Natural and anthropogenic controls of landslides on Vancouver Island

by

Jason Goetz

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Science
in
Geography

Waterloo, Ontario, Canada, 2012

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Empirically-based models of landslide distribution and susceptibility are currently the most commonly used approach for mapping probabilities of landslide initiation and analyzing their association with natural and anthropogenic environmental factors. In general, these models statistically estimate susceptibility based on the predisposition of an area to experience a landslide given a range of environmental factors, which may include land use, topography, hydrology and other spatial attributes. Novel statistical approaches include the generalized additive model (GAM), a non-parametric regression technique, which is used in this study to explore the relationship of landslide initiation to topography, rainfall and forest land cover and logging roads on Vancouver Island, British Columbia.

The analysis is centered on an inventory of 639 landslides of winter 2006/07. Data sources representing potentially relevant environmental conditions of landslide initiation are based on: terrain analysis derived from a 20-m CDED digital elevation model; forest land cover classified from Landsat TM scenes for the summer before the 2006 rainy season; geostatistically interpolated antecedent rainfall patterns representing different temporal scales of rainfall (a major storm, winter and annual rainfall); and the main lithological units of surface geology.

In order to assess the incremental effect of these data sources to predict landslide susceptibility, predictive performances of models based on GAMs are compared using spatial cross-validation estimates of the area under the ROC curve (AUROC), and variable selection frequencies are used to determine the prevalence of non-parametric associations to landslides.

In addition to topographic variables, forest land cover (e.g., deforestation), and logging roads showed a strong association with landslide initiation, followed by rainfall patterns and the very general lithological classification as less important controls of landscape-scale landslide activity in this area. Annual rainfall patterns are found not to contribute significantly to model prediction improvement and may lead to model overfitting. Comparisons to generalized linear models (i.e., logistic regression) indicate that GAMs are significantly better for modeling landslide susceptibility.

Overall, based on the model predictions, the most susceptible 4% of the study area had 29 times higher density of landslide initiation points than the least susceptible 73% of the study area (0.156 versus 0.005 landslides/km^2).

Acknowledgements

I would to thank Prof. Alexander Brenning, my thesis advisor, for providing me with valuable academic guidance and support. Thank you for the opportunities that allowed me to explore the wonderful world of mountains, and pushing me to always do my best.

I thank my committee member, Dr. Richard Guthrie for his (cross-country) geomorphological support of my thesis and previously published research.

Also, I thank the readers of my thesis, Prof. Peter Deadman and Prof. Richard Kelly, as well as Prof. Ian McKenzie for positively contributing to an excellent learning environment in the Department of Geography, which allowed me to enjoy and appreciate my education at the University of Waterloo.

I thank Prof. Thomas Glade, Dr. Rainer Bell and Helene Petschko of the University of Vienna's Geomorphic Systems and Risk Research Group for allowing me to work amongst their team of landslide researchers and teaching me valuable lessons regarding the process of developing and applying landslide susceptibility models for spatial planning.

Thanks to the British Columbia Ministry of Forests, Lands and Natural Resources and the British Columbia Ministry of Environment for providing access to spatial data required to complete my thesis research.

This research was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada through the Alexander Graham Bell Canada Scholarship.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Recent advances in statistical classification methodology have led to innovative approaches for the predictive modeling of landslide susceptibility. These advancements allow us to utilize more flexible modeling techniques to improve predictive mapping of landslide initiation, which can lead to better risk and hazard analysis, reducing the significant negative effects of landslides to transportation and communication infrastructure and, most importantly, human lives.

In general, a landslide can be defined as "the movement of a mass of rock, debris, or earth downslope" (Cruden, 1991). The term 'susceptibility' is applied to maps that estimate the likelihood of landslide initiation (Dai et al., 2002). Susceptibility of an area to landslides is analyzed by determining the spatial probability of slope failure for a range of destabilizing factors (Guzzetti et al., 2006).

There is a wide range of methods used for landslide susceptibility mapping. Traditionally, landside hazards were mapped using geomorphological information that was mainly descriptive. This approach is very subjective to the geomorphologists' interpretation of the landscape. In addition, the reliability of these descriptive based models has been poorly documented. Thus, it is difficult to evaluate the quality of these maps (Guzzetti et al., 1999).

With the advent of geographical information systems (GIS), susceptibility analysis focused more on the use of deterministic approaches that applied physically-based models, such as SHALSTAB and SINMAP, with topographic information (Montgomery and Dietrich, 1994; Pack et al., 1998; Meisina and Scarabelli, 2007). This approach is seen as being more sufficient than traditional descriptive models because its quantitative approach for investigation of instability factors that influence landslides. However, some disadvantages of physically-based models are that they can be too simplistic; the required geotechnical investigation is expensive; and the geotechnical data may have high spatial variability (Carrara, 1993; Guzzetti et al., 1999).

Common today is the application of statistically-based models for landslide susceptibility mapping (Chung and Fabbri, 1999; Guzzetti et al., 2006; Frattini et al., 2010; Sterlacchini et al., 2011; Goetz et al., 2011; Blahut et al., 2010). In general, these models statistically estimate susceptibility based on the predisposition of an area to experience a landslide given range of environmental factors, which may include land use, topography, hydrology and any other related spatial attributes.

1.1 Overview

As the field of geomatics continues to expand, so does the availability of high quality geospatial data. High-resolution digital elevation models (DEM), satellite imagery and many other forms of spatial data have become more and more accessible for researchers and the general public in Canada. Therefore, a door is opened that provides geomatics researchers with the ability to employ new methods and data to solve old problems.

Regarding landslide research, high-resolution DEMs can be utilized for terrain analysis enabling the application of sophisticated landslide modeling approaches. Relationships between landslides and controlling factors can also be investigated with customized land cover classification using now readily available satellite imagery. In addition, access to primary data, such as weather records, allows for researchers to cater available data to their own research needs instead of relying on the availability of 'ready-made' spatial-data products.

This study utilizes accessible spatial data and novel statistical classification methods to explore empirical relationships of landslides to controlling factors on Vancouver Island. A generalized additive model (GAM) is used to observe if there are any non-linear relationships related to the controlling factors of landslides. In addition a GAM is used to produce a landslide susceptibility map for Vancouver Island. Generalized linear models (GLM) are also used in the analysis to provide a reference for model performance for the GAM. Logistic regression, which is a form of a GLM, has been the most common statistical approach for modeling landslide susceptibility (Brenning, 2005).

1.2 Goal and objectives

The purpose of this study is to explore natural and anthropogenic controls influencing landslide initiation at a regional scale for Vancouver Island, British Columbia, Canada. The exploration of natural controls includes topographic, lithologic and climatic factors. Possible anthropogenic controls that are explored include forest land cover, which has be shaped by a long history of forest harvesting, and the influence of the presence of logging roads on landslide initiation. In addition the incremental effects of natural and anthropogenic causes are explored to contribute to building more comprehensive knowledge of landslide initiation on Vancouver Island.

The goal of this study is to use comprehensive knowledge gained from the assessment of landslide controls to create a landslide susceptibility model for all of Vancouver Island using novel statistical classification techniques. The primary objectives to obtain the goal for this research are,

- Acquire and process all necessary spatial data
- Validate the quality of spatial data
- Apply statistical modeling techniques to produce a landslide susceptibility model
- Utilize the landslide susceptibility model to explore in detail the relationship of environmental controls to landslides

The acquisition and processing of spatial data for modeling of landslide susceptibility requires the application and collaboration of research conducted in fields outside of landslide modeling. Some novel techniques of validation of the quality of spatial data and model performance are used to reduce inherent uncertainties in this spatial analysis. In order to explore the controlling factors related to landslide initiation, an adequate model of landslide susceptibility must be produced.

1.3 Motivation of research

This study is expected to have a strong impact for the improvement of landslide hazard and risk modeling in Canada. One of the principles of this study is to present a geomatic analysis of landslide susceptibility in a form that is interpretable and transparent; thus, allowing it to be properly communicated to those interesting in applying these methods for spatial planning purposes. Consequently, the research findings will contribute to Canadian government research and development to reduce landslide hazards while forming collaborative bonds between government and academic institutions.

1.4 Structure of thesis

This thesis presents a landslide susceptibility model for Vancouver Island after first presenting research context to GAM and GLM susceptibility modeling, important controlling factors of landslides, and background knowledge regarding the methods used for the analysis (Chapter 2); a description of the physical geography of Vancouver Island relating to landslide initiation (Chapter 3); a detailed explanation of the various methods used to process or obtain spatial data relevant to the landslide susceptibility analysis; as well the details of susceptibility modeling using GAMs and GLMs (Chapter 4). The results are presented in Chapter 5, which is followed by a discussion on quality of spatial data, geomorphological interpretations and limitations (Chapter 6), and the main conclusions (Chapter 7).

Chapter 2
Research context

2.1 Landslides

In general, landslides are the result of progressive deterioration of slope material by natural geological processes, such as weathering and erosion, over a long period of time (Terlien, 1998). The initiation of landslides depends on the stability of slope material. A slope is stable when the shear stress is less than the shear strength of the slope material. If the shear stress exceeds the shear strength a slope becomes unstable and a landslides occurs (Ritter et al., 2002). A triggering mechanism may be required to cause the immediate initialization of slope movement. Common landslide triggers around the globe include earthquakes, heavy rainfall, rapid snowmelt, volcanic activity, glacial activity and human activity (Sidle and Ochiai, 2006). Other (environmental) factors related to landslide initiation include vegetation cover, lithology, soil type and topography (Kaldova and Rosenfeld, 1998).

The most common landslide classification system used is developed by Varnes (1978). This scheme categorizes landslide type based on the type of slope movement (falls, topples, slides, spreads, flows, and complex) and the type of material (bedrock, coarse soil or fine soil; Table 2.1). Details regarding the specific landslides used in this study (debris flow and debris slides) are discussed in more detail in Section 3.4.

