

# A Landscape Level Analysis of Yellow-Cedar Decline in Coastal British Columbia

by

Claire Evelyn Wooton

BSc. (Hons), University of Stirling, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Geography)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August, 2010

© Claire Evelyn Wooton 2010

# Abstract

Yellow-cedar (*Chamaecyparis nootkatensis*) is currently undergoing a dramatic decline in western North America, with concentrated areas of decline located in southeast Alaska and coastal British Columbia. Recent research suggests that a shift in climate is responsible for the decline and a working hypothesis concerning the role of climate and site specific factors has been proposed. The main objective of this research was to contribute to the understanding of the yellow-cedar decline phenomenon by examining the spatial pattern of the decline and assessing the relations with topographic variables in coastal British Columbia.

The research questions were addressed through a combination of remote sensing and Geographic Information System (GIS) techniques. Sample points were distributed across the landscape according to a stratified sampling scheme and the presence/absence of decline at each point was determined using a forest cover dataset and aerial photograph interpretation. Spatial patterns of topographic factors (e.g. elevation, slope, aspect) were derived from a 25 m digital elevation model of the province. To assess the strength of relations between the distribution of decline and the various environmental predictors, logistic regression and decision-tree models were applied. The lasso technique was used to select a significant set of coefficients and the selection was then validated through bootstrap analysis. Model results indicated that low elevation sites close to the coast, which are more exposed and have more variation in elevation, are more likely to show evidence of decline. The logistic model fit the data well (Nagelkerke  $R^2 = 0.846$ , Hosmer-Le Cessie omnibus test failed to find any evidence of lack of fit) and had high predictive accuracy (AUC = 0.98).

The topographic variables identified by the model influence degree of soil saturation, temperatures and snowpack presence in a forest stand, supporting the proposed associations in the current decline hypothesis. The analysis also highlighted the utility of the lasso logistic model for selecting significant variables and mapping high risk areas for decline. Knowledge of the determinants of the spatial pattern of decline will improve predictability and provide critical information for the conservation and management of yellow-cedar.

# Table of Contents

Abstract.....	ii
Table of Contents.....	iii
List of Tables.....	v
List of Figures.....	vi
List of Abbreviations.....	viii
Acknowledgements.....	x
Chapter 1 - Introduction.....	1
1.1 Motivation for the Study.....	1
1.2 Literature Review.....	4
1.2.1 Biophysical Theory of Forest Decline.....	4
1.2.2 The Case of Yellow-Cedar Decline.....	6
1.3 Research Question.....	12
1.4 Thesis Structure.....	12
Chapter 2 - Study Area and Methods.....	13
2.1 Study Area.....	13
2.2 Data Collection and Processing.....	18
2.2.1 Yellow-Cedar Decline Dataset.....	18
2.2.2. Topographic Variables.....	21
2.3 Statistical Analysis and Modelling.....	30
2.3.1 Exploratory Analysis.....	30
2.3.2 Logistic Regression Analysis.....	31
2.3.3 Analysis of Residuals.....	36
2.3.4 Classification and Regression Trees.....	36
Chapter 3 - Results.....	38
3.1 Yellow-Cedar Distribution.....	38
3.2 Exploratory Analysis of Declining and Healthy Yellow-Cedar Stands.....	41
3.3 Logistic Regression.....	43
3.3.1 Lasso Model.....	45
3.3.2 Model Predictive Success.....	50

3.3.3. Predictive Mapping.....	53
3.3.4 Analysis of Residuals .....	56
3.3.5 Classification and Regression Tree .....	57
Chapter 4 - Discussion .....	61
4.1 Model Interpretation.....	61
4.2 Model Performance .....	64
4.3 Model Representativeness.....	65
4.4 Issues of Data Quality and Spatial Uncertainty .....	67
4.4.1 Vegetation Resource Inventory Layer.....	67
4.4.2 Digitizing Decline Occurrence .....	68
Chapter 5 - Conclusions.....	69
5.1 Key Findings .....	69
5.2 Recommendations for Future Research .....	70
References.....	72

# List of Tables

Table 1: Synopsis of biogeoclimatic variants in study area.....	16
Table 2: A summary of climate for each of the 5 common BEC variants in the study area .....	17
Table 3: Ecological Land Unit (ELU) combination key.....	19
Table 4: Predictor variables .....	24
Table 5: Univariate logistic regression model results.....	44
Table 6: Standardized coefficient estimates computed from all of the data, the mean and SE of the estimates computed from the bootstrap samples (B=1000), and the % of samples in which variable had coefficient at zero .....	49
Table 7: Reported coefficients and significance values from logistic model .....	50
Table 8: Confusion matrix for logistic model (threshold value = 0.7) .....	52
Table 9: Probability of decline occurrence predicted for stands with yellow-cedar component..	53

# List of Figures

Figure 1: Yellow-cedar decline in British Columbia, 2008.....	3
Figure 2: Conceptual model linking global climate change to forest dieback.....	5
Figure 3: Distribution of yellow-cedar & location of North Island-Central Coast Forest District.	7
Figure 4: Working hypothesis for yellow-cedar decline.....	10
Figure 5: Study area location on coastal edge of the North-Island Central Coast Forest District	14
Figure 6: BEC variants found in study area.....	15
Figure 7: Example of colour orthophoto mosaic obtained for study area.....	20
Figure 8: Digitized polygons overlaid on digital elevation model .....	21
Figure 9: Striping artefacts in DEM.....	22
Figure 10: Shape of logistic model .....	32
Figure 11: Receiver-Operating Characteristic (ROC) curve .....	35
Figure 12: Mean elevation values of “all yellow-cedar” stands and “leading yellow-cedar” stands .....	39
Figure 13: Mean slope values of “all yellow-cedar” stands and “leading yellow-cedar” stands	40
Figure 14: Mean aspect values of “all yellow-cedar” stands and “leading yellow-cedar” stands	40
Figure 15: Mean elevation values of declining and healthy yellow-cedar stands .....	42
Figure 16: Mean slope values of declining and healthy yellow-cedar stands.....	42
Figure 17: Mean aspect values of declining (black) and healthy (grey) yellow-cedar stands.....	43
Figure 18: Size of regularized regression coefficients as the absolute sum of regression coefficients increases. ....	46
Figure 19: Lasso shrinkage of coefficients. ....	47
Figure 20: BIC values computed at the values of the penalty parameter where the active set changed .....	48
Figure 21: ROC curve for logistic model .....	51
Figure 22: Sensitivity and specificity curves plotted for different threshold values .....	52
Figure 23: Map depicting likelihood of decline occurrence for yellow-cedar stands within the study area .....	54

Figure 24: Map depicting binary outcomes for predicted decline occurrence using a threshold of 0.7.....	55
Figure 25: Empirical semi-variogram from standardized Pearson residuals .....	56
Figure 26: Full decision tree grown from data.....	58
Figure 27: Cross-validated error as a function of the complexity parameter .....	59
Figure 28: Pruned decision tree .....	59
Figure 29: Estimates of Brier's accuracy scores and Somers' Dxy values obtained using 10-fold cross-validation of a sequence of trees .....	60

# List of Abbreviations

AUC	Area Under the Curve
BC	British Columbia
BCGS	British Columbia Geographic System
BEC	Biogeoclimatic Ecosystem Classification
BIC	Bayesian Information Criterion
CART	Classification and Regression Trees
CDED	Canadian Digital Elevation Data
CRGB	Crown Registry and Geographic Base
CWH	Coastal Western Hemlock
DEM	Digital Elevation Model
ELU	Ecological Land Unit
ESRI	Environmental Systems Research Institute
GIS	Geographic Information System
GLM	Generalized Linear Model
Ha	Hectare
LASSO	Least Absolute Shrinkage and Selection Operator
LRDW	Land and Resource Data Warehouse
MGET	Marine Geospatial Ecology Tools
MH	Mountain Hemlock
MLE	Maximum Likelihood Estimation
MOFR	Ministry of Forests and Range
NICCFD	North Island-Central Coast Forest District
ROC	Receiver Operating Characteristic

TOPEX	TOPographic EXposure
TRIM	Terrain Resource Information Management
TSA	Timber Supply Area
VRI	Vegetation Resources Inventory

# Acknowledgements

I would like to thank the following people for their support, without which this research would not have been possible:

Dr. Brian Klinkenberg, my supervisor, for the direction and insight provided throughout this project. Thank you for your support and invaluable feedback.

Dr. Lori Daniels, my second reader, for providing excellent feedback.

Dr. Dan Moore and Dr. Jiahua Chen, for granting advice concerning the statistical challenges I faced.

Dr. Ze'ev Gedalof at the University of Guelph, who sparked my interest in biogeography and provided a stimulating learning environment, which played a large role in my decision to apply to graduate school.

Stefan Zeglen and the BC Ministry of Forests and Range, for granting me the opportunity to take part in this project.

I would also like to thank my friends:

Lisa, Nez, and Betsy, for the good food, conversation and adventures. Your friendship made this experience a truly memorable one.

Tom, for the forest ecology chats, insight into my project and encouragement provided.

Jill, for the love, light and laughter.

Finally, I would like to extend a huge thank-you to my family and friends back home. I wouldn't be where I am today without your love and support. I simply can't thank you enough.

# Chapter 1

## Introduction

*“Just as the careful study of the geographical variation of the incidence of human disease has provided a powerful heuristic tool in epidemiology,...the study of the spatial structure of forest dieback may be expected to provide a useful guide in the search for causes of forest decline.” (Daoust et al., 1992, pp.156)*

### 1.1 Motivation for the Study

Broad patterns of terrestrial vegetation are largely determined by climate, particularly by mean and variation in annual temperature and precipitation (Box, 1995; Prentice, 1990). This dependency is well illustrated by the climate driven range expansions of many species following the last glacial maximum (Gates, 1990; Parmesan et al., 2005). While terrestrial plant species were generally able to keep pace with climate changes during the Pleistocene-Holocene warming, the rate of climate change was about an order of magnitude less than the projected changes for the 21<sup>st</sup> century (Bush et al., 2004). These impending changes in climate are projected to result in shifts in vegetation distributions at unprecedented rates (Allen and Breshears, 1998; Bush et al., 2004).

Global Climate Models predict that the globally averaged surface temperature will warm by approximately 1.8 to 4° C by the end of the 21<sup>st</sup> century, depending on the greenhouse gas emission scenario (IPCC, 2007b). However, the warming trend is not occurring uniformly across the globe; rather, the greatest warming is occurring at northern latitudes (Gates, 1990; IPCC, 2007b). It is thought that the rate of climate change will shift the climatic boundaries of biomes northwards at a pace which will exceed the rate at which many species can shift their distribution (Gates, 1990). This rapid change could result in extirpations and extinctions in some species as they lose suitable habitat at the southern end of their original distribution (Hansen et al., 2001; Shafer et al., 2001). Hamman and Wang (2006) discovered through their biogeoclimatic envelope study that tree species growing along elevation bands in mountainous terrain may be

particularly vulnerable. These species are predicted to lose potential habitat faster than they gain new habitat and are expected to rapidly decline in frequency at their current elevations (Hamman and Wang, 2006).

A change in climate has been implicated in several region-wide forest diebacks, including the decline of yellow-cedar. Yellow-cedar, (*Chamaecyparis nootkatensis* (D. Don) Spach<sup>1</sup>), also known as Alaska cedar, is a tree of high ecological, commercial, and cultural importance in coastal Alaska and British Columbia (Hennon et al., 2006). The species is currently undergoing the most severe forest decline in western North America (Hennon et al., 1998), with concentrated areas of decline located in southeast Alaska and along the north and central coast of British Columbia (Hennon et al., 1998; Hennon et al., 2005). Recent research into the decline has led to the development of a working hypothesis involving a climate shift and site specific factors (Hennon et al., 2006). The overall objective of this study was to contribute to understanding of the phenomenon by conducting the first landscape-level analysis of the decline in British Columbia. In the remainder of this chapter, I examine climate change as a mechanism for forest declines and summarize previous research into the decline of yellow-cedar (Section 1.2). I have identified key knowledge gaps to provide justification for the research question, which I present in Section 1.3.

---

<sup>1</sup> The taxonomic status of yellow-cedar may change following the discovery of a closely related tree species in northern Vietnam (*Xanthocyparis vietnamensis* Farjon & Hiep). At the next International Botanical Congress in 2011, it will be determined whether yellow-cedar will be transferred to the newly established genus *Xanthocyparis*, or its former genus, *Callitropsis* (Mill & Farjon, 2006).



Figure 1: Yellow-cedar decline in British Columbia, 2008  
Photo credit: T. Maertens

## 1.2 Literature Review

### 1.2.1 Biophysical Theory of Forest Decline

While there is currently a great deal of concern regarding how anthropogenic climate change could affect forests (Winnet, 1998), much uncertainty surrounds both climate change predictions and the likely response of vegetation (Bachelet et al., 2001; Gates, 1990; Hansen et al., 2001). Determining likely responses of vegetation to climate change is difficult due to the complex nature of ecosystem responses (Allen and Breshears, 1998; Winnet, 1998). However, the idea that climate shifts can lead to forest declines directly or indirectly is now widely accepted by forest pathologists (Manion, 1991).

The biophysical theory of Auclair et al. (1992) suggests that episodic, region-wide forest dieback may be driven by climate variation. From a review of forest diebacks in North American hardwoods, Auclair et al. (1992) found evidence indicating that both short-term extreme climate fluctuations, such as severe winter freezing or spring frosts, and long-term shifts in climate, are related to crown dieback events. Temporal correspondence between climate warming in the Northern Hemisphere and forest dieback was found, with the onset of diebacks following periods (3-9 years) of rapid increase in temperature (Auclair et al., 1992). The years of onset for each of the hardwood species considered all had at least one pronounced thaw-freeze event (Auclair et al., 1992). These observed “false spring” patterns led Auclair et al. (1992) to propose that xylem cavitation injury initiated by extreme freezing and/or moisture fluctuations is the mechanism responsible for initiating forest dieback episodes (Figure 2). Freezing events following periods of mild weather can result in cavitation where gas bubbles form when the tissues and fluids freeze, interrupting the xylem flow (Auclair et al., 1992; Hennon and Shaw, 1994). Trees can repair the cavitation if the gases are dissolved, but the damage may be irreversible if frequent thaw-freeze events occur or freezing injury to roots prevents the generation of adequate root pressure to refill the xylem (Hennon and Shaw, 1994; Zhu et al., 2001). The resulting loss of hydraulic conductivity makes the trees susceptible to drought, thereby predisposing them to crown dieback in warm, dry years (Auclair et al., 1992).

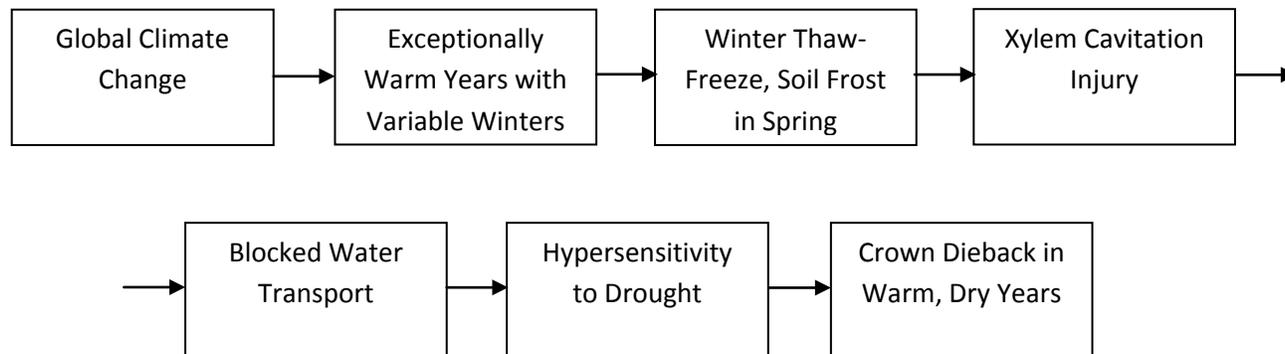


Figure 2: Conceptual model linking global climate change to forest dieback

Source: Auclair et al., 1992

Auclair's (1992) proposed mechanism of decline appears to be supported in studies by Bourque et al. (2005) and Cox and Zhu (2003) investigating yellow birch (*Betula alleghaniensis* Britt.) decline. Bourque et al., (2005) compared historical decline events of yellow birch dieback and decline with past thaw-freeze events, and discovered that there was good temporal and spatial correspondence between the two. Cox and Zhu (2003) concluded in their study that winter thaws predispose yellow birch seedlings to xylem cavitation, freezing of dehardened shallow roots, and shoot dieback.

Several other studies of forest species dieback in North America have also found evidence to suggest that climate variation plays a predisposing role. The decline of red spruce (*Picea rubens* Sarg.) in the Appalachian mountains has been attributed to a loss of equilibrium with its climatic environment (Cook and Johnson, 1989). The climate changes are reported to have exceeded certain physiological thresholds of response and the altered physiology is suspected to have made the species more vulnerable to injury from short-term climatic extremes (Cook and Johnson, 1989). By analysing the relation of growth to weather patterns, Leaphart and Stage (1971) determined that the interaction of climate change with soils of low water-holding capacity triggered the pole blight of western white pine (*Pinus monticola* Dougl.). Recent aspen decline has also been attributed to a combination of above-normal temperatures and moisture stress (Worrall et al., 2007). Extreme weather events such as late winter thaw-freeze events and severe spring frost are also believed to predispose aspen stands to decline, leaving them susceptible to

invasion by insects and pathogens (Frey et al., 2004). The decline of yellow-cedar has also been linked to thaw-freeze events; this evidence is explored in the following section.

## 1.2.2 The Case of Yellow-Cedar Decline

### **Yellow-Cedar Biogeography**

Yellow-cedar has an extensive range, growing along the Pacific Coast of North America from Southern Oregon to Prince William Sound, Alaska (Harris, 1990) (Figure 3). Throughout most of its range, yellow-cedar is limited to high elevations, but north of mid-coastal British Columbia, it grows from near timberline down to sea level (Harris, 1990). Yellow-cedar is abundant in cool, wet boggy sites, and is often less common in better drained more productive sites where it is unable to compete with western hemlock and Sitka spruce (Harris, 1990). While not a competitive species, yellow-cedar is stress-tolerant, which allows it to persist in marginal sites where competition is at a minimum (Hennon and Shaw, 1997). The tree species puts relatively few resources into growth and reproduction, but achieves great longevity, often living a millennium or longer (Hennon and Shaw, 1997). The yellow colour and distinct aroma of the heartwood comes from the powerful natural biocides, such as nootkatin, which protect the tree from infestation and decay (Hennon and Shaw, 1997). The leaves are also protected against insects, containing volatile leaf oils that restrict insect feeding (Hennon and Shaw, 1997).



Figure 3: Distribution of yellow-cedar and location of North Island-Central Coast Forest District  
 Source: Digital representation of yellow-cedar range obtained from USGS, 2006

Despite this defensive ecological strategy, yellow-cedar is known to be declining on over 200,000 ha of undisturbed forest in southeast Alaska, making it the most severe forest decline in western North America (Hennon et al., 1998). Similar patterns of mortality were recently observed in British Columbia, with the greatest concentration located in the North Island – Central Coast Forest District (Hennon et al., 2005). The nature of the dieback is thought to be consistent with the phenomenon in southeast Alaska and has prompted calls for further research into the cause of the decline (Hennon et al., 2005).

### **Past Research**

Research into the decline of this long-lived species began in the early 1980s and a sequence of symptoms was identified. The initial symptom was determined to be fine root death, followed by death of small-diameter roots (Hennon et al., 2006). As the roots start to die, the trees develop thin off-colour crowns and necrotic lesions spread from larger roots up the bole (Hennon et al., 2006). The natural resistance of yellow-cedar heartwood to decay allows dead trees to remain standing for a century or more after death (McDonald et al., 1997). By examining the standing snags it was possible to establish that the decline of yellow-cedars began in about 1880-1900 (Hennon and Shaw, 1997). The age structure of the dying trees and stands indicated that there was no clear pattern between tree age and mortality (Hennon and Shaw, 1994; Hennon and Shaw, 1997). Since the species has great potential longevity, the trees are not dying of senescence (Hennon and Shaw, 1997).

During observational studies it was noted that the decline often appears to be associated with bog (muskeg) or semibog sites and the decline was thought to be spreading (Hennon et al., 1990a). The suspected local spread of decline and the fact that a single species was affected suggested that a pathogenic agent was responsible (Hennon et al., 1990b). Investigations into the decline initially focused on finding a biotic cause, but one by one the suspected agents were ruled out (Hennon and Shaw, 1997). The nature of the decline spread was subsequently investigated and it was determined that the spread occurs as a slow advance along an established ecological and vegetational gradient (Hennon et al., 1990a; Hennon and Shaw, 1994). The mortality tends to originate in bogs and spread upslope along the gradient to better drained communities (Hennon et al., 1990a; Hennon and Shaw, 1994). This indicates that a pathogenic organism is unlikely to

be responsible for the decline, as the spread of such organisms is unlikely to follow physical or plant community gradients that existed prior to mortality (Hennon et al., 1990a).

Attention then shifted to abiotic factors potentially associated with the decline and the association of the decline with wet, poorly drained soils was confirmed (Hennon et al., 2006). However, the relation with soil drainage was determined to be inconsistent, with limited decline occurring on wet sites at higher elevations (Hennon et al., 2006). Air and soil temperature were determined to be stronger risk factors than poorly drained soils (D'Amore and Hennon, 2006).

### **Current Hypothesis for Decline**

These clues led researchers to propose a new, complex hypothesis to explain yellow-cedar decline (Figure 4). According to Hennon et al. (2006), saturated soils create open, exposed canopies that experience soil warming earlier in the spring. This warming triggers the yellow-cedars to dehardens prematurely, making them more susceptible to freezing injury (Hennon et al., 2006). The vulnerability of yellow-cedar to temperature-dependent dehardening was confirmed in a study by Schaberg et al. (2005) that compared the freezing tolerances of yellow-cedar and western hemlock trees in a site affected by yellow-cedar decline. Schaberg et al. (2005) found that yellow-cedar dehardens a greater amount and earlier in the spring than co-occurring western hemlock. These dehardened yellow-cedars were found to be highly vulnerable to root freezing in a study that measured cold tolerance and freezing injury of yellow-cedar seedlings using simulated snow cover (Schaberg et al., 2008). The dehardened roots of yellow-cedar may be damaged by a form of injury similar to that described in Auclair et al.'s (1992) hypothesis of 'xylem injury by cavitation' (Hennon and Shaw, 1994) (see Section 1.2.1).

