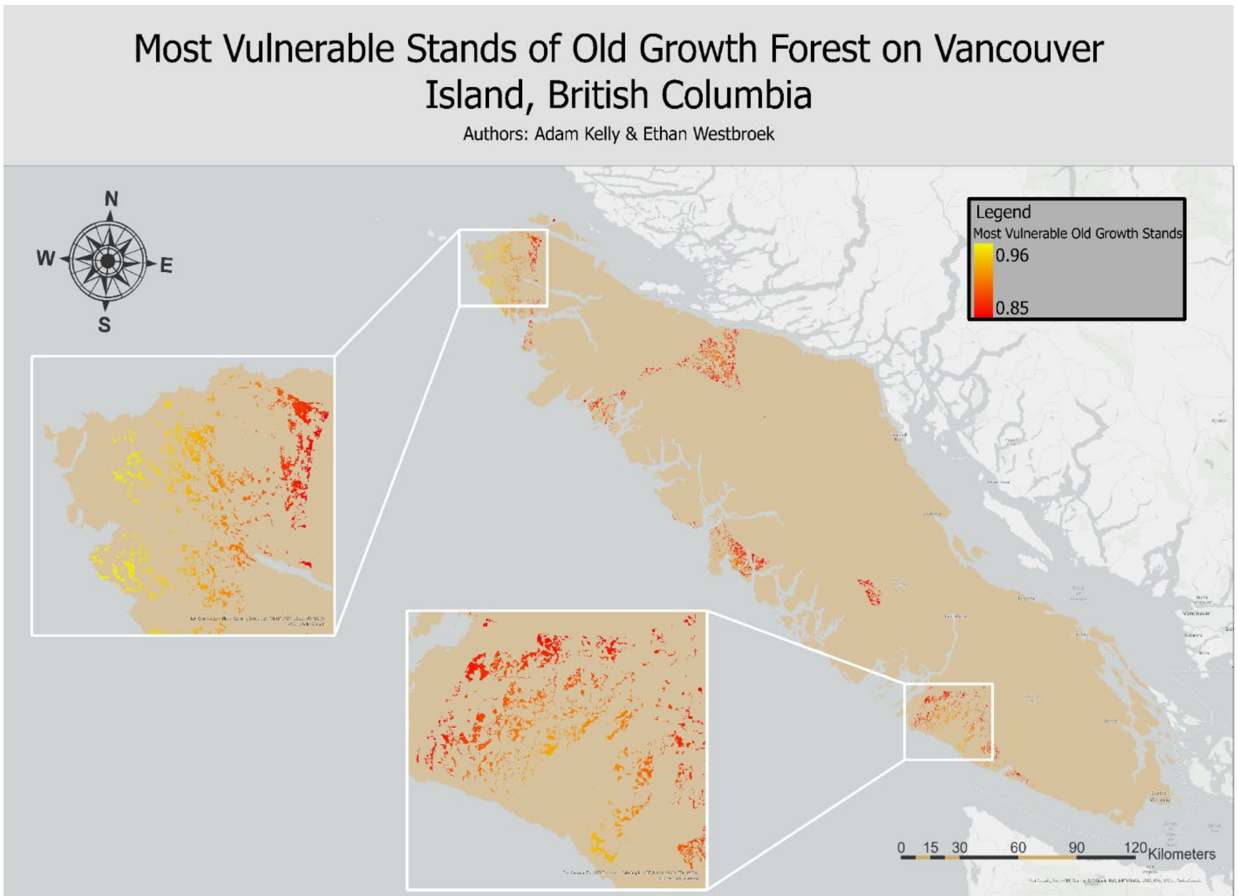
Final Report for GEOG 4480

Assessing Vulnerability in Remaining Old Growth Tree Stands on Vancouver Island, British Columbia, Through a Logistic Regression Model

Adam Kelly & Ethan Westbroek
April 14th, 2023



Abstract:

The deforestation of old growth trees on Vancouver Island and across the entirety of British Columbia is an active resource management issue. As old growth biomes are non-renewable, yet more economical to harvest than second growth stands, there is conflict between those aiming to retain them, and those looking to exploit them. General dissatisfaction from the public with their environment's management led to a strategic review and a recommendation of area based deferrals. Despite these 'calming' measures implemented by the government, logging has still been occurring on Vancouver Island. Various factors that were presumed to create vulnerability of a stand to logging were produced in a raster format to be assessed through logistic regression to determine their influence towards producing deforestation. The factors represented are indicative of the distance to the nearest roads, cities, pipelines, and railways, along with a digital elevation model, and slope angle model. Analysis occurred through an extraction of the variables associated values based on randomized coordinate locations in two separate areas classed as either deforested between 2013 and 2021 or intact through said period. These results were fitted to a logistic regression model that assessed slope, distance to various public infrastructure, and elevation's relationship individually and in one or more combination with each other to the binary response variable of either retained or deforested. The resulting models indicate that distance to cities was the most influential factor in determining old growth deforestation, with minor correlations to pipelines and railways. Yet, the minor correlation between deforestation and existing pipelines is likely attributed more to pipelines correlation with cities than to the actual response variable. Using the results from this model, a probability raster was created which is indicative of the vulnerability of unprotected old growth forests to experiencing deforestation across Vancouver Island. The top 15% most vulnerable stands were classed out of this raster, totalling 46,075 raster cells or 46,075 hectares of old growth stands that are significantly vulnerable to deforestation.

Table of Contents:

Abstract: **2**
Table of Contents: **3**
Introduction: **4**
Problem Context: 4
Research Purpose: 6
Objectives: **6**
Study Area: **7**
Data: **9**
Methods: **10**
Results and Discussion: **14**
Model Assessment 14
Roads 16
Elevation 17
Slope 18
Probability Raster and Old Growth Vulnerability 20
Limitations and Potential Improvements: 25
Conclusion: **27**
References: **28**
Appendices: **32**
Appendix 1: DEM and Slope Preparation 32
Appendix 2: Data Preprocessing 33
Appendix 3: Euclidean Distance Calculations 34
Appendix 4: Change Detection Data Preparation 35
Appendix 5: R Code 36
Appendix 6: Model Output Processing 37

Introduction:

Problem Context:

Old growth forest stands are one of the rarest biomes in Canada, and the Government of British Columbia (B.C.) has been failing to protect them (Raitio & Saarikoski, 2012). Forestry management in B.C. is timber centric, and under this paradigm, the stands are largely treated as inexhaustible (Robertson, 2022). The biome would need to regenerate at a more rapid rate than it currently does for it to ever be considered a renewable resource (Gorley & Merkel, 2020). Treating these stands as renewable when they indeed are not, has led to insufficient and ill-enforced protection of them, leading them on a path towards their ultimate demise. Protecting old growth is key due to its aforementioned irreplaceability, superiority over second growth stands in many environmental metrics, and the non-traditional values contained in its ecosystem.

The alternative to old growth stands is the second-growth stands which replace the old growth which are produced normally through manual replanting but natural regeneration occurs as well. These stands can never reach the level of function the old growth stands had achieved (Clason et al., 2008). Essentially, the second-growth stands reach the competitive exclusionary stage earlier in their biomes successional lifespan. This being due to all the saplings being planted during the same season, there are no canopy gaps which creates a lack of understory development, abbreviating the herb-shrub stage of succession, limiting the number of potential niches available in the biome, and thus these biomes have lower levels of ecosystem services overall (Clason et al., 2008). Two of the most impactful ecosystem services are carbon sequestration and biodiversity protection. As an example, total soil carbon biomass is 2.6-fold (coarse-textured soils) and 3-fold (fine-textured soils) higher in old growth stands than in young second-growth stands (Fredeen et al., 2005). At the same time, northern stands have been found to support seven times as many unique species and a distinctly different community composition for all functional groups (Price et al., 2017). Essentially, old growth ecosystems are much more advanced than second growth in many ecosystem metrics.

Logging is an important industry in British Columbia, as about a quarter of their exports come from timber (Price et al., 2021). The current forest management regimes in British Columbia designed solely around the production of merchantable timber, neglecting the alternative economic returns from these biomes (Robertson, 2022). These returns sometimes classed as NTFP (non-timber forestry products) include capital generated from ecotourism, foraging, and hunting to name a few (Clason et al., 2008).

Many groups of activists, NGOs, and political parties are striving for a reframing of the system that manages British Columbia's old growth biomes. Many of these folks are discontent with the pace and consistency of management. Specifically, recommendation number six from Gorley and Merkel's government-appointed 2020 strategic review of old forests calls for a temporary pause of logging of forests that meet a specific set of criteria, yet many of these deferred forests are still being logged as we speak (Robertson et al., 2022). NGO Stand Earth has been collecting evidence from satellite imagery, advocacy groups, and communities across the province and has documented many areas that were officially deferred, yet logged regardless (Robertson et al., 2022). This failure of governance to protect the forests and the benefits derived from it is why the nature of this analysis focuses on the parameters influencing deforestation. With government intervention historically being one of the largest factor preventing environmental degradation, seeing it fail highlights an increased need to understand the other parameters that are influencing deforestation.

It is apparent that this spatial problem would require a geospatial analysis to reach conclusions about the drivers of the phenomena. Currently, much of the literature on old growth retention and protection is either from ecological (Price, 2017), economic (Clason, 2008), or governance perspectives (Raitio, 2012). Although these fields do often include geospatial and statistical analysis in their study, there is value in performing an analysis that combines aspects from these fields into an overarching geospatial and statistical workflow. Despite reviews like Gorley and Merkel's from 2020 producing spatial outputs like maps, their workflow is still quite different from ours, as statistical analysis of the significance of influential variables was lacking in said report (Gorley & Merkel, 2020). The gap in the research is that nobody has formally assessed through statistics which variables are the most influential in the felling of these habitats.