Table 2.1. Landslide types - an abbreviated version of classification of slope movement types (After Varnes, 1978)

Type of movement		Type of material		
		Bedrock	**Engineering soils**	
			Coarse	*Fine*
Falls		Rock fall	Debris fall	Earth fall
Topples		Rock topple	Debris topple	Earth topple
Slides	**Rotational**	Rock slump	Debris slump	Earth slump
	Translational	Rock block slide; rock slide	Debris block slide; debris slide	Earth block slide; earth slide
Lateral spreads		Rock spread	Debris spread	Earth spread
Flows		Rock flow (deep creep)	Debris flow (soil creep)	Earth flow (soil creep)
Complex slope movements		Combinations of two or more types of movement		

2.2 Rainfall influence on landslides

Rainfall is a hydrological triggering mechanism that can decrease slope stability. The factors that generally control landslide initiation from rainfall are seepage (intensity, duration and infiltration of rainfall), moisture content and antecedent rainfall (Crosta, 1998). Typically, the influence of rainfall on landslides is explored by determining rainfall threshold for initiation (Glade, 2000; Aleotti, 2004; Giannecchini, 2006; Guzzetti et al., 2007). Although rainfall is triggering factor of landslides, it has been used as variable for landslide prediction using integrated empirical approaches (Chang and Chiang, 2009).

Landslides related to hydrological triggering factors are caused by an increase in pore-water pressure on the failure surface, which decreases shear strength and induced slope failure. An increase in pore-water pressure is usually directly related to percolated rainfall. Indirect increase in pore-water pressure is caused by perched water tables, an accumulation of water at the soil bedrock contact or an impermeable soil layer (Terlien, 1998; Sidle and Ochiai, 2006). In general, landslides that are directly triggered by percolating rainfall are referred to as *shallow slides*, which have a maximum depths of 2 m (Terlien, 1998). Shallow soils are generally more susceptible to landslides because infiltrating water can reach an impermeable layer more quickly than deeper soils (Sammori et al., 1993; Sidle and Ochiai, 2006).

Since rainfall is a main trigger of landslides, the spatial distribution of slope failure must reflect, to some extent, the pattern of rainfall. Therefore, knowledge of rainfall patterns is critical for the understanding of landslide distribution.

2.2.1 Mountain topography and rainfall patterns

Due to its high spatial and temporal variability, rainfall is one of the most difficult meteorological parameters to measure (Kidd, 2001). Many studies have attempted to interpolate rainfall; however, the high variability of rainfall makes it extremely challenging to create a single best method. Determining the most appropriate approach for a particular application depends on the local rainfall characteristics and the time and spatial scale of the analysis (Grimes and Pardo-Igúzquiza, 2010). Measurement of rainfall in mountain areas is particularly challenging, especially when trying to estimate extreme rainfall amounts (Krajewski and Smith, 2002). Estimation of rainfall in mountains is difficult because of low-density rain gauge networks that are not always useful for delineating rainfall boundaries in complex terrain.

To examine rainfall distribution in mountain areas it is important to differentiate from flat lands because of the different physical processes involved (Grimes and Pardo-Igúzquiza, 2010). A common (arbitrary) definition of mountains in North America differentiates between hills and mountains if the relief is greater than 600m; this altitudinal change is enough to create different climate conditions and vegetation cover (Thompson, 1964). In general, altitudinal change or elevation is the most common variable to estimate rainfall (Guan et al., 2005); however, the effect of altitude on the vertical distribution of rainfall in mountain areas varies depending on the geographic location (Basist et al., 1994).

The influence of topography on rainfall patterns in mountain areas has been statistically analyzed by Basist et al. (1994). They used a simple linear regression equation to examine the relationship of topographic variables (hillslope, elevation, and orientation to prevailing wind) to mean annual precipitation for study areas that represent different mountain climates across the world. Their analysis found that in some locations the influence of elevation, relating to the orographic effect, had a negative correlation to precipitation. In those cases, it is believed that other topographic features, such as fjords, can have an important role in the process of precipitation. Additionally, Basist et al. (1994) found that the most important topographic factor

relating to the spatial distribution of precipitation is the exposure of a mountain slope to prevailing winds. Daly et al. (1994) also examined the influence of elevation on precipitation and noted that the precipitation-elevation relationship is not simply linear, but may have a log-linear or exponential relationship in different situations. Further observations found the importance of aspect and elevation and their influence on incoming radiation, which also affects precipitation distribution, can vary throughout seasons (Guan et al., 2005; Barry, 2008).

2.2.2 Rainfall interpolation

In geostatistics, rainfall is a common example of a regionalized variable. In general, rainfall amount has a strong correlation over short distances (<10 km) that decreases gradually as distance increase (Grimes and Pardo-Igúzquiza, 2010). Although rainfall is a regionalized variable, there are some difficulties in interpolation using geostatistics. Some of the difficulties are related to rainfall data being heteroskedastic – the variance increases as a function of rainfall amount; the spatial structure of rainfall depends highly on variable weather type and geography; and the typical measurement locations of rain gauge data can be insufficient because the observation locations are usually determined by accessibility (Grimes and Pardo-Igúzquiza, 2010). Although these difficulties exists, studies show that when rainfall estimates are based on a low density rain gauge network, geostatistical interpolation performs better than non-geostatistical techniques that do not consider a pattern of spatial dependence, such as the Thiessen polygon and inverse square distance methods (Creutin and Obled, 1987; Goovaerts, 2000).

There are many different geostatistical techniques that have been used for rainfall interpolation (Krajewski, 1987; Goovaerts, 2000; Guan et al., 2005; Haberlandt, 2007; Grimes and Pardo-Igúzquiza, 2010). The main technique for interpolating rainfall with only rain gauge data is ordinary kriging. However, ordinary kriging does not always provide sufficient results because it is only based on the values and locations of rain gauge data. Therefore, multivariate geostatistical techniques are more commonly used, and have been found to perform better (Goovaerts, 2000; Haberlandt, 2007).

The fundamental objective in geostatistics is to interpolate unknown values of a regionalized variable based on the concept of spatial autocorrelation. In order to understand the spatial

autocorrelation of a variable a semivariogram analysis must be performed. In terms of rainfall, semivariogram analysis is particularly challenging because rainfall events can differ in characteristics such as weather type, intensity, duration, and spatial cover. Thus, it is suggested that a semivariogram should be created based solely on observations related to a specific rainfall event. Additionally, the range determining the correlation pattern of rainfall in mountains is expected to be much shorter than for lowland plains. Therefore, it may be useful to construct a semivariogram for geographical sub-regions that represent different topographic areas, such as lowland plains, hills and mountains (Grimes and Pardo-Igúzquiza, 2010). Anisotropy is another issue to consider for semivariogram analysis because rainfall events usually directionally dependent. However, a study by Haberlandt (2007) found that when using radar as a secondary variable for prediction of rainfall using rain gauge data, there was no significant difference in interpolation performance between isotropic and anisotropic semivariograms. It should also be mentioned that there are numerous methods available for estimating the parameters of a semivariogram model such as general least squares, ordinary least squares, maximum likelihood, and Bayesian analysis. In terms of rainfall interpolation, none of the methods appear to enhance performance over another (Grimes and Pardo-Igúzquiza, 2010).

2.3 Landslides and forest harvesting activities

In general, landslide initiation increases in areas with forest harvesting activities (Swanston and Swanson, 1976; Wu and Mckinnell, 1979; Greenway, 1987; Montgomery et al., 2000; Jakob, 2000; Guthrie, 2002; Rickli and Graf, 2009; Goetz et al., 2011). The main activities that impact slope stability are deforestation and construction of roads (logging roads; Table 2.2).

Forests can influence slope stability through evapotranspiration and root cohesion (Sidle et al., 2006). Evapotranspiration alters the soil moisture regime by reducing the amount of water reaching the soil (Greenway, 1987). Also, interception of precipitation by tree canopies can promote evaporation and reduce the amount of water infiltrating into the soil below (Greenway, 1987). In terms of root cohesion, roots may penetrate through the soil mantle to anchor in bedrock; this effect has a greater influence on slope stability for shallow soils than deeper soils (Wu and Mckinnell, 1979; Greenway, 1987).

Table 2.2. Impacts of engineering activities on factors that influence slope stability in steep forest lands of the Pacific Northwest (Altered from Swantson and Swanson, 1976)

	Engineering activities [a]	
Factors	**Deforestation**	**Logging Roads**
I. Hydrological influences		
A. Water movement by vegetation	Reduce evapotranspiration (-)	Eliminate evapotranspiration (-)
B. Surface and subsurface water movement	Alter snowmelt hydrology (- or +)	Alter snowmelt hydrology (- or +)
		Alter surface drainage network (-)
	Alter concentrations of unstable debris in channels (-)	Intercept subsurface water at roadcuts (-)
		Alter concentration of unstable debris in channels (-)
	Reduce infiltration by ground surface disturbance (-)	Reduce infiltration by roadbed (-)
II. Physical influences		
A. Vegetation		
1. Roots	Reduced rooting strength (-)	Eliminate rooting strength (-)
2. Bole and crown	Reduced medium for transfer of wind stress to soil mantle (+)	Eliminate medium for transfer of wind stress to soil mantle (+)
B. Slope		
• Slope angle		Increase slope angle at cut and fill slopes (-)
		Eliminate mass of vegetation on slope (+)
• Mass on slope	Reduce mass of vegetation on slope (+)	Eliminate mass of vegetation on slope (+)
		Cut and fill construction redistributes mass of soil and rock on slope (- or +)
C. Soil properties		Reduce compaction and apparent cohesion of soil used as road fill (-)

[a] Influence that usually increases slope stability denoted by (+); influence that usually decreases stability denoted by (-)

Forest harvesting activities can change the physical structure of soil by increasing bulk density and compaction, and decreasing organic matter content (Huang et al., 1996; Merino et al., 1998). The chance of landslide initiation can vary depending on the forest type, forest age and diameter of trees; smaller diameter trees and younger forests have been found to have higher chance for landslide initiation (Lee and Min, 2001). Furthermore, clear-cutting, which was the only logging technique used in British Columbia until 1997 (Jakob, 2000), destroys the stabilizing influence of vegetation cover and alters the hydrological regime (Swanston and Swanson, 1976). Evidence of this relationship has been explored by studying root area ratios, the proportion of cross-sectional area to soil cross-sectional area. As expected, the proportion of

roots is much lower in clear-cuts and industrial forests (Schmidt et al., 2001). As a result, the strength contributed by the root for stability of hillslopes is less in logged areas (Wu and Mckinnell, 1979; Schmidt et al., 2001).

