Lily Pogue and Rebecca Risch
EARS 77 – Environmental Applications of GIS
Professor Jonathan Chipman
12 March 2024

Stop, Drop, and Roll: Modeling Wildfire Incidence Risk in Alaska

Abstract

Wildfire severity and frequency of occurrence are dramatically increasing with climate change and rising global temperatures, notably in areas where extensive burns are quite rare. Alaska is facing more wildfires each year, even in "fireproof" areas of the state. The goal of this study is to predictively model wildfire occurrence in Alaska with machine learning (ML). We identified three key factors in wildfire occurrence: land cover type, average summer temperature, and average summer precipitation. To train our ML models, we use historical wildfire polygon data as our ground truth. We decided to experiment with three different types of models: Decision Tree, Logistic Regression, and Random Forest, and compare their performances. All data preparation, and final visualizations, were done in ArcGIS Pro. Lastly, we sourced predicted future summer temperatures and precipitation in Alaska and ran this data through our best performing model, the Random Forest, thereby predicting future areas of Alaska at risk for wildfires. An accurate, predictive wildfire model, with fine-tuned variables and spatial-temporal frames of reference, could greatly benefit firefighters and local communities and prepare them to best handle burns.

Introduction

As global temperatures increase, so does the intensity of extreme weather events. Top news broadcast outlets release daily headlines about "100-year" floods, violent storms, abnormal hot and cold fronts and more, each bringing devastating effects to local landscapes and communities (Hersher, 2023). Wildfires are one such disaster, and each year brings a new round of record-breaking fires across the globe, many of them in places where extensive fires are largely unprecedented. Some more recent examples include the wildfires seen across Europe in

2022, those in Eastern Canada in 2023, and now in Texas, where reports surface daily of the active Smokehouse Creek Fire, the state's largest ever in history (Bogel-Burroughs, 2024). These worsening fires are largely due to increasing temperatures and lengthening fire seasons, phenomena that are seen worldwide but are particularly pronounced near the arctic, where temperatures are rising at twice the average rate globally (Grabinsky, 2020). Future efforts to manage and mitigate these fires will require advanced planning and preparation that will allow communities to respond and adapt to changing fire patterns.

This study aims to train three models to predict likelihood of wildfire occurrence across the state of Alaska in the future. As a state relatively close to the arctic, Alaska is experiencing a profoundly changing climate which is impacting and will continue to impact the state's ecosystems and fire regimes (Grabinsky, 2020). It is predicted that the state will experience larger, more intense, and more frequent wildfires in the near future, even in parts of the state once deemed "fireproof" such as the tundra and recently burned areas (Thiessen, 2022). In attempting to model this, we evaluate three key variables, land cover, predicted future summer precipitation, and predicted future summer temperature, to create a map indicating areas in Alaska that will be at a higher risk of wildfire occurrence than others. This first factor, land cover, is perhaps the most important variable to consider, as it largely determines likelihood of ignition. For instance, in permanently snowy areas or barren areas, it is not possible for a fire to burn, whereas in vegetated areas fire likelihood increases and varies depending on fuel type and availability. For instance, grasslands and coniferous forests are more prone to ignition than broadleaf forests (Zhai et al, 2023). The other two factors that we included in our study are average summer temperatures (in celsius) and average summer precipitation (in millimeters). Fires are more likely to ignite in warmer temperatures, and warm temperatures also help dry out fuel sources, leaving them more prone to ignition. Similarly, we use precipitation data because it too affects the moisture content in fuels and soils – low precipitation or drought conditions contribute to drying out fuel sources and land, and making them prone to ignition (Wasserman and Muller, 2023). In order to create these predictions, we used data indicating where fires have occurred in the past in Alaska as a "ground truth," training our models to recognize specific signatures of land cover, precipitation, and temperature where fires have occurred in the past which they could then use to predict at-risk areas on future precipitation and temperature data.

Our final result is a model that predicts areas at high risk of wildfires based on land cover and future summer temperature and precipitation data. An important distinction to make going forward is that rather than modeling fire *behavior*, which would examine factors such as wind speed and direction, spreading dynamics, etc, our work models risk of fire *occurrence,* or ignition, based on the three main factors. We expect to see an increase in land area that's at risk of wildfire, in line with current trends.

Methods

Data Collection and Preparation

The first step of our project was to collect our data. We sourced our Historical Wildfire polygon data from the USGS, originally provided by the Geospatial Multi-Agency Coordination (GeoMAC). This map, as seen in figure 5, has 1km^2/pixel resolution and provides one polygon for each fire from 2000-2018, representing the areal extent of each given wildfire in Alaska. We converted this polygon map into a raster, and then used the ArcGIS Pro Conditional function to produce a binary raster where pixels represented by the wildfire area received a value of 1, and the pixels in all other areas received a value of 0 which is shown in figure 6.

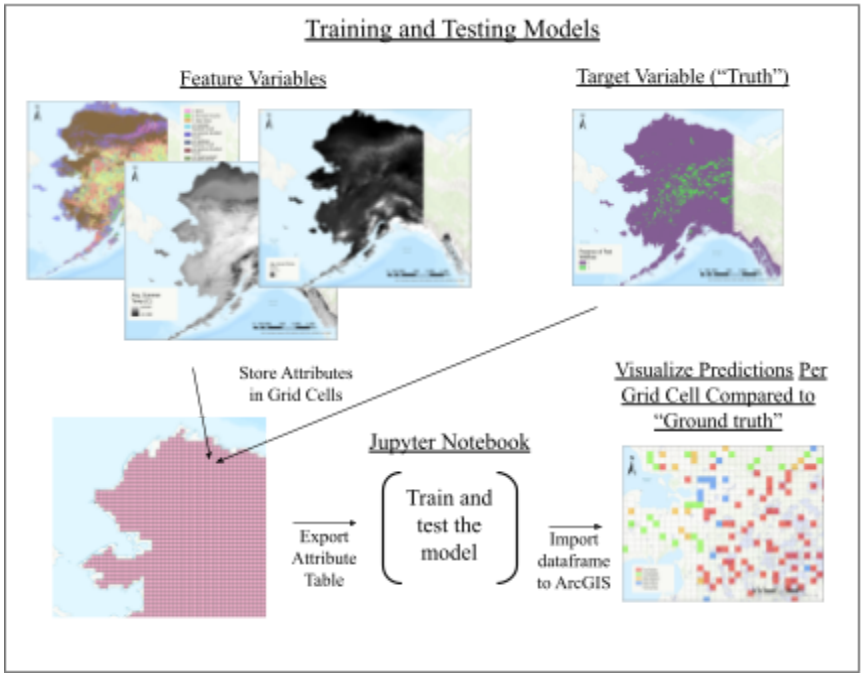
We sourced our Land Cover Type data from the MODIS Terra and Aqua satellite set, which provides five legacy classifications for land cover, of which we used one categorizing broad land cover and vegetation types. This dataset provides a classified map with a resolution of 1km^2/pixel, and was broken up into tiles based off of the MODIS Sinusoidal Tile Grid. We used and mosaiced 13 of these tiles together to cover the area of Alaska. This layer represented 16 different types of land cover over our study area based on the FAO-Land Cover Classification System, which categorizes majority vegetation types (or lack thereof). Some classes include Barren, Deciduous Broadleaf Forests, Mixed Broadleaf/Needleleaf Forests, Open Forests, Sparse Herbaceous, and Srubland/Grassland Mosaics. Details on all of the classes and thresholds for these classes can be seen in figure 2.