Research Purpose:

Our research purpose aims to assess the varying factors influencing deforestation to advance the literature regarding how infrastructure and topography influence deforestation. This purpose aims to aid the citizens of British Columbia by providing data for them to utilize in battle to retain the stands that are the most vulnerable to be clearcut and harvested next. A statistical analysis of the variables that might influence deforestation uncovered and highlighted the relations both between/within variables and from any variable to deforestation. With this knowledge both statistically, and spatially, those fighting for the retention of these biomes will be able to focus their limited resources and mobilize for more effective stewardship outcomes.

Objectives:

Objective #1: Determine variables that may have an impact on deforestation on Vancouver Island, British Columbia.

Objective #2: Determine where old growth stands have been lost in the past. We will use this information to inform the model we will be creating in our future objectives.

Objective #3: Create logistic regression models to determine which variables have the greatest influence on the deforestation of old growth forests.

Objective #4: Use the outputs from the most optimal model to create a probability raster to predict the vulnerability to deforestation on Vancouver Island.

Objective #5: Infer the vulnerability to old growth forests from our probability raster and create a map which shows old growth forest stands at the greatest risk of deforestation based on our model.

Study Area:

The study area of choice is Vancouver Island, a large Island located on the West Coast of British Columbia. The island straddles the border between Universal Transverse Mercator (UTM) zones 9 and 10, and having a UTM based projected coordinate system is important for the meso scale of our study. Further, a UTM projection keeps distances and areas proportional which is essential as many distances will be calculated as part of our workflow. We chose to perform our analysis using UTM zone 10, as it slightly contained more of the area. Communities on the North and West sides of the Island are very dependent on the forestry sector as the primary industry, while larger cities more proximate to the mainland in the South East like Victoria and Duncan boast more diverse economies. Like the rest of Canada, the island was occupied by Indigenous peoples prior to the mid 19th century, and currently they comprise 5% of the total population (Artibise, 2010). Vancouver Island is actually the largest North American Island on the Pacific Coast boasting a total area over 32,000 kilometers squared (Britannica, 2023). The island is the emerging portion of a partially submerged continuation of the Western Cordillera. The island has many fjords running from the interior mountain range that bisect the island. Its highest peak is Golden Hinde sitting at 2200m above sea-level (Artibise, 2010). It is a temperate rainforest biome overall, but precipitation varies from as much as 3m annually on the western side, to as little as 0.8m annually in the southeastern portion of the island (Artibise, 2010). and has many of the same tree species as the rest of the province. It boasts stands of both coniferous and deciduous trees. The most common are *Pseudotsuga menziesii* (Douglas Fir) and *Thuja plicata* (Western Red Cedar). As of 2016, over 90% of the more easily accessible valley bottom ancient forests on the island have already been logged (AFA, 2016). The island and its forestry resources are governed by the provincial government.

Figure 1: Map of Vancouver Island, BC, ON, NAD 1983 UTM Zone 10N

Data:

Table 1: Data used in our analysis

	Name	Source	Scale/Resolution
Roads	Digital Road Atlas (DRA) - Master Partially-Attributed Roads	British Columbia Data Catalogue (2013)	This is a vector dataset of roads across the province of British Columbia.
Pipelines	Pipeline Segments (Permitted)	BC Oil and Gas Commission (2016)	This is a vector dataset of pipelines across the province of British Columbia.
Cities	BC Major Cities Points 1:2,000,000	British Columbia Data Catalogue (2003)	This is a vector dataset that includes all major cities in British Columbia. 1:2,000,000
Railways	National Railway Network - Provincial and territorial pre-packaged NRWN files	Government of Canada Open Data (2021)	This is a vector dataset that includes all major railways in Canada.
Digital Elevation Model (DEM)	Digital Elevation Model for British Columbia - CDED (meters)	British Columbia Data Catalogue - GeoBC Branch.	1:20,000
Vegetation Resource Inventory (training data)	Provincial Forest Inventory 2013, 2021	Provincial VRI Maps and Statistics	This dataset covers the entirety of British Columiba.
Protected Areas	ProtectedForest	Old Growth Strategic Review Spatial Database - Government of BC	100 x 100m cells that run across the entirety of British Columbia.
Old Growth Extent	Big-treed Old Growth	Old Growth Strategic Review Spatial	100 x 100m cells that run across the

		Database - Government of BC	entirety of British Columbia.

Methods:

1. Determine and preprocess the variables influencing deforestation.

This was done through our literature review and ArcGIS Pro. The analysis of peer-reviewed literature, news articles, and interviews was conducted and it was determined that our explanatory variables will be distance to infrastructure (roads, pipelines, cities and railways), slope and elevation. The design of our project makes it so that we have no need to rank the importance of these variables, quite the opposite, our purpose was to compile as many variables as possible so that our statistical model was able to parse out the influential ones. With these variables determined and required data collected, we began to preprocess the data. The data gathered was reprojected into our chosen coordinate system, NAD 1983 UTM Zone 10N and clipped to cover our study area, Vancouver Island. We used our infrastructure data to calculate the distance from roads, cities, pipelines, and railways which is needed to perform a logistic regression. We also processed the BC digital elevation model to only include cells from our study area. These cells were mosaiced together to create a final DEM raster. Finally, we derived a slope raster from our Vancouver Island DEM. The exact workflow and ArcGIS Pro tools we used to execute the preprocessing of our data is shown in Appendix 1: Mosaicing Workflow, Appendix 2: Data Preprocessing, and Appendix 3: Euclidean Distance Calculations.

2. Training the model.

The goal of our logistic regression model is to determine which of the identified variables in objective 1 will be most influential to deforestation. To do this, the model must first have an understanding of what constitutes intact forests and what constitutes deforested areas. We gathered vegetation resource index (VRI) data of B.C. from 2013 and 2021 and reduced them down to our study area. From this data, we created polygons that represented areas which were deforested between 2013 and 2021. We then used a simple random sampling method to produce 100 points that fall within the deforested areas and 100 points that fall outside of the deforested areas. We then created a table which contained all of our 200 points and their accompanying values for each of our variables. A logistic regression models the probability of a binary outcome which in our case, is deforested versus maintained forests. We therefore, used ones to represent points that fall within deforested areas and points that fall within areas of forest retained are assigned a 0. An in depth model of our workflow for this section can be found in Appendix 4: Change Detection Data Preparation.

3. Creating our logistic regression model

After preparing the variables that could be influential towards the deforestation of forests and properly training our model on the parameter values associated with either a deforested or retained response, we created multiple iterations of logistic regression models in R (Table 2). Relations between variables, to each other, and to the response variable were assessed using these 6 different iterations of our logistic regression models, as well as alongside box plots, histograms, Pearson's correlation index, scatter plots and simple linear regression. This iterative process ensured accuracy and provided a robust assessment of each factor's influence on one another and deforestation overall. The p-value output from the logistic regression model was used to determine whether a parameter's influence was statistically significant or not. The p-value is interpreted as the likelihood that random chance would have reproduced the outcome observed, so values closer to 0 are more influential, and anything less than 5% is considered to be independent of stochasticity and statistically significant. After the models were created, the Akaike information criterion (AIC) was utilized to evaluate the efficacy of our models to fit the data efficiently and accurately (Sakamoto, 1986). The AIC gives models with the least complexity that still accurately represents the phenomena the lowest score, meaning it is the best suited. It essentially tests for less useful parameters that do not need to be included for the model to be accurate. The lowest scoring model (M5) was chosen to be the most predictive and was used moving forwards. This model comparison framework addressed our third objective of determining which variables had the largest potential impact on deforestation.

Table 2: Logistic regression models and associated parameters included

Model	Parameters	AIC Score
M1	Cities, Roads, Pipelines, Railways, Elevation, Slope	274.5982
M2	Roads, Elevation, Slope	282.6686
M3	Cities, Pipelines, Railways	272.9550
M4	Cities, Pipelines	271.6305
M5	Cities, Railways	271.0997
M6	Railways, Pipelines	279.5905

4. Manipulation, review and production from regression model output

The output needed for our next objective from our most optimal logistic regression model is the beta coefficients for each of the explanatory variables (Bavaghar, 2015). Although, the beta coefficients in their current state were not representative of the probability we are looking to model. The calculations needed to be made in log odds as probability is not an additive metric, so conversion back is required for interpretation. The reason why we are looking to view our vulnerability map in terms of probabilities is that probability is a statistic that the general public is familiar with. Currently, our response variable is the natural logarithm of the odds. This is a result of the transformations we needed to perform on our model to make it fit our discrete probability classes of 1 being deforested and 0 being retained forest. Converting this output back to probability and then implementing our optimal model to produce a probability raster is done through the following steps.

1. Taking the base log of the equation to revert our ‘linearized’ odds function back to just odds.

$$odds = e^{(b_0 + b_1 X)}$$

2. Rearrange the probability/odds equation to isolate probability.

$$odds = \frac{P}{1 - P}$$

$$P = \frac{odds}{1 + odds}$$

3. Substitute the results from step 1 into the rearranged equation from step 2 in the place of the odds.

$$P = \frac{e^{(b_0 + b_1 X)}}{1 + e^{(b_0 + b_1 X)}}$$

4. Input coefficients from model output as the beta-coefficients, and rasters with its associated beta coefficient within the equation created in step 3 into the raster calculator.
5. The result is a raster which displays the probability of deforestation for each raster cell within our study area.

5. Production of the final product.

Using the results from our fourth objective, we derived the vulnerability of old growth forests on the island. This was done by extracting the values from our probability raster to a new raster via a mask of old growth forest stands. The next step in answering our research question was to perform a reclassification of the masked vulnerability to extract only the cells with values greater than 0.85. This gave us a dataset that was used in creating a map which highlights the old growth forest stands which were most at risk for deforestation based on our model.