In terms of temporal initiation of landslides in forest harvested areas, it has been found there is a time lag from clear-cutting to an increase in landslide frequency. This lag time for increase in landslide frequency has been found to be as long as a few to a dozen years (Swanston and Swanson, 1976; Wu and Mckinnell, 1979; Sidle et al., 2006).

Logging roads can form an imbalance of the strength-stress relationship on a 'natural' hillslope by cut and fill activities and poor construction fills, which lead to alteration of surface and subsurface water flow (Swanston and Swanson, 1976). The causes of these roadside failures were attributed to the lack of full bench road construction and inadequate drainage. The logging roads either do not have sufficient cross drains and ditches or are inactive and have not been re-planted (Swanston and Swanson, 1976; Schwab, 1983).

Full bench roads are constructed typically for slopes that have an inclination that is greater than 60°. The excavated material is either pushed to the downslope location of the road or hauled away. The main problem with pushing the material to the downslope side is leaving it unconsolidated and unstable. Typically, a road constructing practice referred to as *endhauling* is used to maintain slope stability, which refers to the removal of excavated material to an approved waste area (BC Ministry of Forests, 2002).

2.4 Land cover classification

The contribution of forest stand characteristics, which may be related to forestry activities, to the initiation of landslides can be explored using land cover classification maps (Lee and Min, 2001; Goetz et al., 2011). Land cover can be defined as the spatial characterization of natural and anthropogenic features on Earth's surface using remote sensing imagery. Thus in general, land cover maps can be used to summarize some of the biophysical and anthropogenic controls on landslides. Schmidt et al. (2001) suggests that utilizing remote-sensing information, such as mapping canopy structure, can be used as a proxy for characterizing the amount and spatial variation of root cohesion for areas that are susceptible to landslides.

There exists a variety of methods for land cover classification. In general, these methods can be categorized into two approaches: supervised and unsupervised classification. Supervised classification relies on a priori knowledge of cover types, which is used for labeling samples associated with land cover classes for use in a classification technique. In contrast, unsupervised classification does not require prior information about cover types; usually, spectral clusters are formed and labeled by cover type after applying a classifier (Cihlar, 2000). Commonly used supervised approaches include maximum likelihood classification (MLC; Defries and Townshend, 1994; Stefanov et al., 2001; Rogan et al., 2002; Cingolani et al., 2004), *k*-nearest-neighbor classification (KNN; Franco-Lopez et al., 2001; Haapanen et al., 2004; Gjertsen, 2007), artificial neuron networks (ANN; Atkinson and Tatnall, 1997; Pal and Mather, 2003) and decision tree classifiers (DT; Friedl and Brodley, 1997; Rogan et al., 2002; Pal and Mather, 2003). The most commonly used unsupervised approach is the clustering algorithm known as ISODATA (Sader and Winne, 1992; Cohen et al., 1998; Wilson and Sader, 2002; Barnett, 2004).

The standard method when performing supervised image classification for remote sensing is MLC (Arbia et al., 1999; Pal and Mather, 2003). MLC is often not much different, in terms of performance, than methods such as decision tree classifiers and artificial neural networks for land cover classification (Pal and Mather, 2003). The choice of which approach to use is based on previous knowledge of the study area. In general, if the desired land cover classes are already known and there is good knowledge of the where they will occur, in terms of sampling, a supervised method is preferred. An unsupervised method is appropriately applied when mapping a large area that is relatively unknown (Cihlar, 2000).

In performing a classification it is important to consider the number of land cover classes. Typically, the accuracy related to the classification can decrease as the number of classes increases (Cohen et al., 1995).

Past studies have been able to characterize the general age of a forest based on spectral information from a thematic mapper (TM) tassel cap (TC) transformation (Cohen and Spies, 1992; Cohen et al., 1995, 1998). Specifically, the TC transformation can be used to estimate forest cover types (e.g. open canopy forest, or closed forest) using Landsat TM imagery (Cohen et al., 1995).

The TC is an orthogonal transformation of original Landsat data that creates three new components: brightness, greenness, and wetness (Crist et al., 1986). Different forest conditions

can be classified based on spectral response related to brightness, greenness and wetness. In mountainous regions, the greenness and wetness are highly sensitive to changes in topography; however, the TC transformation can still capture most of the spectral variability required to distinguish between cover types (Cohen and Spies, 1992; Cohen et al., 1995).

Cohen et al. (1995) describe how the TC transformation can be used to classify forest conditions into exposed surface (or recently logged forest), open canopy forest, semi-open canopy forest and closed canopy forest. This study focused on forests located in the Pacific Northwest region of the United States (just south of Vancouver Island). The majority of spectral variation to distinguish between forest conditions can be captured with brightness and greenness (Cohen and Spies, 1992). A severely disturbed forest stand (e.g., clear-cut) has, relatively, the lowest wetness, greenness, and the highest brightness values. Semi-open forest (increasing green vegetation), captures higher wetness and greenness values. Closed forest stands are associated with the highest greenness, relatively high wetness and moderate brightness values.

2.5 Landslide susceptibility with GAMs and GLMs

Empirical techniques for landslide susceptibility modeling, such as logistic regression analysis, can provide insight regarding the presence or absence of a response variable (e.g., landslide initiation) to changing predictor variable values (e.g., environmental factors). Since the predictor variables for landslide initiation can be represented spatially, it is possible to predict the spatial distribution of landslide susceptibility. Logistic regression, a generalized linear model (GLM), is a method for prediction of a binary response variable (e.g., the presence or absences of landslides) by utilizing the logit transformation (more detail in Section 4.5.1). Logistic regression has been found to perform more adequately than machine-learning models, such as support vector machines and tree classifiers, which are more likely to overfit to the data (Brenning, 2005). However, GLMs lack the ability to properly represent non-linear effects that are known to exist in many environmental geomorphological analyses (Phillips, 2003; Brenning et al., 2007; Brenning, 2009; Goetz et al., 2011). Only recently have these nonlinear effects been modeled using nonlinear regression techniques, such as generalized additive models (GAM), for geomorphological distribution models in complex terrain (Brenning, 2009) and landslide

susceptibility models (Goetz et al., 2011; Park and Chi, 2008). A GAM is an extension of a GLM that can represent covariates as linear or non-linear (Hastie and Tibshirani, 1986).

The relationships of the landslides with the environmental factors can be drawn with inference by assessing the relative contributions each factor has for explaining landslide initiation. The primary purpose of using a GAM or GLM is to learn more about the processes causing landslides by creating a model that describes the dependence of landslides on the environmental factors. The subsequent knowledge gained can be used to better predict landslide susceptibility using a set of environmental factors (Hastie and Tibshirani, 1990).

2.5.1 Uncertainty

It is very well understood that the quality of the data used for modeling will be reflected in the final product, as the computer axiom states, "garbage in, garbage out". Data quality is a term that can be used to describe the general capacity of data to assist in an analysis. Key components of data quality are uncertainty and suitability. Relating to physical geography, uncertainty can be defined as, "an expression of our inability to resolve a unique, causal, world either in principle or in practice" (Brown, 2004). This definition can be directed to attempts, by scientists, to model the world around; specifically, uncertainty is used to describe what we do not know. Suitability is the ability of data to be adequately applied to a specific problem. For example, data created for one purpose may not be appropriate for applying to another.

There are many elements of uncertainty. Rowe (1994) describes these elements in terms of temporal, structural, metrical and translational uncertainties. Temporal uncertainty is related to likelihood of a future event occurring; we may understand a probability of a future event, but how confident can we be in our prediction (Rowe, 1994)? Structural uncertainty is related to what and how many parameters or variables are selected to model a situation and its complexities (Rowe, 1994). Metrical uncertainty is related to the accuracy and precision of measurements of values of variables or parameter attributes (Rowe, 1994). Translational uncertainty is a combination of the other uncertainties, which relates to the ability to explain the uncertainties and how they are interpreted (Rowe, 1994).

These concepts can be applied to geographical models, including modeling of landslide susceptibility, to better identify, understand and describe the range of model uncertainties in an

attempt to reduce translational uncertainty; as a result, decision-makers may more confidently apply a supporting model.

2.5.2 Spatial data

Spatial data collected for predictive modeling of landslides contains information that characterizes the general conditions influencing initiation of an event (Chung and Fabbri, 1999). Van Westen et al. (2005) suggests that the spatial data required for landslide susceptibility, hazard and risk assessment can be subdivided into the following groups: landslide inventory data, environmental factors, triggering factors, and elements at risk. Since this is a landslide susceptibility modeling study, it will focus on the environmental factors and triggering factors of landslides.

A landslide inventory is a collection of spatial and temporal information corresponding to individual classes of landslides. Typically, a landslide hazard study should begin by making a landslide inventory (van Westen et al., 2008). Issues regarding data quality can be summarized by topics of scale, accuracy and precision and temporal relevance. In terms of scale, aerial and satellite image interpretation is the most common technique for mapping landslides (Tribe and Lier, 2004). High-resolution optical images (e.g., Landsat TM/ETM+, SPOT) are useful for mapping many large landslides. However, investigation of a single landslide event should rely on very high-resolution imagery (e.g., QuickBird, IKONOS). However, the high costs of very high-resolution imagery may considerably limit the use, particularly for multi-temporal analysis (van Westen et al., 2008).