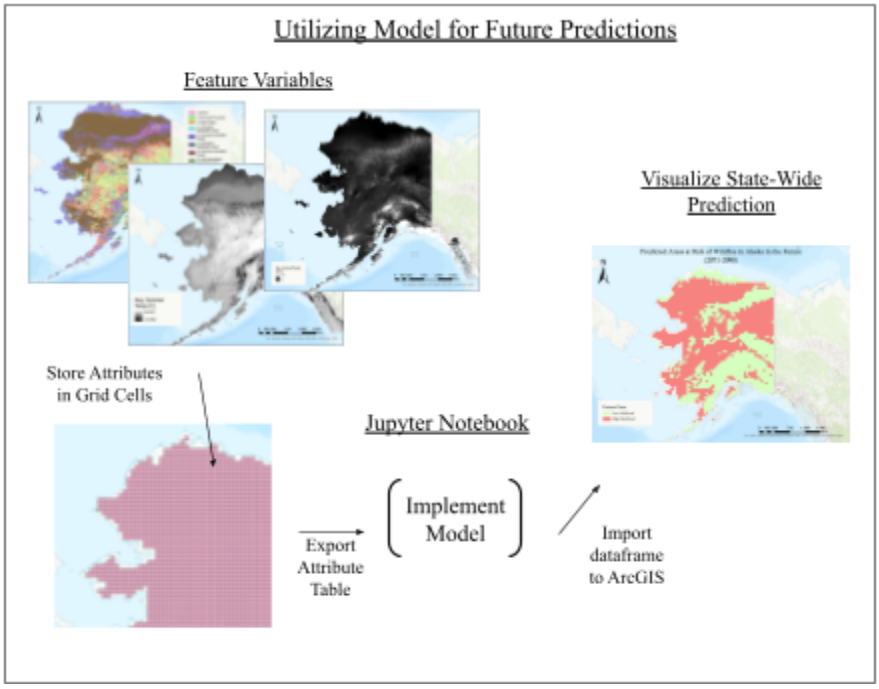
We sourced our Average Summer Temperature (ºC) (figure 3) and Average Summer Precipitation (mm) (figure 4) raster layers from the US Department of Agriculture Forest Service, who compiled their maps from SNAP: Scenarios Network for Alaska and Arctic Planning. These data sets are also 1km2 per pixel in resolution. Summer months are defined as June through August, and are the months considered to be fire season in Alaska (as with much of

the Northern Hemisphere). The data for our wildfire risk prediction was sourced from this same site, where they have predicted summer temperature and precipitation datasets (seen in figures 12 and 13).

Finally, to align our layers, we reprojected each layer to the North American Datum 1984 (2011) (equal area) Albers projected coordinate system for Alaska, as it's the most recent projection for the State that preserves accurate area, which is important in this context as we aim to compare areal extent of past and future risk areas. In addition, to ensure all of our data covered the exact same, detailed extent of our study area, we masked each layer into the shape of Alaska.

Once our data had been gathered and prepared, we used a fishnet grid covering the extent of Alaska with cell sizes of 50 x 50 km in order to spatially relate each variable to another. This allowed us to assign a value from each of our data layers to each individual grid cell, all of which could then be exported as a table where each grid cell (identifiable by an Object ID) was assigned different attributes. For instance, each individual cell contained a measure of the mean summer temperature, mean summer precipitation, majority land cover class, and binary value indicating the presence of a past fire or not. Once this table had been created, we were able to import it as a CSV into the computing software Jupyter Notebook in which we did the bulk of our analysis.





Analysis

After importing data and all necessary packages and tools for machine learning into a Jupyter Notebook file (the main one being scikit-learn), we assigned our feature and target variables as X (an array with all three of our feature variables) and y. Next, we split this data into training and testing splits, with 80% of the grid cells assigned to the training set, and 20% assigned to the testing set. We then initialized our first classifier, the Decision Tree, and performed a Grid Search function to determine the best hyperparameters (for example, maximum depth of the decision tree)for our tree based on our data. We then fit our training data to that classifier with the optimized hyperparameters, which completed the initializing and training process of this first model. Next, we predicted the occurrence of wildfires off of the X test data to be able to compare the model's predictions to the ground truth, and thereby assess the model's performance. To do so, we calculated the accuracy metric, confusion matrix, and true and false positive rates. We then merged the model's predictions back onto the X test data and the corresponding Object IDs to have one final table to visualize in ArcGIS Pro. We repeated the above steps, starting with initializing the models, for two additional types of models, Logistic Regression and Random Forest. These models used different hyperparameters than the Decision Tree but were completed with the same process. We also created a diagram to visualize our decision tree, (as seen in figure 7 and 8) using a library called graphviz.

With tables representing each models' predictions of the test set's fire occurrences, we visualized the models' performances in ArcGIS Pro. The following process was repeated for each models' table: First, we joined the table onto the fishnet. We then added a field, titled "Error" and calculated that field conditionally to compare the true fire occurrence binary of the test set and the model's predictions. For each grid cell, if there was no difference between the ground truth and the model's prediction, error equaled zero, but if the model predicted incorrectly, and gave a different fire binary value than the ground truth, error equaled one. Then, we displayed both variables, the ground truth fire binary value and the error value. We colored each outcome differently for the best visual representation of both the ground truth and the model's predictions. The display symbology is as follows:

- Red: True Positive (Fire Binary = 1, Error = 0)
- Green: True Negative (Fire Binary = 0, Error = 0)
- Yellow: False Negative (Fire Binary = 0, Error = 1)
- Blue: False Positive (Fire Binary = 1, Error = 1)

We have three maps with this symbology, each representing our different models' outputs: Decision Tree, Logistic Regression, and Random Forest.

After visualizing our model's performance on historical data, we wanted to predict fires that have not yet occurred. We found raster layers that represented predicted future average summer temperature and precipitation (from 2071-2090) from the same source as our historical temperature and precipitation data – the USDA Forest Service (SNAP). We brought in the data into ArcGIS Pro and prepared the layers and sourced the Zonal Statistics tables with the same methods as we used on the historical data. Because our Random Forest model performed the best, as deemed by its accuracy score, we used it to predict future fires. We ran the model on the future data table, which outputted fire predictions, again in the form of binary. We repeated the same processes of joining the predictions back onto the object IDs and feature variables, exporting the final csv, and joining the table onto our fishnet. Because we have no ground truth of wildfires that have not yet occurred, we could not compare the predictions to said ground truth and assess the model's performance/accuracy with this new data. Therefore, we could not calculate an "error" field. Instead, we simply represented predicted future wildfire extent in red, and the rest of the land area in green (see Figure 14). We also overlaid this result onto land cover to better understand the correlation between land cover and wildfire occurrence (See Figure 15).

With all of our final visualizations and accuracy metrics, we were able to compare model performances and discuss the broader implications of each output.

Results

Model	Accuracy	True Positive/False Positive
Decision Tree	0.8011	0.735/0.148
Logistic Regression	0.8047	0.761/0.161
Random Forest	0.8102	0.761/0.161

Preliminary results after training and testing our three models can be seen in the table above, and show relatively similar accuracies across methods with the random forest model performing slightly better (0.683% higher accuracy) than the other two. Interestingly, the Logistic Regression model performed slightly better than the Decision tree. For the Random Forest and Logistic Regression models, both have the same True Positive Rate (TPR) of 0.761 and False Positive Rate (FPR) of 0.161, suggesting that they have similar abilities to correctly identify areas of past fires. The Decision Tree model has a slightly lower TPR of 0.735, but also a lower FPR of 0.148 compared to the other models. This indicates that the Decision Tree model is relatively less effective at correctly identifying fire locations compared to the other models, but has a lower false alarm rate. Visualizations of these predictions can be seen in figures 9 - 11 which show the models correct and incorrect identifications of wildfires compared to the "truth," or actual past fires. These predictions are done on the 20% of the dataset reserved for testing, so only 548 of the 2737 total grid cells show predictions. In the end, because overall accuracy is our primary consideration, the Random Forest model is the one we chose to do our future prediction. However, because it has a similar accuracy rate and a lower false positive rate, the Decision Tree model could also be a good option for proceeding with prediction.

