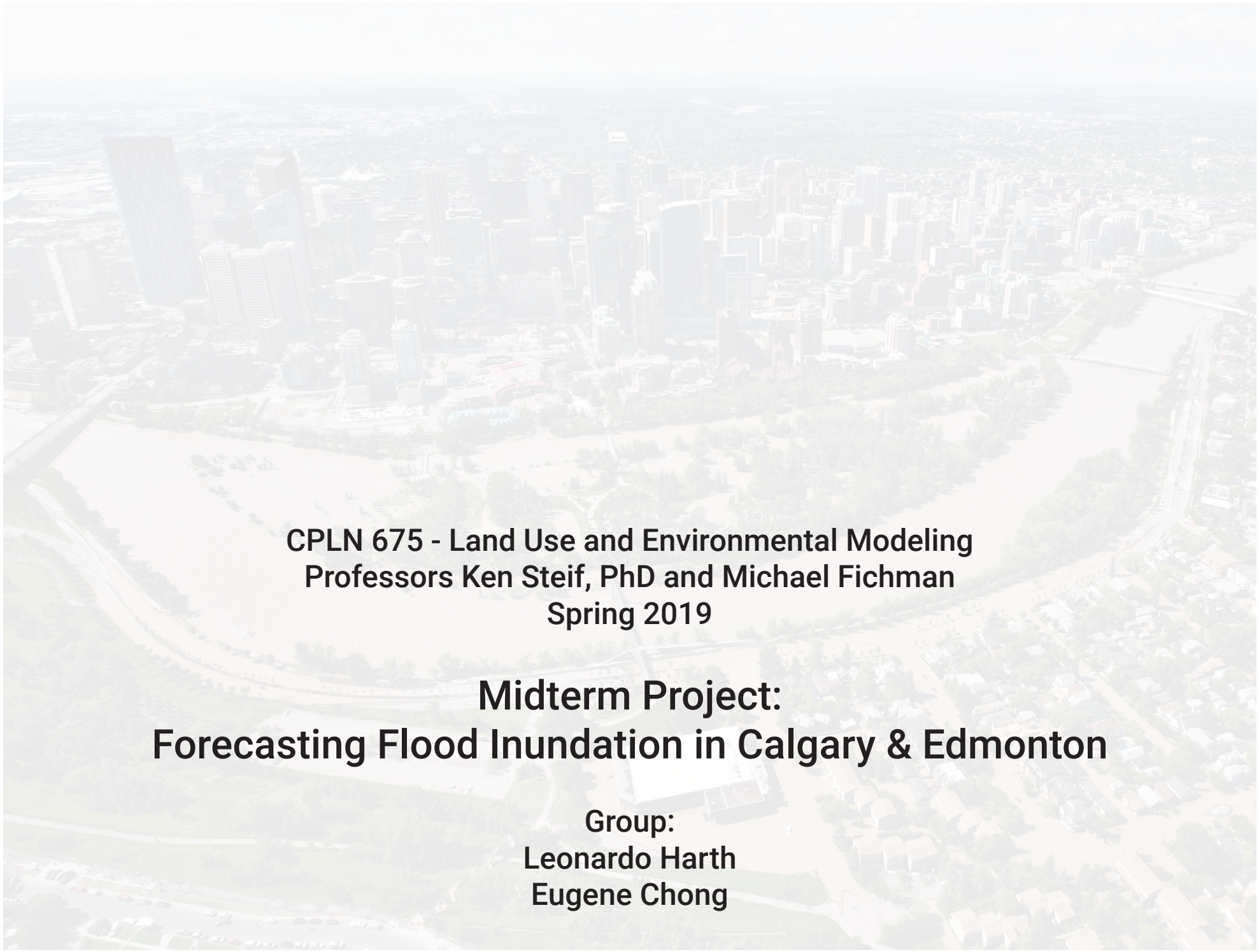


CPLN 675 - Land Use and Environmental Modeling
Professors Ken Steif, PhD and Michael Fichman
Spring 2019

Midterm Project:
Forecasting Flood Inundation in Calgary & Edmonton

Group:
Leonardo Harth
Eugene Chong

MODEL FEATURES

We used Calgary's elevation data to build flow length, stream order, and waterflow accumulation. We removed flow length, however, because it was not significant.