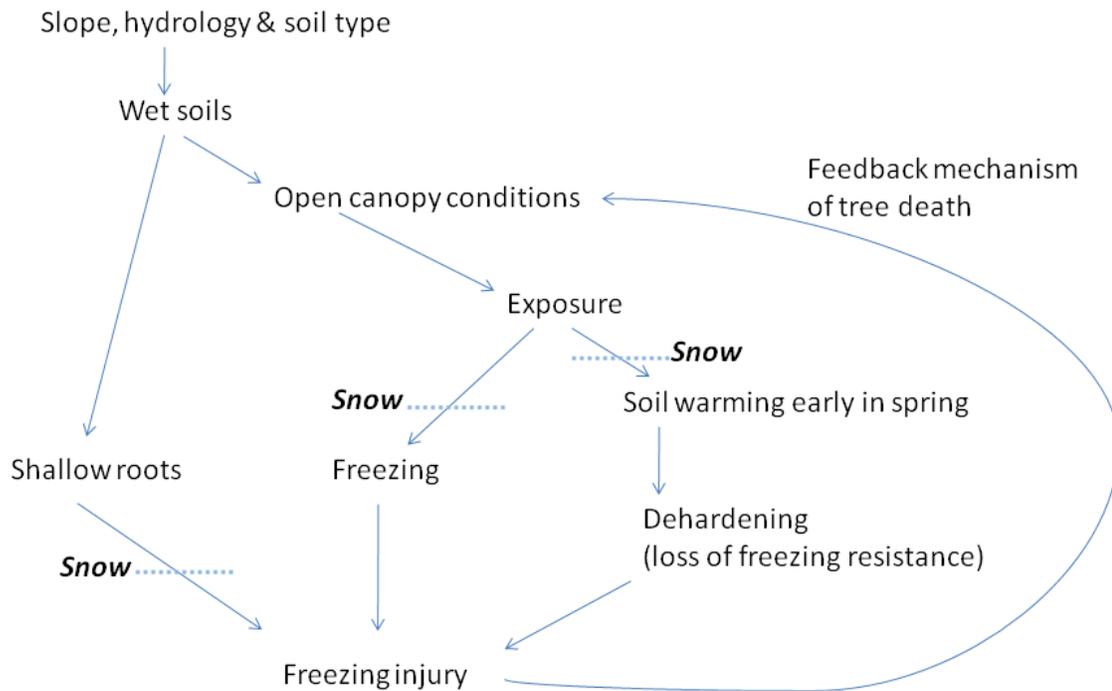


Figure 4: Working hypothesis for yellow-cedar decline. Dashed lines represent the protection from freezing injury that snow cover appears to offer yellow-cedar roots.

Source: Adapted from Hennon et al. 2006

Snow appears to protect yellow-cedar roots against this potential freezing injury by preventing soil warming (Hennon et al., 2006). However, there has been a reduction in the snowpack at lower elevations since the end of the Little Ice Age, which coincided with the onset of decline (Hennon et al., 1990a; Hennon et al., 2006) The loss of this protective snowpack increases susceptibility of fine roots to freezing because trees growing in saturated soils have a shallow root system and saturated soils offer poor insulation (D'Amore and Hennon, 2006; Hennon et al., 1990a; Hennon and Shaw, 1994). At higher altitudes the trees have a greater cold tolerance and a more persistent winter snowpack making freezing injury less likely (Hennon et al., 1990a; Hennon and Shaw, 1994; Schaberg et al., 2005). This theory would account for the apparent elevational limit to the decline observed at study sites in southeast Alaska. Mapping of the decline at two study sites in the Alexander Archipelago of the Alaska Panhandle (57° N) revealed that the decline is concentrated at lower elevations, with very little decline present

above 300 m, and more decline occurs on warm aspects (south and southwest) (Hennon et al., 2006).

The slow spreading pattern of mortality may be explained by the feedback mechanism of tree death (D'Amore and Hennon, 2006). When the trees die they create openings in the canopy at the perimeter of the decline patches, which exposes soils with better drainage to warming, which leads to more tree death (D'Amore and Hennon, 2006; Hennon et al., 1990a; Hennon et al., 2006).

It would appear that the onset of yellow cedar decline has developed due to regional climate warming and decreased winter snowpack at low elevations, along with thaw-freeze events (D'Amore and Hennon, 2006). Hennon et al. (2006) suggested that yellow-cedar perhaps migrated to these lower elevation sites during the Little Ice Age (approximately A.D. 1450 to 1850) when these sites would have had a more consistent snowpack, but the warming trend has left these open, exposed canopies vulnerable due to a loss of protective snow. Yellow-cedar seems to be particularly susceptible due to its unique physiological traits (greater proportion of roots in shallow zone, limited cold tolerance, precocious root activity) that combine with particular site factors (low elevation sites, warm aspects) to predispose the tree to decline. The shift in climate may represent the environmental trigger responsible for the decline and suggests that the dieback may expand if warming trends continue (Hennon et al., 2006).

Researchers in British Columbia have noted that the decline appears to be most concentrated at lower elevations, around 300-400 m, and is frequently found on south facing aspects (Hennon et al., 2005). These observations seem to be congruent with the decline hypothesis. However, limited investigations into the decline observed in British Columbia have been made. Further knowledge of where the decline is occurring and whether the pattern fits the decline hypothesis is required.

## 1.3 Research Question

While observational studies and recent experimental findings support the hypothesis that yellow-cedar decline is a result of shallow root freezing due to a lack of snow persistence, more research is required to further elucidate the mechanism for the decline. The decline situation in British Columbia is of particular interest as it is not clear whether the pattern of decline matches that observed in Alaska and it presents an opportunity to study the southernmost extent of the decline.

This research represents the first landscape-level analysis of the decline in British Columbia and aims to examine how the decline is related to the landscape. The specific research question addressed by this thesis is:

*How does yellow-cedar decline relate to topographic variables at the landscape-scale?*

Examining the spatial distribution of the decline and its relation to the landscape will enable the identification of site factors predisposing trees to decline. The insight gained from this research will allow for a more thorough understanding of the mechanism and provide insight into the distribution of the dead and dying trees in British Columbia. Knowledge of the determinants of the spatial pattern of decline will improve predictability and provide important information for conservation and resource management. If the proposed hypothesis for yellow-cedar decline is correct, then it indicates how a loss of equilibrium with climate can have dramatic impacts on a forest ecosystem. Ultimately this research will provide insight into how a climatic-driven forest decline manifests across the landscape.

## 1.4 Thesis Structure

This thesis is comprised of five chapters. In the following chapter (Chapter 2) I provide a description of the study area, data collection process, and methods of analysis and model development. In Chapter 3 I present the results of the data analysis and model building. I discuss how the results of the models address the major research question identified in Section 1.3 in Chapter 4. Finally, in Chapter 5 I have summarized the key research findings and identified opportunities for future research.

# Chapter 2

## Study Area and Methods

### 2.1 Study Area

My study site encompasses an area of approximately 800,000 ha on the coastal edge of the North Island-Central Coast Forest District (NICCFD) (51°39'N, 127°18'W) (Figure 5). This location is where the greatest acreage of yellow-cedar decline has been observed in British Columbia (Westfall and Ebata, 2008). The forest district is divided into the Mid-Coast Timber Supply Area (TSA) and the Kingcome TSA. The Mid-Coast TSA is comprised of 2.2 million hectares and is situated on the central coast of British Columbia. This TSA stretches from Cape Caution in the south to Sheep Passage in the north, with Tweedsmuir Provincial Park providing the eastern boundary (BC MOF, 1999b). The Kingcome TSA covers approximately 1.14 million hectares including much of Northern Vancouver Island (BC MOF, 1999a). On the mainland, the Kingcome TSA encompasses the area between Knight Inlet in the south and east, to Cape Caution in the northwest and to Tweedsmuir Park in the northeast (BC MOF, 1999a).

The TSAs combine to form an expansive area with variable and rugged terrain, and diverse ecological communities (MOFR, 2009). The most westerly portion of the forest district consists largely of low-lying islands with relatively low forest productivity (BC MOF, 1999b). Further inland the terrain is mountainous, lined with productive forests at the lower elevations, and non-productive forests (alpine and subalpine) and ice fields at higher elevations (BC MOF, 1999b). Due to the diversity of terrain and climatic conditions, the forest district contains seven different biogeoclimatic ecosystem classification (BEC) variants (Figure 6).

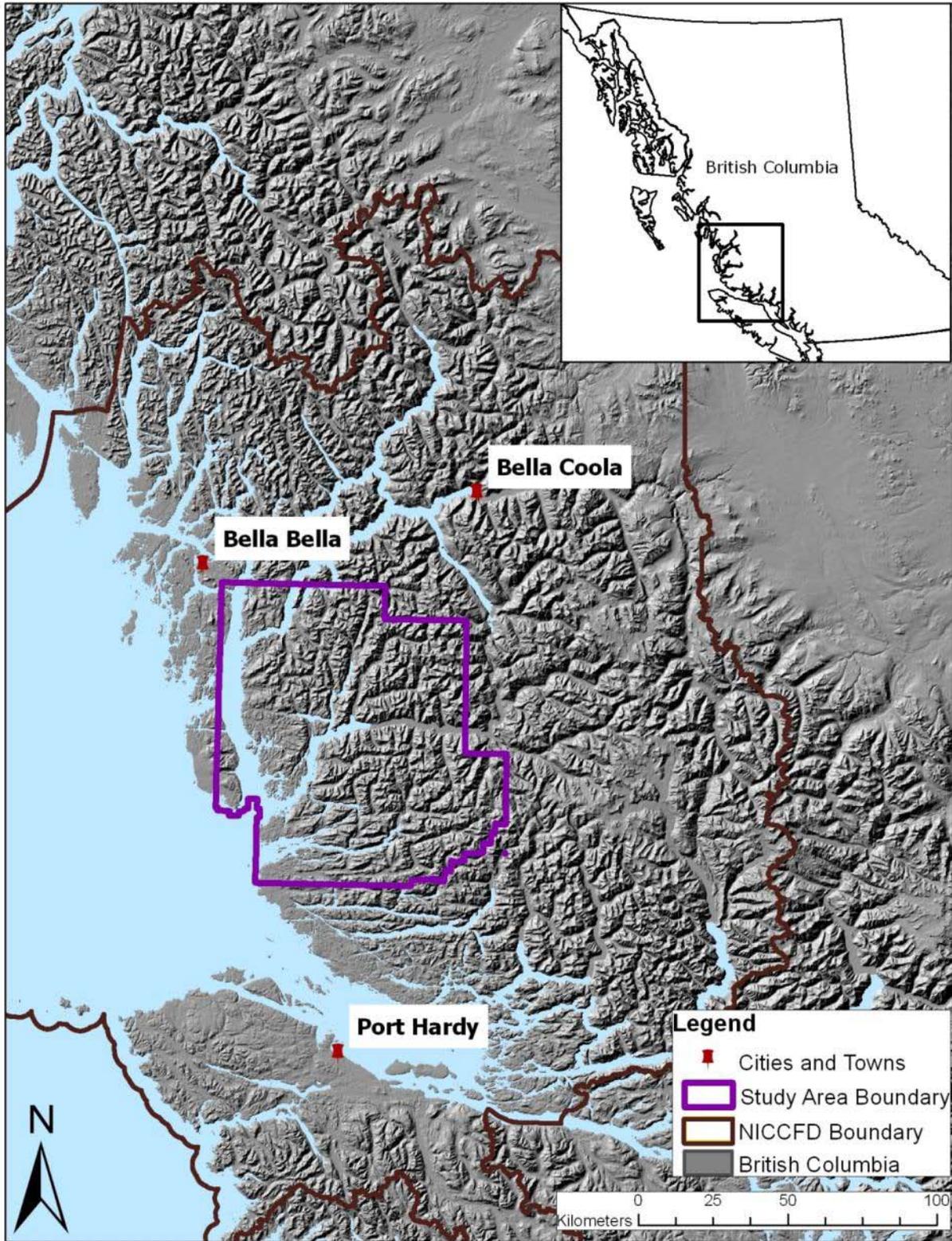


Figure 5: Study area location on the coastal edge of the North-Island Central Coast Forest District

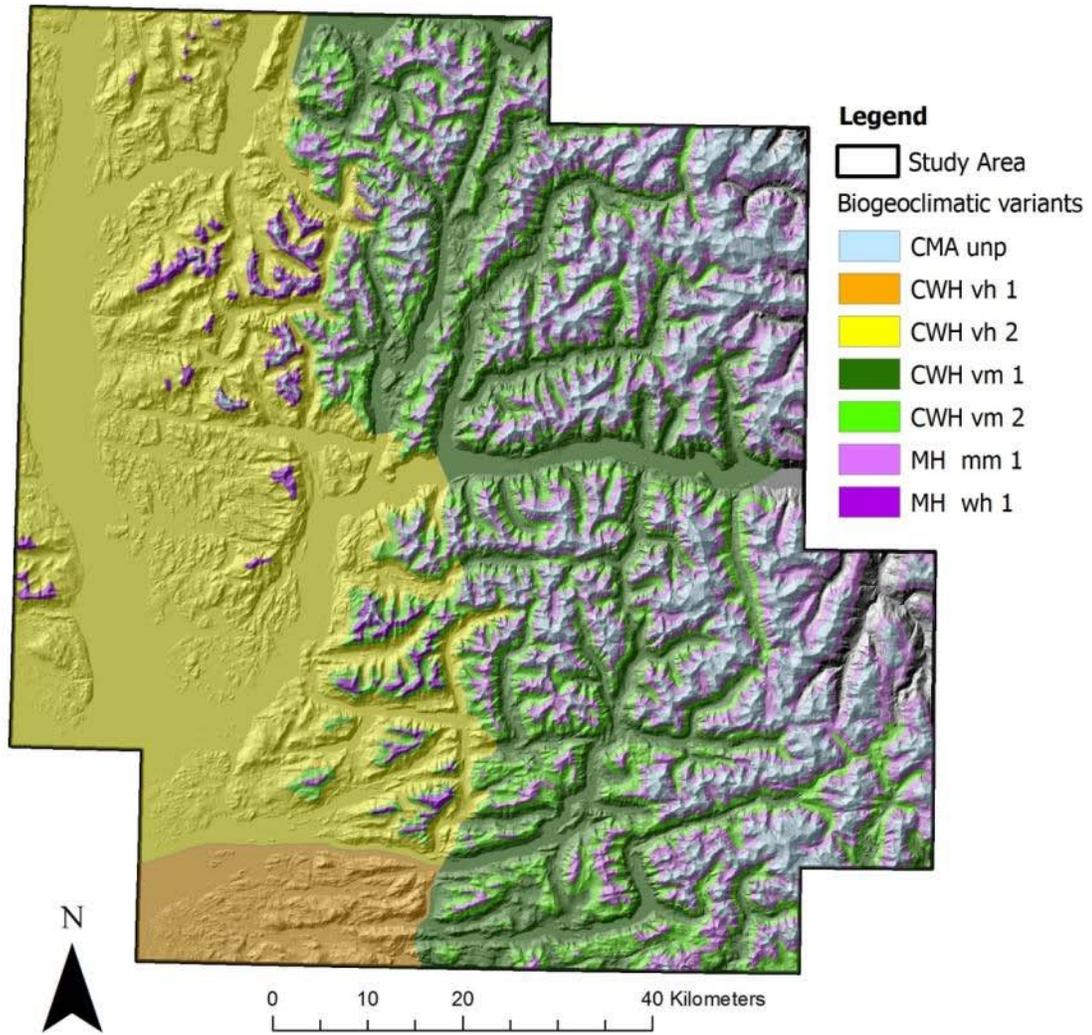


Figure 6: BEC variants found in study area

My research will focus on the coastal edge of the forest district, where yellow-cedar is a dominant species in some areas. In this region, yellow-cedar grows in two biogeoclimatic zones: the Coastal Western Hemlock (CWH) zone and the Mountain Hemlock (MH) zone. The most common biogeoclimatic variants are listed in table 1, and a summary of the climate for each of the major variants is provided in table 2. The CWH zone is the most productive forest ecosystem in Canada and occurs at low to mid elevations, mainly west of the coastal mountains (BC MOF, 1999b; Meidinger and Pojar, 1991). In the southern and mid-coast portion of British Columbia, the CWH zone occupies elevations from sea level to 900 m on windward slopes (1050 m on leeward slopes), and is restricted to lower elevations in the north (Meidinger and Pojar, 1991). The CWH zone is characterized by high rainfall, cool summers and mild winters (BC MOF, 1999b; Meidinger and Pojar, 1991). The MH zone is typically the subalpine forest above the CWH zone, occupying elevations of 900-1800 m in the south and 400 to 1000 m in the north (Meidinger and Pojar, 1991). This subalpine forest has long, cool, wet winters with a persistent snowpack that results in a short growing season (Meidinger and Pojar, 1991). Yellow-cedar is widely distributed throughout the CWH zone, frequently occurring in the very wet maritime and hypermaritime subzones. At low elevations, yellow-cedar typically occurs as a scattered component of mixed species forests. The species reaches its greatest concentration at middle and higher elevations, reaching its greatest importance in the MH zone (MHmm1 & MHwh1 variants) (Meidinger and Pojar, 1991; Zeglen, 2004).

Table 1: Synopsis of biogeoclimatic variants in study area

<b>BEC code</b>	<b>BEC variant</b>
CWHvh1	Coastal Western Hemlock southern very wet hypermaritime
CWHvh2	Coastal Western Hemlock central very wet hypermaritime
CWHvm1	Coastal Western Hemlock submontane very wet maritime
CWHvm2	Coastal Western Hemlock montane very wet maritime
MHwh1	Mountain Hemlock windward wet hypermaritime
MHmm1	Mountain Hemlock windward moist maritime
CMAunp	Coastal Mountain-heather Alpine undifferentiated and parkland

Table 2: A summary of the climate (1960-1991) for each of the 5 common BEC variants in the study area

<b>Biogeoclimatic unit</b>		<b>CWHvh1</b>	<b>CWHvh2</b>	<b>CWHvm1</b>	<b>CWHvm2</b>	<b>MHmm1</b>
<b>Name of reference station</b>		Estevan Point	Ethelda Bay	Haney Loon Lake	Tunnel Camp	Grouse Mt. Resort
<b>Reference elevation (m)</b>		7	8	354	671	1128
<b>Mean annual precipitation (mm)</b>	Range ref. stn.	3120 (2009 – 3943)	2682 (1555 – 4387)	2850 (2760 – 2850)	3186 (1532 – 4218)	2565 (2565 – 2954)
<b>Mean annual temperature (°C)</b>	Range ref. stn.	9.1 (5.4 - 9.4)	8.3 (7.0 - 10.1)	No data	7.7 (6.7 - 8.5)	4.6 (4.6 - 5.0)
<b>Extreme minimum temperature (°C)</b>	Range ref. stn.	-13.9 (-7.5 - 17.22)	-19.4 (-8.9 - 22.8)	No data	-16.7 (-11.1 - 24.4)	18.5 (-18.5 - 26.7)
<b>Extreme maximum temperature (°C)</b>	Range ref. stn.	28.9 (22.8 - 37.8)	34.4 (27.8 - 41.1)	No data	29.4 (23.4 - 33.3)	29.0 (29.0 - 33.3)
<b>Growing degree-days &gt; 5 (°C)</b>	Range ref. stn.	1607 (818 – 1722)	1633 (1313 – 2011)	No data	1319 (1148 – 1485)	933 (919 – 933)
<b>Frost –free period (days)</b>	Range ref. stn.	229 (163 – 265)	199 (165 – 252)	No data	160 (156 – 272)	126 (125 – 126)

(From Green and Klinka, 1994) Mt. = mountain, ref. stn. = reference stations.

## 2.2 Data Collection and Processing

### 2.2.1 Yellow-Cedar Decline Dataset

The main aim of my research is to identify relations between areas of declining yellow-cedar trees and topographic factors. To analyze the pattern of decline in the study area, locations of healthy and declining yellow-cedar trees and topographic variables were required. The BC Ministry of Forests and Range (MOFR) has produced a spatial dataset detailing known locations of yellow-cedar decline from aerial surveys in 2006 and 2007. However, these decline patches were determined to be inaccurate in terms of their spatial location and therefore were not appropriate for use in this study. The pattern of decline on the landscape was instead determined through a combination of GIS and remote sensing analysis.

A Vegetation Resources Inventory (VRI) layer was downloaded from the Land and Resource Data Warehouse (LRDW) and used to identify the location of yellow-cedar stands in the study area. This dataset is comprised of spatial layers and associated textual attributes detailing the location and vegetation composition of forest inventory units. There are concerns surrounding the reliability of the VRI data, particularly relating to the species composition estimates and consistency of data and delineation standards (Interfor and Timberline, 2008). For these reasons, polygons interpreted before the last major inventory in 1998, some of which date back to the 1950s, were excluded from analysis. Of the original set of polygons of interest, approximately 86% were retained following the exclusion of the older inventory data.

The “polygons of interest” in the VRI data were defined as polygons in which yellow-cedar was the leading species (minimum species composition of 30% according to gross volume estimates). The analysis was restricted to areas where yellow-cedar dominates as mortality is not easily detected on images where it is only a minor component of the stand. The distribution of leading yellow-cedar stands on the landscape was then compared to that of all stands with a yellow-cedar component in order to determine whether the stands occur in similar locations in terms of elevation, slope and aspect. This information was used to inform the sampling scheme.

A stratified sampling scheme was created to ensure that the range of environmental gradients under which yellow-cedar grows, and thus may be declining, were captured. Environmental

gradients across the entire study area of elevation, slope and aspect were grouped into discrete classes and then combined to create 36 ecological land units (ELUs), a classification scheme I developed (Table 3). This layer of ELUs was then clipped to the extent of the leading yellow-cedar layer. Using Hawth's Tools for ArcGIS (Beyer, 2004), three points were randomly distributed in each of the 36 ELUs giving a total of 108 sample points. When generating the sample points a minimum distance of 1000 m was used to separate the points, and only one point was allowed to fall in any one particular leading yellow-cedar polygon.

Table 3: Ecological Land Unit (ELU) combination key

<b>Slope Class</b>	<b>Code</b>	<b>Elevation Class</b>	<b>Code</b>	<b>Aspect Class</b>	<b>Code</b>
< 20°	10	0 – 400 m	100	NW – NE 315° - 45°	1000
20 - 40°	20	400 – 800 m	200	NE – SE 45° - 135°	2000
> 40°	30	800 – 1400 m	300	SE – SW 135° - 225°	3000
				SW – NW 225° - 315°	4000

The most suitable imagery available for the study area was digital colour orthophotographs and corresponding stereopairs, which were taken in the summer of 2006 (Figure 7). Digital orthophotos are digital images of aerial photographs that have been geometrically corrected to remove variations in scale and displacement. The orthophotos were obtained from the Crown Registry and Geographic Base (CRGB) as orthophoto mosaics (with 0.5 m pixel size), which are orthophotos that have been merged or tiled together, then clipped to British Columbia Geographic System (BCGS) 1:20000 map grid. The stereo pairs were also acquired from the CRGB and these images were viewed with PurVIEW ArcGIS Desktop extension which permits stereo-viewing in the ArcMap environment.



Figure 7: Example of a colour orthophoto mosaic obtained for study area. In this mosaic the orthophotos for BCGS map sheet 92M.044 have been tiled together and clipped to the BCGS 1:20,000 grid. Draney Inlet is the large body of water visible in the SE region.

At each sample point, the health of the forest stand was determined by visually inspecting the corresponding areas on the orthophotos and stereopairs. Stands were considered to be declining . When no signs of yellow-cedar decline were detected, the boundary of the leading yellow-cedar polygon containing the sample point was digitized. In instances where yellow-cedar decline was observed, the full extent of the decline was digitized. This approach meant that an area less than or greater than the original yellow-cedar polygon in which the point was placed could be digitized. Stereo pairs were used in the image interpretation stage as they permitted a more detailed examination of the tree crowns, whereas the digitizing was performed on the orthophotos. After digitizing all 108 polygons, this layer was overlaid onto the digital elevation model (DEM) (Figure 8) and the ArcGIS zonal statistics tool was used to obtain summary statistics of all topographic variables that were derived from the DEM (see section 2.2.2).