In spatial databases, the accuracy and precision of locating and defining the boundaries of landslides is critical for the further investigation of landslide hazards. The uncertainty in the mapping of landslides may be related to the expertise of the image interpreter or surveyor and a lack of sufficient historical data, such as precise location, time and classification of a landslide (Carrara, 1993; van Westen et al., 2008). Issues regarding time and location can be overcome by acquiring imagery most recent to the event under study. Otherwise, it is possible that some landslides may be missed because changing environmental factors, such as land-use (Carrara, 2008); some researchers have relied on further field investigation and interviews with local residents to improve the reliability of landslide information (Carrara et al., 2003).

The temporal relevance of the landslide inventory is very important for modeling susceptibility. In general, most landslide susceptibility models are based on an approach that relies on precedent conditions to characterize areas susceptible to landslides. However, there is high temporal and spatial uncertainty associated with these methods when applied outside of the precedent conditions used to establish the model. As a result, it is common that landslide hazard assessment is not based on precedent conditions, but on theoretically determined causative factors (Dai and Lee, 2002). Further complicating this issue is that it can be difficult to obtain a temporal database of landslides. To obtain this data, one must map landslides after a particular triggering event (Guzzetti et al., 1999). Therefore, a researcher is often left to analyze susceptibility using a landslide inventory where the dates of the slides are only roughly understood, which makes it difficult to draw confident conclusion regarding the conditions that initiated the slides.

Exploring the spatial relationship of environmental factors to landslide locations can be used to assess landslide susceptibility. In an empirical model, the environmental factors are usually represented by spatial variables of a model. Triggering factors are also important for understanding where landslides will occur. A susceptibility model can integrate triggering factors into its analysis by representing them as independent spatial variables. For example, the influence of precipitation on sliding may be explored by examining maps of average annual rainfall. The accuracy and precision of spatial data for landslide factors is just as important as the landslide inventory. Often, not enough attention is given in landslide literature to describe the sources of error and uncertainties related to data acquisition and manipulation (Guzzetti et al., 1999).

Another issue with environment factors is the classic statistical problem of finding a close association between variables without a process-based relationship (Gritzner et al., 2001), since trying to capture all the variables involved in complex geomorphological processes can be very difficult and time consuming. In practice, it may be favorable to accept the associations between variables in return for prediction improvement. This is especially important in statistical modeling, where there is flexibility in the input data to provide variables representing, or serving as proxies for, observed factors that influence sliding.

The temporal variation of landslide factors is a major limitation in landslide susceptibility modeling. As discussed above for landslide inventory data, predictive landslide models assume

that landslides in the future will occur under the present or past relationships used in an analysis. For factors such as bedrock lithology, structure and morphology, these assumptions may be correct. However, temporally variable data, such as land-use and precipitation patterns, are much more difficult to apply these stationary assumptions (Dai et al., 2002). Furthermore, strong assumptions that the conditions are stationary, as opposed to allowing for temporal variability, may lead to erroneous predictions (Guzzetti et al., 1999).

Training and validation of landslide susceptibility models depends on the quality of inventory data and the established relationships to environmental factors and triggering factors (Dai et al., 2002). Thus, it is important that the spatial data issues are transparently discussed.

2.5.3 Model assessment

Assessment of predictive models can be simply understood as a quality assessment of model performance. Without some sort of assessment of model performance, the susceptibility model is practically useless for decision-makers (Chung and Fabbri, 2003). The quality of a susceptibility model can be assessed in terms of the reliability, robustness, degree of fitting, and prediction skill (Guzzetti et al., 2006). The prediction skill can be determined by using a suitable performance measure and estimation method. In order to assess the reliability and robustness of a model further methods of evaluation are required and will be discussed below.

The performance of landslide susceptibility models depends on inputs, especially the landslide inventory. The spatial pattern of susceptibility maps can vary depending on the different inventories (Blahut et al., 2010). Thus, an evaluation method should be selected to account for the possible variability in results relating to the selection of training and test data. Holdout method, random subsampling, *k*-fold cross-validation, and bootstrap are common methods for producing estimations of error (Brenning, 2005; Hand, 1997).

The holdout method is the simplest of the error estimation techniques. This method randomly partitions the data set into a *test* set and *training* set; usually, the training set consists of two-thirds of the data set. The training set is used to build the model, while the test set is set aside to estimate the accuracy.

Random subsampling is simply the holdout method for repeated k times. The overall estimation of model performance is based on the average of the error estimates calculated for the repetitions.

In order to utilize the entire data set for training and testing the model, the modeling technique can be evaluated using k-fold cross validation. This method divides the data set randomly into k subsets of equal size. The model is trained using $k - 1$ subsets and tested on the remaining subset. This process is repeated while rotating the position of the test set k times.

The bootstrap is resampling-based method that can utilize an entire data set. The bootstrap method is a non-parametric estimation of error that draws independent samples with replacement from available data. It can be repeated k times and the overall error estimations can be measured as the median of error estimates from each repetition. In addition, the repeated resampling can be used to provide estimates of variability in performances.

A fundamental problem with typical model validation sampling methods, which rely on spatially random selection of samples for test and training sets, is the lack of the ability to use a landslide distribution sample for modeling and apply it for the general distribution of landslides in an area (Brenning, 2005). This problem arises if the samples from the training and test set are only separated by small distances. Consequently, the error estimates may be overoptimistic due to the spatial dependencies between the two sets. Brenning (2005) proposed that this issue can be overcome by using a *spatial* cross validation method, where the training and test sets are spatially partitioned. This approach has been applied successfully in Brenning et al. (n.d.) where it is confirmed that simple spatial random samples produced overoptimistic error estimates.

Guidelines for acceptable model performance are a subject matter that is seldom looked at in landslide susceptibility modeling literature. Guzzetti et al. (2006) attempted to explain the requirements for acceptable model performance. They suggest that an overall degree of model fit greater than 75% is ‘acceptable’ and 80% is ‘very satisfactory’. In their case, degree of model fit is defined by the percentage of correctly classified landslides (true positives). If the model fit is greater than 90% then the degree of model fit becomes questionable; the model predictions may be too specific to the original landslide inventory, which may be a case of over-fitting.

The proposed guidelines for model performance by Guzzetti et al. (2006) may be appropriate for hard-classifiers that simply predict the presences or absence of landslide initiation. However, many predictive modeling techniques, such as GLMs and GAMs, predict the

probability of a landslide event to occur (Brenning, 2005). The performance of these probabilistic methods can be estimated by measuring the sensitivities and specificities of specific probabilities to predict landslide initiation (Brenning, 2005).A receiver-operating characteristic curve (ROC) can be used to represent the estimates of sensitivities and specificities (Zweig and Campbell, 1993). ROCs are a plot of sensitivity (y-axis) and specificity (x-axis). In terms of landslide analysis, sensitivity is the percentage of correctly classified landslide points and specificity is the percentage of correctly classified non-landslide points. The overall model performance can be determined by calculating the area under the ROC curve (AUROC), which is a method that does not depend on the spatial density of landslides. In addition to the AUROC, sensitivity at high specificity can be calculated to assess the ability of a model to predict landslide initiation with detail. In general, the area delineated as unsafe or unstable should be small to reflect the typical low density of landslides (Goetz et al., 2011).

2.6 Summary

Modern landslide susceptibility analysis investigates the predisposition of landslides to occur based on controlling environmental factors. These relationships can be modelled using novel statistical techniques, such as the GAM, that allow for modeling of nonlinear relationships.

There are many inherent uncertainties associated with geospatial analysis using multiple data sources. Appropriate model assessment is vital for communicating these model uncertainties to improve confidence in interpretation of model results.

Rainfall and land cover, related to forest harvesting, are important control on landslide initiation. There are a variety of methods that can be used for model patterns of rainfall. However, geostatistical techniques are shown to produce the most promising results. Land cover classification is a long studied topic with many available methods that usually perform similar. One such method, MLC, is a standard approach that has been proven to produce regular adequate results.

Data acquisition and processing of important landslide controls and incorporating them into landslide susceptibility modeling can be used to explore, enhance and confirm existing knowledge of processes leading to landslides.

Chapter 3

Physical Geography of Vancouver Island

3.1 Physiography and geology

Vancouver Island (31 788 km^2) is located off the west coast mainland of British Columbia, Canada. The west coast of the island is defined by its fjord landscape, while the central valley contains many "finger lakes" formed from deep glacial scouring. Most of Vancouver Island is made up of a mountain range referred to as the Vancouver Island Ranges, which are a sub-range of the Insular Mountains that run along the Pacific Coast and include the Queen Charlotte Mountains. The Vancouver Island Ranges have peaks of approximately 1000 m to 2200 m above sea level. The island landscape has been heavily modified during Pleistocene glaciation (Muller, 1977).

The lithology of the landscape can be generalized into formations of igneous (intrusive and volcanic), metamorphic and sedimentary rock (Figure 3.1). Central Vancouver Island is occupied by volcanic and intrusive rocks that follow the Vancouver Island Ranges (Muller, 1977). Sedimentary rocks are found mainly along the coastal Nanaimo lowlands located along south-eastern coastline. This area, the Nanaimo formation, is made up of undivided sedimentary rocks. Other sedimentary rocks, such as limestone deposits, can also be found in the Quatsino formation that runs parallel to the Vancouver Island Ranges from north of the Holberg Inlet to Nootka Sound.