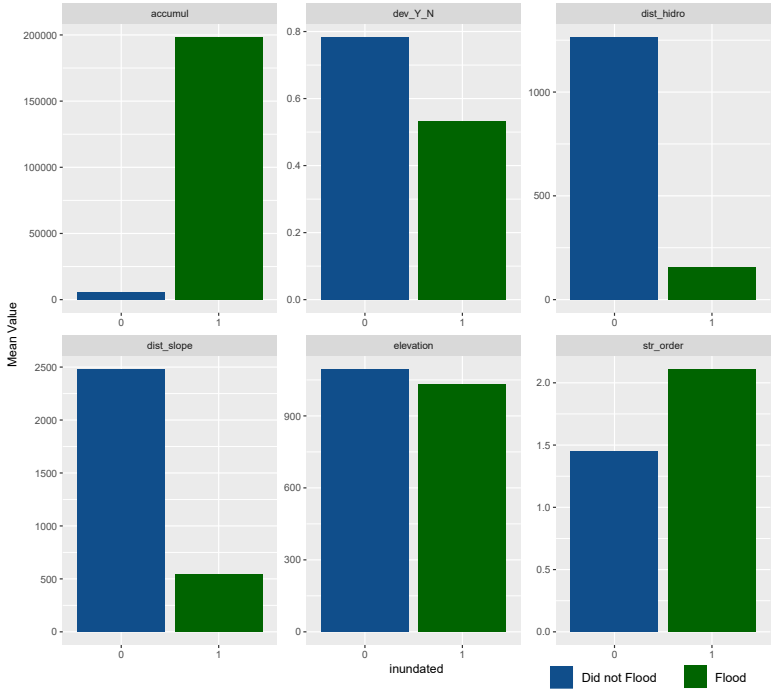
- Stream Order: Is the nearest stream the highest order stream?

- Waterflow accumulation: Highest accumulation in the fishnet cell.

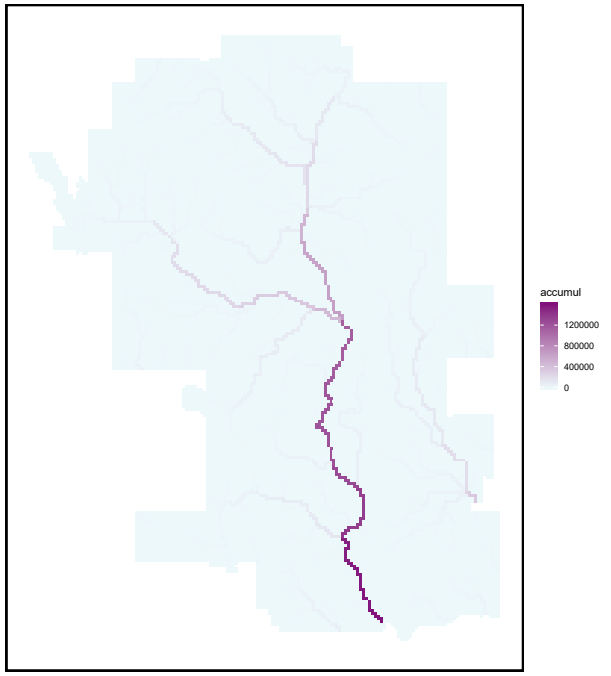
- Distance to Hydrology: Average distance to nearest water feature.

- Elevation: Average cell elevation minus the lowest in the city.

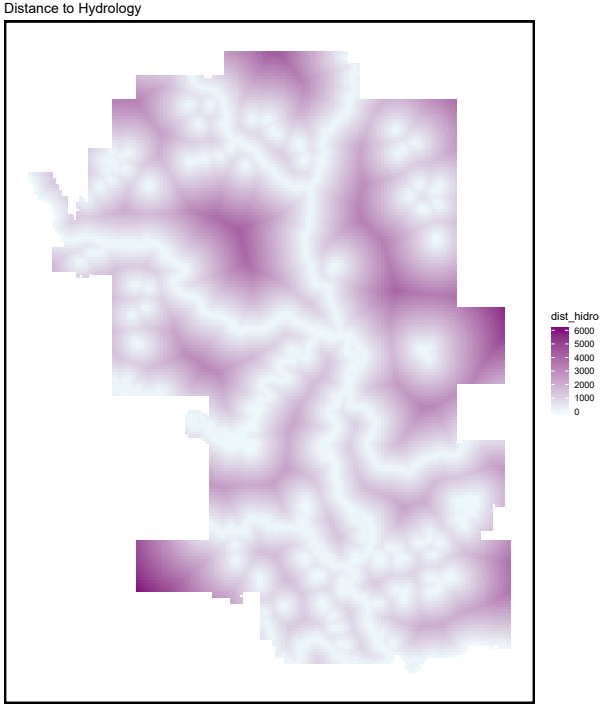
- Developed: Is the majority of a fishnet cell developed land?