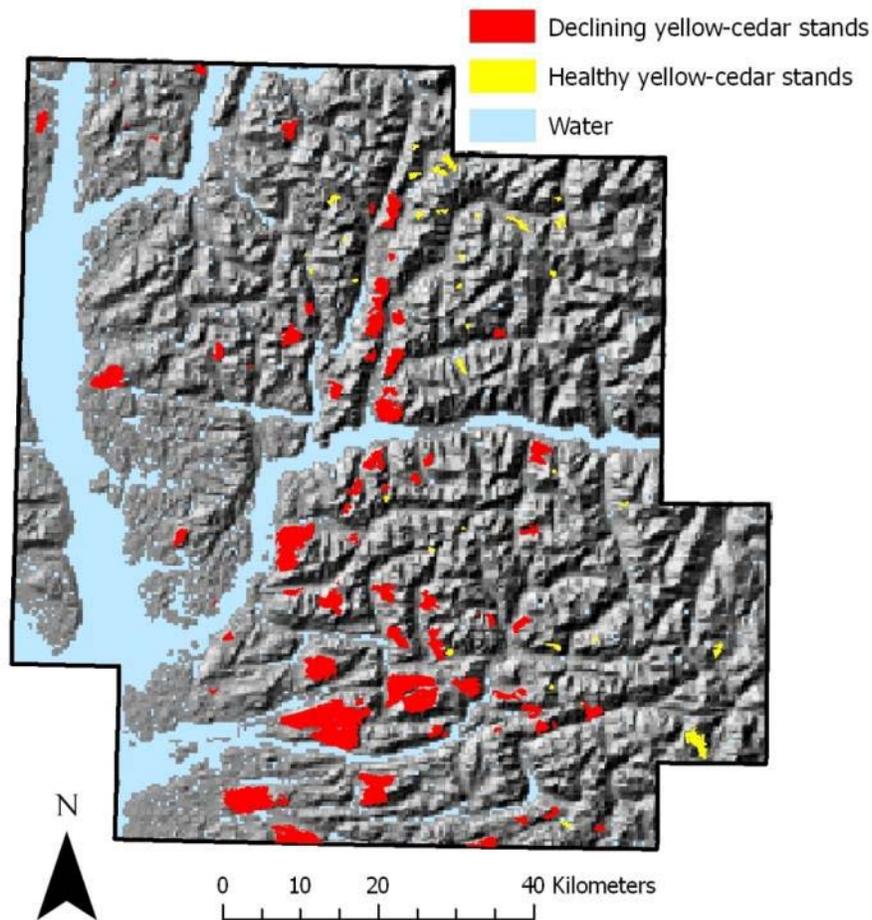


Figure 8: Digitized polygons overlaid on digital elevation model

### 2.2.2. Topographic Variables

To create a set of topographic variables, a DEM for British Columbia was first obtained from the CRGB. This regularly gridded DEM has a 25 m cell resolution and was derived from the 1:20,000 Terrain Resource Information Management (TRIM) data. Several terrain parameters based on the first or second derivative of the DEM surface, such as slope, aspect and surface curvature were then produced. When inspecting the resulting grids for the various parameters and the hillshaded TRIM-DEM (sun elevation angle: 45°, azimuth: 315°), a quilted pattern was discovered (Figure 9). These stripes are recognizable as production artefacts due to their systematic pattern and lack of correspondence with natural features observed on aerial photographs. The presence of these artefacts affects the accuracy of the DEM and thus the accuracy of derived products (Oimoen, 2000).

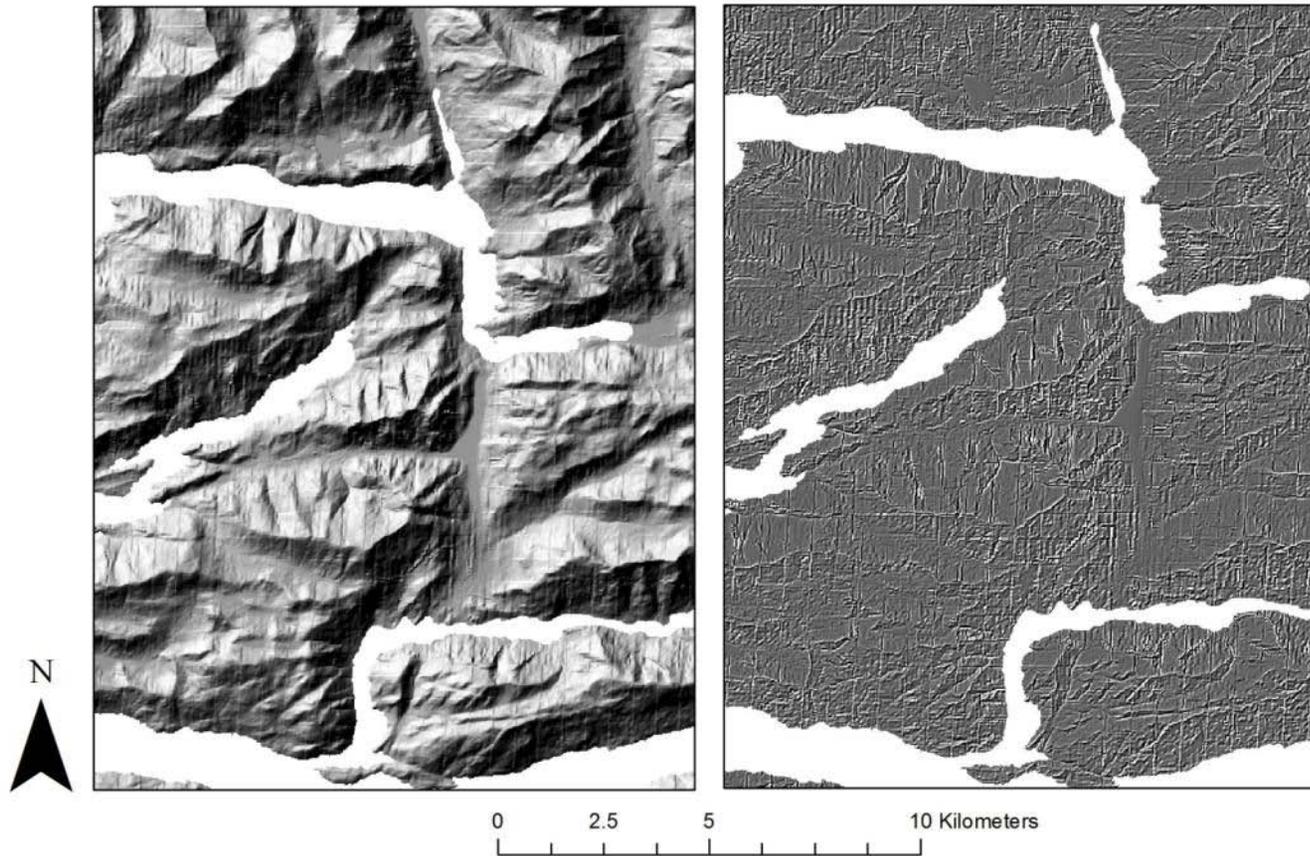


Figure 9: Striping artefacts in DEM (left – striping visible on hillshade, right – pronounced striping visible on curvature layer)

To address this problem, a set of BC Canadian Digital Elevation Data (CDED) files was downloaded from GeoBase and used to obtain a DEM free of production artefacts for the study area. The CDED files are distributed in ASCII format and have a .dem file extension. To produce a useable DEM it was necessary to batch convert all the CDED files from .dem to raster file format and then mosaic all the elevation rasters. The grid, which was originally in a geographic coordinate system with Decimal Seconds as the unit, was then projected in Albers Equal Area Conic Projection (standard projection for British Columbia) and bilinear resampling was applied to achieve a 25 m cell size. The resulting DEM file was determined to be free of striping artefacts.

The DEM was used to produce maps of elevation, slope, aspect, exposure, topographic wetness, solar insolation, terrain shape and distance to coast (Table 4). Terrain variables were used as surrogate predictors for climate in this study because the sparse distribution of climate stations and presence of highly variable micro-climates created by the mountainous terrain means that accurate climate data is not available. The variables were created using ArcGIS version 9.2<sup>®</sup> software (ESRI, 2006). Descriptions of the variables, and their presumed impact on climate/environmental conditions, are presented below.

Table 4: Predictor variables

<b>Variable</b>	<b>Variable Name</b>	<b>Definition</b>
<b>Mean elevation</b>	Elev_mean	Average elevation
<b>Variance in elevation</b>	Elev_var	Square of the standard deviance
<b>Elevation kurtosis</b>	Elev_krt	Homogeneity
<b>Elevation skew</b>	Elev_skw	Massiveness (inverse)
<b>Elevation relief</b>	Elev_relief	Local relief ( $z_{\max}-z_{\min}$ )
<b>Mean slope</b>	Slope_mean	Steepness
<b>Slope standard deviation</b>	Slope_stddev	Heterogeneity
<b>Slope kurtosis</b>	Slope_krt	Modality
<b>Slope skew</b>	Slope_skw	Limitation
<b>Mean aspect</b>	Aspectsw_mean	Measure of “southwesterlyness”
<b>Potential solar radiation</b>	Solarspring_mean	Solar radiation over winter/spring
<b>Terrain Shape Index</b>		Measure of terrain shape
<b>Topographic Wetness Index</b>	Twi_mean	Measure of soil moisture
<b>Mean distance from coast</b>	Coastdist_mean	Distance to coast (m)
<b>Mean topographic exposure</b>	Tpx3000_mean	Distance limited (3 km) topographic exposure (TOPEX)

### **Elevation**

Elevation, in meters above sea level, was obtained directly from the DEM. Elevation directly affects temperature; higher elevations are strongly correlated with colder ambient temperatures.

### **Slope**

A slope raster (in degrees) was generated for the study area using the Slope tool from the Spatial Analyst toolbox. This function measures the maximum rate of change in elevation from each cell to its neighbours. Conceptually, a plane is fitted to the z-values of a 3x3 window of cells

surrounding the center cell and the slope of the plane is calculated (ESRI, 2006). Slope, in combination with aspect, affects the amount of solar insolation a surface receives.

## **Aspect**

Aspect values were calculated using the Aspect tool from the Spatial Analyst toolbox. The tool uses the elevation grid and measures the direction of the maximum rate of change from each cell to its neighbours. In other words, the tool identifies the direction the slope faces. Aspect values are provided as a compass bearing from 0-359 degrees for each cell.

In order to use this measure in the preliminary analysis of the location of yellow-cedar decline, it was necessary to calculate mean aspect of the polygons manually in order to handle the circular nature of the data. Sine and cosine grids were generated from the aspect grid and the zonal statistics tool was used to create a sum of each of these values for every polygon. The arc tangent of the summed sine and cosine values was then calculated to determine the correct mean aspect. Returned values were converted to degrees and expressed in a 0-360 range.

For use in the multivariate statistical models, aspect was transformed from a circular variable into a linear measure relevant to vegetation. The transformation produces a continuous variable that reflects the differences in exposure to solar radiation. South-west facing slopes commonly have the greatest potential evapotranspiration due to high solar radiation load and land surface temperature, whereas north-eastern slopes are the least exposed and tend to be coolest (Austin, 2005). The equation was based on Beers (1966), but modified so that south-west facing slopes have a value of 0 and north-east slopes have a value of 1. The aspect grid was first converted to radians and then the following formula was applied to provide an index of “southwesterlyness”:

$$(Cos (.7845 - [aspect\ grid\ in\ radians] + 1) / 2) \quad (1)$$

## **Solar Radiation**

An insolation map for the study area was produced using the area solar radiation analysis tool in the Spatial Analyst Extension. The total amount of incoming solar radiation (direct + diffuse) over the late winter / spring months was calculated for the raster surface with outputs in unit watt

hours per square meter. February 1<sup>st</sup> was chosen as the starting date for calculating solar insolation because recent studies suggest that this is the earliest that yellow-cedar may begin dehardening under ambient thaw conditions (Schaberg et al., 2005, 2008). An end date of May 1<sup>st</sup> was selected to correspond with the work of Beier et al. (2008) who determined that hard frosts were not found in instrumental records after this date. Solar insolation influences soil and air temperatures and potential evapotranspiration levels.

### **Terrain Shape Index (TSI)**

A measure of terrain shape was derived using a routine provided by Zimmerman (2000). The routine follows the method of McNab (1989) for quantifying the local convexity of concavity of a terrain surface by comparing elevations at each pixel to the mean elevations in neighbouring pixels. Positive terrain shape values indicate locally concave deviations from the surrounding terrain (e.g. a ravine), while negative values indicate locally convex topography (e.g. a ridge or hummock) (McNab, 1989). Plains and homogeneous slopes exhibit values close to zero (McNab, 1989). This measure is strongly correlated with net curvature and the values are treated as unitless.

The Zimmerman (2000) routine calculates the terrain shape for different spatial scales using the following calculation:

$$tsi_x = dem - focalmean(dem, circle, radius(x)) \quad (2)$$

In other words, the terrain shape at scale x is equal to the difference between the elevation values at each cell on the DEM minus the mean elevation of the cells in a circular window of size x. The calculation is repeated for each radius value supplied and the terrain shape values for the spatial scale providing the most extreme deviations from a homogenous surface are retained (Zimmerman, 2000). The index of terrain shape is thought to be associated with the soil moisture regime and accumulation of leaf litter (McNab, 1989; McNab and Loftis, 2002).

## **Topographic Wetness Index (TWI)**

An index of topographic wetness was obtained using the TauDEM (Terrain Analysis Using Digital Elevation Tools) toolbar plug-in (Tarboton, 2008) for ArcGIS. Several steps are involved in the production of a topographic wetness grid, the first of which entails the filling of pits (or sinks) in the DEM. Pits are cells in the DEM that are completely surrounded by cells with higher elevation and therefore have an undefined drainage direction (Tarboton, 2008; ESRI, 2006). Pits are generally considered to be artefacts that disrupt the flow of water across the DEM so they are removed by raising their elevation to match the boundary cell with the lowest elevation (Tarboton, 2008; ESRI, 1996).

This filled elevation grid was then used in the calculation of the flow direction grid and slope grid. Both grids were calculated using the  $D_{\infty}$  method of Tarboton (1997) in which the flow direction is assigned based on the steepest downward slope among eight triangular facets. This method allows the flow direction to be proportionally allocated to one or two down-slope elements in the cardinal and/or diagonal directions. The  $D_{\infty}$  algorithm is also used in calculating the accumulating area upslope of each location to produce a contributing catchment area grid.

The wetness index function in TauDEM takes the ratio of the slope and contributing catchment area. This function is the inverse of the commonly used wetness index of Beven and Kirkby (1979) (see equation 1). The contributing catchment area is used as the denominator to avoid divide by 0 errors when slope is 0 (Tarboton, 2008).

$$twi = \ln(a/\tan\beta) \quad (3)$$

where  $a$  is local upslope area draining through a certain point per unit contour length and  $\tan\beta$  is local slope gradient

## **TOPEX (distance-limited)**

TOPEX (TOPographic Exposure) is a measure of the topographical modification of local wind speeds and is calculated by summing the angles from a given location to the visible horizon for the eight main compass directions (Wilson, 1984). Hannah et al. (1995) modified this original measure to permit negative TOPEX values and limit the distance over which TOPEX was

calculated, thus reducing the overestimation of the shelter provided by distant hills and differentiating between hills and plateaus (Quine and White, 1998). A TOPEX raster was generated using a script developed by the Windthrow Research Group at UBC, which uses the method of Hannah et al. (1995), to calculate distance limited topographic exposure. As in the study by Bian and Walsh (1993), a distance limit of 3-km was chosen for the study because this approximates the distance between the mountain ridges.

### **Distance to Coast**

To create a map of distance to coast, a layer detailing the linear coastlines of British Columbia was downloaded from the CRGB. The Euclidean distance tool in the Spatial Analyst toolbox was then used to obtain a raster containing the measured distance from each cell to the coast in meters. The distance to coast measure reflects the degree to which an area is likely influenced by the moderating effects of the ocean. Coastal locations have narrower annual and diurnal temperature ranges than continental regions (Snow, 2005). Continental climates also have relatively low annual precipitation amounts compared to the maritime regions (Snow, 2005).

### **Summary Statistics**

In order use the above topographic variables in the multivariate analysis, mean values were calculated for each of the digitized polygons. However, in assigning a single value to represent a polygon there is a loss of information and the “representativeness” of this value is questionable, particularly in cases where the polygons are neither small, homogeneous nor particularly compact (de Smith et al., 2009). To gain more information about the “topographic conditions” of the polygons, statistics based on the first four moments of the elevation and slope distribution were calculated, two variables that are hypothesised to hold an important role in controlling where yellow-cedar decline occurs (see table 4). These statistics have been determined to be useful for terrain characterization in previous studies (Evans, 1998; Guth, 2006). The free, open-source program GRASS-GIS (GRASS Development Team, 2010) was used to calculate these statistics.

One additional variable was calculated to further characterize the topography within the polygons. This descriptor is the “elevation relief-ratio” which expresses the relative proportion of upland to lowland within a sample region (Rasemann et al., 2004). The formula was developed by Wood and Snell (1960), and is defined as:

$$E = ([AvgZ - MinZ]/[MaxZ - MinZ]) \quad (4)$$

where  $Z$  = elevation in meters above sea level.

## 2.3 Statistical Analysis and Modelling

### 2.3.1 Exploratory Analysis

Differences between the distribution of forest stands with a leading yellow-cedar component and all stands with a yellow-cedar component were first visually determined by plotting the relative areas of the stands against topographic variables (elevation, slope and aspect). Similar plots were then created for healthy and declining yellow-cedar stands to visually assess the relations between decline presence and topographic variables. The non-parametric Mann-Whitney U test was used to determine whether the “all yellow-cedar” stands and “leading yellow-cedar” stands, and the healthy and declining stands, are significantly different in terms of elevation and slope.

To evaluate differences between samples in terms of aspect, circular statistics were used to assess the distributions of the stands. For each sample the mean vector was calculated as the mean of all aspects. The mean aspect has two properties: direction (the mean angle,  $\bar{\alpha}$ ) and length ( $r$ ) (Zar, 1984). The parameter  $r$  is a unitless measure of concentration ranging from 0 (high dispersion) to 1 (high concentration) (Zar, 1984). Angular deviation was also calculated to measure dispersion and it ranges from a minimum of zero to a maximum of  $81.03^\circ$  (Zar, 1984). The nonparametric Rayleigh test (Batschelet, 1981) was used to assess whether the sampled populations are uniformly distributed around a circle (implying no mean direction).

### 2.3.2 Logistic Regression Analysis

To test the strength of the associations between yellow-cedar decline and each of the predictor variables created, logistic regression was used to analyze the data. Logistic regression is a form of generalized linear modeling that is appropriate for data sets where the outcome is binary ( $Y = 0$  or  $1$ , e.g. decline is absent or present) or ordinal (e.g. high severity, low severity, absent). The logistic model specifies that the probability or likelihood of an event is related to a set of variables  $[x_1 \ x_2 \ \dots \ x_k]$  in the following way:

$$E[Y|x] = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}} \quad (5)$$

where  $E[Y|x]$  represents the conditional mean of  $Y$  given  $x = [x_1 \ x_2 \ \dots \ x_k]'$ ,  $\beta_0$  represents the unknown intercept parameter, and  $\beta = [\beta_1 \ \beta_2 \ \dots \ \beta_k]'$  are the slope parameters that represent the effect of the predictor variables on the probability of an event (Elswick et al., 1997).

Applying the logit transformation (the log of the odds that an event occurs) the model takes the form:

$$\ln\left(\frac{E[Y|x]}{1 - E[Y|x]}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (6)$$

which is often written as:

$$\text{logit}(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (7)$$

Thus the logistic regression is linear in the logit, but when transformed back from the logit to the probabilities, the relation between  $x$  and the probabilities is sigmoidal (Figure 10) (Menard, 2001).

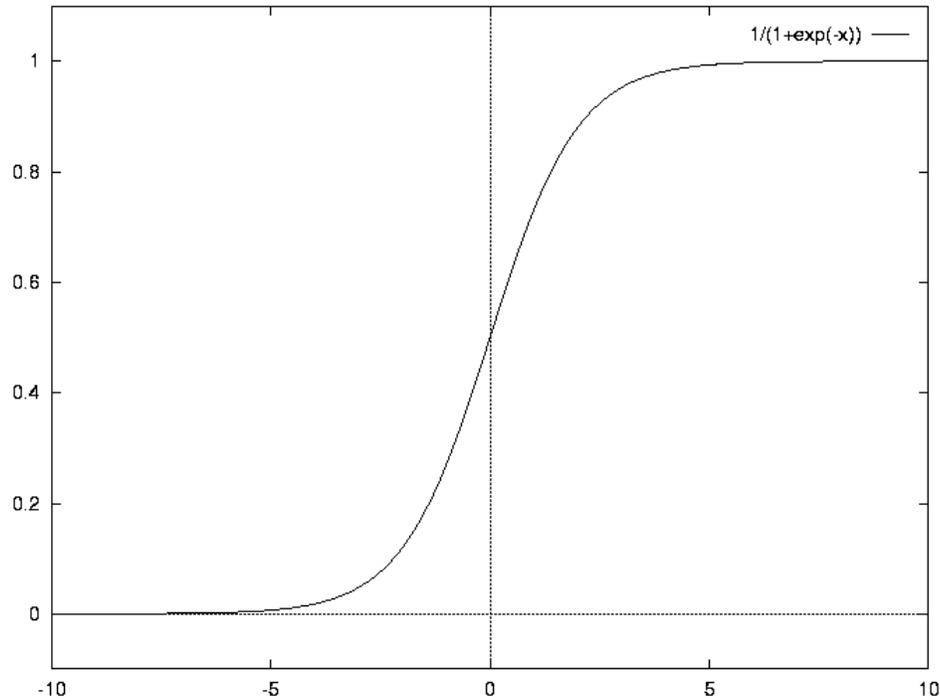


Figure 10: Shape of logistic model

In order to find the parameter estimates for the logistic regression equation, the maximum likelihood estimation (MLE) must be used when one or more of the predictor variables are measured on the interval or ratio scale (Menard, 2001). The maximum likelihood techniques are used to identify the set of parameters that maximize the probability (log-likelihood function in logistic regression) of the observed values of Y, given the values of the predictor variables and parameters  $\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$  (Menard, 2001). The estimation is an iterative process involving computer-implemented numerical algorithms that provide parameter values once the solution converges.

Prior to fitting the logistic model, collinearity between predictor variables was assessed using Pearson's product moment correlation coefficient. When a pair of predictor variables was found to be strongly associated ( $r > 0.7$ ), only one was used in the model fitting procedure. A full model was then fit in R (R Development Core Team, 2010) using the lrm function in the Design package (Harrell, 2009). Assumptions of linearity were assessed using the Box-Tidwell test and,

for those predictor variables determined to exhibit nonlinearity in the logit, quadratic forms were included. Interaction terms were then created and assessed for significance.