The general slope stability characteristics can be characterized by lithologic rock classes. The intrusive rocks on Vancouver Island typically support the steep slopes of the Insular Mountains. They are coarse grained and made up of durable minerals (quartz, feldspars) that are relatively resistant to weathering, but still subject to mechanical breakdown. Slope stability is controlled by lines of weakness. The most unstable areas are major joints or faults that are

adjacent to gullies or valleys (Pike et al., 2010). Typically, mass movement on these slopes results in coarse, bulky colluvial slopes. The volcanic rocks, such as basalts, are finer grained. This rock class is highly subject to mechanical weathering, such as freeze-thaw, because water can easily infiltrate into the rock joints. Also, the minerals in basalts are particularly susceptible to chemical weathering. Landslides may occur on layered volcanic sequences that have a layer of severely weathered rock material or clay residues (Pike et al., 2010). Metamorphic rocks can be extremely resistant to weathering, while sedimentary rocks are subject to solution by acidic water; as a result, karst formations that are common on central and northern Vancouver Island, are subject to collapse, which can form steep depressions in the landscape (Pike et al., 2010).

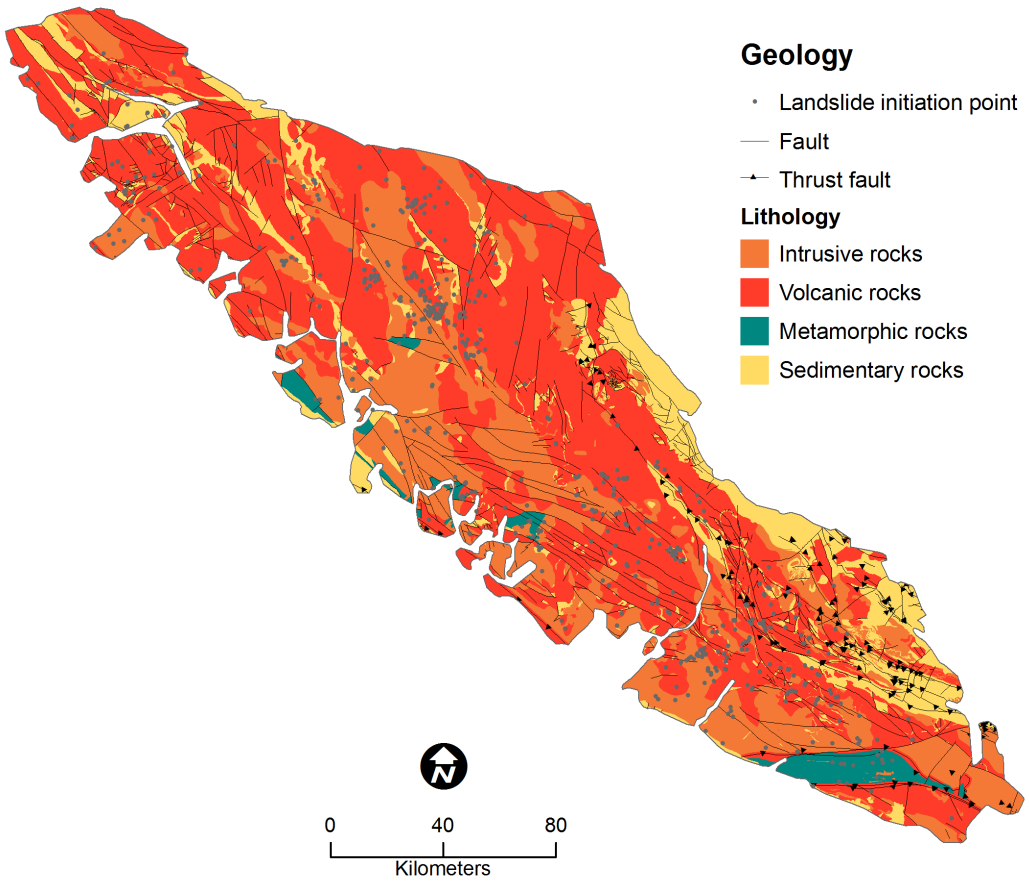

Figure 3.1. Surface geology map of Vancouver Island (Data from Massey et al., 2005)

3.2 Climate

The west coast of Vancouver Island has some of the highest annual precipitation amounts in Canada (McKenney et al., 2006). The pattern of rainfall is typical of coastal mountain ranges located on the Pacific Ocean. The mean annual precipitation ranges from 800-1200 mm along the east coast (mountain shadow) and increases towards the west coast (windward side) to more than 3000 mm (Figure 3.2; McKenney et al., 2006). Precipitation on Vancouver Island is most abundant from early autumn to midwinter. During this period the prevailing wind is from the south-southwest (Basist et al., 1994).

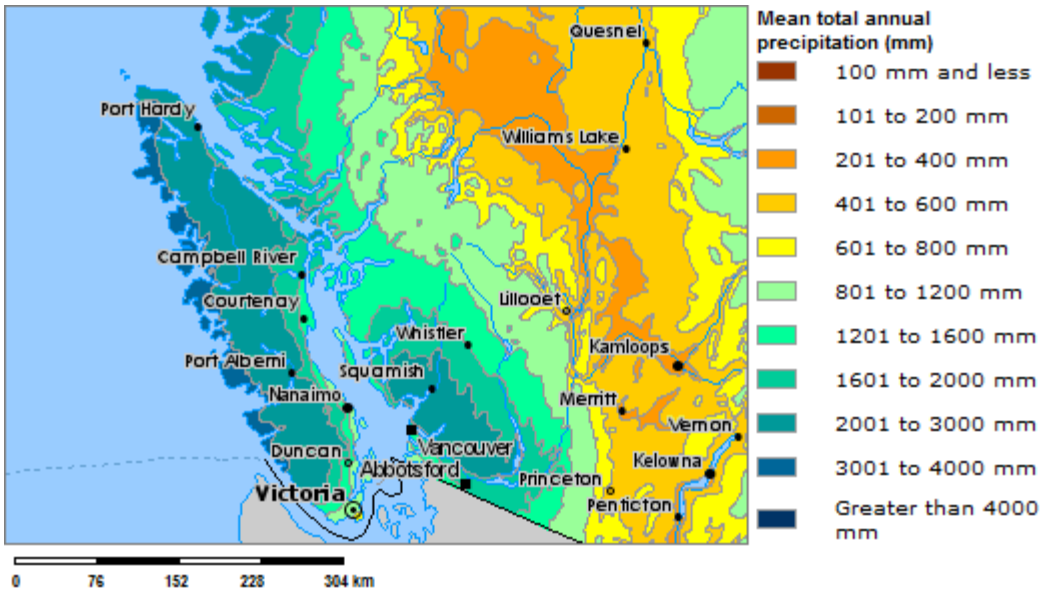

Figure 3.2. The mean total annual precipitation from 1971 to 2000 (McKenney et al., 2006; altered from Natural Resources Canada, 2012)

In general, patterns of rainfall are controlled by land surface type, topography, surrounding oceans, large-scale circulations and thermodynamic conditions. The El Niño Southern Oscillation (ENSO) influence patterns of rainfall occurring in British Columbia. ENSO is a climatic pattern in the tropical Pacific Ocean that occurs every 2-7 years, and lasts from 12-15 months (Biggs, 2003). The El Niño period, which is associated with ENSO, causes warming of sea-surface temperature in the eastern tropical Pacific Ocean. The consequent moist tropical air that collects over the western United States causes British Columbia to experience warmer air

and greater rainfall. In contrast, the other period of ENSO, La Niña, is related to much cooler air temperature and less rainfall in British Columbia (Biggs, 2003).

3.3 Biogeoclimatic zones

Biogeoclimatic zones, which have been developed by the British Columbia Ministry of Forests, are useful for summarizing geographical areas that have similar climate, soil and vegetation characteristics (Meidinger and Pojar, 1991). Vancouver Island is comprised of 4 of these zones: Coastal Western Hemlock, Mountain Hemlock, Coastal Douglas-fir, and Alpine Tundra (Figure 3.3).

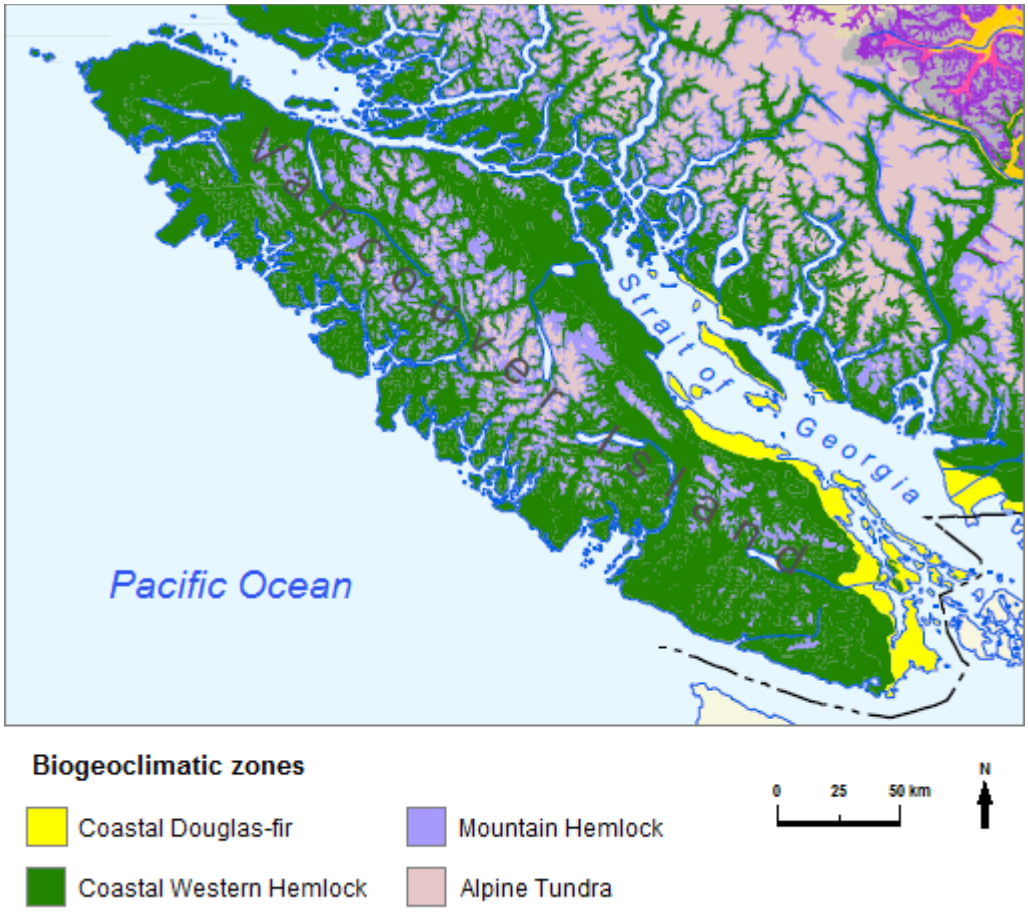

Figure 3.3. Biogeoclimatic zones of Vancouver Island (altered from British Columbia Ministry of Forests, 2012)