After implementing and comparing (numerically and visually) the results of our models, we selected the Random Forest Model to use in predicting future areas at risk of fire in Alaska

based on the future summer precipitation and temperature data. This data represents predicted average values for the years 2071 - 2090, and therefore our risk predictions (seen in figure 14) are for that future timeframe. The model has identified areas (in red) which have similar characteristics (in terms of precipitation, temperature, and landcover) to areas where past fires have taken place and thus will be prone to wildfires in the future as temperatures and precipitation levels change. As this result represents a future prediction, we are unable to measure accuracy. However, predictions' alignments with overall trends help measure its success.

Discussion

Our findings, as seen in figure 14, matching our expectations, show a large increase of area at risk of wildfires in many parts of the state. We can compare this to past fire extents, which as seen in figure 5 stretch across the central and eastern part of the state extending down slightly to the southwestern corner. Fire risk area in our future prediction increases across much of this same area, while also extending further to the southwest and up along the western coast, and notably covering a large swath of the top of the state to the north. These areas at risk not only cover a larger portion of the already-prone center of the state, which is made up largely of open forests (yellow) and sparse forests (pale pink) but also areas with relatively few fires in the past where the dominant land cover classification is shrubland/grassland mosaic (brown). This is visualized below in figure 15, where one can see that risk areas align closely with these land cover types, particularly in the north and along the west coast where we see fire risk extending to shrubland/grassland (tundra) areas which have seen relatively few fires in the past. The tundra in Alaska stretches around the western and northern coast (figure 16) and is characterized by a lack of trees, high winds, and very little rain or snow. Historically, average yearly temperatures in this region remain low, with summer temperatures very rarely reaching above 50℉ ("Tundra…"). However, this average is rising as global temperatures change, and the tundra is not only increasing in temperature but is increasing faster than much of the rest of the state; average annual temperature change in the tundra has increased 6.2℉ from 1970 to 2019, compared to the rest of the state which increased between 2.5 and 4.5 ℉ (Grabinski and McFarland, 2020). Our model's prediction thus aligns with other estimates that in northern and western Alaska, warm and dry conditions have and will contribute to an increase in wildfire frequency, and it's

estimated by some that "by 2100, the Alaskan tundra may experience twice as much total burned area and up to four times more frequent burns compared to historical records" (Grabinski and McFarland, 2020, pg 1).

There are of course some limitations to our model, the first being the number of variables used to access ignition risk. Wildfires are incredibly complex processes, and many factors are at play in determining their ignition such as humidity, soil and vegetation moisture, and fuel structure and distribution. We chose to use three comprehensive variables to broadly account for those factors, for example using used precipitation data as an proxy for humidity levels and vegetation moisture content, making the assumption that given Alaska climate, areas with low precipitation could be presumed to have low moisture levels in terms of humidity and vegetation moisture content, both of which are important factors in wildfire ignition. Now that the methodology has proved successful, it may be worth re-training our model or a similar one on these more specific variables.

In a similar vein, our analysis was done per 50 x 50 m grid cell, meaning that attributes for each cell were averaged over 2,500 m^2. While this provided a suitable level of analysis for the purposes of calculating risk state-wide, there is room for refinement through finer-level analysis, especially in areas where localized risk assessment is most important, such as around critical infrastructure.

Conclusion

Our findings provide insight on the use of machine learning models to calculate risk. The success of our three models at accurately (~80%) predicting historical fires points to its capability to capture broad patterns of wildfire risk in the future, a fact that's supported by the alignment of our model's predictions to known environmental factors, such as warming and drying trends in Alaska's tundra. These findings also underscore the urgency of addressing climate change impacts on wildfire frequency and severity, as temperatures will continue to rise and wildfire mitigation efforts will become ever more important. In this sense, our model holds promise for practical applications for wildfire and risk management efforts, as insights on at-risk areas can help inform emergency response and strategic planning, as well as resource allocation. In a broader sense, this work provides an example of how the use of predictive modeling with machine learning can help communities become more resilient to escalating climate risks and natural disasters.

Acknowledgements

We would like to thank Professor Jonathan Chipman, Aletha Spang, Victoria Halvorson, and Lexi Farnell for their support throughout this project, whether by answering many questions, advising us on our methods, or helping us search for data sources.

Works Cited

"Alaska's Changing Wildfire Environment." *University of Alaska Fairbanks; International Arctic Research Center."* https://uaf-iarc.org/alaskas-changing-wildfire-environment/.

Bogel-Burroughs, Nicolas, Jacey Fortin, and Anna Betts. 2024. "Texas Fires Span Over 1.2 Million Acres. Here Is What We Know." The New York Times. https://www.nytimes.com/article/texas-smokehouse-creek-fire.html.

Dobilas, Saul. 2021. "CART: Classification and Regression Trees for Clean but Powerful Models." Medium. https://towardsdatascience.com/cart-classification-and-regression-trees-for-clean-but-powerful-models-cc89e60b7a85.

Grabinski, Z. & H. R. McFarland. 2020. Alaska's changing wildfire environment [outreach booklet]. Alaska Fire Science Consortium, International Arctic Research Center, University of Alaska Fairbanks.

Hersher, Rebecca. 2023. "Climate change makes heat waves, storms and droughts worse, climate report confirms." NPR. https://www.npr.org/2023/01/09/1147805696/climate-change-makes-heat-waves-storms-and-droughts-worse-climate-report-confirm.

Oxford, Mariah. 2022. “Wildfire Hazard map Gives Residents Important Risk Information.” University of Alaska Anchorage. https://www.uaa.alaska.edu/news/archive/2022/05/Wildfire-hazard-map-residents-risk-information.cshtml.

Thiessen, Mark. 2022. “Massive wildfires burn through ‘fireproof’ regions in Alaska.” Public Broadcasting Service. https://www.pbs.org/newshour/nation/massive-wildfires-burn-Through-fireproof-regions-in-alaska.

“Tundra Ecosystems.” Alaska Department of Fish and Game. https://www.adfg.alaska.gov/index.cfm?adfg=tundra.extent

Wasserman, T. N., & Mueller, S. E. (2023). Climate influences on future fire severity: a synthesis of climate-fire interactions and impacts on fire regimes, high-severity fire, and forests in the western United States. Fire Ecology, 19(1), 43.

“Wildfires and Climate Change.” Center For Climate and Energy Solutions. https://www.c2es.org/content/wildfires-and-climate-change/#:~:text=Wildfire%20risk%20depends%20on%20a,climate%20variability%20and%20climate%20change.

Zhai, Jun, Zhuo Ning, Ram Dahal, and Shaoyang Yang. 2023. "Wildfire Susceptibility of Land Use and Topographic Features in the Western United States: Implications for the Landscape Management" Forests 14, no. 4: 807. https://doi.org/10.3390/f14040807.