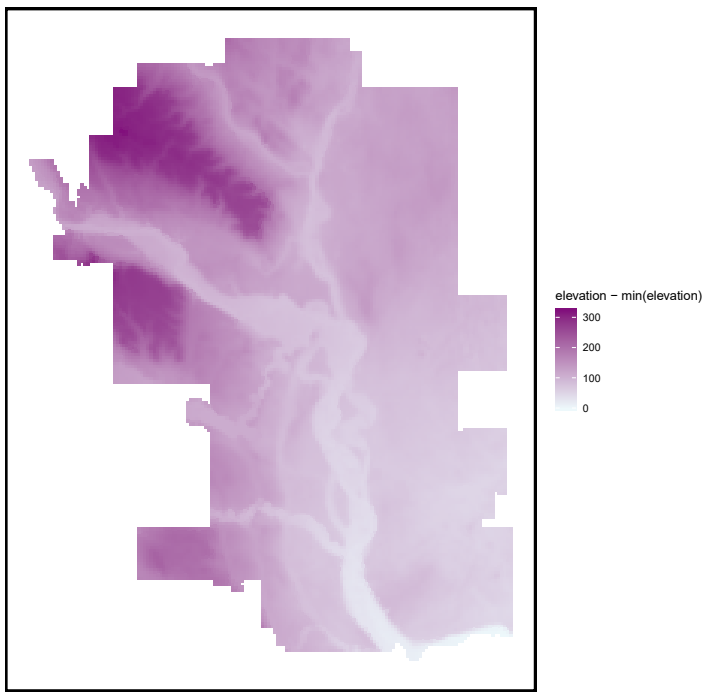
Waterflow Accumulation



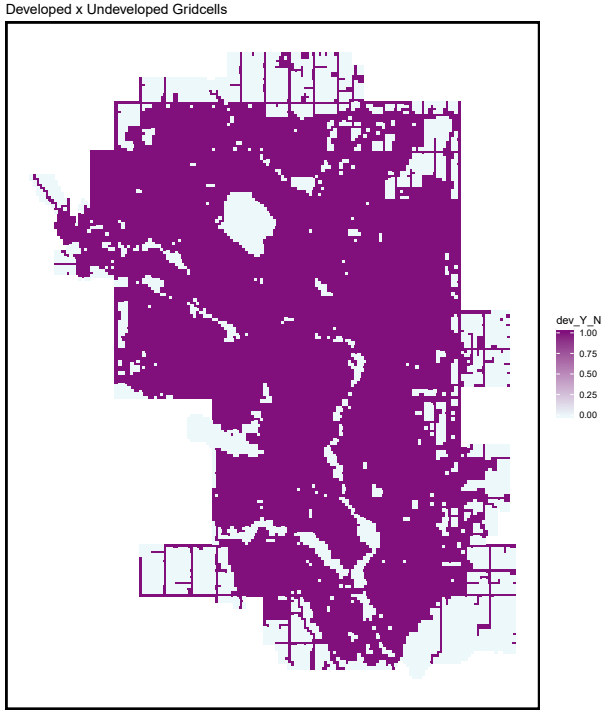
Distance to Hydrology



Gridcells elevation

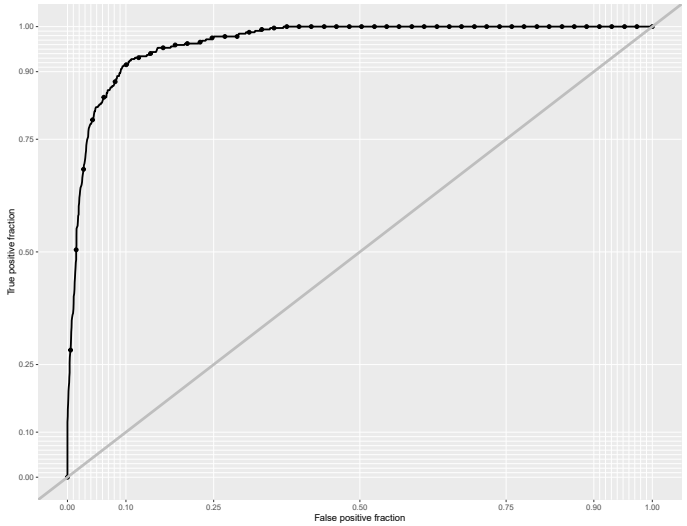


Developed x Undeveloped Gridcells



LOGISTIC REGRESSION

ROC CURVE



	Dependent variable:
	Inundation
elevation	-0.033***
	(0.002)
dist_slope	-0.001***
	(0.0001)
dist_hidro	-0.005***
	(0.0003)
dev_Y_N	-0.657***
	(0.104)
accumul	0.00000***
	(0.00000)
str_order	-0.554***
	(0.128)
Constant	3.483***
	(0.230)
Observations	15,134
Log Likelihood	-1,460.220
Akaike Inf. Crit.	2,934.440
Note:	*p<0.1; **p<0.05; ***p<0.01

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 6102  195
         1   65  124

               Accuracy : 0.9599