Results and Discussion:

Model Assessment

M1 contained all of our six variables and was used to determine the variables which were statistically significant in influencing the instance of whether an area had retained its stands or had its stands deforested (Table 2). The p-value was utilized to determine whether a parameter was statistically significant or not. There was not a statistically significant correlation between roads, elevation or slope, but roads (22.5%) were a bit stronger correlated than the later two (74.87% and 67.95%). Cities (Figure 2) held the strongest correlation, at about 0.1%. Pipelines and railways still held a correlation with our response variable as well though, with a 5.78% and 6.74% chance that their correlation could have been reproduced by random chance.

Table 3: Model M1

	Estimate	**Std. Error**	**Z - Value**	**P - Value**
Intercept	-1.264e+00	4.477e-01	-2.824	0.00475
Cities	4.851e-05	1.484e-05	3.268	0.00108
Roads	3.148e-05	2.596e-05	1.213	0.22529
Pipelines	-2.473e-05	1.352e-05	-1.829	0.06735
Railways	2.106e-05	1.110e-05	1.897	0.05778
Elevation	-1.550e-04	4.839e-04	-0.320	0.74864

Slope	-5.372e-03	1.300e-02	-0.413	0.67950

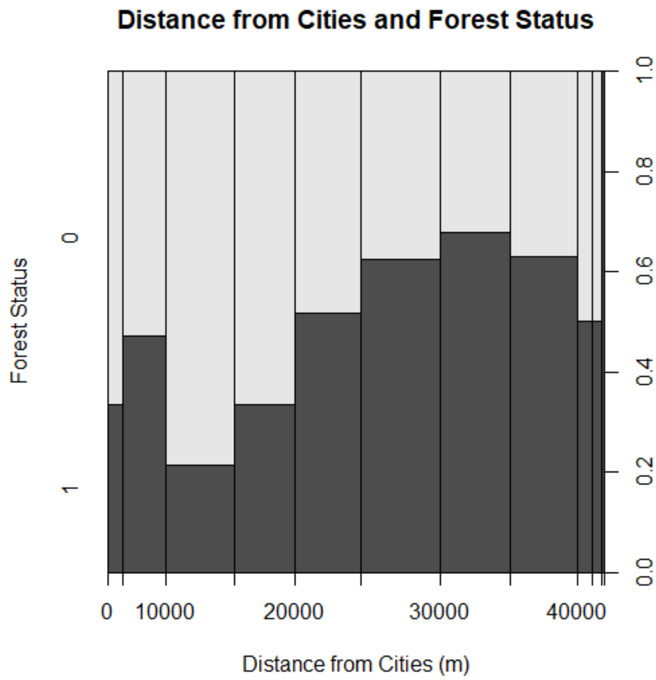


Figure 2: This binary histogram shows the distance from cities and probability of a forest stand being deforested. On the y-axis, 1 refers to areas that were deforested and 0 refers to areas that have not been deforested. The width of each bar represents the number of our randomly sampled points which fall in each distance category. For example, this graph shows that a small amount of points fall within 0 and 5000 meters of cities. This histogram shows that deforestation increases with distance from cities.

Table 4: Model M2

	Estimate	**Std. Error**	**Z - Value**	**P - Value**
Intercept	-1.088e-01	2.690e-01	-0.405	0.686
Roads	3.271e-05	2.192e-05	1.492	0.136
Elevation	2.160e-04	4.535e-04	0.476	0.634
Slope	-7.557e-03	1.258e-02	-0.601	0.548

Our next model (M2) (Table 3) aimed to verify the lack of correlation between the 3 uncorrelated parameters The resulting analysis showed that indeed none of the three variables have a p-value less than 10%, and cannot statistically confirm the existence of any correlation. There are a few reasons why we hypothesize these variables held little correlation:

Roads

Before we conducted our analysis, we predicted that resource roads would be a major factor in deforestation on Vancouver Island. Therefore, we were surprised to see that our model proved the opposite of what we expected. We believe that the main reason for the poor correlation between resource roads and deforestation is due to the high density of roads throughout the entire island. A large percentage of our random points fell within a distance of 5 km of a resources road, which can be seen in figure 3. As a consequence of this, there was no correlation to be observed since almost every deforested area was in close proximity to a road.

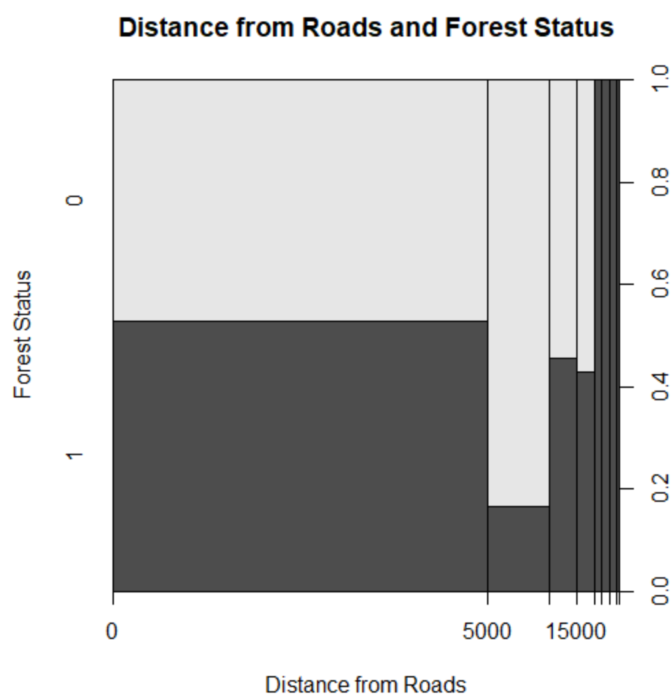


Figure 3: This binary histogram shows the distance from roads and probability of a forest stand being deforested. On the y-axis, 1 refers to areas that were deforested and 0 refers to areas that have not been deforested. The width of each bar represents the number of our randomly sampled points which fall in each distance category. This histogram shows that a very large portion of our random points fall within 5000 meters of a road. This histogram does not show a clear relationship between distance to roads and deforestation.

Elevation

There are two main reasons we theorize that deforestation and elevation do not correlate. The first, is the fact that as elevation increases tree size decreases (Malizia et al., 2020). Therefore, there is less economic value in harvesting forests at higher elevations. Furthermore, the average elevation of Vancouver Island is relatively low when compared to mainland BC. 75% of BC has an elevation equal to or greater than 1000 (GISGeography, 2023). As seen in figure 4, the average elevation of Vancouver Island is much lower.

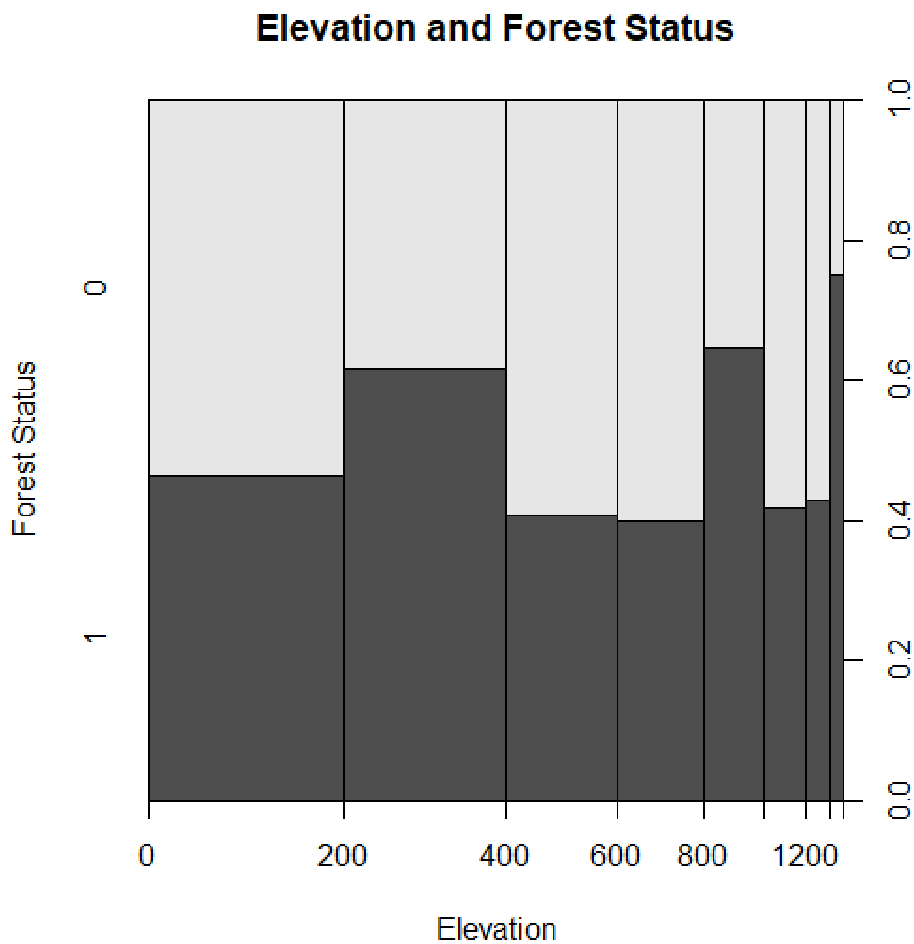


Figure 4: This binary histogram shows elevation in meters and probability of a forest stand being deforested. On the y-axis, 1 refers to areas that were deforested and 0 refers to areas that have not been deforested. The width of each bar represents the number of our randomly sampled points which fall in each elevation category. This histogram shows that there is not a clear relationship between elevation and deforestation.