Automated stepwise procedures such as those found in multiple regression are often used for selecting predictor variables for inclusion or removal from a logistic regression model (Hair et al., 1998; Menard, 2001; Vayssières et al., 2000). However, several criticisms concerning automated stepwise selection of variables have been made (Harrell, 2001; Menard, 2001; Reineking and Schroder, 2006; Steyerberg et al., 1999; Steyerberg et al., 2001; Whittingham et al., 2006). Computer-controlled stepwise selection procedures will attempt to identify the most parsimonious model for the given data set by drawing on the random variations in the data and ultimately produce results that are particular to the dataset and thus difficult to replicate in other data samples (Menard, 2001). This greedy nature of the selection techniques means that they are numerically unstable and often result in model overfitting (Steyerberg et al., 1999). Penalized estimation techniques, such as lasso regression, address these issues by selecting variables in a less greedy fashion than forward selection/backward deletion (Park and Hastie, 2006; Steyerberg, 2009).

The lasso (“least absolute shrinkage and selection operator”) technique imposes a constraint on the sum of the absolute values of the parameter estimates, shrinking some coefficients to zero (Tibshirani, 1996). By excluding the predictor variables that have a regression coefficient of zero, the lasso effectively performs a form of variable selection. To estimate the regression coefficients of standardized coefficients, the lasso minimizes the log-likelihood subject to  $\sum_j |\beta_j| \leq t$ , where  $t \geq 0$  is a tuning parameter which determines the amount of shrinkage applied the model (Steyerberg, 2009; Tibshirani, 1996). The lasso parameter is scaled so that it varies over a grid between 0 and 1 ( $s = t / \sum_j |\beta_j|$ ), and  $s$  may be interpreted as a standardized shrinkage factor (Steyerberg, 2009). At  $s = 1.0$ , the constraint has no effect and the lasso solution is just the maximum likelihood estimates (Steyerberg et al., 2001). Decreasing the value of  $s$  results in shrinkage of the solution towards 0, and some coefficients may be set exactly equal to 0 (Tibshirani, 1996).

The lasso can be used for generalized linear models such as the logistic or Cox model. In this study the lasso implementation of logistic regression was applied using the lasso predictor-corrector algorithm of Park and Hastie (2006), available through the R package `glmnet` (Park

and Hastie, 2007). The predictor variables were first standardized to have mean of 0 and variance of 1. The optimal tuning parameter was estimated by studying the BIC (Bayesian Information Criterion). The selected value of the regularization parameter was then used to select a set of coefficients. According to Hastie et al. (pp.91, 2009), the estimates of the non-zero coefficients provided by the lasso shrinkage technique are biased towards zero, and the estimates don't necessarily converge to the true values as the sample size grows. This bias was reduced by following the recommendation of Hastie et al. (pp. 91, 2008), which involves using the lasso to identify the set of non-zero coefficients, and then fitting an unrestricted linear model to the selected set of features.

Having chosen a final model, the overall goodness of fit was assessed using the Hosmer-Le Cessie test (Hosmer et al., 1997) and the rescaled  $R^2$  of Nagelkerke (1991). The Nagelkerke measure is a pseudo- $R^2$  for logistic regression that is rescaled to the expected 0-1 range, with higher values indicating better model fit.

The predictive accuracy of the model was then assessed using the area under the curve (AUC) of the receiver-operating characteristic (ROC) plot, and associated measures of sensitivity and specificity. In order to get binary outcomes ( $Y = 1$  or  $0$ ) from the logistic model, a probability threshold must first be applied. For any threshold level (e.g. probability = 0.5), it is possible to compute two measures of prediction success, the *sensitivity* and the *specificity*. The sensitivity is the proportion of true "positives" correctly predicted, and the specificity is the proportion of true "negatives" correctly predicted. The ROC curve is a plot of the proportion of true positives against the proportion of false positives (the complement of the specificity) at a series of thresholds for a positive outcome (Rossiter and Loza, 2008). A model would ideally have high sensitivity and specificity, even at low thresholds, so the plotted curve would lie close to the y-axis and approach the top border of the graph (Rossiter and Loza, 2008) (Figure 11). A less accurate model would lie closer to the diagonal line, which represents the random case in which the chance of a true positive is equal to that of a false positive at any threshold (Rossiter and Loza, 2008) (Figure 11). The ROC curve can be summarized using the area under the curve (AUC). The optimum threshold value for a model (predicts most of the true positives with few false positives) can be identified from the ROC plot and applied to obtain binary outcomes.

Resulting prediction accuracy for a threshold value can then be assessed in terms of commission and omission errors displayed in a confusion matrix.

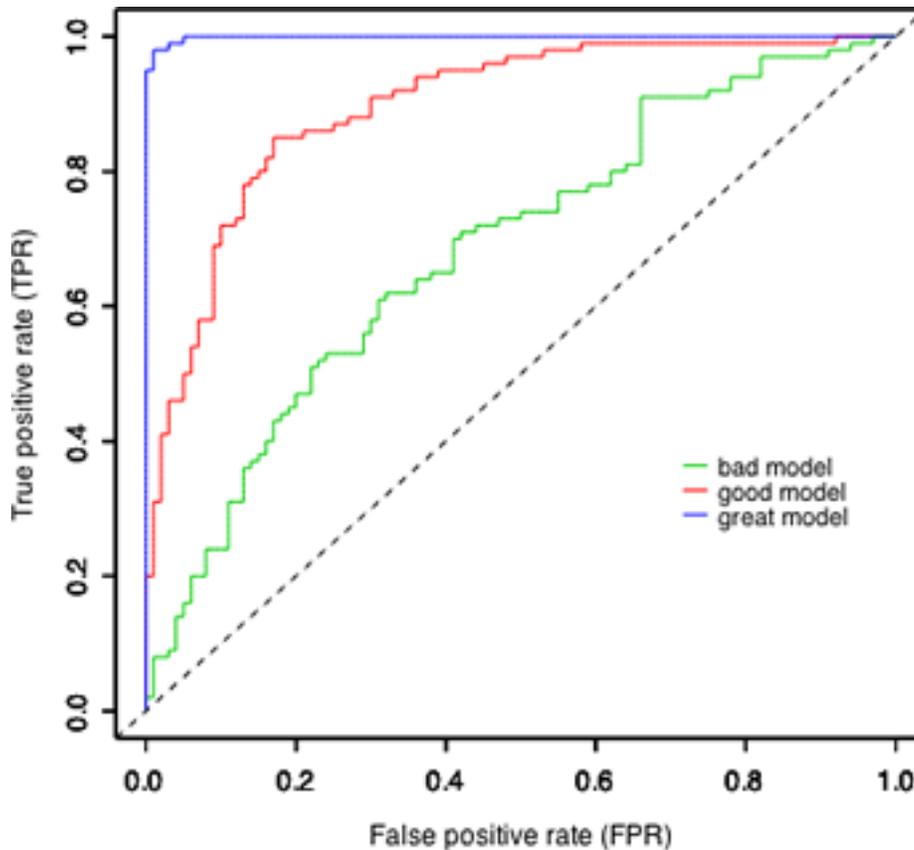


Figure 11: Receiver-Operating Characteristic (ROC) curve

Due to the small size of the data set, it was not possible to split the data into a training set and validation set and test set, as is often done in data rich situations. Instead, the variable selection and estimated coefficients were validated through a bootstrap procedure.

Following selection of the optimal model, the fitted relation was projected into geographical space in order to produce a probability surface for the decline of yellow-cedar. Using the Marine Geospatial Ecology Tools (MGET), an open source geoprocessing toolbox that connects ArcGIS with the statistical package R, maps of the probability of decline occurrence and binary presence/absence maps were produced.

### 2.3.3 Analysis of Residuals

Residual analysis was used to confirm that the model fits the data well across the entire set of observations and to identify outliers and cases that potentially have a large influence on the parameters of the model (Menard, 2001). Having addressed assumptions of linearity and additivity in the model building process, the final step was to address the assumption of independence of observations. While the spatial dependency exhibited by environmental variables such as elevation, topographic wetness, temperature, etc. may account for much of the spatial patterning observed in yellow-cedar decline, remaining spatial dependence could arise from unmeasured or misspecified variables or processes (Miller, 2005). So although spatial autocorrelation can be exploited as a source of information on patterns, structure and processes (Franklin, 1995; Gould, 1970), statistical issues can also arise if it is not accounted for in the model (Legendre, 1993).

An empirical semi-variogram was plotted to test for autocorrelation in the Pearson residuals of the logistic model. Semi-variograms provide a quantitative measure of spatial correlation between two near-by values. The semi-variance is calculated by averaging half the squared difference between pairs of observations separated by the so-called lag distance (Kleinschmidt et al., 2000). When spatial autocorrelation is present, semi-variance values will be smaller for low lag distances, indicating that values that are close together in space are more alike (less variable) than those further apart (Kleinschmidt et al., 2000).

### 2.3.4 Classification and Regression Trees

Decision trees (also known as Classification and Regression Trees, or CART) provided a more visually intuitive means of modeling the data. Decision tree models use a non-parametric, probabilistic machine-learning method, which generates a set of rules to classify a response variable based on the values of the predictor variables (Franklin, 1995). The models provide a useful means of elucidating structure in the data and expressing complex relations that are non-linear, non-additive or hierarchical in nature (Elith et al., 2006; Miller and Franklin, 2002).

The rpart package (Therneau and Atkinson, 2010) in R was used to build a classification tree. The original tree was grown with a minimum node size of 5 samples as the stop value. To achieve a more parsimonious model, this tree was pruned back to a size that minimized cross-validated error. According to a rule suggested by Breiman et al. (1984), the best tree was deemed to be the smallest tree such that its estimated error rate was within one standard error of the minimum. This tree was then validated through 10-fold cross-validation using the validate.tree function in the Design package (Harrell, 2009). The validation was performed to assess the degree of variation in the sizes of the best tree and whether the selected tree is atypical.

# Chapter 3

## Results

### 3.1 Yellow-Cedar Distribution

Visual examination of the distribution of leading yellow-cedar stands and comparisons with the distribution of all stands with a yellow-cedar component provided information for the stratified sampling design (Figures 12 – 14) (in the figures, the data were normalized to show percent of stand area rather than number of stands). The distribution of “all yellow-cedar” stands among elevations was significantly different from the distribution of the “leading yellow-cedar” stands ( $p < 0.01$ , Figure 12). The greatest proportion of stands with a leading yellow-cedar component occurs at higher elevations, with the 500 – 1000 m range containing the majority (c.70%) of the sampled area. The distribution of “all yellow-cedar” stands is less peaked, with approximately 50% of the total area occurring in the 500 – 1000m range and 30% in the 0 – 400 m range. For both “all yellow-cedar” and “leading yellow-cedar” stands, only a very small percentage of the total area occurs above 1200 m (Figure 12).

Although the slope angles of majority of “all yellow-cedar” and “leading yellow-cedar” sites occur within the 24°- 40° range, the distributions among slope angle classes were significantly different ( $p < 0.001$ , Figure 13). A greater proportion of the “all yellow-cedar” group occurs on the sites with slope angles  $< 24^\circ$ , while the “leading yellow-cedar” group dominates in the steeper slope classes (Figure 13).

Similarities were observed between the mean aspect values of “all yellow-cedar” and “leading yellow-cedar” stands. The majority of both “all yellow-cedar” and “leading yellow-cedar” stands occur on slopes that have an aspect between 90 – 270° (east – west). Only a small proportion of the total area of “all yellow-cedar” stands occur on northeast facing slopes and less than 2% of the total area of “leading yellow-cedar” stands occur on this slope aspect (Figure 14). For both “all yellow-cedar” stands (mean angle = 205°,  $z = 120.02$ ,  $r = 0.11$ ,  $n = 9522$ ,  $p < 0.05$ ) and “leading yellow-cedar” stands (mean angle = 190°,  $z = 54.41$ ,  $r = 0.13$ ,  $n = 3448$ ,  $p < 0.05$ ) the

Rayleigh test was significant, indicating that the distributions departed nonrandomly from uniformity. However, the Rayleigh's test assumes a unimodal direction, which is not observed in Figure 14 (Zar, 1984). The lack of a unimodal direction combined with the low  $r$  values and high angular deviation values (c.  $75^\circ$  for both “all yellow-cedar” and “leading yellow-cedar”) indicate that the results are unreliable due to the low concentration of the data (Morellato et al., 2010).

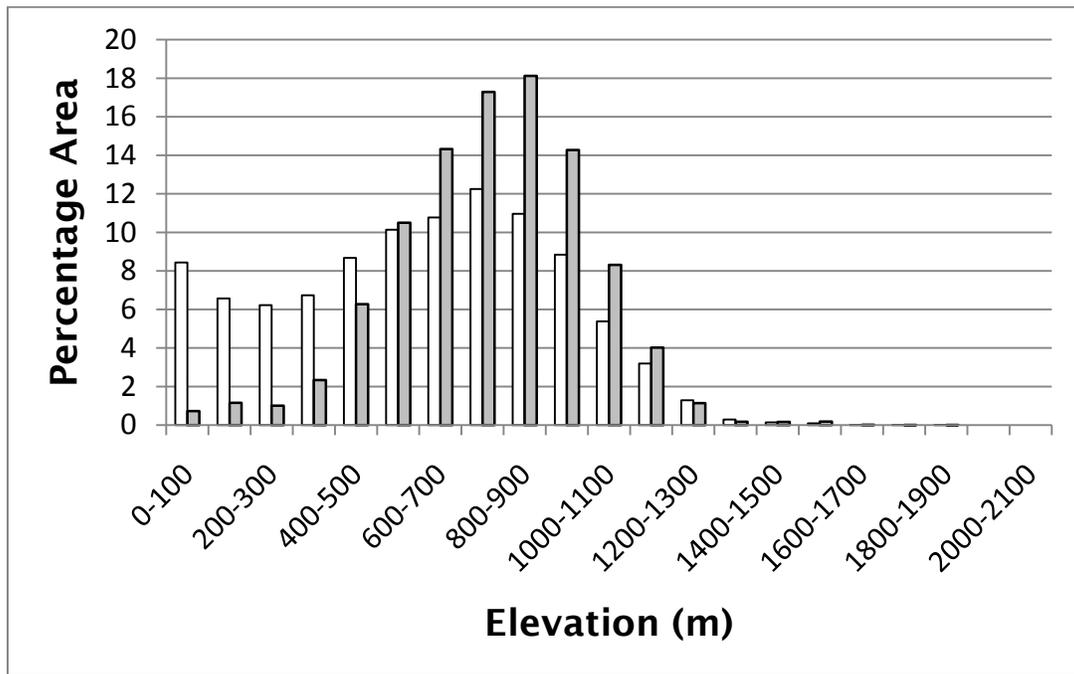


Figure 12: Mean elevation values of “all yellow-cedar” stands (white) and “leading yellow-cedar” stands (grey) (normalized by area). (Statistically significant difference between samples,  $p$ -value =  $<0.001$ )

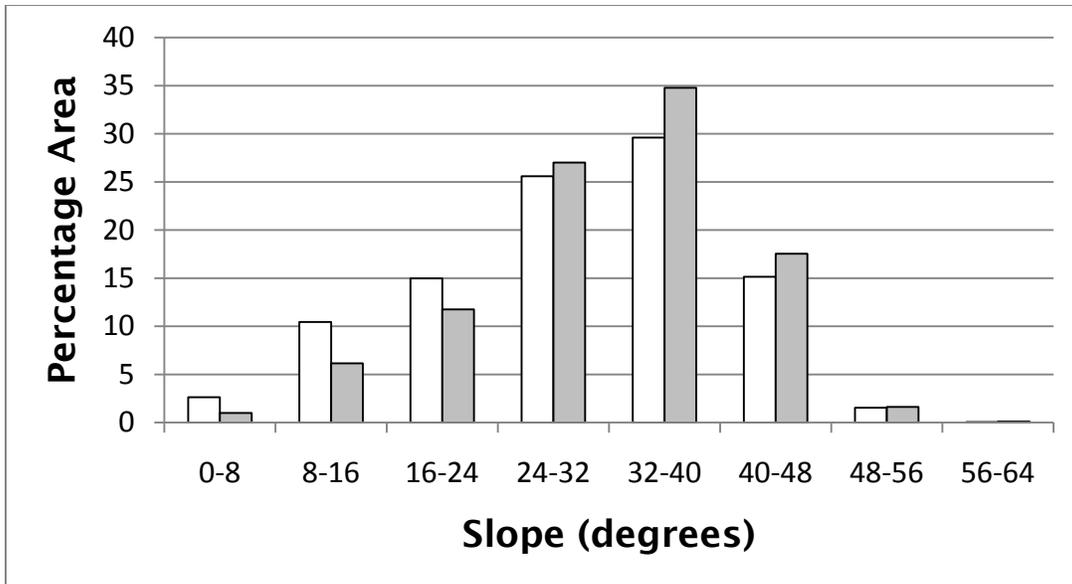


Figure 13: Mean slope values of “all yellow-cedar” stands (white) and “leading yellow-cedar” stands (grey) (normalized by area). (Statistically significant difference between samples,  $p$ -value =  $<0.001$ )

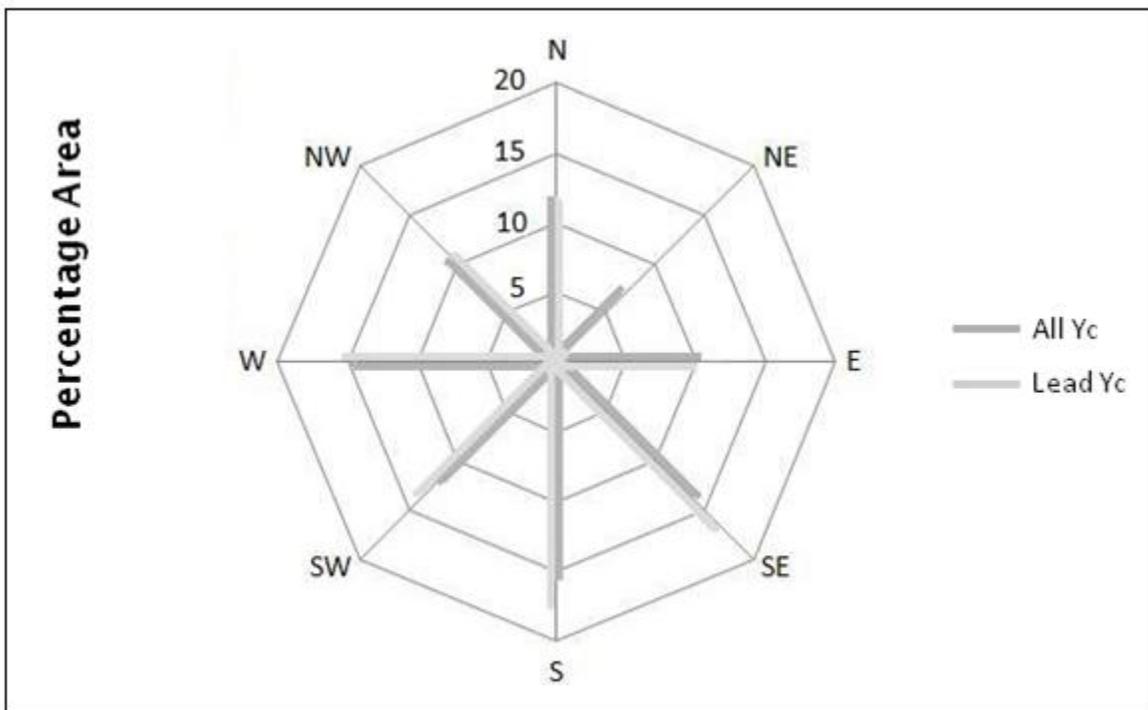


Figure 14: Mean aspect values of “all yellow-cedar” stands (dark grey) and “leading yellow-cedar” stands (light grey) (normalized by area)

## 3.2 Exploratory Analysis of Declining and Healthy Yellow-Cedar Stands

Comparisons of the distributions of declining yellow-cedar stands and healthy yellow-cedar stands reveals that the two samples exhibit differences in terms of elevation ( $p < 0.001$ ), slope ( $p < 0.006$ ) and aspect (Figures 15 – 17) (in the figures, the data were normalized to show percent of stand area rather than number of stands). Dieback is most common at lower elevations with more than 90% of decline area sampled occurring at 200-700m. The majority (>75%) of the area of healthy yellow-cedar stands sampled occurs at elevations above 800 m (Figure 15). A greater proportion of the declining stands occur on gentler slopes compared to healthy stands. While approximately 70% of the total area of both sampled declining and healthy areas fall on slopes of 24 – 40°, approximately 25% of total decline area occurs on slopes 16-24°, whereas over 20% of total healthy yellow-cedar area sampled occurs on slopes >40° (Figure 16).

The majority of the sampled decline area occurs on slopes that have southwest-west facing aspects (c. 55% of total decline area sampled has a mean aspect value between 202.5 - 292.5°). Very little decline occurs on northeast – south facing aspects (<20 % total sampled decline area falls on slope aspects from 22.5 - 202.5°). The majority of healthy yellow-cedar stands, in comparison, occur on south southeast facing slopes (c. 50% of total area of sampled healthy yellow-cedar occurs on slope aspects from 112.5 - 202.5°) (Figure 17).

For both declining stands (mean angle = 274°,  $z = 1.41$ ,  $r = 0.14$ ,  $n = 74$ ,  $p > 0.05$ ) and healthy stands (mean angle = 175°,  $z = 1.19$ ,  $r = 0.19$ ,  $n = 34$ ,  $p > 0.05$ ) the Rayleigh test was not significant, indicating that the samples have uniform circular distributions. However, the low  $r$  values and high angular deviation values (c. 75° for declining yellow-cedar and 73° healthy yellow-cedar) again indicate that the results are unreliable due to the low concentration of the data (Morellato et al., 2010). The distributions of the healthy and declining stands are clearly not uniform (Figure 17).