Figures

Figure 1: Land Cover in Alaska

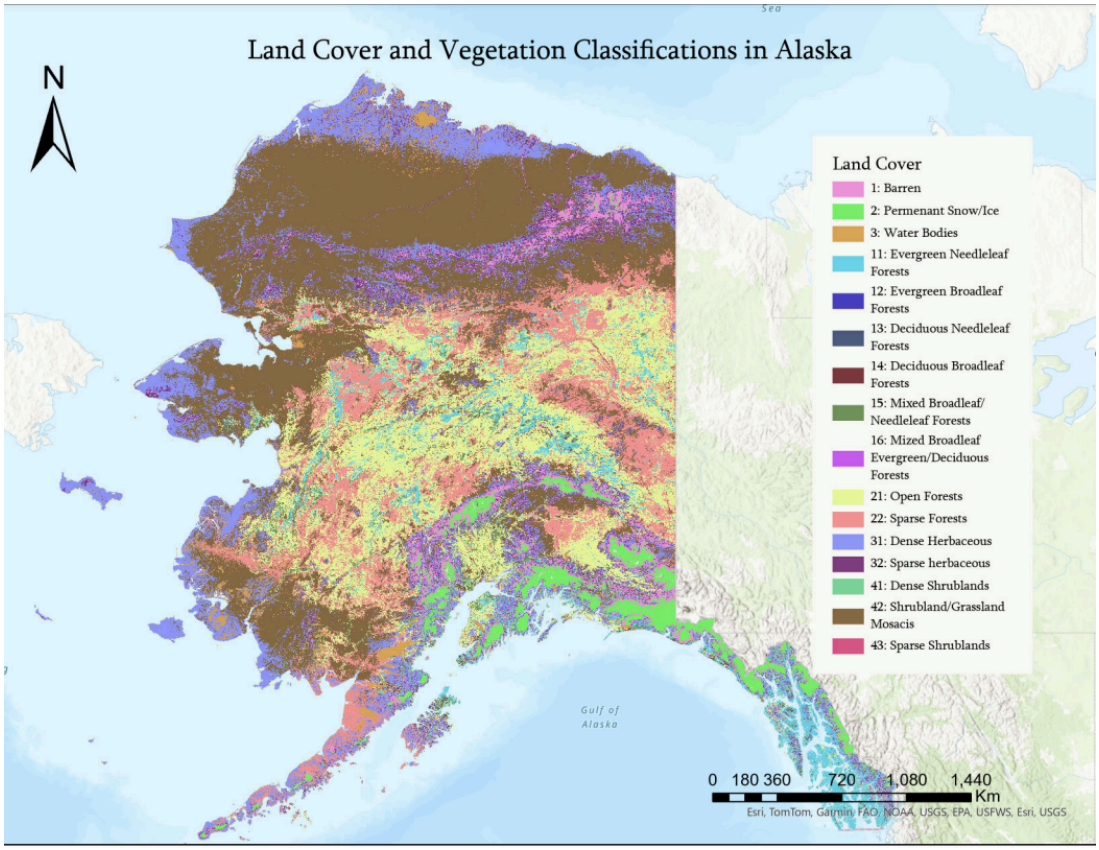


Figure 2: Land Cover Classes and Thresholds

Table 8: FAO-Land Cover Classification System land cover (LCCS1) legend and class definitions.

Name	Value	Description
Barren	1	At least of area 60% is non-vegetated barren (sand, rock, soil) or permanent snow/ice with less than 10% vegetation.
Permanent Snow and Ice	2	At least of area 60% is covered by snow and ice for at least 10 months of the year.
Water Bodies	3	At least 60% of area is covered by permanent water bodies.
Evergreen Needleleaf Forests	11	Dominated by evergreen conifer trees (>2m). Tree cover >60%.
Evergreen Broadleaf Forests	12	Dominated by evergreen broadleaf and palmate trees (>2m). Tree cover >60%.
Deciduous Needleleaf Forests	13	Dominated by deciduous needleleaf (larch) trees (>2m). Tree cover >60%.
Deciduous Broadleaf Forests	14	Dominated by deciduous broadleaf trees (>2m). Tree cover >60%.
Mixed Broadleaf/Needleleaf Forests	15	Co-dominated (40-60%) by broadleaf deciduous and evergreen needleleaf tree (>2m) types. Tree cover >60%.
Mixed Broadleaf Evergreen/Deciduous Forests	16	Co-dominated (40-60%) by broadleaf evergreen and deciduous tree (>2m) types. Tree cover >60%.
Open Forests	21	Tree cover 30-60% (canopy >2m).
Sparse Forests	22	Tree cover 10-30% (canopy >2m).
Dense Herbaceous	31	Dominated by herbaceous annuals (<2m) at least 60% cover.
Sparse Herbaceous	32	Dominated by herbaceous annuals (<2m) 10-60% cover.
Dense Shrublands	41	Dominated by woody perennials (1-2m) >60% cover.
Shrubland/Grassland Mosaics	42	Dominated by woody perennials (1-2m) 10-60% cover with dense herbaceous annual understory.
Sparse Shrublands	43	Dominated by woody perennials (1-2m) 10-60% cover with minimal herbaceous understory.
Unclassified	255	Has not received a map label because of missing inputs.

Figure 3: Average Current Summer (June - August) Temperatures in Alaska (C)