Slope

We believe that no correlation was seen between deforestation and slope because of the distribution of slope values on Vancouver Island. A BC law from 2003 states that no forestry equipment can be operated on a slope which exceeds 50%. This means that no slope greater than 45 degrees can be correlated to deforestation in BC. Figure 5 shows the distribution of slope values in our random sample and that the majority of these values are less than 45 degrees.

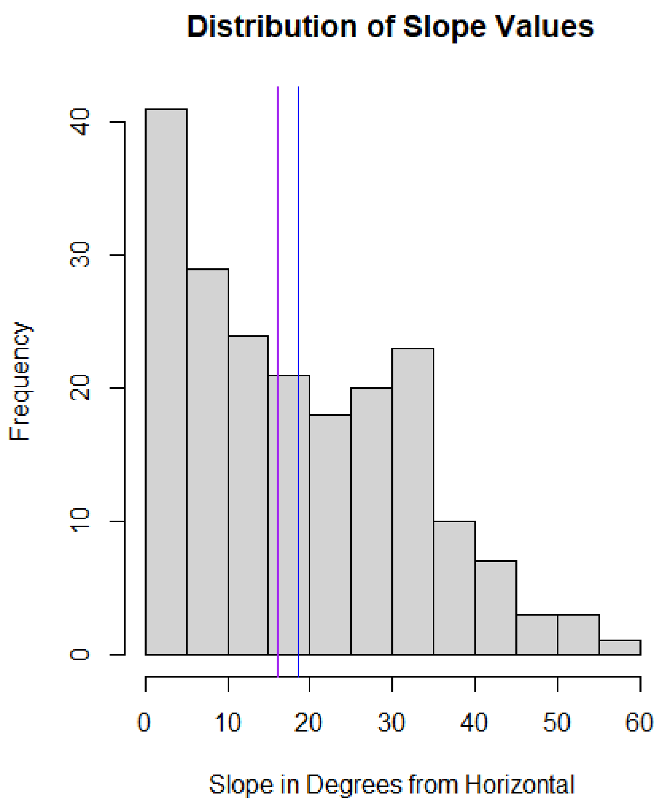


Figure 5: This histogram shows the distribution of slope values in degrees for our randomly sampled points. The chart displays that the majority of our points have a slope value less than 40 degrees.

Modeling the variables that were statistically significant (M3) and assessing parameter interrelatedness (Table 2) showed which were truly the most significant, and gave insight as to why. The results of a Pearson’s correlation matrix showed that railways and pipelines are extremely correlated to each other on the island (Table 4). This relationship can be seen further in figure 6. Despite pipelines being more correlated to cities than railways were according to table 2, producing models M3 (Table 5), M4 (Table 6), M5 (Table 7) and M6 (Table 8) confirmed that railways had greater statistical significance towards the response variable of forest retention. Although utilizing more variables creates a stronger model in theory (M3), utilizing variables that do not add any new information is not beneficial (Sakamoto, 1986), and thus the AIC determined using the model with just two parameters (M5) would be the best course of action.

Table 5: Pearson correlation matrix of cities, railways and pipelines. Shows the close relationship between railways and pipelines.

	Cities	**Railways**	**Pipelines**
Cities	1.00	0.20	0.23
Railways	0.20	1.00	0.98
Pipelines	0.23	0.98	1.00

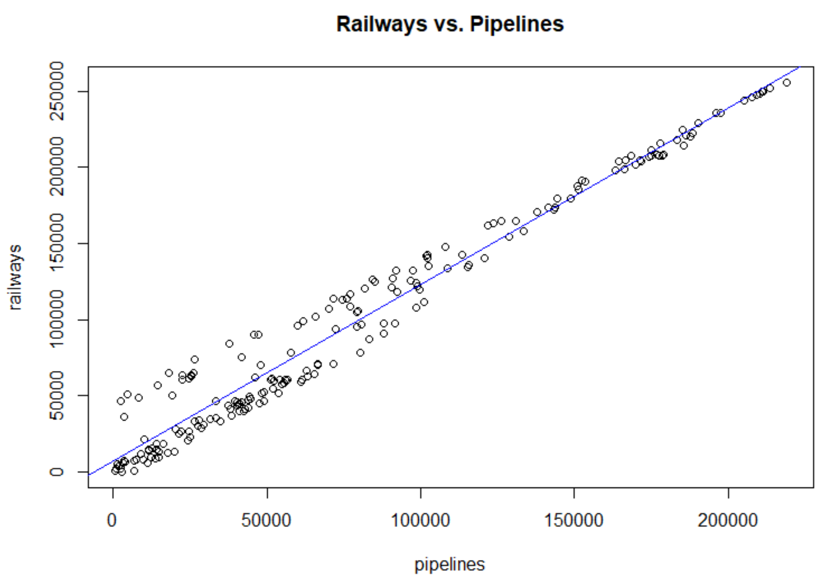


Figure 6: The distance of our 200 points from railways vs. pipelines. This shows that the difference in distance to pipelines and distance to railways is very small.

Table 6: Model M3

	Estimate	Std. Error	Z - Value	P - Value
Intercept	-1.367e+00	4.080e-01	-3.350	0.000809
Cities	4.639e-05	1.436e-05	3.229	0.001242
Pipelines	-2.030e-05	1.287e-05	-1.577	0.114737
Railways	1.873e-05	1.081e-05	1.734	0.082990

Table 7: Model M4

	Estimate	Std. Error	Z - Value	P - Value
Intercept	-1.167e+00	3.867e-01	-3.019	0.00254
Cities	4.308e-05	1.420e-05	3.035	0.00240
Pipelines	1.588e-06	2.426e-06	0.655	0.51262

Table 8: Model M5

	Estimate	Std. Error	Z - Value	P - Value
Intercept	-1.226e+00	3.949e-01	-3.105	0.00190
Cities	4.255e-05	1.412e-05	3.014	0.00258
Railways	1.998e-06	2.043e-06	0.978	0.32832

Table 9: Model M6

	Estimate	Std. Error	Z - Value	P - Value
Intercept	-3.427e-01	2.424e-01	-1.414	0.157
Railways	1.372e-05	1.034e-05	1.327	0.185
Pipelines	-1.280e-05	1.224e-05	-1.046	0.296

Probability Raster and Old Growth Vulnerability

Armed with the knowledge that railways and cities create the strongest model (M5), a prediction raster was calculated from the correlative coefficients for each respective variable. Our prediction raster (Figure 7), shows which areas of Vancouver Island are the least likely to be deforested based on their relative proximity to cities and pipelines. The areas least likely to be deforested are in the North-West, and generally traced the entire West Coast of the Island.

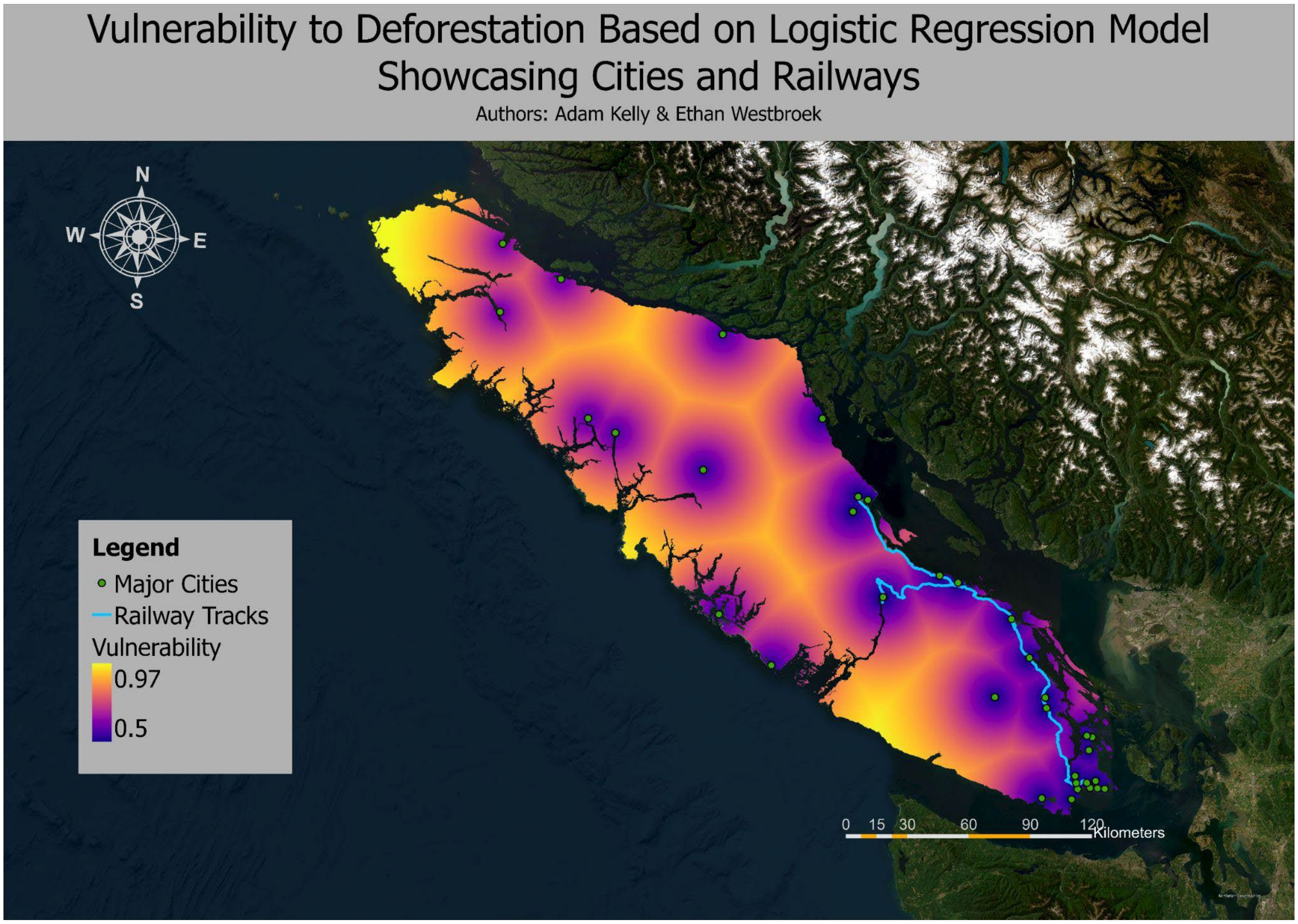


Figure 7: Probability raster showing the vulnerability to deforestation produced using a logistic regression model considering proximity to cities and railways.