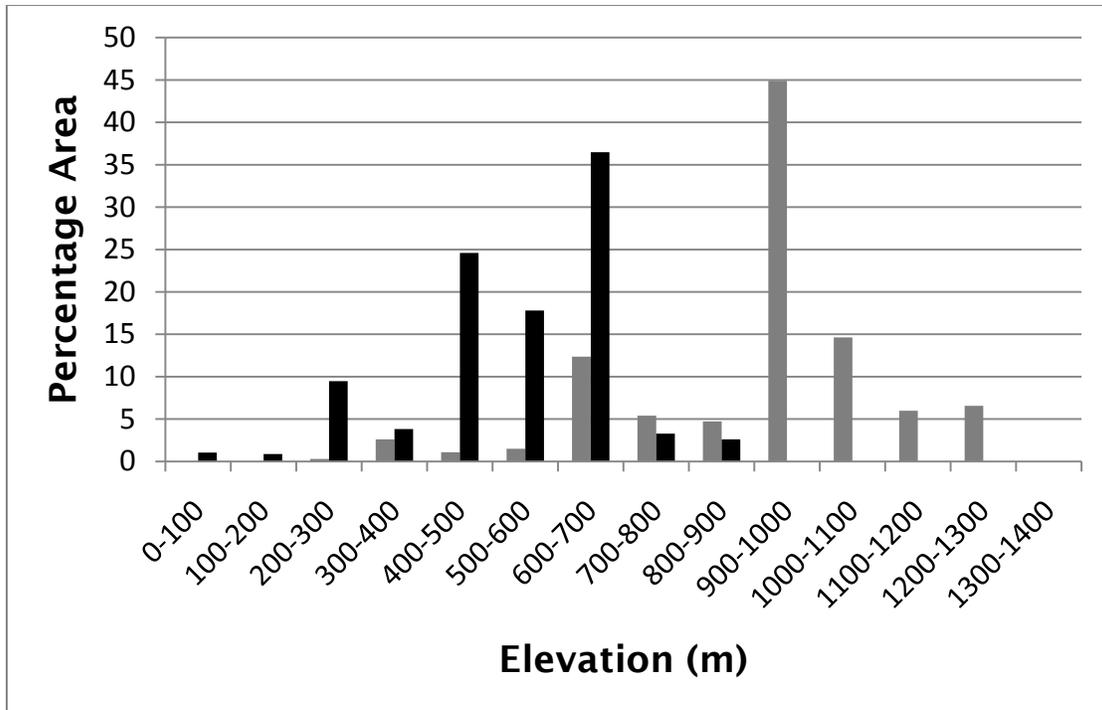


Figure 15: Mean elevation values of declining (black) and healthy (grey) yellow-cedar stands (normalized by area). (Statistically significant difference between samples,  $p\text{-value} = <0.001$ )

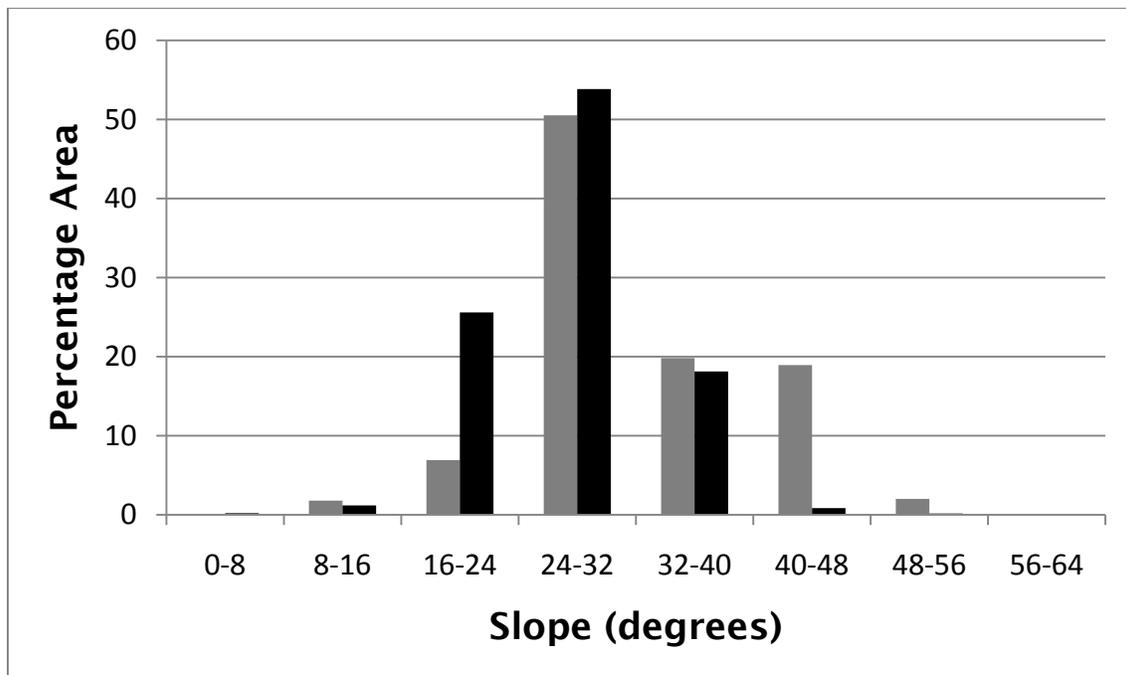


Figure 16: Mean slope values of declining (black) and healthy (grey) yellow-cedar stands (normalized by area). (Statistically significant difference between samples,  $p\text{-value} = <0.006$ )

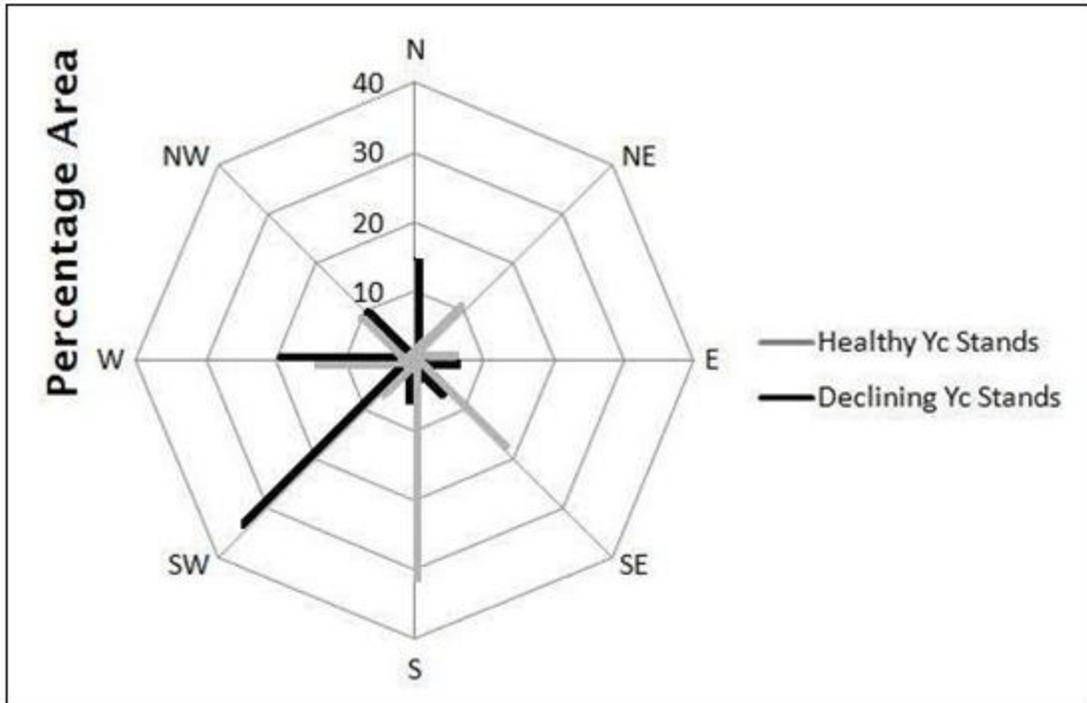


Figure 17: Mean aspect values of declining (black) and healthy (grey) yellow-cedar stands (normalized by area).

### 3.3 Logistic Regression

Prior to building the multivariate logistic regression model, univariate models were built for each of the predictor variables and each variable was assessed for non-linearity in the logit (Table 4). Six variables were excluded from subsequent model building after non-linearity was detected and no adequate transformation was found (Table 5). All of the excluded terms had an  $R^2$  value of less than 0.033 and p-values  $> 0.1$  (Table 5). In the model, squared terms were included for mean elevation, mean topographic wetness index and mean slope to account for observed non-linearities and possible unimodal response curves. A correlation matrix of the 12 predictor variables showed that Pearson correlation coefficients for all pairs of variables were  $< 0.7$  and, thus, all were included in the model. After building a full model with these 12 variables, 3 multiplicative (interaction) pairwise terms based on biophysical principles were built and tested for significance. The interaction terms consisted of the following: elevation\*aspect,

elevation\*distance from coast, and exposure\*distance from coast. However, none of these terms proved to be candidates for inclusion in the final model as all had p-values >0.05.

Table 5: Univariate logistic regression model results

<b>Variable</b>	<b>d2</b>	<b>R<sup>2</sup></b>	<b>p</b>	<b>c</b>	<b>Dxy</b>	<b>Non-linear</b>
<b>Elev_mean</b>	0.3798	0.529	0	0.863	0.726	✓
<b>Elev_mean_sq</b>	0.4282	0.58	0	0.863	0.726	
<b>Coastdist_mean</b>	0.3272	0.47	0	0.873	0.746	
<b>Elev_var</b>	0.1159	0.189	0.001	0.723	0.446	
<b>Tpx3000_mean</b>	0.0993	0.163	0.003	0.696	0.392	
<b>Slope_mean</b>	0.061	0.103	0.0042	0.662	0.324	
<b>Slope_mean_sq</b>	0.0662	0.111	0.0028	0.662	0.324	
<b>Slope_stddev</b>	0.0439	0.075	0.015	0.64	0.281	
<b>Tw_i_mean</b>	0.0409	0.07	0.19	0.606	0.213	✓
<b>Tw_i_mean_sq</b>	0.0625	0.105	0.0037	0.607	0.214	
<b>Solarspring_mean</b>	0.0192	0.033	0.1081	0.584	0.169	✓
<b>Slope_krt</b>	0.0163	0.028	0.1814	0.545	0.089	✓
<b>Slope_skw</b>	0.0142	0.025	0.1722	0.551	0.103	✓
<b>Tsi_mean</b>	0.0119	0.021	0.206	0.581	0.161	✓
<b>Aspectsw_mean</b>	0.008	0.014	0.3001	0.546	0.091	
<b>Elev_relief</b>	2.00E-03	0.004	0.6006	0.529	0.057	
<b>Elev_krt</b>	9.00E-04	0.002	0.7276	0.476	-0.048	✓
<b>Elev_skw</b>	2.00E-04	0	0.8758	0.493	-0.014	✓

Note: Greyed out rows indicate variables that were excluded from future model building.

d2 = proportion of null model deviance explained

R<sup>2</sup> = Nagelkerke R2 index

p = p-value (parameter significance)

c = c index (area under ROC curve)

D<sub>xy</sub> = Somers' Dxy rank correlation between a variable x and a binary (0-1) variable y

### 3.3.1 Lasso Model

Following identification of variables appropriate for the modeling procedure, the lasso logistic regression path was trained over all of the data. The `glm` package in R uses an algorithm that implements the predictor-corrector method to estimate the coefficients at important junctions as the size of the penalty varies. These important junctions are points in the coefficient path at which variables enter or leave the active set. In Figure 18 the coefficient paths achieved are plotted, with vertical lines marking the points at which the active set is modified. Changes in the active set of variables as the size of the penalty ( $\lambda$ ) is varied are shown in Figure 19. The optimum penalty was estimated by studying the BIC estimated for each junction (Figure 20). The lowest BIC (81.97) suggests an optimal selection of 8 predictors (step 9, penalty = 0.72). However, in the spirit of parsimony, the next lowest BIC value was selected (82.64), as it is very close to the smallest value. This BIC value suggests a selection of 5 predictors (step 6, penalty = 4.28). At this step, topographic variables with non-zero coefficients include: mean elevation squared, mean distance from coast, variance in elevation, mean TOPEX and standard deviation in slope.

The selected set of coefficients was then validated through a bootstrap analysis using the BIC (Table 6). Of the variables selected, three had non-zero coefficient values for >97 % of bootstrap samples (mean elevation squared, mean distance from coast and variance in elevation). Mean TOPEX had a zero coefficient value in 9.2% of samples, and standard deviation in slope had a zero value in 25.6% of samples. The proportion of samples with zero value coefficients varied from c. 20% to c. 60% for the other variables.

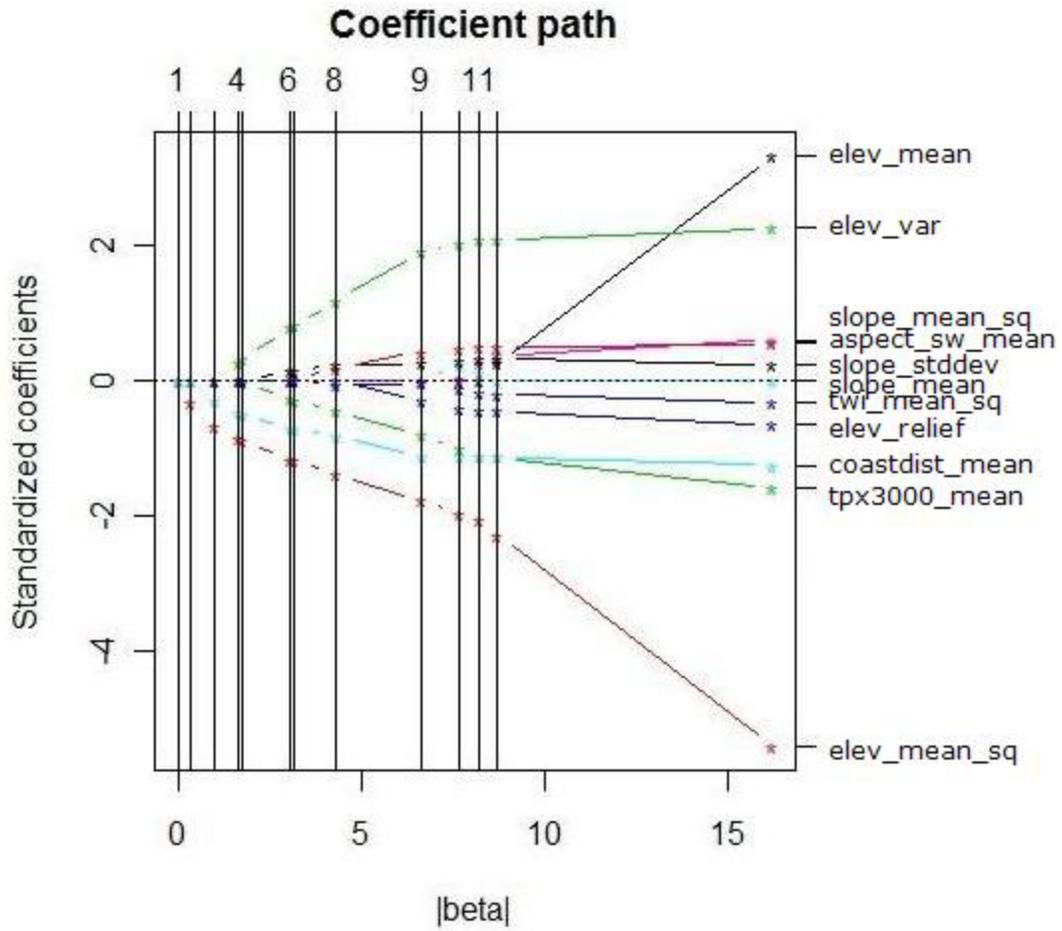


Figure 18: Size of regularized regression coefficients as the absolute sum of regression coefficients increases. Each curve represents a regularized coefficient (labelled on the right) according to the absolute sum of the regression coefficients. Vertical lines indicate where the active set was modified (in this case, where a variable entered the model)

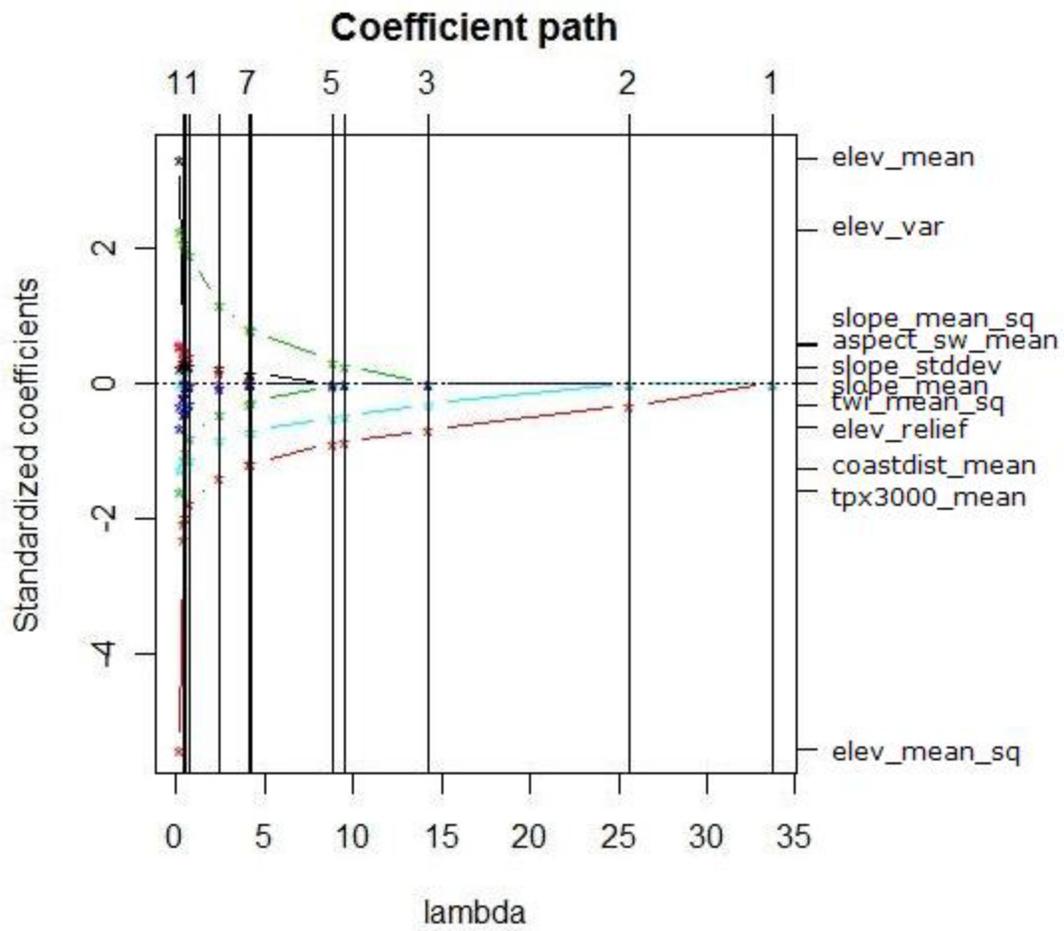


Figure 19: Lasso shrinkage of coefficients. Each curve represents a regularized coefficient (labelled on the right) as a function of the lasso penalty parameter. Vertical lines indicate where the active set was modified (in this case, where a variable leaves the model)

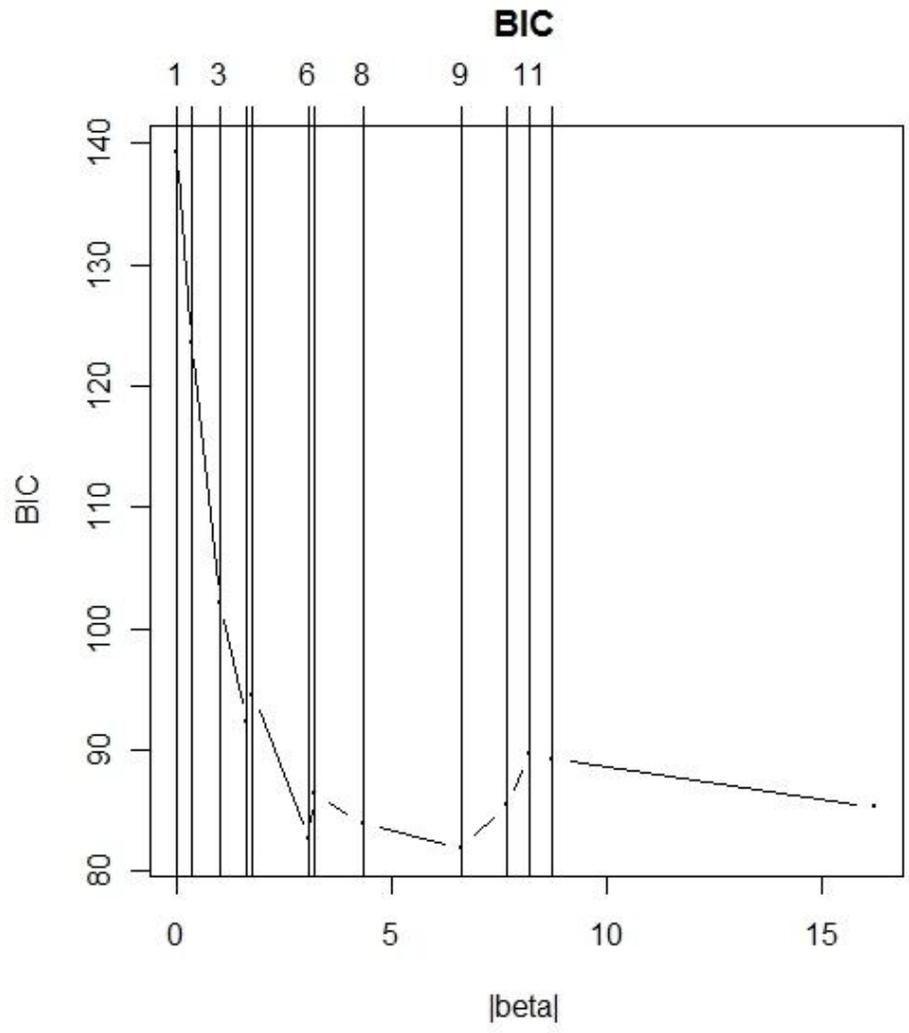


Figure 20: BIC values computed at the values of the penalty parameter where the active set changed

Table 6: Standardized coefficient estimates computed from all of the data, the mean and SE of the estimates computed from the bootstrap samples (B=1000), and the % of samples in which variable had coefficient at zero

<b>Variable</b>	$\hat{\beta}$	Mean $\hat{\beta}^b$	SE $\hat{\beta}^b$	Num. Zero / $\beta$
<b>Elev_mean</b>	0	34.387	2.008	57.4
<b>Elev_mean_sq</b>	-1.784	-46.143	2.418	0.2
<b>Elev_var</b>	1.896	15.845	0.759	2.0
<b>Elev_relief</b>	-0.312	-6.366	0.369	26.5
<b>Slope_mean</b>	0	-20.008	1.813	54.0
<b>Slope_mean_sq</b>	0	22.278	2.030	56.2
<b>Slope_stddev</b>	0.249	2.113	0.353	25.6
<b>Aspectsw_mean</b>	0.412	4.250	0.288	19.4
<b>Tpx3000_mean</b>	-0.813	-11.773	0.657	9.2
<b>Twi_mean</b>	0	7.265	1.119	51.1
<b>Twi_mean_sq</b>	-0.047	-8.451	0.914	38.8
<b>Coastdist_mean</b>	-1.125	-8.194	0.383	1.6

An unrestricted logistic model was then fit to the data, the coefficients of which are shown in table 6. The variable ‘standard deviation of slope’ was excluded from the final model as its inclusion resulted in error indicating the presence of a singular information matrix. The presence of singular estimated covariance matrix is likely caused by multicollinearity among the predictor variables (Lesaffre and Marx, 1993).

Table 7: Reported coefficients and significance values from logistic model

<u>Variable</u>	<u>Coefficient</u>	<u>S.E.</u>	<u>Wald Z</u>	<u>P</u>
<b>Intercept</b>	3.62E+00	2.17E+00	1.67	0.0953
<b>elev_mean</b>	1.67E-02	8.38E-03	1.99	0.0467
<b>elev_meansq</b>	-1.96E-05	7.35E-06	-2.67	0.0077
<b>coastdist_mean</b>	-2.44E-04	9.41E-05	-2.6	0.0094
<b>elev_var</b>	0.000178	6.11E-05	2.91	0.0036
<b>tpx3000_mean</b>	-3.38E-02	1.66E-02	-2.03	0.0421

From examining the signs of the reported coefficients (Table 7), it was determined that as elevation increases the odds of decline being present decreases, similarly the odds decrease with increasing distance from coast. There is an increase in the odds of decline presence with increasing variation in elevation and increasing topographic exposure. Therefore, the model seems to suggest that low elevation sites close to the coast, which are more exposed and have more variation in elevation, are more likely to show evidence of decline.