                 95% CI : (0.9549, 0.9646)
    No Information Rate : 0.9508
    P-Value [Acc > NIR] : 0.0002743

                  Kappa : 0.4687
 Mcnemar's Test P-Value : 1.242e-15

            Sensitivity : 0.38871
            Specificity : 0.98946
         Pos Pred Value : 0.65608
         Neg Pred Value : 0.96903
             Prevalence : 0.04918
         Detection Rate : 0.01912
   Detection Prevalence : 0.02914
      Balanced Accuracy : 0.68909

       'Positive' Class : 1
```

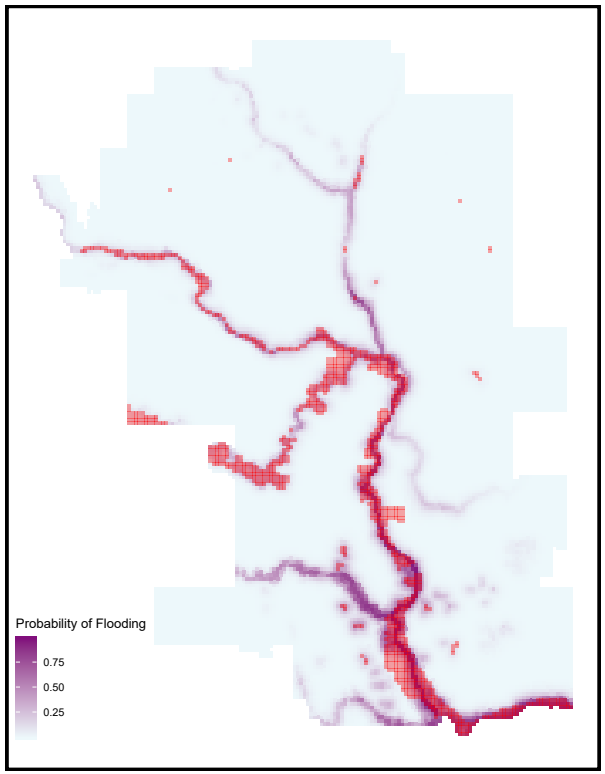
CROSS VALIDATION METRICS

```
Generalized Linear Model

21620 samples
    6 predictor
    2 classes: '0', '1'

No pre-processing
Resampling: Cross-Validated (100 fold)
Summary of sample sizes: 21403, 21404, 21404, 21404,
Resampling results:

  Accuracy   Kappa
  0.9576813  0.4520274
```

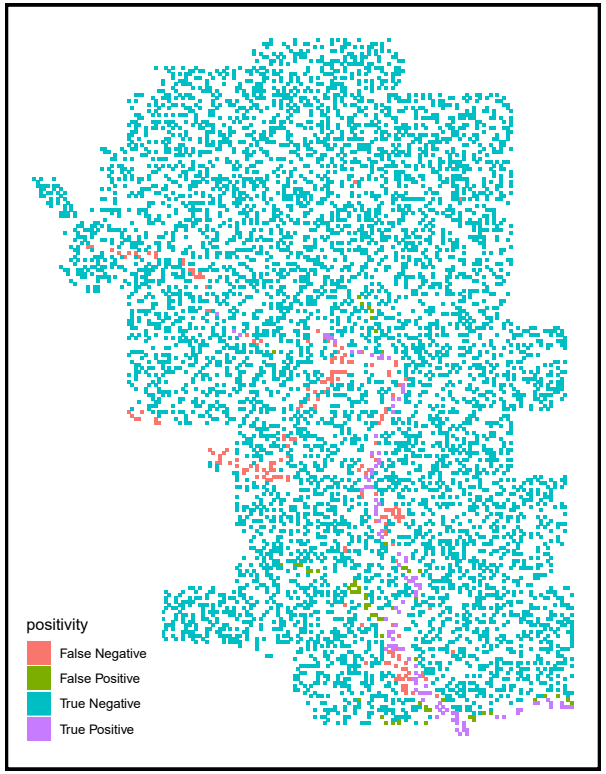
All of the selected independent variables were significant in the model. However, the model does not perform well predicting flooding in Calgary. The confusion matrix above shows the the model often fails to predict where flooding it actually happened (only a 39% true positive rate).

Its high accuracy (96%) is misleading since it performs well predicting the cell grids that did not flood would not flood. Since flooding is a rare event, it is easy to predict where it did not happen, and since the vast majority of cells did not flood, our overall accuracy is inflated. So, despite its high accuracy, our model is wanting in its purpose.

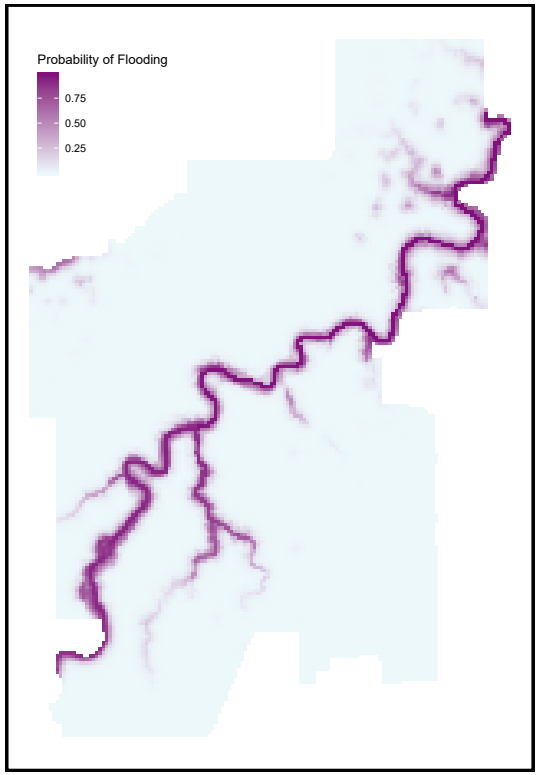
PREDICTIONS



Predicted x Reality - Actual Flooded cells in orange



Model Predictions for Calgary Test Set



Model Predictions for Edmonton

As discussed on the previous page, our model does not perform well predicting flood inundation in Calgary. The first map shows the actual flooding in orange and our predicted flooding in purple. Our model predicts flood inundation along the waterways, but it fails to predict flooding in areas that extend further past the banks of the Bow River, missing significant amounts of the inundated areas.

One possible explanation is the way the model interprets distance to hydrological features and flow accumulation. There may be an excessive difference between the grid cells with high flow accumulation and those nearby (which drain to the river).

Another possible explanation is that the many of the inundated grid cells are water surfaces, and this might be skewing the model's results. The model's predictions on the test set (second map) reinforce the findings discussed in the first paragraph.

We then applied the model to the City of Edmonton, north of Calgary. It features similar elevation, and it also has a river that flows through it. Like Calgary, Edmonton also features a massive amount of urban sprawl. Our model presents similar patterns in Edmonton, predicting flooding all along the waterways, but generally failing to extend far beyond the riverbanks.

Some other methods of wrangling the data that we collected, such as aggreating variables in different ways, may also improve the model and obtain better predictions.