This heatmap (Figure 7) of a raster cell's probability of being either deforested or having its stands retained is representative for any cell on the land, and does not include any spatial filtering of where the old growth stands actually are located. The subsequent two maps showcase two stages of filtering, first towards only cells classified as old growth forest based on the B.C. governments strategic review map 2 titled *Big-Treed Old Growth* (Gorley & merkel, 2020) (Figure 8), and second an exclusion of any stands that are considered protected based on the additional *ProtectedForest* data from the same directory (Figure 9).

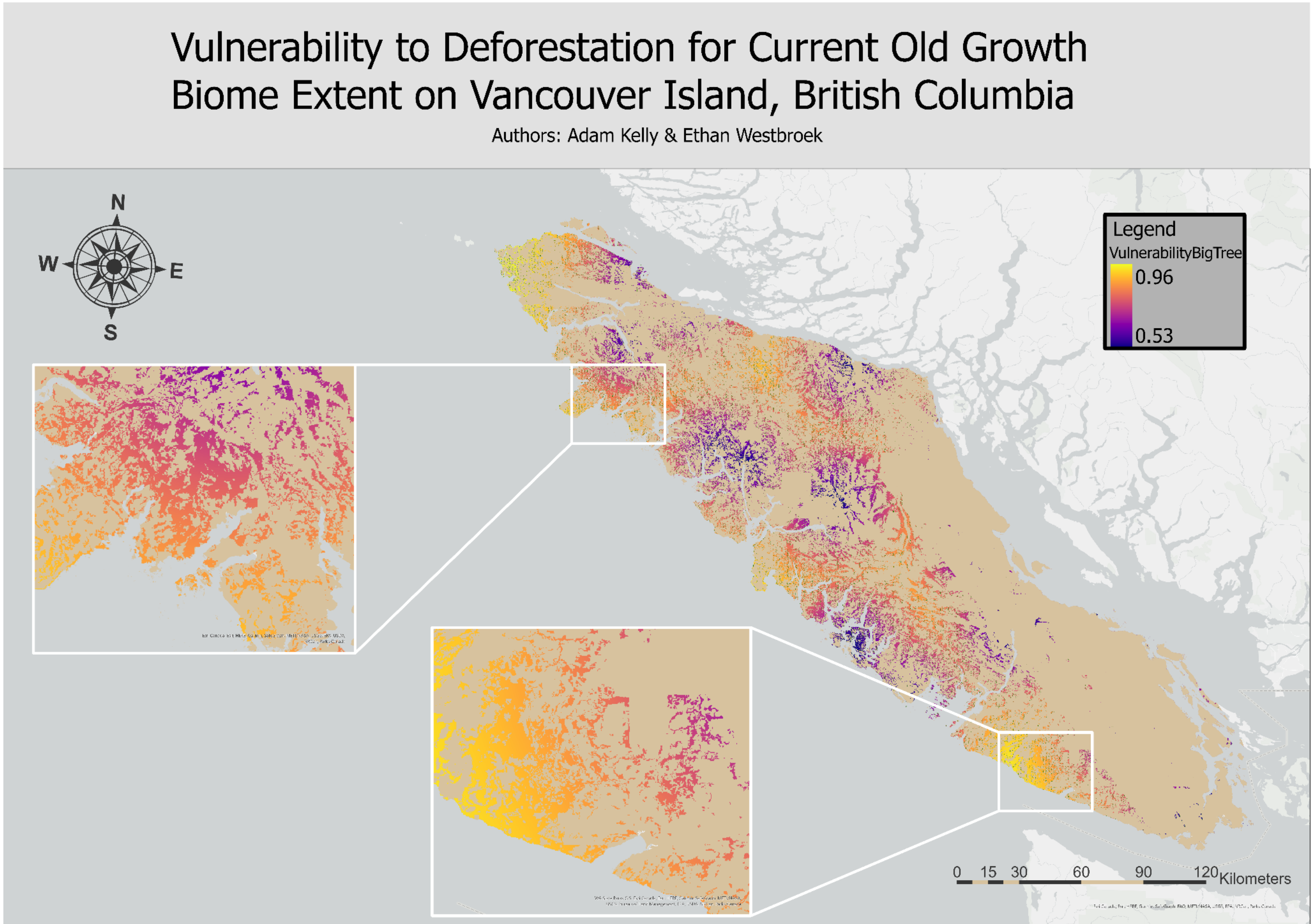


Figure 8: Vulnerability to deforestation for current old growth biome extent on Vancouver Island, British Columbia.

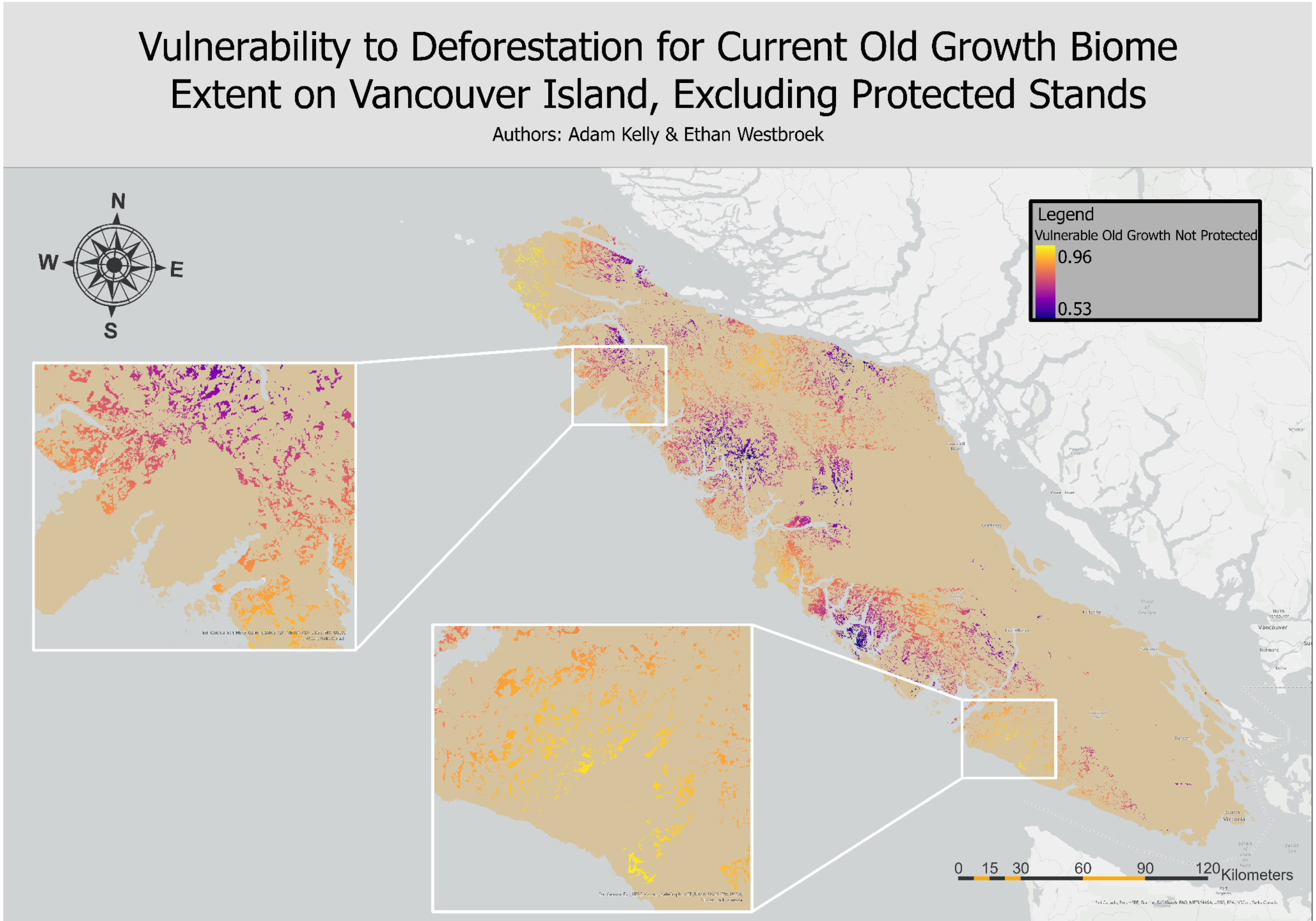


Figure 9: Vulnerability to deforestation for current old growth biome extent on Vancouver Island, British Columbia, excluding protected old growth stands.

These maps effectively showcase the reclassifications of our vulnerability model to only portray the vulnerability for the raster cells that qualify as both old growth and currently not under the protection of any spatial conservation method.