### 3.3.2 Model Predictive Success

In terms of goodness-of-fit statistics and predictive accuracy, the logistic model is very successful. The Nagelkerke  $R^2$  value of 0.846 indicates that a high proportion of the variation is accounted for. In addition, the Hosmer-Le Cessie omnibus test (Hosmer et al., 1997) failed to find any evidence of lack of fit at the 95 percent confidence level ( $p$ -value > 0.05). Inspecting the ROC plot reveals that the model has high predictive accuracy, the plotted line runs very close to the upper left corner and the AUC value is 0.98 (Figure 21). This value may be interpreted according to the scale proposed by Swets (1988), which suggests that AUC values above 0.90 show ‘excellent’ discriminating ability.

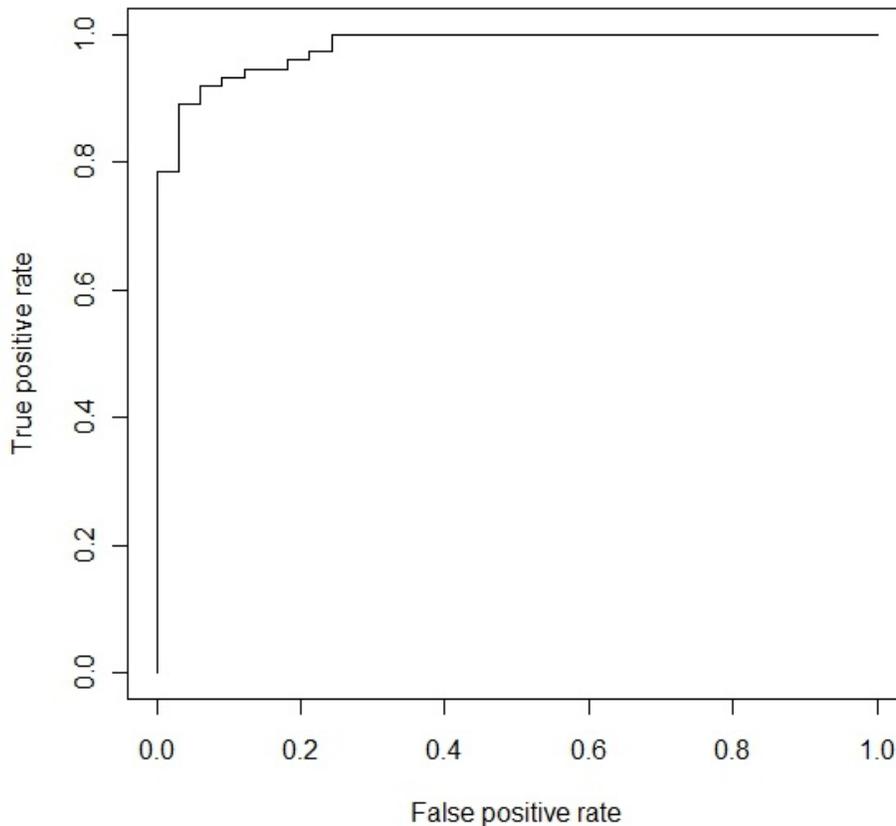


Figure 21: ROC curve for logistic model. True positive rate = sensitivity, false positive rate = specificity

While the ROC curve permits assessment of model accuracy without defining a threshold, a threshold value is required to produce a map of predicted decline presence. An optimal threshold value of 0.7 was chosen by selecting the point that maximized sensitivity and specificity simultaneously (Figure 22). The confusion matrix generated using this threshold value shows that 0.7 level results in high precision for identifying both healthy and declining stands (Table 7). A cutoff value was chosen assuming that both sensitivity and specificity are equally important because locations where yellow-cedar is healthy will likely be of almost as much interest as those where it is declining from a management perspective. As the threshold was lowered, more declining stands were correctly identified as such, but fewer healthy stands were correctly identified. The opposite was true for increasing the threshold, resulting in an underprediction of decline. However, the overall percentage of stands correctly classified was very similar over the range of threshold values examined.

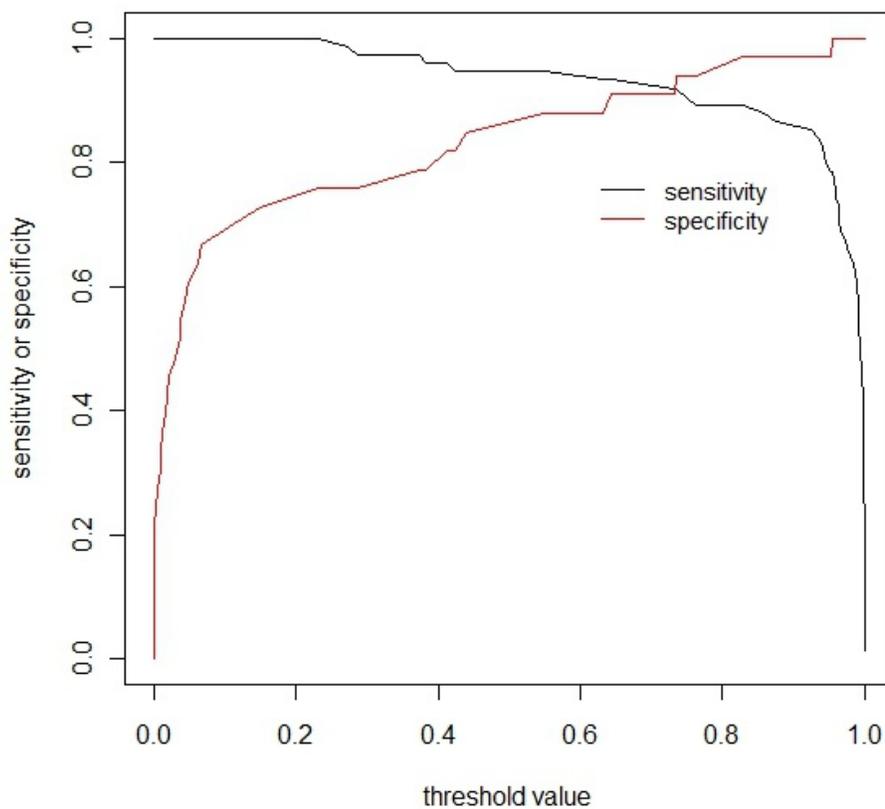


Figure 22: Sensitivity and specificity curves plotted for different threshold values

Table 8: Confusion matrix for logistic model (threshold value = 0.7)

		Predicted		
		Decline Outcome		% Correct
Observed		Healthy	Declining	
Decline Outcome	Healthy	30	3	91
	Declining	6	68	92
Overall Percentage				92

### 3.3.3. Predictive Mapping

Having selected a threshold value of 0.7, the logistic regression equation was realised in the GIS environment using the MGET geoprocessing toolbox (Roberts et al., 2010). This tool permitted predictions of decline occurrence to be made for all stands in the study area with a leading yellow-cedar component. Values of mean elevation, mean elevation squared, mean distance from coast, variance in elevation and mean topographic exposure were calculated for each of the stands and then converted to raster layers before being combined as defined by the logistic equation. The resulting map of predicted decline occurrence is shown in figure 23. A large proportion of the stands with a yellow-cedar component was predicted as having a high probability of decline occurrence (55.5% of total area) and approximately one quarter of the area was predicted as low probability (23.9% of total area) (Table 9). Binary predictions of decline occurrence using the 0.7 cutoff value are illustrated in figure 24.

Table 9: Probability of decline occurrence predicted for stands with a yellow-cedar component. Figures in parentheses represent areas within the leading yellow-cedar stands.

<b>Probability of decline occurrence</b>	<b>Percentage of total surface area</b>	<b>Area (ha)</b>
<b>0.00 – 0.20</b>	23.9 (30.3)	45741 (26316)
<b>0.20 – 0.40</b>	6.1 (8.0)	11598 (6958)
<b>0.40 – 0.60</b>	5.7 (7.2)	10966 (6222)
<b>0.60 – 0.80</b>	8.7 (10.1)	16679 (8755)
<b>0.80 – 1.00</b>	55.5 (44.4)	106188 (38508)

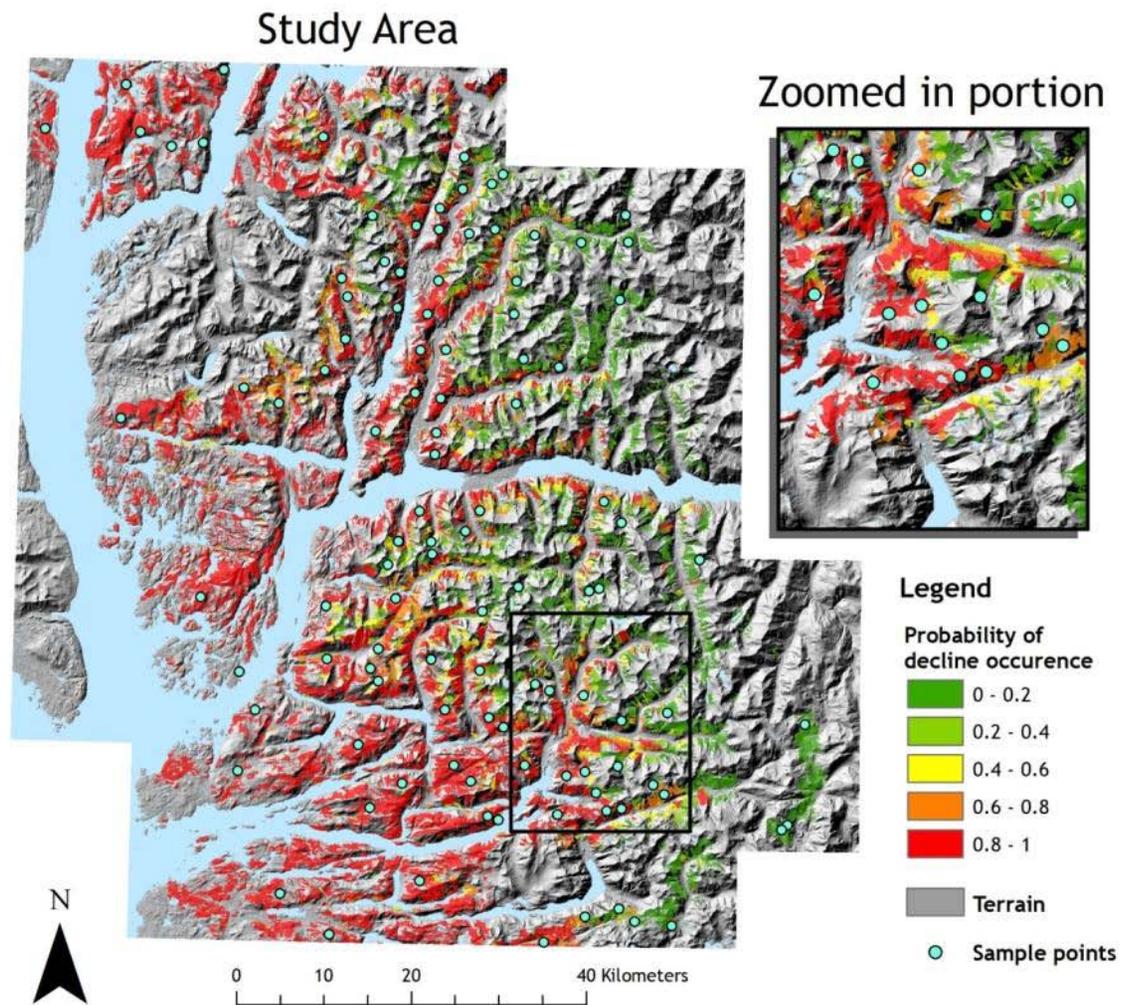


Figure 23: Map depicting likelihood of decline occurrence for yellow-cedar stands within the study area

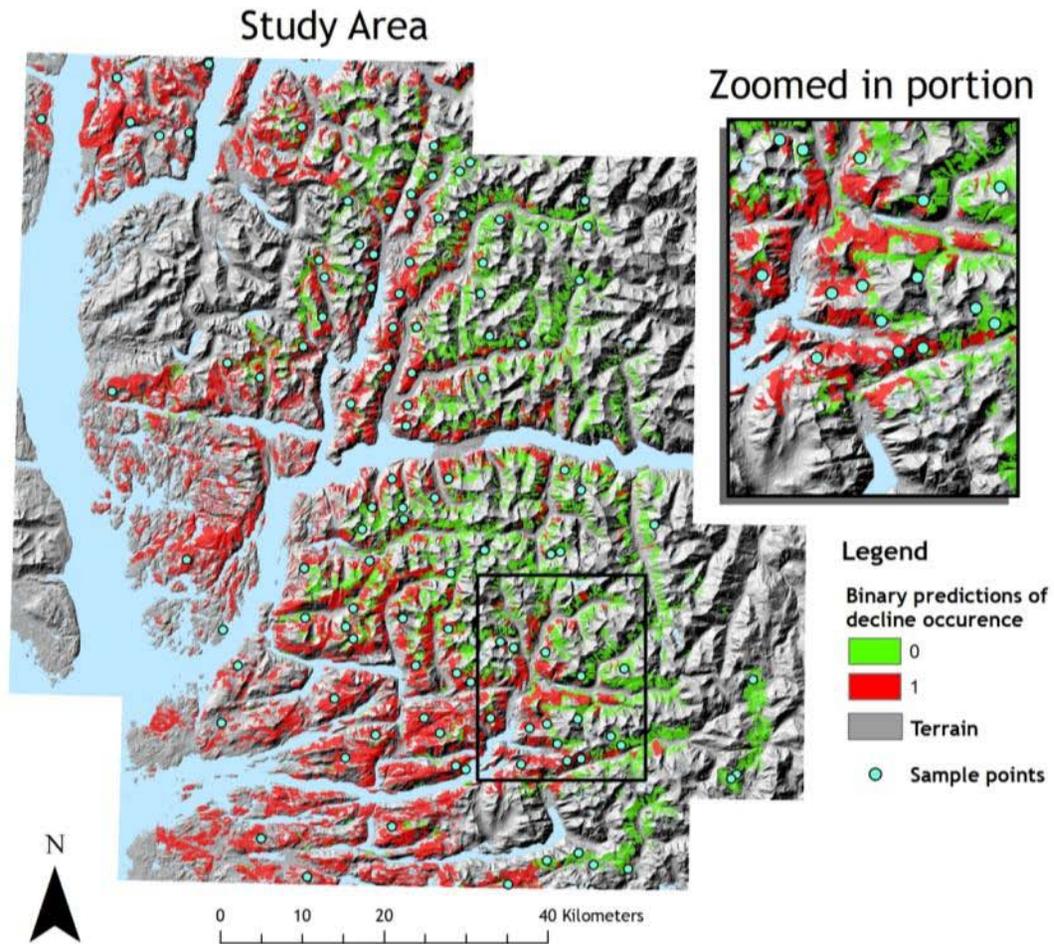


Figure 24: Map depicting binary outcomes for predicted decline occurrence using a threshold of 0.7

### 3.3.4 Analysis of Residuals

A small number of outliers and leverage points were identified when analysing the residuals; however, the removal of only one point from the model was felt to be justified due to high uncertainty concerning digitized boundaries of the sample. From examining the plot of semi-variance against distance (lag = 2000 m), it is clear that the residuals exhibit some spatial correlation beyond that accounted for by the logistic model (Figure 25). The spatial correlation appears to occur over ranges <60 km (Figure 25).

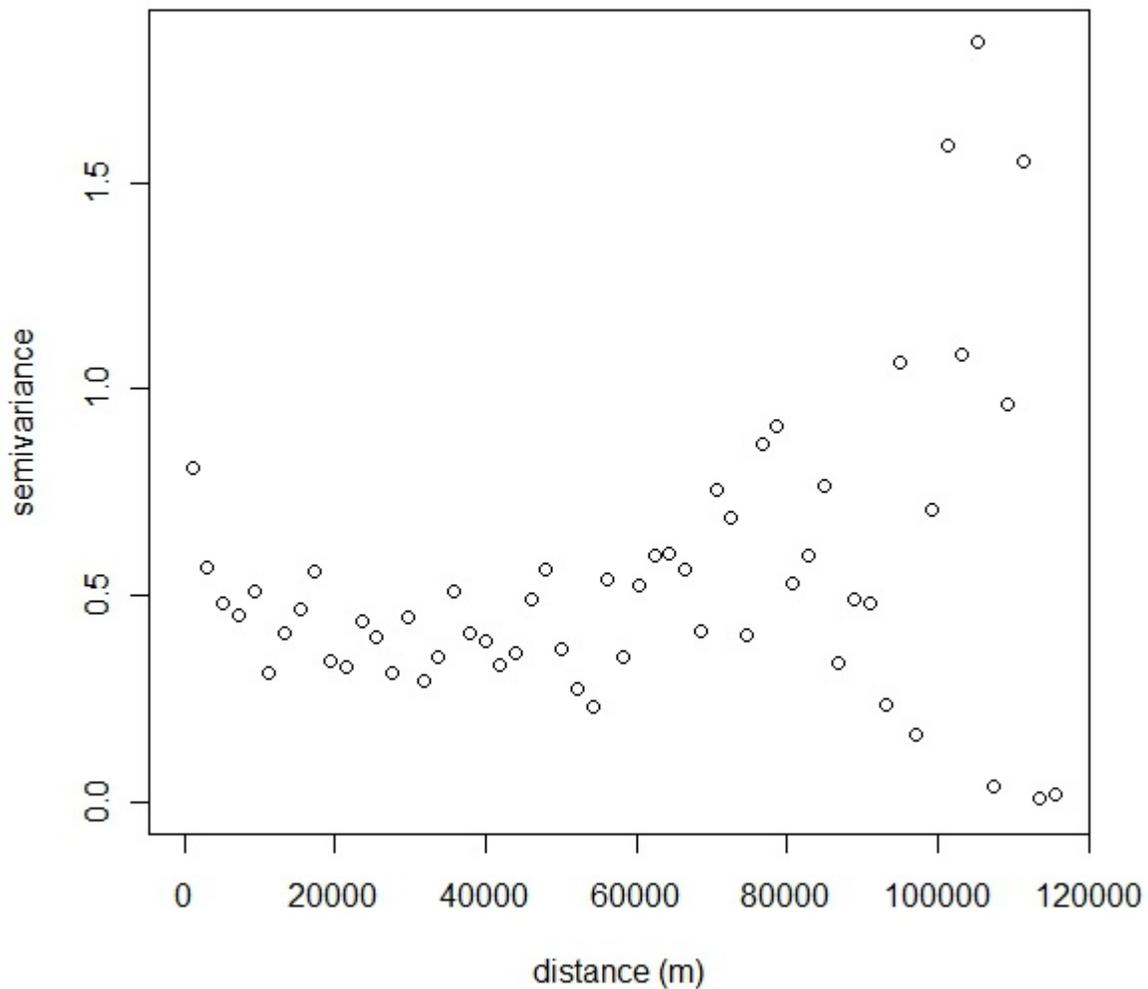


Figure 25: Empirical semi-variogram from standardized Pearson residuals (lag distance = 2000 m)

### 3.3.5 Classification and Regression Tree

The classification tree utilised four of the nine variables provided to the model: mean elevation, mean slope, mean TOPEX, and mean topographic wetness index (Figure 26). The full tree had six terminal nodes (5 splits). Results suggest that sites below 840 m, with high topographic exposure, and a high topographic wetness index are declining. The model is likely overfit as it is separating small numbers of samples based on small differences in the data. A more robust prediction was made by pruning the tree using cross-validation and a cost-complexity measure (Figure 27). The pruned tree has 2 nodes resulting from the split using a mean elevation of 840 m. Cross-validation of tree size selection showed that trees with one split have the best discrimination ability (Figure 28). The varying levels of predictive accuracy for differing tree size are shown in Figure 29. A tree with one split is determined to have higher predictive accuracy as the Somers' Dxy rank correlation measure is low and the Brier Score value is high (Figure 29).