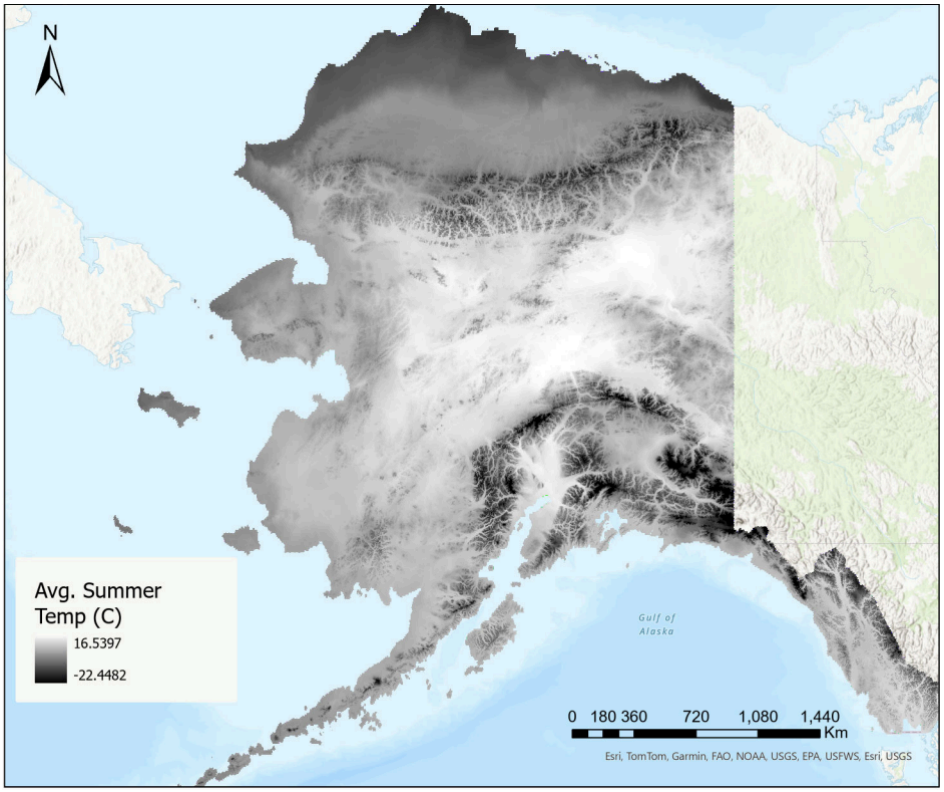


Figure 4: Average Summer (June - August) Precipitation (mm) in Alaska

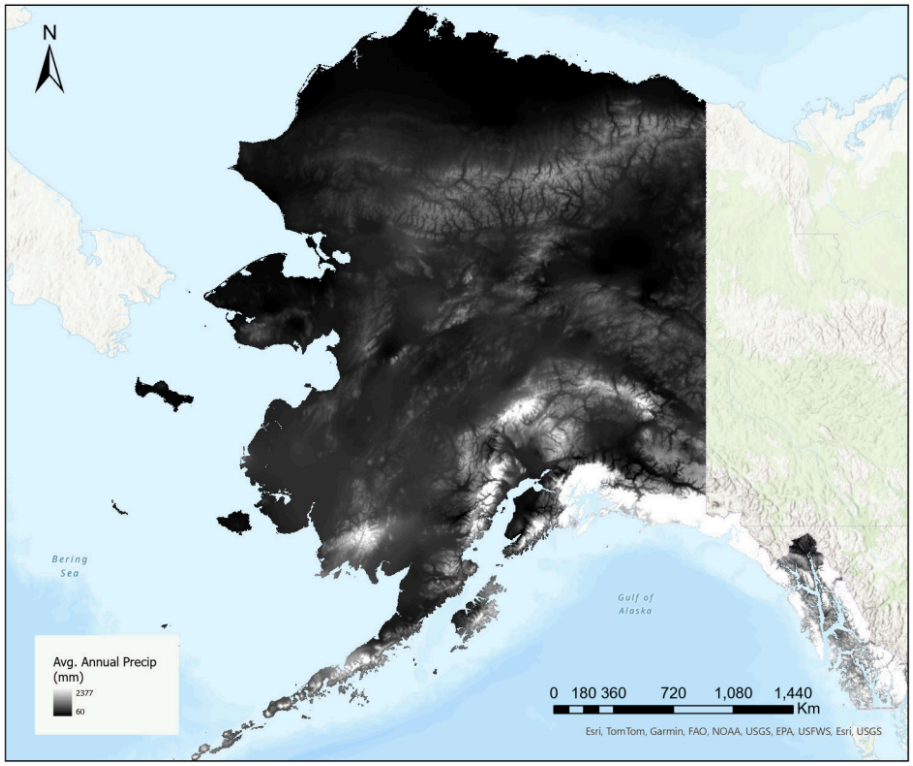


Figure 5: Extents of Historical Wildfires in Alaska (2000-2018)