Vancouver Island is predominantly covered by the Coastal Western Hemlock (CWH) zone (Figure 3.3). This zone is prevalent from sea level to elevations from 900 m (windward slopes) to 1050 m (leeward slopes; Figure 3.4; Meidinger and Pojar, 1991). The average monthly temperature of 8°C ranges from 5.2°C to 10.5°C. Mean annual precipitation ranges from 1000 mm to <4400 mm. The amount of precipitation occurring as snowfall can be as little as 15% in southern regions (Meidinger and Pojar, 1991). The most common tree species in this zone is the western hemlock, but also includes Douglas-fir and amabilis fir, which can be found at upper elevations (Meidinger and Pojar, 1991; Krajina, 1969). It is common to find red alder wide spread in disturbed sites (e.g. logged areas) in the CWH zone (Meidinger and Pojar, 1991).

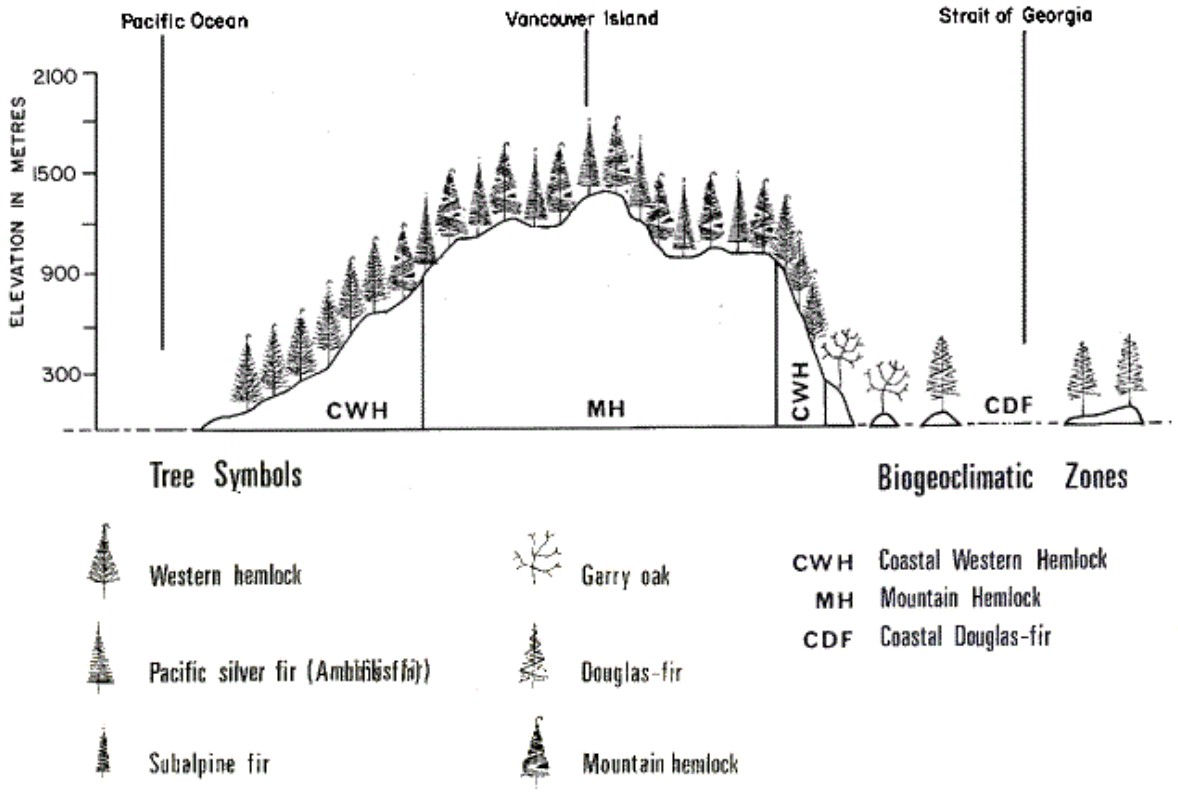

Figure 3.4. Pattern of vegetation across Vancouver Island (altered from Krajina, 1969)

The Mountain Hemlock (MH) zone, which is found at elevations between 900 m to 1800 m a.s.l., is located on the Insular Mountains of Vancouver Island above the CWH zone (Figure 3.3). The mean annual temperature ranges from 0°C to 5°C. The mean annual precipitation, which can

occur 20% to 70% as snowfall, varies from 1700 mm to 5000 mm (Meidinger and Pojar, 1991). The forest cover of the MH zone is predominately comprised of mountain hemlock, amabilis fir and yellow cedar. At higher elevations forest cover thins out because of a shorter growing season, increased duration of snow, and cooler temperatures (Meidinger and Pojar, 1991).

The Alpine Tundra (AT) zone is located along the highest mountain peaks of the Insular Mountains on Vancouver Island at elevations above 1650 m (Meidinger and Pojar, 1991). This is the coldest zone, with mean annual temperature from -4°C to 0°C. The mean annual precipitation of 700 mm to 3000 mm typically occurs as snowfall (Meidinger and Pojar, 1991). This zone is predominately treeless.

The Coastal Douglas-fir (CDF) zone is located in the rainshadow along a small segment of the southeast Vancouver Island adjacent to the Strait of Georgia (Figure 3.3 and Figure 3.4). This zone occurs mostly below 150 m a.s.l. (Meidinger and Pojar, 1991). It is the warmest zone with a mean annual temperature form 9.2°C to 10.5°C. The mean annual precipitation ranges from 647 mm to 1263 mm and predominately occurs as rainfall (Meidinger and Pojar, 1991). The CDF zone has experienced heavy logging during the early 20th Century, with old growth forests remaining only in parks (Meidinger and Pojar, 1991). The most common tree species is Douglas-fir; however, the tree cover type varies significantly across the zone, which is believed to be related to human disturbances (Meidinger and Pojar, 1991).

3.4 Common landslide types

On Vancouver Island, the most common triggering mechanism is precipitation and snow melt. Also, seismic activity has been known to cause landslides in this area (Hodgson, Ernest, 1946; Mathews, 1979; Rogers, 1980; VanDine and Evans, 1992).

The most common landslide types on Vancouver Island are debris slide and debris flows. In general, flows can be defined as a landslide that consists of individual movement of particles with a moving mass (Dikau et al., 1996). Debris flows usually occur on slopes that are made up of a thin layer of unconsolidated material. Thus, flows are composed of an assortment of fine material (sand, silt, and clay), coarse material (gravel and boulders), and organic material that is mobilized into a slurry moving down slope. The movement of a flow generally follow the path of

an existing channel or gully, where deposition from previous events may be picked up as an addition to the current flow of material (Dikau et al., 1996).

Slides refer to movement of material along an identifiable shear surface (Dikau et al., 1996). They are grouped by type of movement as rotational or translational. Rotational slides have a somewhat rotational movement along an axis parallel to the ground surface. The sliding (slumping) occurs along a concavely upward failure surface (Varnes, 1978). In contrast, translational slides are more or less a planar failure that is influenced by discontinuities such as faults, bedding planes, thrusts and deposits (Dikau et al., 1996).

Debris slides on Vancouver Island are typically shallow, occur on steep slopes and move rapidly. Initiation, which is responsive to rainfall, of debris slides occurs typically in concave hollows or seepage zones: an area of hillslope where seepage is concentrated. These are both areas where hydrostatic pressure may increase. Debris slides usually occur in till, colluvium and fluvially deposited sediments; also they can occur in Folisols (upland organic soils), which are typically located in the north and central coast of Vancouver Island (Pike et al., 2010). Debris flows on Vancouver Island are predominately triggered by an initial failure of a debris slide, which can occur on the gully sidewall or headwall (Brayshaw and Hassan, 2009). Also, debris flow initiation is much more likely to occur in steep channels than low gradient channels (Brayshaw and Hassan, 2009). In general, the amount of collected sediment in a channel relates to the size of slope failure required to initiate a flow; small slope failures can trigger debris flows in channels with little sediment collection, and large slope failures are required to initiate debris flows in channels with greater sediment collection. As a result of this relationship, it seems to be easier for debris flows to be triggered after a debris flow event that leaves little in-channel sediment collection (Brayshaw and Hassan, 2009). However, since the frequency of the debris flow events is increased in channels with low sediment collection, the magnitude of the event also decreases.

3.5 Forest harvesting

Vancouver Island has a history of landslides relating to forest harvesting activities (Rollerson, 1992; Rollerson et al., 1998; Jakob, 2000; Guthrie, 2002; Guthrie and Evans 2004; Chatwin; 2005, Guthrie, 2005). Logging roads appear to have a greater influence on landslide

activity than clear-cutting (Swanston and Swanson, 1976; Guthrie, 2002; Schwab, 1983). Guthrie (2002) examined the impact of logging roads and deforestation on landslide density (landslides per unit area) of three watersheds on Vancouver Island: Macktush Creek, Artlish River and Nahwitti River. It was found that in general logging activities substantially increased the number of landslides that occurred. Some watersheds experienced up to 16 times more landslides following forest harvesting. Logging roads have been found to increase landslide density up to 94 times compared to 'naturally' forested areas (Guthrie, 2002). Jakob (2000) found in the Clayqout Sound that 49% of the landslides were related to logging activities. In particular, it was found that more areas are affected by landslides in logged areas because of larger landslide densities – not larger landslides. In comparison, an earlier study had similar findings in Rennell Sound on Queen Charlotte Island, British Columbia, which is located just northwest of Vancouver Island (Schwab, 1983). Following a large rain storm in Rennell Sound, Schwab (1983) investigated factors, other than rainfall, that contributed to landslides on the island. It was found that the frequency of landslides per unit area is less in forested areas, greater in clear-cut areas, and the greatest near logging roads; however, it was noted that newly constructed roads had a reduction in landslide initiation.