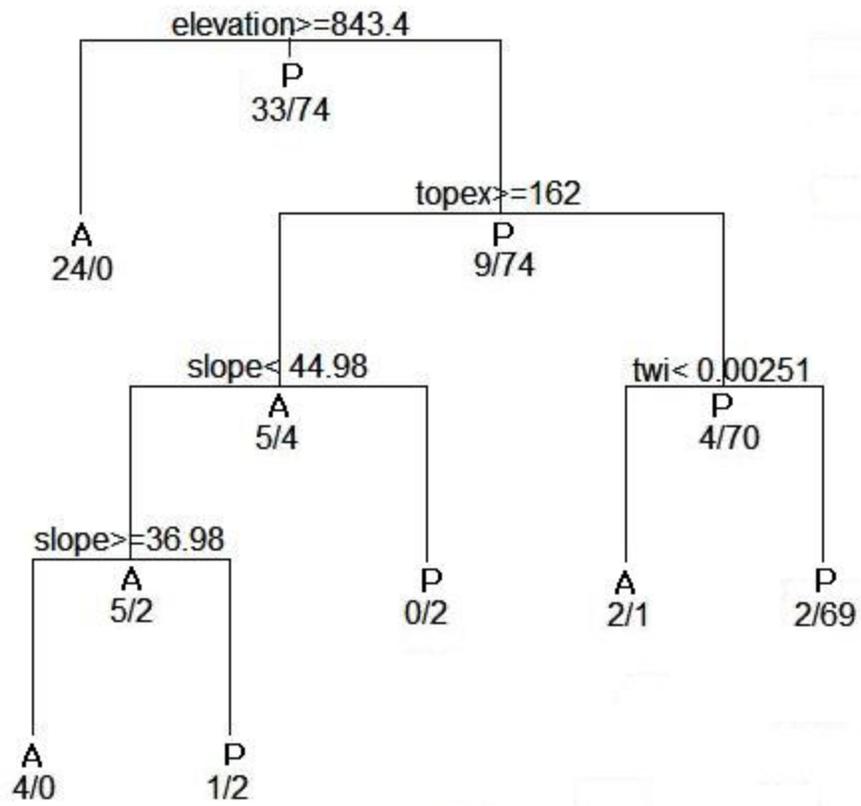


Figure 26: Full decision tree grown from data

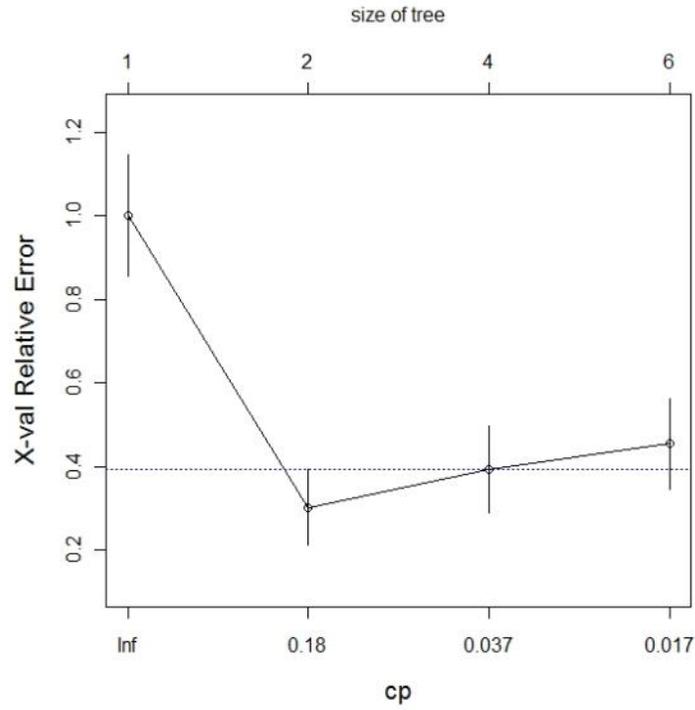


Figure 27: Cross-validated error as a function of the complexity parameter

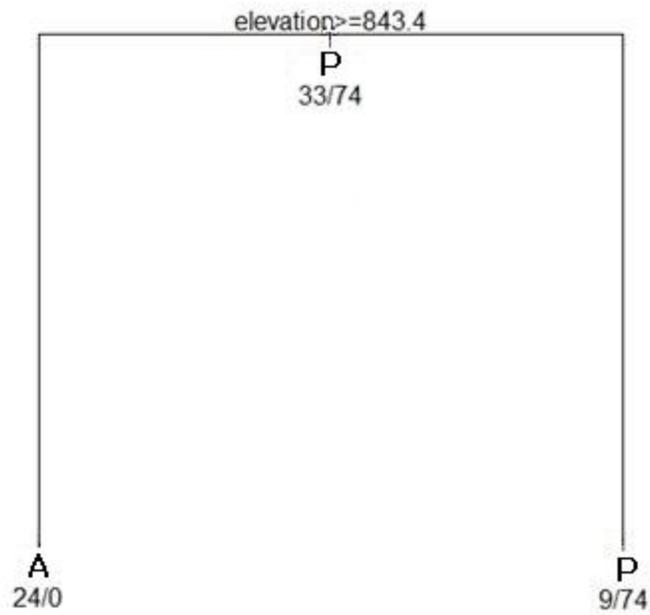


Figure 28: Pruned decision tree

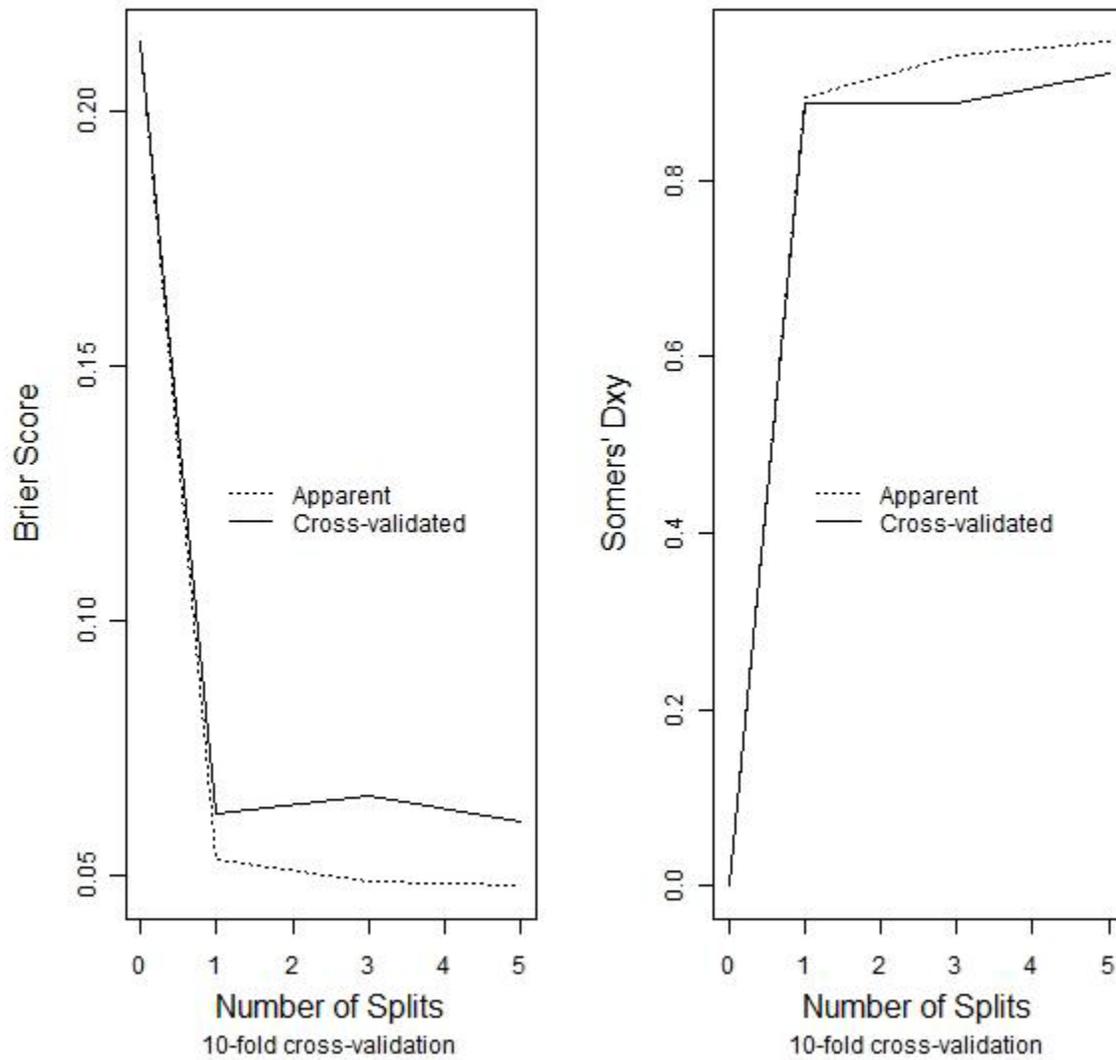


Figure 29: Estimates of Brier's accuracy scores and Somers' Dxy values obtained using 10-fold cross-validation of a sequence of trees

# Chapter 4

## Discussion

### 4.1 Model Interpretation

The main objective of my study was to contribute to the understanding of the yellow-cedar decline phenomenon by examining the spatial pattern of the decline and assessing relations with topographic variables. Having determined that the healthy and declining yellow-cedar stands exhibit differences in terms of elevation, slope and aspect (Section 3.2), statistical modeling provided a means of examining the relations with topographic variables in greater depth. The results of the logistic model suggested that four of the supplied predictor variables are strongly related to the presence of decline. These variables include: mean elevation, variance in elevation, distance from coast and topographic exposure. By examining the signs of the reported coefficients (Table 7) and considering the working hypothesis for the decline (Figure 4), the results can be interpreted in a more biologically meaningful way and examined in terms of how they might relate to the decline hypothesis.

In the model results, the likelihood of decline presence is seen to increase as elevation decreases, which supports the findings of Hennon et al. (2006) and Beier et al. (2008). The results may be explained by considering the factors that afford higher elevation stands greater protection from freezing injury than those at low elevations. At higher elevations, thaw periods are both less frequent and less severe due to the colder ambient temperature (Beier et al., 2008). The strength and retention of the root hardening is also greater at higher elevations, meaning that yellow-cedars at these sites retain their cold tolerance throughout the freeze thaw cycles occurring at lower elevations (Beier et al., 2008). The presence of snow cover contributes to retention of cold hardiness and also protects the roots by preventing soil warming. However, there has been a loss of protective snow cover at the lower elevations since the end of the Little Ice Age and, thus, these low-elevation yellow-cedar stands may not have adequate snow cover to insulate soils and roots (Beier et al., 2008; Hennon et al., 2006).

The observed increase in probability of decline occurrence with increasing variance in elevation may be related to the proposed effects of drainage on yellow-cedar health. A high level of variation in elevation within a stand may suggest that there is a greater likelihood of cold air and water accumulating in depressions within the stand. Observational studies in Alaska have indicated that both soil drainage and soil and air temperature are risk factors for yellow cedar decline (Hennon et al., 2006). Soil saturation creates stands with open canopies and shallow roots, which triggers the trees to dehardening prematurely by permitting soil warming in early spring (Hennon et al., 2006).

The negative coefficient value provided for the TOPEX variable suggests that sites with greater physical exposure are more likely to be declining (negative TOPEX indicates high topographic exposure). Sites with low TOPEX scores have greater topographic exposure to wind and are found along mountain tops and ridges. The increased physical exposure may result in less snow accumulation due to local windiness and, perhaps, less canopy cover due to stunted tree growth. These factors, combined with the exposure that creates rapid temperature shifts and frost periods, may result in freezing damage (Wittwer, 2004).

Decreasing distance from coast is another factor noted to increase the probability of decline occurrence. The degree of exposure and proximity to coast likely exert a large influence on the extreme temperatures experienced by sites such as winter minimums and summer maximums (Ashcroft et al., 2008). Sites with greater influence from coastal climatic moderators typically have milder winters, suggesting that there may not be a continual snowpack leaving the yellow-cedars more vulnerable to freeze-thaw cycles. The distance to coast parameter may also, to some extent, reflect cold air drainage at night. The coastal boundary was digitized such that many of the larger coastal inlets were traced and thus measuring distance to coast may be capturing the proximity to cold air ponds. Distance to coast was measured using the Euclidean distance, but perhaps this measure should be modified to measure in the direction of the prevailing winds (Ashcroft et al., 2008).

This coastal effect and influence of elevation on decline distribution can be related to the presence of different biogeoclimatic zones. Of the 108 sample sites, 90 were determined to be in the Coastal Western Hemlock (CWH) zone, and 18 in the Mountain Hemlock (MH) zone. All of the sample sites in the MH zone are healthy, which account for 52 % of all healthy sites. The MH zone is known for having a persistent, deep winter snowpack (Meidinger and Pojar, 1991), suggesting that yellow-cedar trees in this zone may be protected from freezing injury.

There are two CWH subzones within the study area that contain sample points, the CWH very wet maritime (CWH vm) subzone and the CWH very wet hypermaritime (CWH vh) subzone. A mixture of healthy and declining stands were found within the very wet maritime areas, but only declining sites (19 sites) were found in the hypermaritime subzone. The hypermaritime subzones occur on the outer mainland coast and low-lying islands and with this decreased continentality comes greater rainfall (Meidinger and Pojar, 1991) (Table 2). The lower elevation poorly drained soils of the hypermaritime subzone may combine to predispose yellow-cedar stands to decline.

The full regression tree identified mean elevation, TOPEX, topographic wetness index and mean slope as being important variables in determining the location of yellow cedar decline. Similar to the logistic model, declining sites were determined to have low mean elevation and low TOPEX scores. The small number of cases used to define threshold values of the predictor variables after the second split (the rule identified for TOPEX), however, suggests that great uncertainty accompanies these values and the model is likely overfit. Indeed, when the model is pruned back using Breiman et al.'s (1984) cost-complexity measure, only elevation is retained in the model. An elevation threshold value of 840 m appears to be the most effective value for splitting the declining and healthy yellow-cedar stands. This threshold is very close to the elevational boundary between the CWH and MH zones (approximately 800 m) in my specific study area.

## 4.2 Model Performance

While the use of a penalized logistic regression model permitted the identification of variables with the greatest explanatory power, in order for the model to be useful from a forest planning perspective it is necessary to take the model performance and robustness into account. The predictive accuracy of the final model, as evaluated by ROC analysis, was high (AUC 0.98), indicating that the model fits the data well. However, to fully assess the adequacy of the model to predict yellow-cedar decline occurrence, an independent data set is required for comparison. Unfortunately the limited amount of data available for the current study meant that the selection of predictor variables could only be validated through a bootstrap analysis. From analysis of the bootstrap output it was determined that the predictor variables selected for the final logistic model were also selected by the lasso algorithm in the majority of the bootstrap samples (all had non-zero coefficient values in >90 % of samples). However, the number of variables with non-zero coefficient values varied greatly between samples, as did the size of the coefficients. This finding suggests that the model results are sensitive to small changes in the data and the reported coefficients should be used with caution. In order to have confidence in the model results, more data is required for further model development and evaluation in an attempt to elucidate consistent trends.

Despite enforcing a minimum distance rule of 1000 m between sampling points, the residual analysis showed the presence of some spatial dependence that the model failed to take into account. This finding also suggests that returned coefficients should be interpreted with caution as spatial autocorrelation can cause exaggerated estimates of significance for explanatory variables (Miller and Franklin, 2002). It is recommended that future modeling efforts assess whether the addition of more predictor variables explain the observed spatial autocorrelation, or whether a separate explanatory term that accounts for the spatial dependence (a spatial autocovariate term) is required.

### 4.3 Model Representativeness

The predictive mapping output (using a threshold of 0.7) suggests that over half of the yellow-cedar stands in the study area have a high probability (0.8-1.0) of decline occurrence. The maps show that these stands occur at low elevations along the coast and inlets. The predictive maps combined with interpretation of the model coefficients indicates that yellow-cedar conservation efforts should be focused in sheltered, high elevation areas that have good drainage and are further inland. Applying the logistic model in a GIS environment is a useful means of visualizing the model results and informing management decisions. The ability to vary the model threshold and thus adjust the probability surface to suit the management requirements is a major advantage of this modeling approach (McDermid and Smith, 2008).

When applying the logistic model to the GIS dataset, predictive outcomes were generated for all stands with a yellow-cedar component. Probabilistic outcomes were generated for these polygons despite having only sampled stands that have a leading yellow-cedar component. The representativeness of the chosen sample points both within and beyond the geographical extent of the study site is an important consideration. From the analysis of the distribution of “all yellow-cedar stands” versus “leading yellow-cedar stands” (Section 3.1), it was identified that although the two groups were similar in terms of slope and aspect, a much greater proportion of “all yellow-cedar” stands occur at lower elevations. The stratified sampling design ensured that these lower elevations were represented in the sample. By creating ecological land units for sampling the range of environments observed and possible interaction effects were captured.

The use of indirect variables in this study, however, means that the final model can only be applied within a limited geographical extent (Gibson, 2004; Guisan and Zimmerman, 2000; Austin, 2002). Indirect variables, as defined by Austin and Smith (1989), are variables that have no direct physiological influence on the growth and health of plants (e.g. elevation, slope, aspect, topographic exposure, distance from coast etc.). Direct gradients, in contrast, are environmental variables that have direct physiological impact, but are not consumed (temperature, pH). Resource gradients deal with matter and energy used by plants (light, nutrients, water, CO<sub>2</sub> and O<sub>2</sub>). The indirect, topographic variables used in this study are useful in that they are easily

generated from DEM for large study areas. However, the final model cannot be applied out with the geographic extent of the study area because in a different location the same topographic position can experience a different combination of direct and resource gradients (Guisan and Zimmerman, 2000). This observation was termed the “law of relative site constancy” by Walter and Walter (1953). This law is exemplified by comparing the elevation threshold above which yellow-cedar decline does not appear identified in this study to that determined from studies in southeast Alaska. At this study’s location in British Columbia, yellow-cedar decline was not observed at locations above 840m, whereas at one study site in southeast Alaska decline is generally restricted to elevations below 150 m (Hennon et al., 2010). This example illustrates how differences in conditions between locations, in this case a 6° latitude difference, can be compensated for by adjusting the topographic position, in this instance elevation.

Environmental variables potentially more directly linked to the decline (an index of topographic wetness, potential solar radiation) were included in the original set of model parameters (Section 2.2.2., Table 4), but neither of these variables was found to be significant. In the case of the topographic wetness index (TWI), it is possible that errors in the digital elevation model resulted in an inaccurate parameter that failed to capture the true variation in topographic moisture. Van Niel et al. (2004) found that the TWI is more sensitive to errors in a DEM than the more indirect variable ‘topographic position’.

Potential solar radiation may not have been a statistically significant predictor of decline occurrence because the study failed to take cloud cover or stand canopy cover into consideration (Ashcroft et al., 2008). As mentioned in the description of the working hypothesis for the decline (Section 1.2.2), sparse canopy cover is thought to result in soil warming early in the spring, which causes the yellow-cedars to lose their cold tolerance and become vulnerable freezing injury (Hennon et al., 2010). However, using canopy cover as a predictor variable in this study would potentially confound model results (Ashcroft et al., 2008). That is, an undesirable situation would be created in which knowledge of decline occurrence was required in order to determine canopy cover, and knowledge of canopy cover was required to estimate decline occurrence.

## 4.4 Issues of Data Quality and Spatial Uncertainty

The term spatial uncertainty was defined by Goodchild (2008, p. 480) as “...the difference between contents of a spatial database and the corresponding phenomena in the real world.” All of the spatial datasets used in this study (aerial photographs, digital elevation models, topographic variables, VRI layer and digitized decline layer) are representations of the real world, and as such there will invariably be differences between the datasets and the real characteristics they allegedly represent (Goodchild, 2008). These differences typically arise when datasets are generated using processes involving approximation, measurement error or generalization (Goodchild, 2008). The use of the VRI layer, digitized decline layer and digital elevation model have all likely introduced spatial uncertainty into the analysis and thus the model results.

### 4.4.1 Vegetation Resource Inventory Layer

As mentioned in Section 2.2.1, there is concern surrounding both the spatial accuracy of the digitized boundaries of the forest stands in the VRI layer and the quality of the reported attributes. Species composition and other forest attributes change with time and therefore those stands that have not been re-inventoried in recent years will likely have greater inaccuracies. An attempt was made to reduce the level of uncertainty introduced to the study by excluding the oldest inventory data from the analysis. A new VRI layer is currently being produced for the study area, which is intended to “provide users of the data with more confidence in the delineation and estimation of tree and vegetation attributes using photo interpretation and provide more accurate and consistent spatial (polygon) and attribute data” (Interfor et al., 2008, p. *i*). The photo interpretation stage is expected to provide information on the location of yellow-cedar dieback (Interfor et al., 2008). The improved inventory information is expected to be available in the spring of 2012 and will provide better decision support for land managers and stakeholders.

#### 4.4.2 Digitizing Decline Occurrence

There is also uncertainty present concerning the boundaries and classification of the decline occurrence polygons. The availability of digital orthophotos and stereo pairs permitted a detailed examination of the forest stands, which allowed a more spatially accurate delineation of the decline than that achieved using aerial sketch mapping. However, the ability to determine the presence of decline occurrence was complicated by varying proportions of yellow cedar within stands. In stands where there was only a small yellow-cedar component it was difficult to discern whether the stand is healthy or if scattered mortality was present. The identification of yellow-cedar decline was also confounded by the presence of spike-top western redcedar (*Thuja plicata*) (Hennon, 2005). Western redcedar commonly exhibits a dead leader, which can make the identification of dead or dying yellow-cedars difficult due to the visual similarities between the two and the tendency for the two species to co-occur (Zeglen et al., 2008). Confidence in the digitizing of the decline was improved by determining whether stands in which western redcedar was present, but yellow-cedar was not, were healthy, and whether decline occurred in areas without western redcedar. Stands were also examined to see if the amount of decline present varied with differing proportions of yellow-cedar present. Despite these efforts, it is possible that both errors of omission and commission have occurred and the study would benefit from a ground truthing programme.

# Chapter 5

## Conclusions

### 5.1 Key Findings

The logistic lasso model appears to provide an excellent means of analysing the relation between yellow-cedar decline and topographic variables. By providing the model with a range of topographic variables suspected to play a role in determining the pattern of decline on the landscape, it was possible to identify a subset of variables that were strongly related to decline occurrence. Plausible reasons for the relations uncovered between the decline and the topographic variables were established by interpreting the results in a biologically meaningful manner and relating pattern to process. The model results indicated that low elevation sites close to the coast, which are more exposed and have more variation in elevation, are more likely to be declining. These topographic variables influence the degree of soil saturation, the temperatures and the snowpack presence in a forest stand. The findings support the hypothesised mechanism of decline, which emphasises the role of soil drainage, soil and air temperature and loss of snow cover in predisposing yellow-cedar trees to decline by increasing susceptibility to freezing injury.

The model produced in the study was determined to have high discrimination ability, which permitted the separation of declining and healthy sites according to the topographic values of the stands. The utility of such a model was demonstrated by applying it to geographic space to create a map depicting the likely occurrence of decline across the study site. Such a map could be interpreted as a risk-map for yellow-cedar decline and used to guide yellow-cedar management efforts in the study area.

Despite the great predictive success of the model, there are several issues that suggest model results should be interpreted with caution. First of all, the results of the model validation show that the number of parameters selected and the size of the model coefficients varies between

bootstrap samples, indicating that the model is sensitive to small changes in the dataset. Another issue is the presence of spatial dependence observed in the model residuals. Concerns also exist regarding the potential for errors of omission and commission in the digitizing of the yellow-cedar decline.

The decision tree produced in the analysis provided an easily interpretable means of visualizing the important variables and corresponding thresholds for separating healthy yellow-cedar sites from declining ones. However, the small size of the dataset meant that the decision tree was pruned back to a single split, which proved to emphasize the importance of elevation in determining decline presence.

Ultimately, this research provided insight into the relations between yellow-cedar decline and topographic variables. Additionally, evidence was found that supports the proposed associations in the current decline hypothesis. The analysis also highlighted the utility of the lasso logistic model for selecting significant parameters and mapping high risk areas for decline. The identification of these high risk areas could be used to guide yellow-cedar salvage operations. The model results suggest that yellow-cedar conservation and active regeneration efforts would best be focused in high elevation areas that are further inland, sheltered and well drained. Finally, this investigation highlighted the extensive nature of the decline phenomenon in coastal British Columbia. The expected future climate warming in British Columbia suggests that the decline may expand further, emphasising the need for the timely implementation of a management plan.

## 5.2 Recommendations for Future Research

While this study has provided insight into the location of decline occurrence in coastal British Columbia and the strength of relations with topographic variables, it has also raised further questions and highlighted the need for more investigations.

In order to assess the stability of the trends identified, further data on the location of decline occurrence is required. Collecting more data from within the geographical extent of the current study area would permit full validation of the model. Conducting external validation of the

model (i.e. collecting data from out with the geographical extent for validation purposes), would allow the identification of consistent trends across geographical space. Performing further internal, as well as external, validation would ensure that the trends identified are not spurious results from this particular dataset and would provide better decision support for yellow-cedar management plans.

In addition to further efforts to fully elucidate the spatial pattern of decline, investigations into the temporal aspect would deepen understanding of the decline process. If yellow-cedar decline is expanding upslope (Beier et al., 2008; Hennon et al., 2006) then earlier dates for the onset of mortality should be found at lower elevations. Adjusting the spatial scale of analysis would perhaps assist in identifying the sites where decline is initiated and from where it proceeds to spread. Using mean values to represent topographic conditions in a polygon may have masked some of the topographic complexity that is influencing the pattern of decline. Akashi and Mueller-Dombois (1995) found in their study of Hawaiian rainforest dieback that the micro-topography had a large role in determining the location of dieback initiation as well as the shape and intensity of the finer dieback patterns. Examining the topography of the individual yellow-cedar stands in greater detail, particularly the slope gradient, shape of the terrain (concave or convex) and the relative slope position, may lead to better identification of the combination of site factors that are likely to predispose yellow-cedar to decline.

In future modeling efforts, further investigations should be made into the shape of the decline response curves. When adding variables to the GLM it was necessary to assess the term for linearity and determine whether a higher order term should be included. Second order polynomial terms, which account for unimodal response curves, were the highest terms added to the model. Third order polynomial terms and higher may have captured skewed or bimodal functions that would have resulted in significance for a variable, however the resulting response curves are often spurious and difficult to interpret in a biologically meaningful manner (Guisan and Zimmerman, 2000). The use of generalized additive models (GAMs) would be better for data exploration as they allow a wider range of response curves to be modelled and, because the response curves are data driven, no decisions have to be made regarding the level of polynomial term to include for each variable (Guisan and Zimmerman, 2000; Yee and Mitchell, 1991).