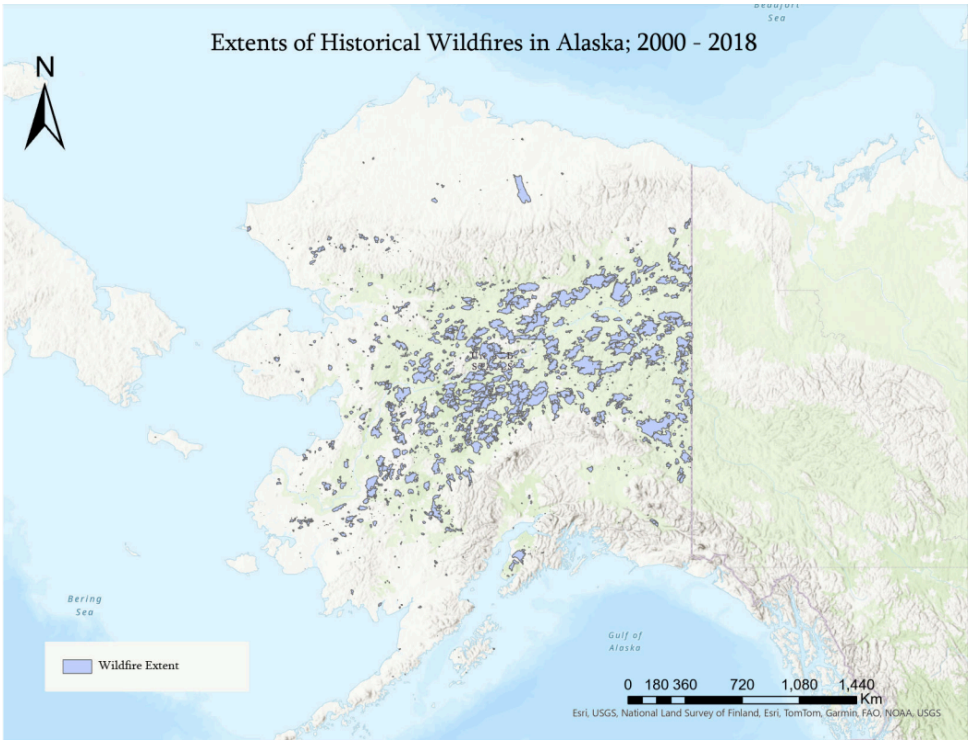


Figure 6: Binary Historical Fire Layer

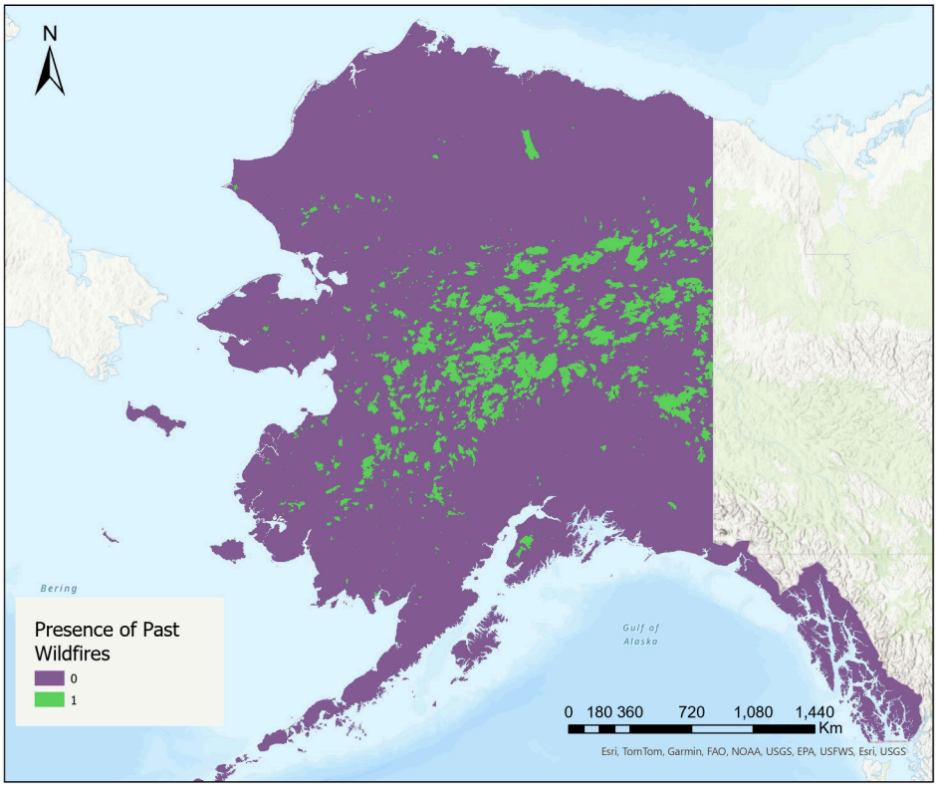


Figure 7: Visualization of Our Decision Tree

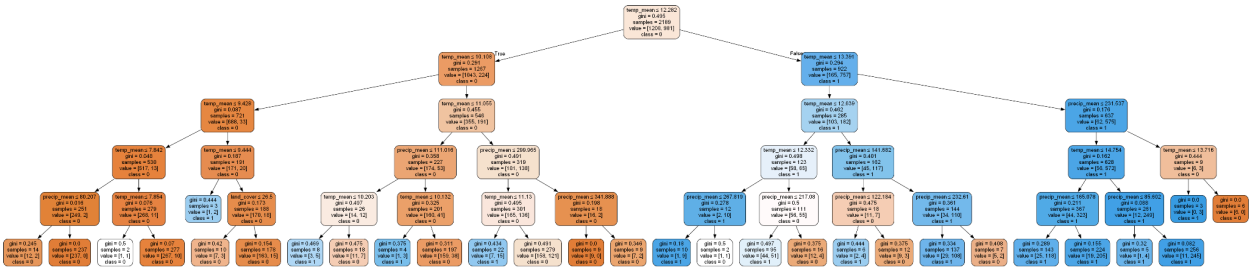


Figure 8: Segment of Decision Tree

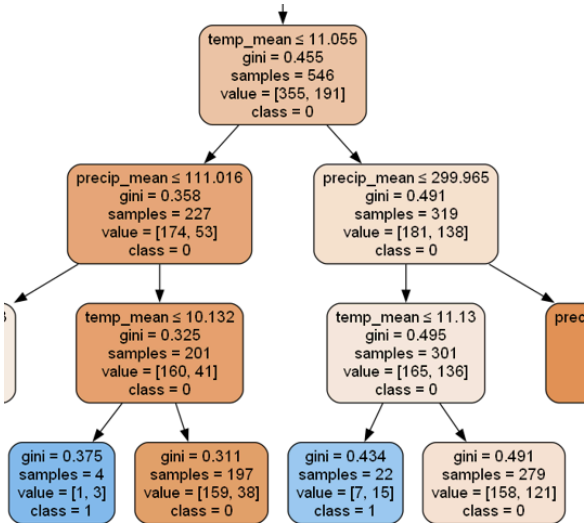


Figure 9: Results of Prediction on Test Dataset; Decision Tree

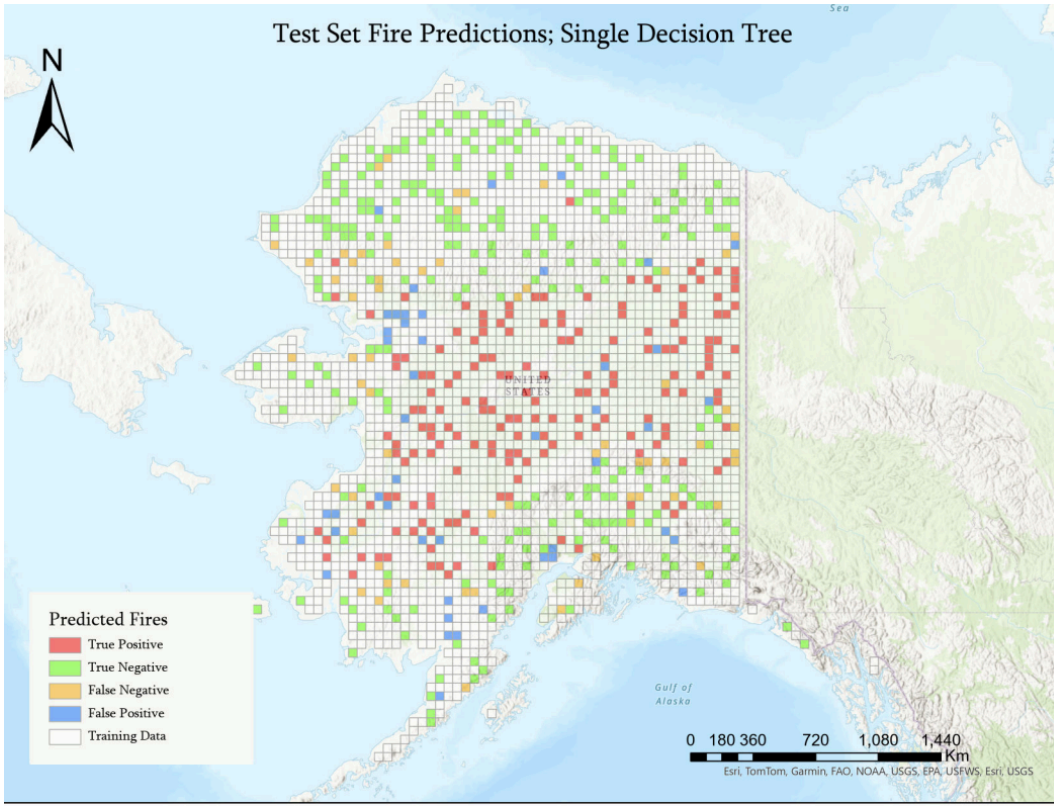


Figure 10: Results of Prediction on Test Dataset; Logistic Regression

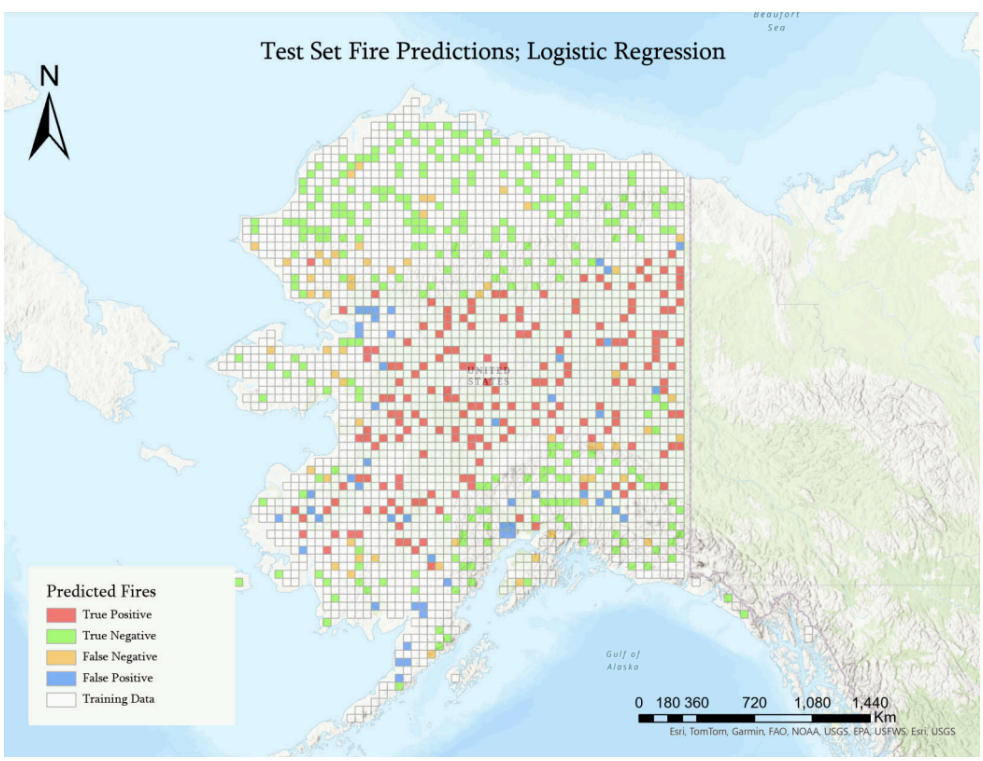