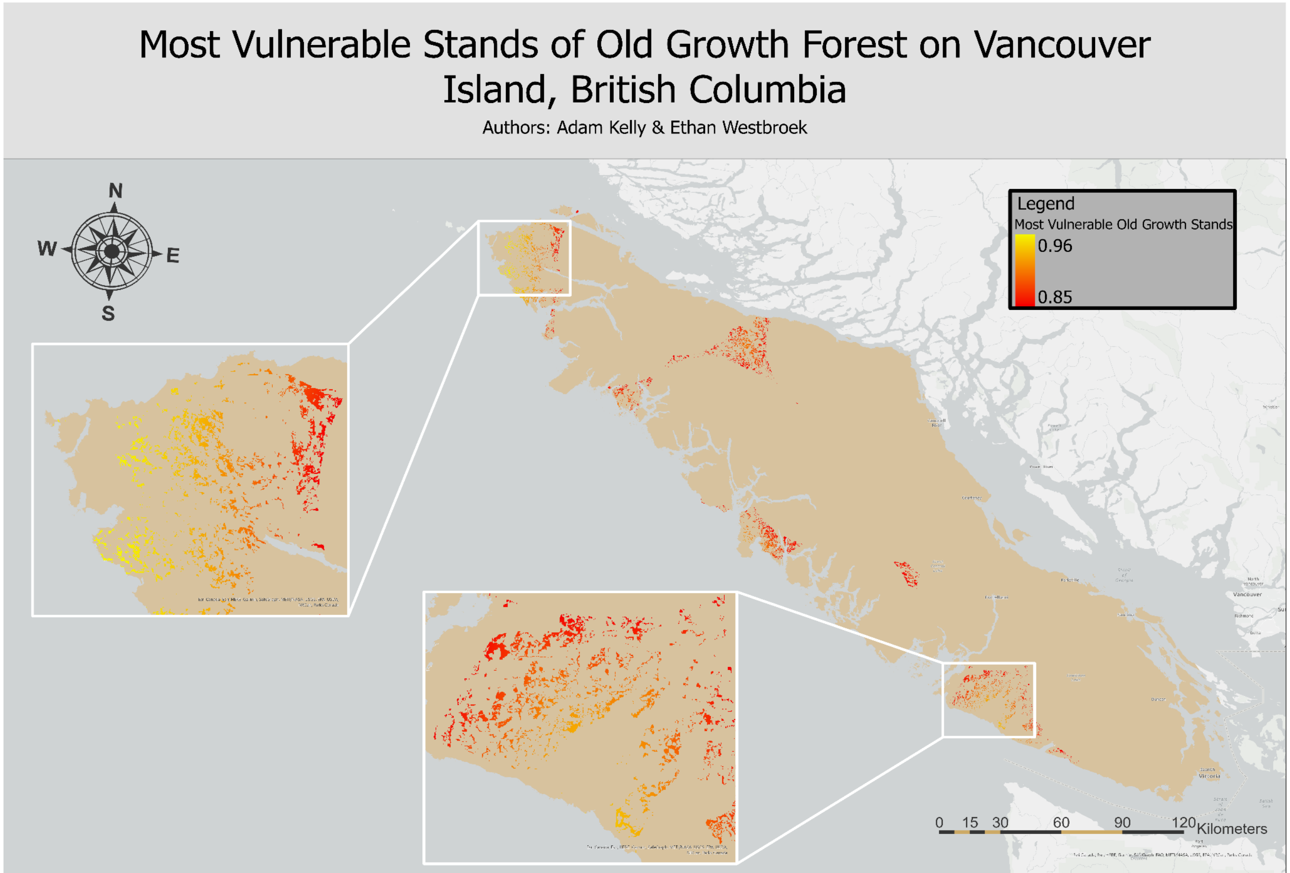


Figure 10: The most vulnerable old growth stands to deforestation on Vancouver Island.

Figure 10 represents the 46,076 hectares of old growth that is considered extremely vulnerable. Spatially, there are a few hotspots on the island that have clumps of these extremely vulnerable stands. These are found mainly along the Western coast of the island. An interesting result is that the two areas that have the greatest extent of vulnerable forests, had extensive protected areas surrounding them. This is indicative that the governance methods used to determine the various protected areas included in the *PortectedForest* layer have been able to recognize both the value and vulnerability of these stands already. The expansion of these protected areas to include periphery stands based on our analysis, would result in the retention of many more hectares of irreplaceable old growth.

Limitations and Potential Improvements:

The use of a simple random sample did give us an accurate distribution of our variables. However, as much of Vancouver Island has been logged already (AFA, 2016), the areas being deforested are at the fringes, such as high elevations, far from infrastructure, and steep graded slopes. Therefore, we believe it would have been beneficial to change our primary sampling method from random sampling to stratified random sampling. This switch would increase the robustness by ensuring these extreme ranges with very few data points are included due to the stratification of our sampling. Secondly, performing a secondary analysis that uses a sampling method which favors sampling these extreme values much more heavily such as, a disproportionate stratified random sample would give insight into these extreme instances. We believe a switch to these two sampling methods would solve the issue of the more extreme values being equally represented, when we want them represented at a greater rate, as this is where the deforestation is likely occurring (AFA, 2016).

Another limitation of our model would be related to the selection of our parameters. We were able to pre-process and prepare 6 different factors that we hypothesized were influential to deforestation. As much as these parameters were important, there were many potential additional parameters not included that might have added to the strength of our model.

The first set would be more spatial data about logging operations. A set of distance accumulation rasters based on logging mills, storage facilities, and headquarter locations for logging operators could provide more context as to how easy a firm might access a stand. Additionally, a spatial analysis of all current logging licenses, and their correlation to their respective companies infrastructure could have been prudent and revealed again answers to questions of a firm's accessibility to the stands it has rights to harvest.

The next set of variables that could have been included in our analysis would be more ecological variables. In our analysis, we are treating all old growth stands as the same. In reality, there are a variety of characteristics about old growth forests that differentiate one stand from another, and thus logging companies will have differential preferences for which they would prefer to extract from. These characteristics might be total stand/height, which in combination with stand density is a good indicator of the total quantity of timber in any given area.

The last parameter that regrettably was not included was spatial data regarding Indigenous sovereignty and land rights. These areas vary in legal recognition, enforcement capacity and formality, but irregardless of their interpretation by the colonial government or logging corporations, they have an impact on how land is managed. The inclusion of these spatial extents could have added additional complexity and robustness to our model.

Conclusion:

The vulnerability of old growth forests on Vancouver Island, B.C. is obviously an important issue. This analysis aimed to determine the main variables which influence deforestation on the island. Through logistic regression modeling, we determined that Model 5 which included railways and cities was the most accurate at modeling deforestation on Vancouver Island. After creating a probability raster and comparing it to old growth extent, we found that a surprisingly large amount of old growth stands that the model determined to be most vulnerable, were already classified as protected forests by the government of British Columbia. This is quite a compelling result, however, there is still a large amount of unprotected old growth forests which lie in vulnerable areas. We found that 46,067 hectares of unprotected old growth stands are highly vulnerable to deforestation. We believe that these results would be helpful in determining new priority deferral areas. Finally, it would be beneficial to further this analysis by including more explanatory variables and to further examine the influence that more extreme values may have on deforestation on Vancouver Island.

References:

AFA. (2016). *Myths & Facts*. Ancient Forest Alliance. Retrieved March 10, 2023, from https://ancientforestalliance.org/ancient-forests/myths-facts/

Artibise, A. F. J. (2010). *Vancouver Island*. The Canadian Encyclopedia. Retrieved March 10, 2023, from https://www.thecanadianencyclopedia.ca/en/article/vancouver-island

BC Oil and Gas Commission (2016) Pipeline Segments (Permitted) [Vector]. Scale undetermined; generated by BC Oil and Gas Commission. https://data-bcogc.opendata.arcgis.com/datasets/359b7e14fafa4abc84ff873bc55015fb_0/explore?location=50.016254%2C-125.264501%2C9.83 (Feb 16th, 2023).

British Columbia Data Catalogue (2014) - GeoBC Branch. Digital Elevation Model for British Columbia CDED. [Raster - 1:50,000]. Geo BC Branch. https://catalogue.data.gov.bc.ca/dataset/7b4fef7e-7cae-4379-97b8-62b03e9ac83d

British Columbia Data Catalogue (2003) BC Major Cities Points 1:2,000,000 [Vector]. Scale undetermined; generated by GeoBC Branch.https://catalogue.data.gov.bc.ca/dataset/bc-major-cities-points-1-2-000-000-digital-baseline-mapping (Feb 16th, 2023).

British Columbia Data Catalogue (2013) Digital RoadAtlas (DRA) - Master Partially-Attributed Roads [Vector]. Scale undetermined; generated by GeoBC Branch. https://catalogue.data.gov.bc.ca/dataset/digital-road-atlas-dra-master-partially-attributed-roads (Feb 16th, 2023).

Britannica. (2023, January 27). *Vancouver island*. Encyclopædia Britannica. Retrieved March 10, 2023, from https://www.britannica.com/place/Vancouver-Island

British Columbia Regulations. (2010). *Occupational Health and Safety Regulation: Forestry Operations*. Occupational Health and Safety Regulation. Retrieved March 31, 2023, from https://www.bclaws.gov.bc.ca/civix/document/id/loo81/loo81/296_97_21

Clason, A. J., Lindgren, P. M. F., & Sullivan, T. P. (2008). Comparison of potential non-timber forest products in intensively managed young stands and mature/old growth forests in south-central British Columbia. Forest Ecology and Management, 256(11), 1897–1909. https://doi.org/10.1016/j.foreco.2008.07.013

Fredeen, A. L., Bois, C. H., Janzen, D. T., & Sanborn, P. T. (2005). Comparison of coniferous forest carbon stocks between old growth and young second-growth forests on two soil types in central British Columbia, Canada. Canadian Journal of Forest Research, 35(6), 1411–1421. https://doi.org/10.1139/x05-074

GISGeography. (2023, February 16). *British Columbia Map*. GIS Geography. Retrieved March 31, 2023, from https://gisgeography.com/british-columbia-map/

Gorley, A., & Merkel, G. (2020, April 30). Old-growth Strategic Review. A New Future for Old Forests. Retrieved November 17, 2021, from https://www2.gov.bc.ca/assets/gov/farming-natural-resources-and-industry/forestry/stewardship/old-growth-forests/strategic-review-20200430.pdf.