# References

- Akashi, Y., and Mueller-Dombois, D. (1995). A landscape perspective of the Hawaiian Rain Forest dieback. *Journal of Vegetation Science*, 6(4):449-464
- Allen, C.D., and Breshears, D.D. (1998). Drought-induced shift of a forest-woodland ecotone: rapid landscape response to climate variation. *Ecology*, 95:14839-14842
- Ashcroft, M.B., Chisholm, L.A., and French, K.O. (2008). The effect of exposure on landscape scale soil surface temperatures and species distribution models. *Landscape Ecology*, 23:211-225
- Auclair, A.N.D., Worrest, R.C., Lachance, D., and Martin, H.C. (1992). Climatic perturbation as a general mechanism of forest dieback. In: Manion, P. D. and Lachance, D. (eds). *Forest Decline Concepts*, APS Press. St. Paul, Minn. pp. 38-58
- Austin, M.P. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157:101-118
- Austin, M. P. (2005). Vegetation and environment: discontinuities and continuities. In: van der Maarel, E. (eds). *Vegetation Ecology*. Blackwell Science Ltd. Oxford, United Kingdom. pp. 52-84
- Austin, M. P., and Smith, T.M. (1989). A new model for the continuum concept. *Vegetatio*, 83:35-47
- Bachelet, D., Neilson, R.P., Lenihan, J.M., and Drapek, R.J. (2001). Climate change effects on vegetation distribution and carbon budget in the United States. *Ecosystems*, 4:164-185
- Batschelet, E. (1981). *Circular statistics in biology*. Academic Press, London, U.K.
- BC MOF (1999a). *Kingcome timber supply area analysis report*. [Online] Available: <http://www.for.gov.bc.ca/hts/tsa/tsa33/tsr2/analysis.pdf> [Accessed: 22<sup>nd</sup> August 2010]
- BC MOF (1999b). *Midcoast timber supply area analysis report*. [Online] Available: <http://www.for.gov.bc.ca/hts/tsa/tsa19/tsr2/analysis.pdf> [Accessed: 22<sup>nd</sup> August 2010]
- Beers, T.W., Dress, P.E., and Wensel, L.C. (1966). Aspect transformation in site productivity research. *Journal of Forestry*, 64:691-692
- Beier, C.M., Sink, S.E., Hennon, P.E., D'Amore, D., and Juday, G.P. (2008). Twentieth-century warming and the dendroclimatology of declining yellow-cedar forests in southeastern Alaska. *Canadian Journal of Forest Research*, 38(6):1319-1334
- Beven, K. J., and Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1):43-69

- Beyer, H. L. (2004). *Hawth's Analysis Tools for ArcGIS*. [Online] Available: <http://www.spatial ecology.com/htools> [Accessed: 9<sup>th</sup> August 2010]
- Bian, L., and Walsh, S.J. (1993). Scale dependencies of vegetation and topography in a mountainous environment in Montana. *The Professional Geographer*, 45(1):1-11
- Bourque, C.P.A, Cox, R.M., Allen, D.J., Arp, P.A., and Meng, F.R. (2005). Spatial extent of winter thaw events in eastern North America: historical weather records in relation to yellow birch decline. *Global Change Biology*, 11:1477-1492
- Box, E.O. (1995). Factors determining distributions of tree species and plant functional types. *Vegetatio*, 121:101-116
- Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth International Group, California, USA
- Bush, M.B., Silman, M.R., and Urrego, D.H. (2004). 48,000 years of climate and forest change in a biodiversity hot spot. *Science*, 303:827-829
- Cook, E.R., and Johnson, A.H. (1989). Climate change and forest decline: a review of the red spruce case. *Water, Air, and Soil Pollution*, 48:127-140
- Cox, R.M. and Zhu, X.B. (2003). Effects of simulated thaw on xylem cavitation, residual embolism, spring dieback and shoot growth in yellow birch. *Tree Physiology*, 23:615-624
- D'Amore, D.V., and Hennon, P.E. (2006). Evaluation of soil saturation, soil chemistry, and early spring soil and air temperatures as risk factors in yellow-cedar decline. *Global Change Biology*, 12:524-545
- Daoust, G., Anseau, C., Theriault, A., and van Hulst, R. (1992). Pattern and process of maple dieback in southern Quebec (Canada). In: Manion, P.D. and Lachance, D. (eds). *Forest Decline Concepts*. APS Press, St. Paul, Minn. pp. 155-167
- de Smith, M. J., Goodchild, M.F., and Longley, P.A. (2009). *Geospatial Analysis: A comprehensive guide to principles, techniques, and software tools*. 3rd Edition. Troubador, London. [Online] Available: <http://www.spatialanalysisonline.com> [Accessed: 9<sup>th</sup> August 2010]
- Elith, J., Graham *et al.* (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29:129-151
- Elswick, R.K., Schwartz, P.F., and Welsh, J.A. (1997). Interpretation of the odds ratio from logistic regression after a transformation of the covariate vector. *Statistics in Medicine*, 16(15):1695-703
- ESRI (2006). *ArcMap Version 9.2*. Environmental Systems Research Institute (ESRI), Redlands CA

- Evans, I. S. (1998). What do terrain statistics really mean? In: Lane, S.N, Richards, K.S. and Chandler, J.H. (eds). *Landform Monitoring, Modelling and Analysis*. J. Wiley, Chichester. pp. 119–138.
- Franklin, J. (1995). Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19(4):474-499
- Frey, B.R., Lieffers, V.J., Hogg, E.H., and Landhäusser, S.M. (2004). Predicting landscape patterns of aspen dieback: mechanisms and knowledge gaps. *Canadian Journal of Forest Research*, 34:1379-1390
- Gates, D.M. (1990). Climate change and forests. *Tree Physiology*, 7:1-5
- Gibson, I.A., Wilson, B.A., Cahill, D.M., and Hill, J. (2004). Spatial prediction of rufous bristlebird habitat in a coastal heathland: a GIS-based approach. *Journal of Applied Ecology*, 41:213-223
- Goodchild, M. F. (2008). Imprecision and Spatial Uncertainty. In: Shekhar, S. and Xiong, H. (eds). *Encyclopedia of GIS*. Springer, New York
- Gould, P.R., (1970). Is statistix inferens the geographical name for a wild goose? *Economic Geography*, 46:439-448.
- GRASS Development Team (2010). *Geographic Resources Analysis Support System (GRASS)*. Software, Version 6.4.0. Open Source Geospatial Foundation. [Online] Available: <http://grass.osgeo.org> [Accessed: 22<sup>nd</sup> August 2010]
- Green, R.N., and Klinka, K. (1994). A field guide for site identification and interpretation for the Vancouver Forest Region. Land Management Handbook No. 28. BC. Min. For., Victoria, BC. 285 pp.
- Guisan, A., and Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135:147-186
- Guth, P.L. (2006). Geomorphometry from SRTM: Comparison to NED. *Photogrammetric Engineering & Remote Sensing*, 72(3):269-277
- Hair, J.E., Anderson, R.E., Tatham, R.L., and Black, W.C. (1998). *Multivariate Data Analysis*. 5<sup>th</sup> Edition. Prentice Hall. Upper Saddle River, New Jersey
- Hamann, A., and Wang, T. (2006). Potential effects of climate change on ecosystem and tree species distribution in British Columbia. *Ecology*, 87:2773-2786
- Hannah, P., Palutikof, J.P., and Quine, C.P. (1995). Predicting windspeeds for forest areas in complex terrain. In: Coutts, M.P. and Grace, J. (eds). *Wind and Trees*. Cambridge University Press. pp. 113-129

- Hansen, A.J., Neilson, R.P., Dale, V.H., Flather, C.H., Iverson, L.R., Currie, D.J., Shafer, S., Cook, R., and Bartlein, P.J. (2001). Global change in forests: Responses of species, communities, and biomes. *BioScience*, 51(9):765-779
- Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. Springer-Verlag, New York, NY
- Harrell, F.E. Jr. (2009). *Design: Design Package*. R package version 2.2-0 [Online] Available: <http://CRAN.R-project.org/package=Design> [Accessed: 22<sup>nd</sup> August 2010]
- Harris, A.S. (1990). *Chamaecyparis nootkatensis* (D. Don). Spach: Alaska-cedar. In: *Silvics of North America, Volume 1: Conifers*, Agric. Hanb. 654. U.S. Department of Agriculture, Forest Service, Washington, DC. pp 97-102 [Online] Available: [http://na.fs.fed.us/spfo/pubs/silvics\\_manual/Volume\\_1/chamaecyparis/nootkatensis.htm](http://na.fs.fed.us/spfo/pubs/silvics_manual/Volume_1/chamaecyparis/nootkatensis.htm) [Accessed: 9<sup>th</sup> August 2010]
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2<sup>nd</sup> Edition. Springer-Verlag, New York, NY
- Hennon, P.E., Hansen, E.M., and Shaw III, C.G. (1990a). Dynamics of decline and mortality of *Chamaecyparis nootkatensis* in southeast Alaska. *Canadian Journal of Botany*, 68:651-662
- Hennon, P.E., Shaw III, C.G., and Hansen, E.M. (1990b). Dating decline and mortality of *Chamaecyparis nootkatensis* in southeast Alaska. *Forest Science*, 36(3):502-515
- Hennon, P.E., and Shaw III, C.G. (1994). Did climatic warming trigger the onset and development of yellow-cedar decline in southeast Alaska? *European Journal of Forest Pathology*, 24:399-418
- Hennon, P.E., and Shaw III, C.G. (1997). The enigma of yellow-cedar decline – What is killing these long-lived, defensive trees? *Journal of Forestry*, 95(12):4-10
- Hennon, P.E., Shaw III, C.G., and Hansen, E.M. (1998). Reproduction and forest decline of *Chamaecyparis nootkatensis* (yellow-cedar). in southeast Alaska, USA. In: Laderman, A.D. (ed). *Coastally Restricted Forests*. Oxford University Press, New York. pp. 54-69
- Hennon, P.E., D'Amore, D.V., Zeglen, S., and Grainger, M. (2005). *Yellow-cedar decline in the North Coast Forest District of British Columbia*. Res. Note RN-549. Portland, OR: U.S. Dep. Agric., Pacific Northwest Research Station. pp.20
- Hennon, P.E., D'Amore, D., Wittwer, D., Johnson, A., Schaberg, P., Hawley, G., Beier, C., Sink, S., and Juday, G. (2006). Climate warming, reduced snow, and freezing injury could explain the demise of yellow-cedar in southeast Alaska, USA. *World Resource Review*, 18(2):427-450

- Hennon, P.E., D'Amore, D.V., Witter, D.T., and Lamb, M.B. (2010). Influence of forest canopy and snow on microclimate in a declining yellow-cedar forest of southeast Alaska. *Northwest Science*, 84(1):73-87
- Hosmer, D.W., Hosmer, T., Le Cessie, S., and Lemeshow, S. (1997). A comparison of goodness of-fit tests for the logistic regression model. *Statistics in Medicine*, 16:965-980
- Interfor (International Forest Products Ltd.) and Timberline (Timberline Natural Resource Group Ltd.). (2008). *Mid Coast Timber Supply Area, Tree Farm License 25 Block 5 & Tree Farm License 39 Block 7. Vegetation Resources Inventory Photo Interpretation Project Implementation Plan (Phase I VPIP)*. VRI Reports and Publications, BC Ministry of Forests and Range. [Online] Available: [http://www.for.gov.bc.ca/hts/vri/reports&pub/tsa\\_vpips/midcoasttsa\\_vripi\\_vpip.pdf](http://www.for.gov.bc.ca/hts/vri/reports&pub/tsa_vpips/midcoasttsa_vripi_vpip.pdf) [Accessed: 9<sup>th</sup> August 2010]
- Interfor (International Forest Products Ltd.), Timberline (Timberline Natural Resource Group Ltd.), and the Ministry of Forests and Range. (2008). *Central Coast LRMP Area. Vegetation Resources Inventory Strategic Inventory Plan (VSIP)*. VRI Reports and Publications, BC Ministry of Forests and Range. [Online] Available: [http://www.for.gov.bc.ca/hts/vri/reports&pub/tsa\\_vsips/centralcoast\\_vri\\_vsip.pdf](http://www.for.gov.bc.ca/hts/vri/reports&pub/tsa_vsips/centralcoast_vri_vsip.pdf) [Accessed: 9<sup>th</sup> August 2010]
- IPCC (2007a). *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden P.J. and Hanson, C.E. (eds.), Cambridge University Press, Cambridge, UK, pp. 976
- IPCC (2007b). Summary for Policymakers. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M. and Miller, H.L. eds. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA
- Kleinschmidt, I., Bagayoko, M., Clarke, G.P.Y., Craig, M., and Le Sueur, D. (2000). A spatial statistical approach to malaria mapping. *International Journal of Epidemiology*, 29:355-361
- Leaphart, C.D. and Stage, A.R. (1971). Climate: a factor in the origin of the pole blight disease of *Pinus Monticola* Dougl. *Ecology*, 52(2):229-239
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74(6):1659-1673
- Lesaffre, E. and Marx, B.D. (1993). Collinearity in generalized linear regression. *Communications in Statistics – Theory and Methods*, 22(7):1933-1952
- Manion, P.D. (1991). *Tree disease concepts*. Prentice Hall, Englewood Cliffs, N.J.

- McDermid, G.J., and Smith, I.U. (2008). Mapping the distribution of whitebark pine (*Pinus albicaulis*). in Waterton Lakes National Park using logistic regression and classification tree analysis. *Remote Sensing*, 34(4):356-366
- McDonald, K.A., Hennon, P.E., Stevens, J.H., and Green, D.W. (1997). *Mechanical properties of salvaged dead yellow-cedar in southeast Alaska: Phase 1*. Res. Pap. FPL-RP-565. USDA Forest Serv., Forest Prod. Lab., Madison, WI. pp. 9
- McNab, W.H. (1989). Terrain Shape Index: Quantifying Effect of Minor Landforms on Tree Height. *Forest Science*, 35(1):91-104
- McNab, W.H., and Loftis, D.L. (2002). Probability of occurrence and habitat features for oriental bittersweet in an oak forest in the southern Appalachian mountains, USA. *Forest Ecology and Management*, 155:45-54
- Meidinger, D.V., and Pojar, J. (1991). *Ecosystems of British Columbia*. B.C. Ministry of Forests, Victoria, BC. Special Report No. 6. 330 pp. [Online] Available: <http://www.for.gov.bc.ca/hfd/pubs/Docs/Srs/SRseries.htm> [Accessed: 22<sup>nd</sup> August 2010]
- Menard, S. (2001). *Applied Logistic Regression Analysis*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-106. Thousand Oaks, CA: Sage.
- Mill, R.R., and Farjon, A. (2006). Proposal to conserve the name *Xanthocyparis* against *Callitropsis* Oerst.(Cupressaceae). *Taxon*, 55(1):227-238
- Miller, J. (2005). Incorporating spatial dependence in predictive vegetation models: residual interpolation methods. *The Professional Geographer*. 57(2):169-84
- Miller, J., and Franklin, J. (2002). Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling*. 157(2-3): 227-247
- MOFR (2009). North Island-Central Coast Forest District Profile. [Online] Available: <http://www.for.gov.bc.ca/dni/Profile.htm> [Accessed: 18<sup>th</sup> April 2009]
- Morellato, L.P., Alberti, L.F., and Hudson, I.L. (2010). Applications of circular statistics in plant phenology: a case studies approach. In: Hudson, I.L., and Keatley, M.R. (eds.) *Phenological research: methods for environmental and climate change analysis*. Springer, Dordrecht. pp. 339-359
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*. 78:691-692
- Oimoen, M.J. (2000). An effective filter for removal of production artifacts in U.S. Geological Survey 7.5-minute digital elevation models. *Proceedings of Fourteenth International Conference on Applied Geologic Remote Sensing, Las Vegas, Nevada, November 6-8, 2000*. Veridian ERIM International, Ann Arbor, Michigan, pp. 311-319

- Park, M.Y., and Hastie, T. (2006). *L<sub>1</sub> regularization path algorithm for generalized linear models*. Technical report, Stanford University, Stanford.
- Park, M.Y., and Hastie, T. (2007). *glmpath: L1 Regularization Path for Generalized Linear Models and Cox Proportional Hazards Model*. R package version 0.94 [Online] Available from: <http://cran.r-project.org/web/packages/glmpath/index.html> [Accessed: 22<sup>nd</sup> August 2010]
- Parmesan, C., Gaines, S., Gonzalez, L., Kaufman, D.M., Kingsolver, J., Townsend Peterson, A., and Sagarin, R. (2005). Empirical perspectives on species borders: from traditional biogeography to global change. *OIKOS*, 108:58-75
- Prentice, K.C. (1990). Bioclimatic distribution of vegetation for general circulation model studies. *Journal of Geophysical Research*, 95(D8):11,811-11,830
- Quine, C.P., and White, I.M.S. (1998). The potential of distance-limited topex in the prediction of site windiness. *Forestry*, 71(4):325-332
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. [Online] Available from: <http://www.R-project.org> [Accessed: 22<sup>nd</sup> August 2010]
- Rasemann, S., Schmidt, J., Schrott, L., and Dikau, R. (2004). Geomorphometry in mountain terrain. In: Bishop, M.P., Shroder, J.F. (eds). *GIS & Mountain Geomorphology*. Springer, Berlin. pp. 101-145
- Reineking, B., and Schroder, B. (2006). Constrain to perform: Regularization of habitat models. *Ecological Modelling*, 193:675-690
- Roberts, J.J., Best, B.D., Dunn, D.C., Treml, E.A., and Halpin, P.N. (2010). Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environmental Modelling & Software*, 25:1197-1207 [Online] Software available: <http://code.nicholas.duke.edu/projects/mget> [Accessed: 22<sup>nd</sup> August 2010]
- Rossiter, DG., and Loza, A. (2008). *Technical Note: Analyzing land cover change with logistic regression in R*. Technical Report ITC, Enschede, NL. Version 2.01, pp. 69 [Online] Available: [http://www.itc.nl/personal/rossiter/pubs/list.html#pubs\\_m\\_R](http://www.itc.nl/personal/rossiter/pubs/list.html#pubs_m_R) [Accessed: 9<sup>th</sup> August 2010]
- Schaberg, P.G., Hennon, P.E., D'Amore, D.V., Hawley, G.J., and Borer, C.H. (2005). Seasonal differences in freezing tolerances of yellow-cedar and western hemlock trees at a site affected by yellow-cedar decline. *Canadian Journal of Forest Research*, 35:2065-2070
- Shafer, S.L., Bartlein, P.J., and Thompson, R.S. (2001). Potential changes in the distributions of western North America tree and shrub taxa under future climate scenarios. *Ecosystems*, 4:200-215

- Snow, R. (2005). Continental climate and continentality. In: Oliver, J.E. (ed.) *Encyclopedia of world climatology*. Springer, The Netherlands. pp. 303-305
- Steyerberg, E.W. (2009). *Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating*. Springer, New York
- Steyerberg, E.W., Eijkemans, M.J.C., and Habbema, J.D.F. (1999). Stepwise selection in small data sets: A simulation study of bias in logistic regression analysis. *Journal of Clinical Epidemiology*, 52(10):935-942
- Steyerberg, E.W., Eijkemans, M.J.C., Harrell, F.E., and Habbema, J.D.F. (2001). Prognostic modeling with logistic regression analysis: In search of a sensible strategy in small data sets. *Medical Decision Making*, 21:45-56
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240:1285-1293
- Tarboton, D. G. (1997). A new method for the determination of flow directions and contributing areas in grid digital elevation models. *Water Resources Research*, 33(2):309-319
- Tarboton, D.G. (2008). *Terrain Analysis Using Digital Elevation Models (TauDEM), version 3.1*. [Online] Available: <http://hydrology.neng.usu.edu/taudem/taudem3.1/index.html> [Accessed: 22<sup>nd</sup> August 2010]
- Therneau, T.M, and Atkinson, B. (2010). rpart: Recursive Partitioning. R package version 3.1-46 [Available from: <http://cran.r-project.org/web/packages/rpart/index.html>]
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1):267-288
- USGS (2006). *Tree species distribution maps for North America*. [Online] Available: <http://esp.cr.usgs.gov/data/atlas/little/> [Accessed: 22<sup>nd</sup> August 2010]
- Van Niel, K.P., Laffan, S.W., and Lees, B.G. (2004). Effect of error in the DEM on environmental variables for predictive vegetation modelling. *Vegetation Science*, 15:747-756
- Vayssières, M.P, Plant, R.E., and Allen-Diaz, B.H. (2000). Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, 11(5):679-694
- Walter, H., and Walter, E. (1953). Das Gesetz der relativen Standortskonstanz: Das Wesen der Pflanzengesellschaften. *Berichte Deutsche Botanische Gesellschaft*, 66:228-236.
- Westfall, J., and Ebata, T. (2008). *Summary of forest health conditions in British Columbia. British Columbia*. British Columbia, Ministry of Forests and Range. Forest Practices Branch
- Whittingham, M.J., Stephens, P.A., Bradbury, R.B., and Freckleton, R.P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, 75: 1182-1189

- Wilson, J.D. (1984). Determining a topex score. *Scottish Forestry*, 384:251-256
- Winnet, S.M. (1998). Potential effects of climate change on U.S. forests: a review. *Climate Research*, 11:39-49
- Wittwer, D., Hennon, P., D'Amore, D., and Lamb, M. (2007). *Yellow-cedar decline: key landscape features and snow modeling of a climate-induced forest decline on a dormant volcano*. USFS Poster Presentation. [Online] Available: [http://www.fs.fed.us/r10/spf/fhp/cedar/pubs/EM\\_cedar\\_mt\\_edgcumbe\\_year\\_2\\_v2\\_web\\_ual.pdf](http://www.fs.fed.us/r10/spf/fhp/cedar/pubs/EM_cedar_mt_edgcumbe_year_2_v2_web_ual.pdf). [Accessed: 19th April 2009]
- Wood, W.F., and Snell, J.B. (1960). *A Quantitative system for classifying landforms* (Technical Report ER-124). U.S. Quartermaster Research & Engineering Center, Natick, MS.
- Worrall, J., Egeland, L., Eager, T., Mask, R., Johnson, E., Kemp, P. And Shepperd, W. (2007). Sudden aspen decline in southwest Colorado: site and stand factors and a hypothesis on etiology. *Western International Forest Disease Work Conference*, 55:1-4
- Yee, T.W., and Mitchell, N.D. (1991). Generalized additive models in plant ecology. *Journal of Vegetation Science*, 2(5):587-602
- Zar, J.H. (1984). *Biostatistical analysis*. 2<sup>nd</sup> Edition. Prentice-Hall, Inc. Englewood Cliffs, NJ
- Zeglen, S. (2004). Forest health projects: Yellow-cedar decline. In: Westfall, J. (ed.) *2004 Summary of forest health conditions in British Columbia*. British Columbia, Ministry of Forests and Range, Forest Practices Branch
- Zeglen, S., Hodge, J., Heppner, D., and Burleigh, J. (2008). *Coast Forest Region: 2008-10 Coastal Timber Supply Areas Regional Forest Health Overview*. British Columbia, Ministry of Forests and Range.
- Zimmerman, N.E. (2000). *Tools for analyzing, summarizing, and mapping of biophysical variables: Topographic position mapping routines*. [Online] Available: <http://www.wsl.ch/staff/niklaus.zimmermann/progs.html> [Accessed: 15th August 2010]
- Zhu, X.B., Cox, R.M., Meng, F.-R., and Arp, P.A. (2001). Responses of xylem cavitation, freezing injury and shoot dieback to a simulated winter thaw in yellow birch seedlings growing in different nursery culture regims. *Forest Ecology and Management*, 145(3):243-253