Figure 11: Results of Prediction on Test Dataset; Random Forest

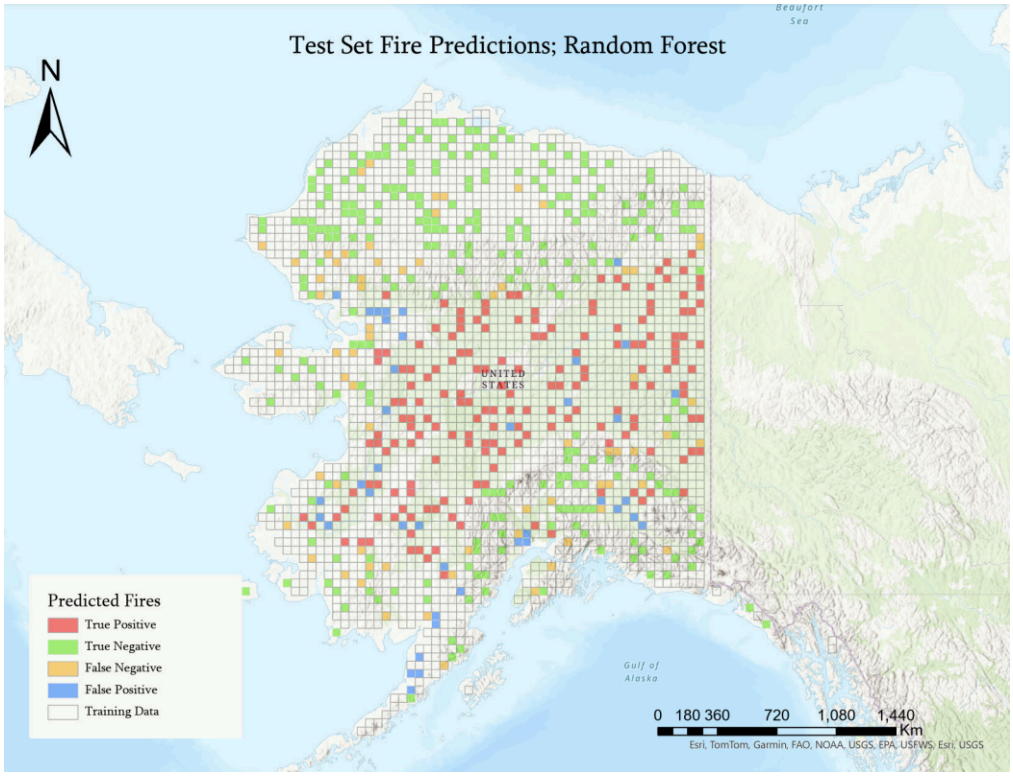


Figure 12: Predicted Future Summer Temperature (June - August) in Alaska (C)

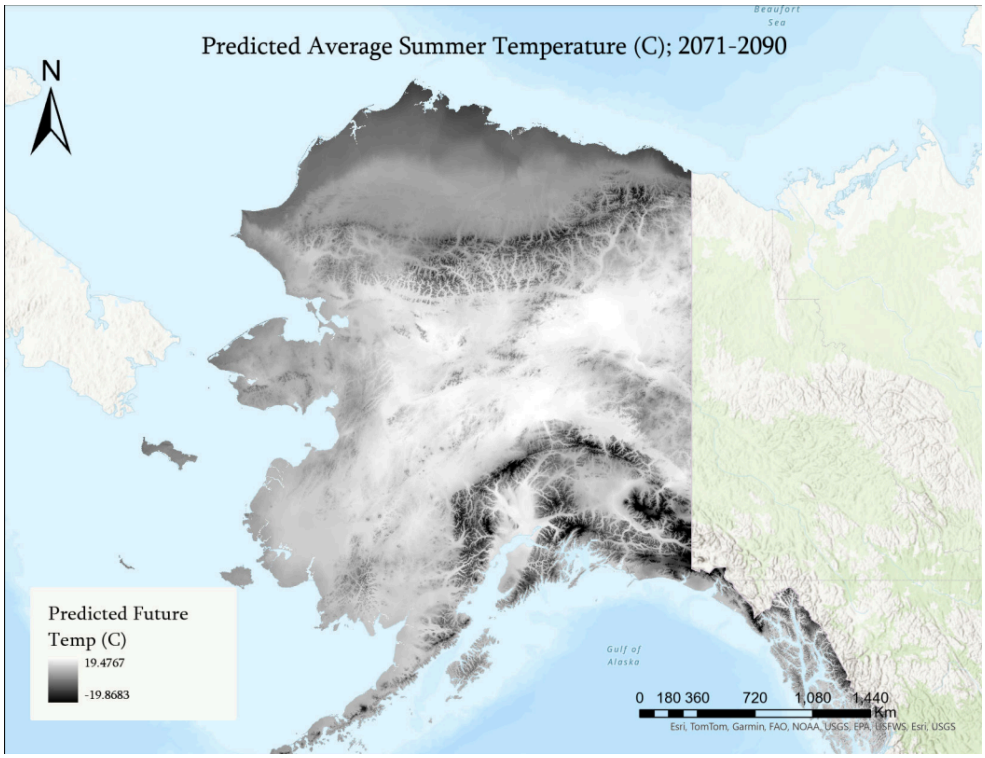


Figure 13: Predicted Future Summer Precipitation (June - August) in Alaska (mm)

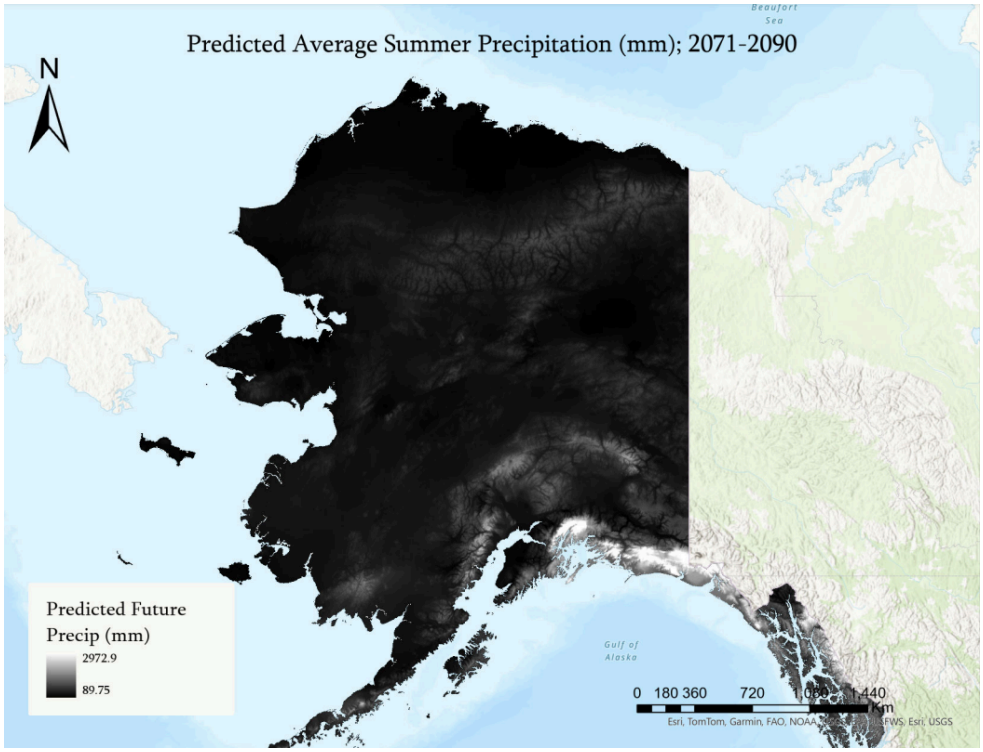


Figure 14: Final Results; Future Wildfire Prediction for Alaska based on Predicted Summer Precipitation and Temperature for 2071 - 2090