3.6 Mapping landslide susceptibility

Forestry management policy and practices in British Columbia have been established to reduce associated increases in landslide activity (Chatwin, 2005). Thus, there is a need to provide landslide hazard information for forest practices (Schwab and Geertsema, 2008). Typically in British Columbia, areas that were more prone to landslides were mapped using a heuristic approach based on the knowledge of professional geoscientists (Chatwin, 2005). This approach relied on the British Columbia Terrain Classification System, which categorizes units of terrain that have similar slope, surficial material and slope morphology (e.g., curvature; Howes and Kenk, 1997; Schwab and Geertsema, 2008). The quality of this classification is limited by the subjectivity involved in the drawing of terrain unit boundaries (Rollerson et al., 1998). Furthermore, some of these maps assigned susceptibility to landslides based only on a qualitative description of landslide activity in each terrain unit (Chatwin, 2005). More recently the terrain mapping approach has been modified to classify landslide susceptibility using quantitative

methods that define landslide susceptibility based on the presence and density of landslides in each terrain unit (Rollerson et al., 2002; Chatwin, 2005; Guthrie, 2005).

Statistical methods for modeling landslide susceptibility on Vancouver Island have also been studied (Chung et al, 2001; Goetz et al, 2011). Chung et al. (2001) applied a Bayesian probability method using topographic attributes derived from a digital elevation model, surface geology, and biogeoclimatic zones. Goetz et al. (2011) investigated enhancing landslide susceptibility modeling by integrating physically-based landslide models, with a GAM that utilized terrain attribute information and land use characteristics related to forest harvesting. Both studies highlighted the ability to improve the detail of landslide prediction by using statistical methods that classify susceptibility for individual cells in a raster dataset.

3.7 Summary

The landscape of Vancouver Island is dominated by mountains and lithology predominately composed of intrusives and volcanics. Annual precipitation can vary from 800-1200 mm on the leeward side of the Insular Mountains to >3000 mm on the windward side. Precipitation is most abundant from mid-autumn to early winter, which is related to winds coming from the south-southeast direction. The most common tree species is western hemlock. The treeline is around 1650 m a.s.l.

Debris slides and debris flows are the most common type of landslide occurring on Vancouver Island. Initiation of these landslides is typically associated to rainfall. In addition, difference in mechanical slope failure can be partially attributed to different lithology classes. Also, the impacts of the forestry and logging activities on landslide initiation have been well documented. Past studies show freshly-cut forest and logging roads are associated with a higher frequency of landslide initiation on the island.

Landslide susceptibility maps have been applied for the management of forest practices in British Columbia to reduce the associated impacts on landslide activity. A variety of methods have been applied to predict landslide susceptibility, which include qualitative as well as quantitative approaches.

Chapter 4

Methods

4.1 Landslide inventory

A landslide inventory of 639 debris flow and debris slides polygons mapped for landslides that occurred in the winter of 2006-2007 was provided by the British Columbia Ministry of Environment. The majority of these slides have been cited as occurring during a storm on 15 November, 2006 (Guthrie et al., 2010b).

These landsides were mapped from 5 m spatial resolution SPOT satellite imagery by using an automated change detection method comparing scenes from the summer (May to September) of 2006 to the summer of 2007, which is the season when landslides are least frequent (Guthrie et al., 2010b). Some errors in building this inventory using change detection include positional shift and alignment errors, shadows and cloud cover associated to the SPOT imagery (Guthrie et al., 2010b).

Initiation points digitized from the landslide polygons were used for the subsequent landslide susceptibility analysis (Figure 4.1). The location of an initiation point was digitized where the main scarp may be expected. These initiation points were used to approximate the environmental conditions that led to slope failure.

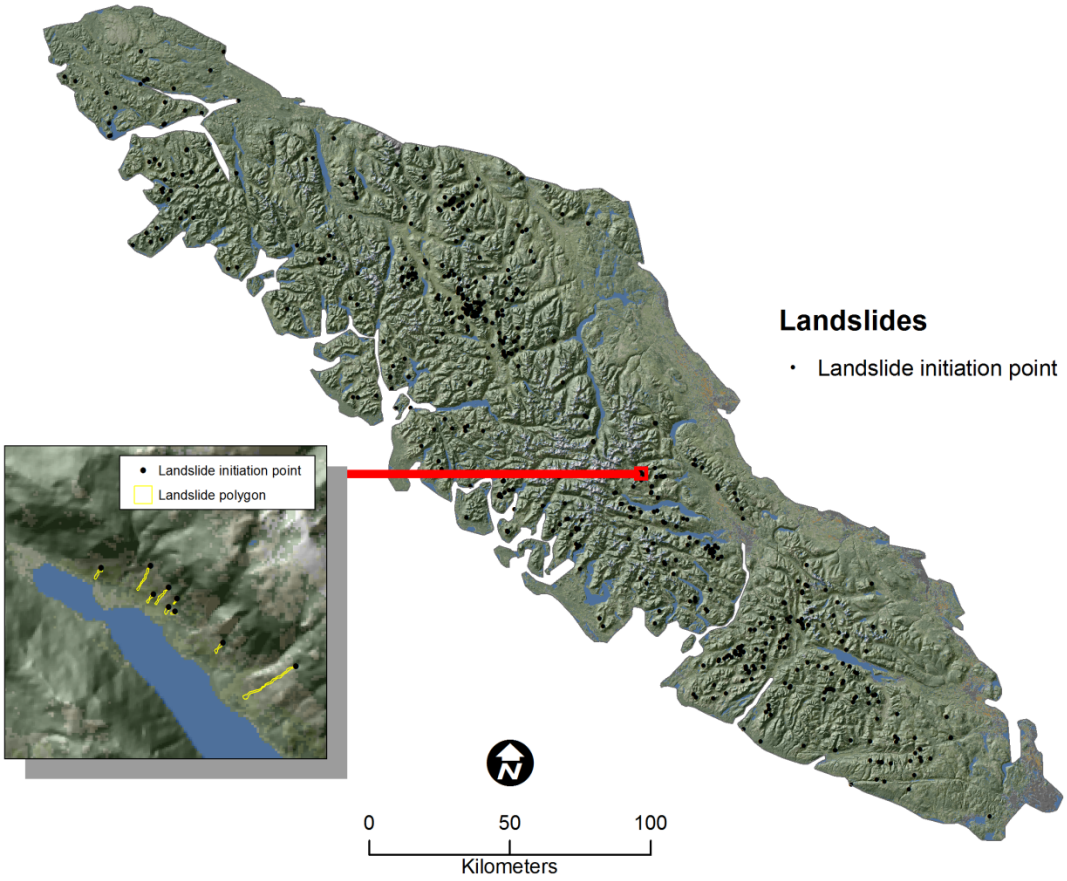

Figure 4.1. 2006-2007 landslides (debris slides and debris flows) on Vancouver Island

4.2 Rainfall interpolation

The purpose of the rainfall interpolation is to explore the relationship between locations of landslide initiation to patterns of rainfall. The influence of rainfall on 2006-2007 landslides was investigated by comparing rainfall interpolation for different temporal scales: two weeks of rainfall leading up to an extreme weather event within the temporal scale of the landslide inventory (two weeks); the winter months pertaining to estimated temporal span of the slides (winter); and the annual rainfall for 2006 (annual). The months included for winter rainfall were October 2006 to February 2007; March is not included because many weather stations across Vancouver Island were shutdown indefinitely that month. The association of rainfall with landslide initiation for these time periods were compared to determine which temporal scale

provides the 'best' information for prediction of landslide susceptibility given this landslide inventory.

4.2.1 Weather station data

Weather station data was compiled from different sources in order to have sufficient coverage of records of rainfall accumulation across Vancouver Island. The data was provided by government sources: Environment Canada, BC Ministry of Transportation and BC Ministry of Forest, Lands, and Natural Resource Operations. The National Climate Data and Information Archive of Environment Canada is available freely online and houses daily and monthly rainfall accumulation values (Environment Canada, 2011). The BC Ministry of Transportation provides hourly rainfall accumulation for stations that are mainly adjacent to major highways; it is also available freely online (BC Ministry of Transportation and Infrastructure, 2011). The Wildfire Management Branch of BC Ministry of Forests, Lands, and Natural Resource Operations provided hourly rainfall data (BC Wildfire Management Branch, 2011).

Altogether, 53 stations across Vancouver Island were used as a basis for the rainfall interpolation (Figure 4.2). These stations range in elevation from 0 m to 580 m above sea level. Since some of the rainfall data had gaps in data, a threshold for accepting a weather station for interpolation was decided. The threshold for removing a station was if 3 or more days were missing (weeks), or 3 or more months were missing (winter and annual). Consequently the stations that were not included in analysis were: two weeks – Menzies Camp; winter – Mesachie DL and Saanichton CDA; annual – Mesachie DL, Saanichton CDA, TS Effinghanm, TS San Juan, and TS Naka Creek.

The data from the BC government was available in hourly rainfall and Environment Canada data was available for daily amounts. Thus, these data sets were aggregated into weekly and monthly rainfall by calculating the sum of rainfall for a given period.