Government of Canada Open Data (2021) National Railway Network - Provincial and territorial pre-packaged NRWN files [Vector]. Scale undetermined; generated by Government of Canada. https://open.canada.ca/data/en/dataset/ac26807e-a1e8-49fa-87bf-451175a859b8/resource/e9aabb53-6145-4a98-b0f5-c10270e9aa58 (Feb 16th, 2023).

Guthrie, R. H. (2002). The effects of logging on frequency and distribution of landslides in three watersheds on Vancouver Island, British Columbia. *Geomorphology*, *43*(3-4), 273-292.

Malizia, A., Blundo, C., Carilla, J., Osinaga Acosta, O., Cuesta, F., Duque, A., Aguirre, N., Aguirre, Z., Ataroff, M., Baez, S., Calderón-Loor, M., Cayola, L., Cayuela, L., Ceballos, S., Cedillo, H., Farfán Ríos, W., Feeley, K. J., Fuentes, A. F., Gámez Álvarez, L. E., … Young, K. R. (2020). Elevation and latitude drives structure and tree species composition in Andean forests: Results from a large-scale plot network. *PLOS ONE*, *15*(4). https://doi.org/10.1371/journal.pone.0231553

Old Growth Strategic Review Spatial Database (2021) - Government of BC. Big-treed Old Growth [Raster]. Scale undetermined; generated by the Government of BC. https://www.for.gov.bc.ca/ftp/HTS/external/!publish/OldGrowthStrategicReview/Recommendation6/TechnicalAdvisoryPanel/Spatial_Data/OGSR_TAP_Raster_Data/Additional_Data/ (Feb 16th, 2023).

Old Growth Strategic Review Spatial Database (2021) - Government of BC ProtectedForest [Raster]. Scale undetermined; generated by the Government of BC.

https://www.for.gov.bc.ca/ftp/HTS/external/!publish/OldGrowthStrategicReview/Recommendation6/TechnicalAdvisoryPanel/Spatial_Data/OGSR_TAP_Raster_Data/Additional_Data/ (Feb 16th, 2023).

Raitio, K., & Saarikoski, H. (2012). Governing old growth forests: The interdependence of actors in Great Bear Rainforest in British Columbia. Society & Natural Resources, 25(9), 900–914. https://doi.org/10.1080/08941920.2011.642462

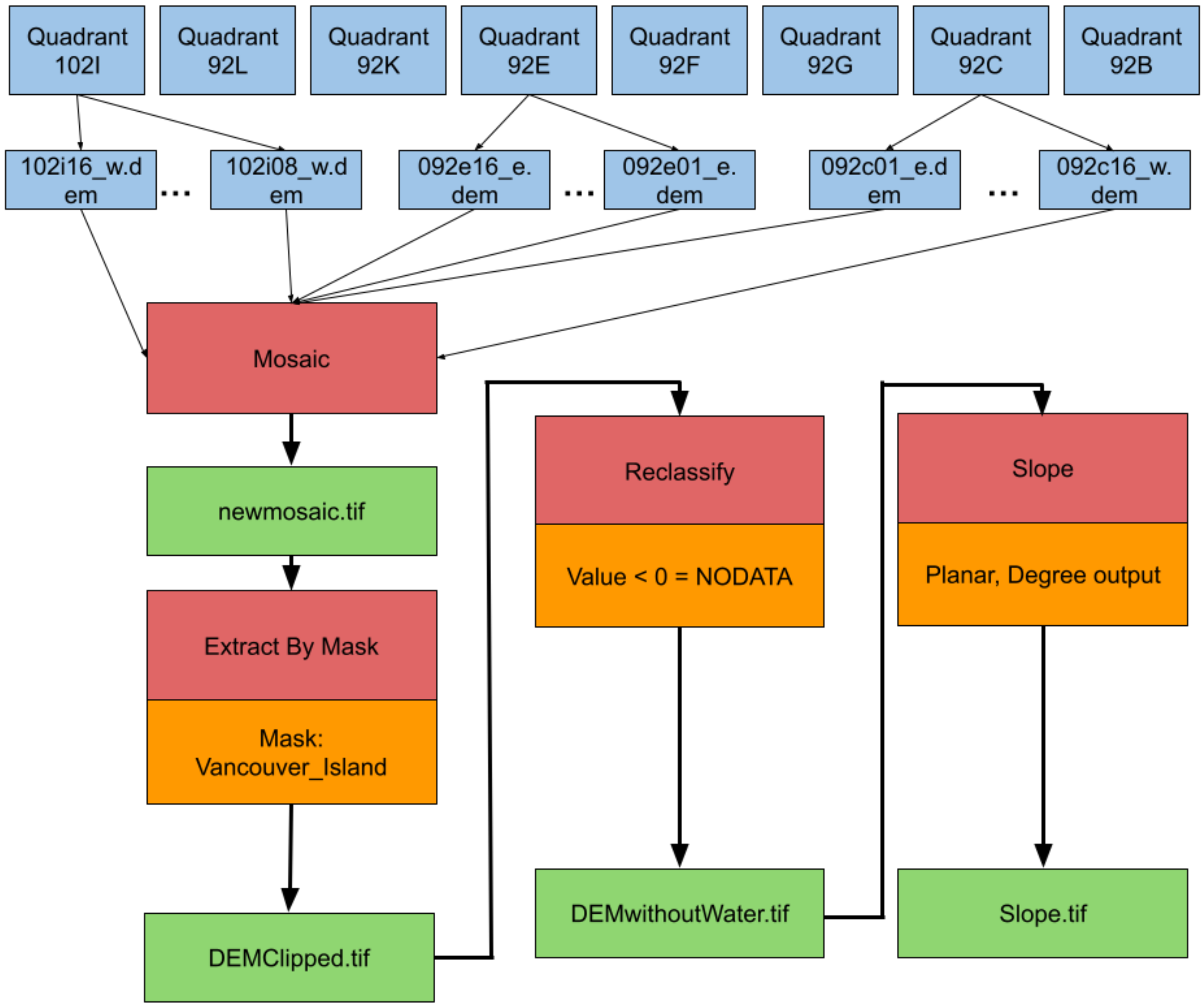
Robertson, A., Hansen, T., Robertson, R., McDowell, L., & Walton, E. (2022). Stand Earth - Tall Talk. Retrieved 2023, from https://old.stand.earth/sites/stand/files/tall_talk_final.pdf

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, *81*(10.5555), 26853.

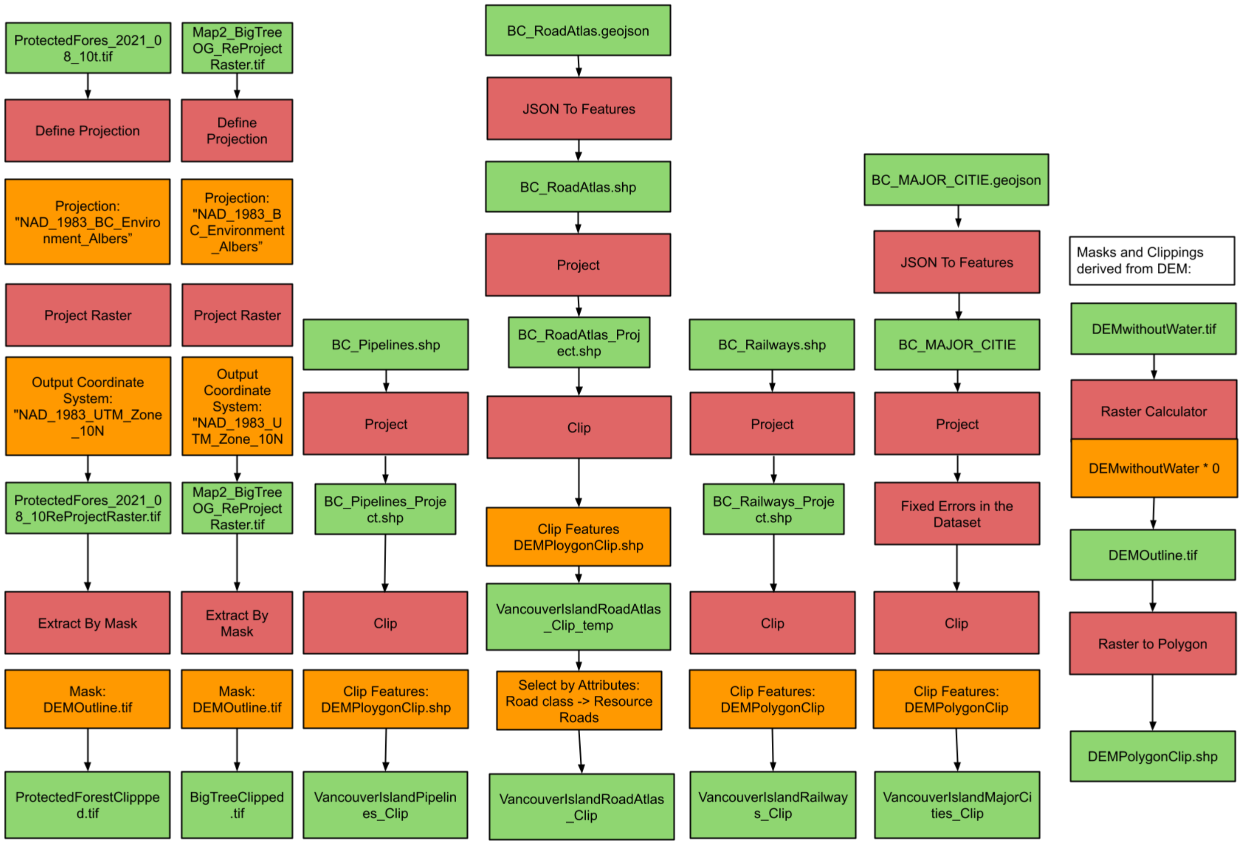
VIM. (2015). *Vancouver Island mushrooms - tree identification*. Vancouver Island Mushrooms - Tree Identification. Retrieved March 10, 2023, from https://www.westcoastforager.com/resources/identify-bc-tree-types#h.l8ehmauamd92