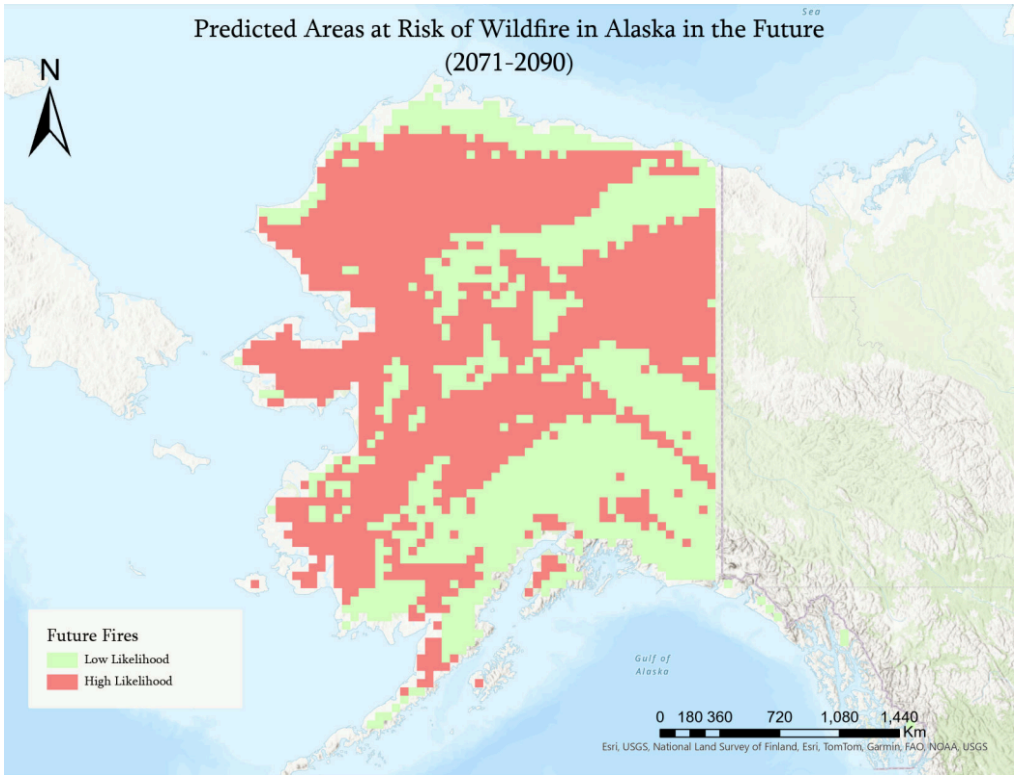


Figure 15: Predicted Future Fires Overlayed on Alaska's Land Cover

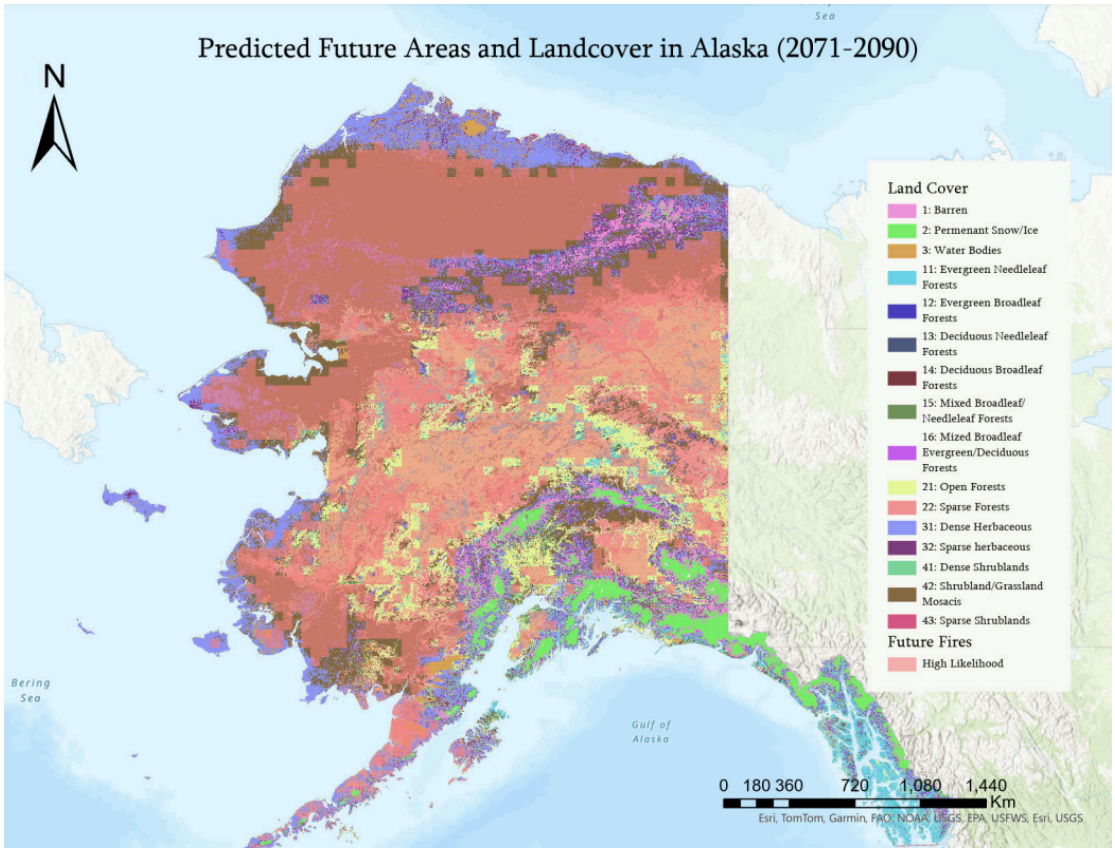
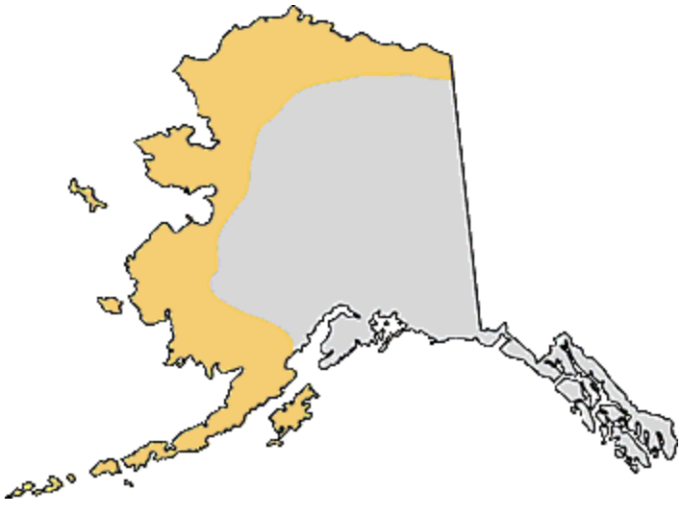



Figure 16: The Broad Extent of Alaska's Tundra Ecosystem