Appendices:

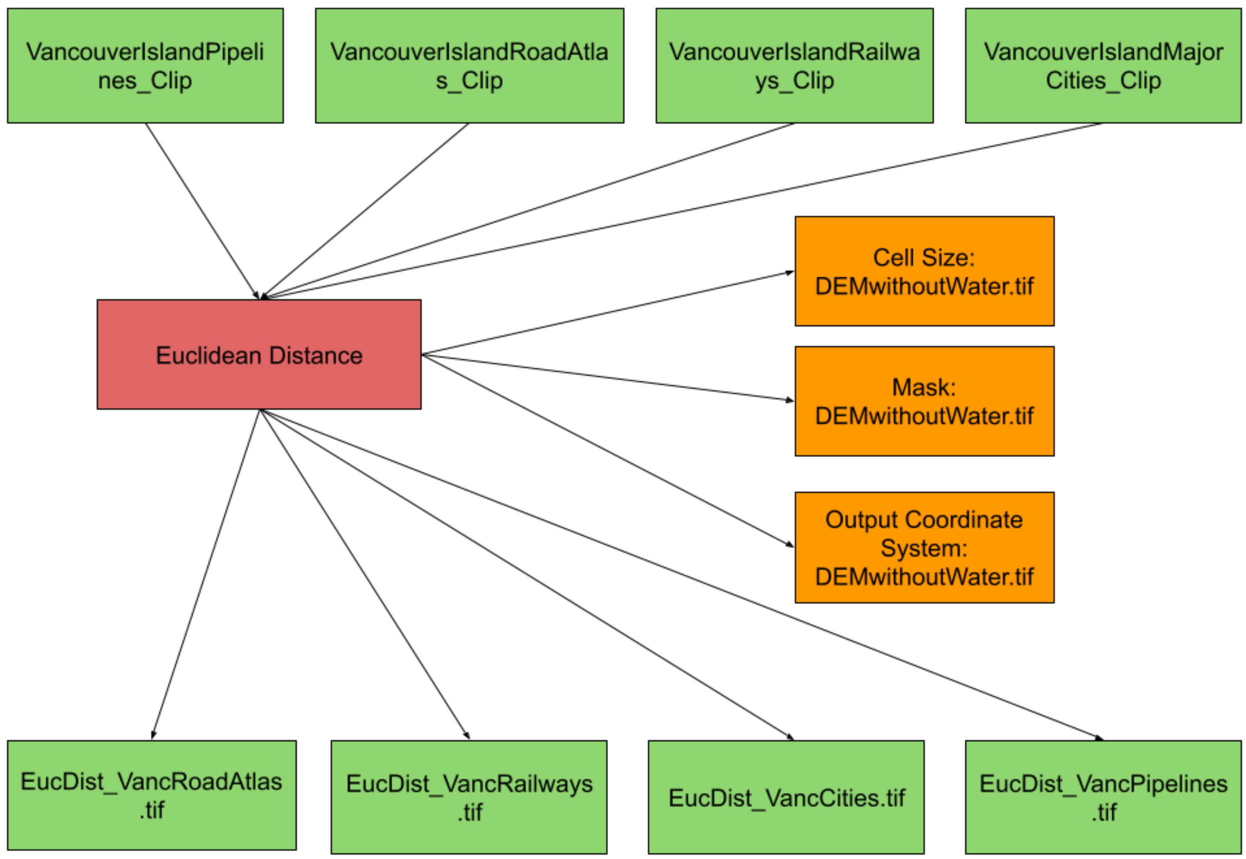
Appendix 1: DEM and Slope Preparation



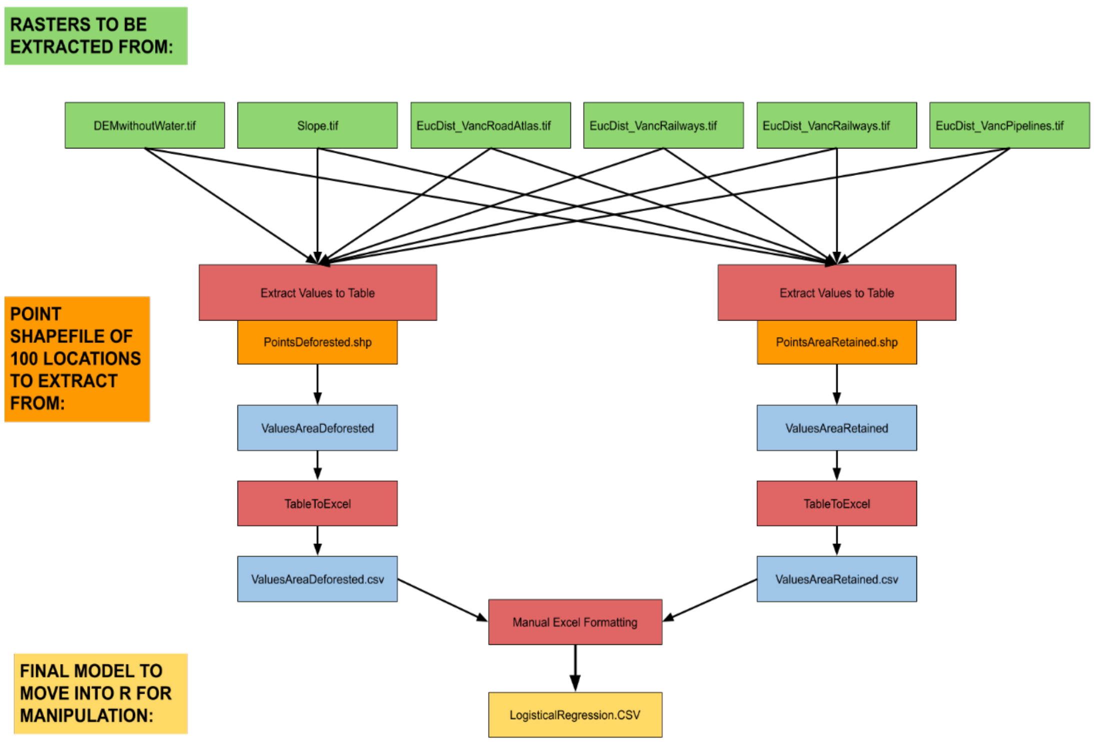
Appendix 2: Data Preprocessing



Appendix 3: Euclidean Distance Calculations



Appendix 4: Change Detection Data Preparation



Appendix 5: R Code

```r
#### Logistic Regression for anlaysis of deforestation on Vancouver Island ####

setwd("C:/Users/akelly15/Downloads/GEOG4880_0323/GEOG4480/")
logreg <- na.omit(read.csv("LogisticalRegressionCSV.csv"))
names(logreg)
logreg$Status <- factor(logreg$Status)

## look at some relationships
plot(Status ~ cities, ylab = "Forest Status", xlab = "Distance from Cities (m)", main = "Distance from Cities and Forest Status", data = logreg)
plot(Status ~ roads, data = logreg, ylab = "Forest Status", xlab = "Distance from Roads", main = "Distance from Roads and Forest Status")
plot(Status ~ elevation, data = logreg, ylab = "Forest Status", xlab = "Elevation", main = "Elevation and Forest Status")

## histograms
hist(logreg$slope, main = "Distribution of Slope Values", xlab = "Slope in Degrees from Horizontal")
abline(v = mean(logreg$slope), col = 'blue')
abline(v = median(logreg$slope), col = 'purple')
abline(v = sd(logreg$slope), col = 'grey')
mean(logreg$slope)
median(logreg$slope)
sd(logreg$slope)

### fit a logistical regression model
# model 1: Forest status as a function of all parameters
logit <- binomial(link = "logit")

model1 <- glm(Status ~ cities + roads + pipelines + railways + elevation + slope, family = logit, data = logreg)
summary(model1)

# model 2: Forest status as a function of seemingly uncorrelated parameters
model2<- glm(Status ~ roads + elevation + slope, family = logit, data = logreg)
summary(model2)

# check correlation matrix between variables showing correlation
cor(cbind(logreg$cities, logreg$railways, logreg$pipelines))

plot(railways ~ pipelines, main = "Railways vs. Pipelines", data = logreg)
abline(lm(logreg$railways ~ logreg$pipelines, data = mtcars), col = "blue")

# model 3: forest status as a function of cities, pipelines and railways
model3<- glm(Status ~ cities + pipelines + railways, family = logit, data = logreg)
summary(model3)

# model 4: forest status as a function of cities, pipelines
model4<- glm(Status ~ cities + pipelines, family = logit, data = logreg)
summary(model4)

# model 5: forest status as a function of cities, railways
model5<- glm(Status ~ cities + railways, family = logit, data = logreg)
summary(model5)

# model 6: forest status as a function railways, pipelines
model6<- glm(Status ~ railways + pipelines, family = logit, data = logreg)
summary(model6)

# model comparison using the Akaike Information Criterion (AIC)
AIC(model1, model2, model3, model4, model5, model6)
coef(model5)
```

13:1 Logistic Regression for anlaysis of deforestation on Vancouver Island R Script

Appendix 6: Model Output Processing