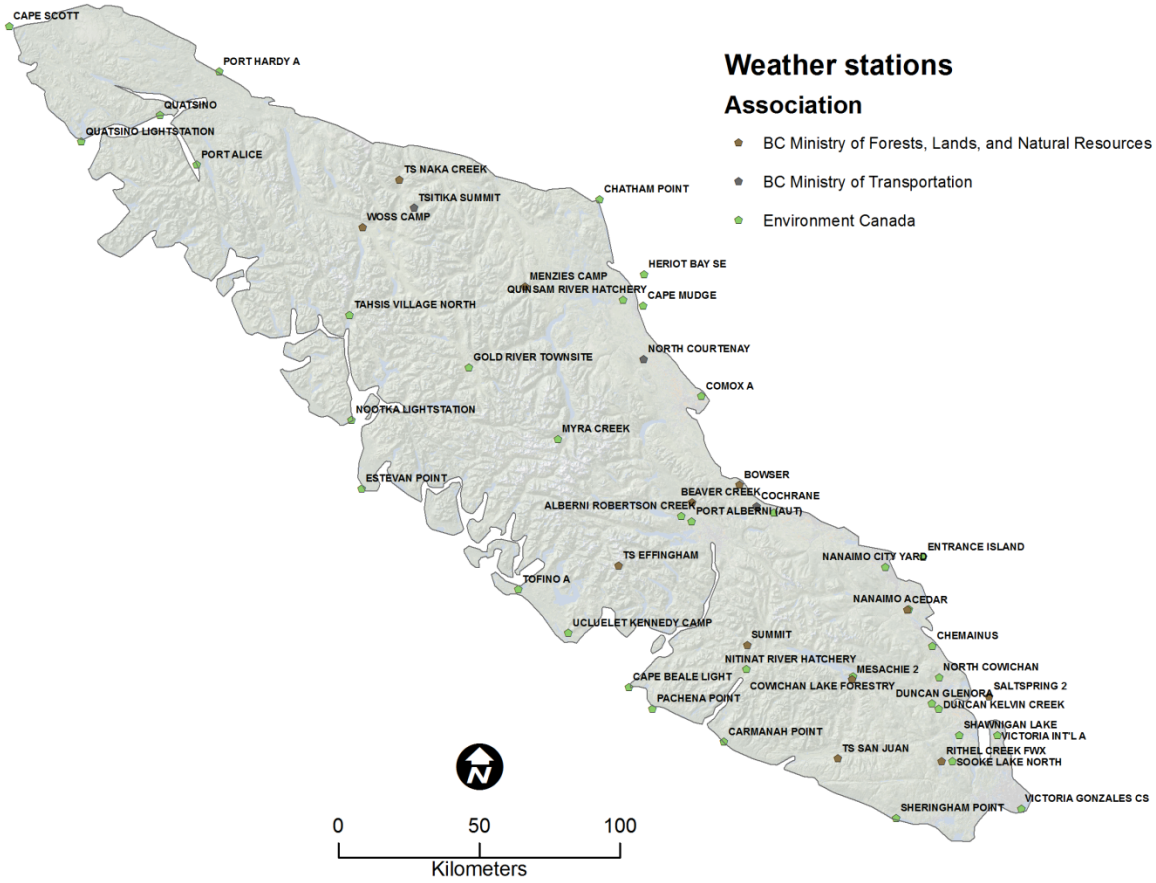

Figure 4.2. Map of weather stations and related associations used for interpolation of rainfall: BC Ministry of Transportation; BC Ministry of Transportation and BC Ministry of Forest, Lands, and Natural Resource Operations; and Environment Canada.

4.2.2 Storm analysis

The rainfall event chosen for this study was an extreme storm that occurred around November 15, 2006. Southwestern Vancouver Island was hit the hardest in terms of high rainfall intensity and accumulation that resulted in major flooding and landslides. This storm exceeded the Rainfall Frequency Atlas of Canada 100-year return period daily rainfall records for Vancouver Island (Forest Practices Board, 2009). It has been found that there was an increase in landslide activity following the large rainstorm event (Forest Practices Board, 2009). A detailed analysis of this weather event and the related landslides has also been completed by Guthrie et al. (2010). The highest recorded daily rainfall for this event was 126.4 mm at Port

Alberni. An analysis of maximum daily rainfall for weather stations across Vancouver Island from 2004 to 2007 confirms the November 2006 event as having the highest amount of rainfall in that period. Maximum daily rainfall amounts were examined using the Canadian Daily Climate Data (CDCD; Environment Canada, 2011), which is compiled and distributed freely online by the Meteorological Service of Canada of Environment Canada. It is also important to note there was an El Niño phase from July 2006 to February 2007, which would typically result in greater expected rainfall for that season.

4.2.3 Digital elevation model and scale analysis

Elevation data, which was used only for the analysis of rainfall interpolation, was derived from a mosaicked 3" (approx. 80 m spatial resolution) DEM that was provided by Canadian Digital Elevation Data (CDED). Tiles of this CDED DEM can be obtained online from GeoBase, which is initiated by federal, provincial and territorial governments of Canada to provide easy access to geospatial information (GeoBase, 2011).

The association between elevation and rainfall is most apparent at spatial scales from 5 km to 10 km (Daly et al., 2008). This scale generally captures the effects of air movements around topographic obstacles (Daly et al., 1994; Funk and Michaelsen, 2004; Sharples et al., 2005). The spatial scale for the interpolation of rainfall in this study follows the same methods used by Daly et al. (2008) – an 800 m resolution DEM is up-scaled from the 80 m CDED and filtered with a Gaussian method using a circular neighbourhood having a 7 km radius. The correlation between rainfall and elevation at different resolutions was explored to ensure that an appropriate DEM was selected for rainfall interpolation. The resolutions of DEMs explored were 80 m, 800 m, 1000 m and 5000 m. The DEMs coarser than 80 m resolutions were scaled up using a Gaussian method. Spearman's correlation coefficient was used to measure the variation of correlations of different filter radius (1 km to 10 km) to rainfall accumulation during the 2006 November storm.

4.2.4 Geostatistical interpolation

An analysis comparing a variety of geostatistical interpolation methods was completed to determine an approach for interpolating rainfall for Vancouver Island for each temporal scale.

Ordinary kriging (OK), universal kriging (UK), and ordinary cokriging (OCK) were compared. In the models for UK and OCK elevation was used as an additional variable to rainfall amount. Thus, OK was the only univariate method and acts as a base for comparison of model performance. Since all of the above methods require a semivariogram model to represent spatial autocorrelation, experimental semivariograms were created and fitted with the appropriate semivariogram model for each method.

The spatial variation in spatial structures can be represented by the semivariogram, which is a measure of dissimilarity between observations. An experimental semivariogram $\hat{\gamma}(h)$ is formed by using the function in the form (Goovaerts, 1997):

$$\hat{\gamma}(h) = \frac{1}{2|N_h|} \sum_{(i,j)\in N_h} \left[z(u_i) - z(u_j)\right]^2, \quad (1)$$

where N_h denotes the set of pairs of observations i, j separated by the vector h, and $z(u)$ is a corresponding realization of a random variable $Z(u)$ (rainfall accumulation) for point u in the domain under study.

OK is a linear estimator, meaning it uses a linear combination of neighbouring values (Goovaerts, 1997):

$$Z^*_{\mathrm{OK}}(u_0) = \sum_{i=1}^{n} \lambda_i^{\mathrm{OK}} Z(u_i), \quad (2)$$

where λ_a^{OK} represents the calculated weights for OK. Z has been observed at n locations $(u_1, \ldots, u_n)$ in the study domain; Z is a random field with a constant, but unknown mean m; the semivariogram γ of Z is known; and the semivariogram is stationary: the mean and variance do not change in time or space.

Since in some situations the stationary condition is violated (e.g. mean rainfall depends on elevation), non-stationary methods such as UK are used in this analysis. UK uses a spatial linear model in the form (Goovaerts, 1997),

$$Z_{\mathrm{UK}}(u) = \mathbf{b}^{\boldsymbol{T}}\mathbf{f}\,(u) + e(u) = \sum_{i=1}^{n} b_i f_i(u) + e(u), \tag{3}$$

where $e(u)$ is the residual from the drift that is spatially dependent with zero mean, $\mathrm{b}^{\boldsymbol{T}}\mathrm{f}\,(u)$ is a deterministic trend. UK does require the secondary variable $f_i(u)$ to be collocated with a significant number of data points. This method does not require estimation of cross-semivariograms or regression analysis of the variables (Ahmed and De Marsily, 1987). Also, universal kriging is generally less sensitive than OK to the semivariogram fitting approach and produces smaller absolute errors (Haberlandt, 2007). The usefulness of the secondary variables depends on the correlation to the regionalized variable, rainfall (Ahmed and De Marsily, 1987).

OCK is multivariate extension of OK (Goovaerts, 1997),

$$Z^{*}_{\mathrm{OCK}}(u) = \sum_{i_1=1}^{n_1(u)} \lambda_{i_1}^{OCK}(u)\, Z_1\left(u_{i_1}\right) + \sum_{i_2=1}^{n_2(u)} \lambda_{i_2}^{OCK}(u)\, Z_2\left(u_{i_2}\right) \tag{4}$$

where the primary data (rain gauge) weights $\lambda_{i_1}^{OCK}$ are constrained to sum to one and the secondary data (elevation) weights $\lambda_{i_2}^{OCK}$ are constrained to sum to zero (Goovaerts, 1997). An advantage of cokriging is that it requires fewer assumptions than other multivariate methods such as UK (Ahmed and De Marsily, 1987). In addition, it has been found to produce adequate results when spatial correlation is present and if the there is a high correlation between the collocated variables (Ahmed and De Marsily, 1987). Like UK, OCK also requires a significant number of common data points (Ahmed and De Marsily, 1987).

Since there were only 53 weather stations, global interpolation, which is the use of all observations for predicting u_0, was used for all of geostatistical methods. The parameters (nugget, sill and range) of the semivariogram were estimated by iteratively reweighted least squares (Goovaerts, 1997). More detailed information regarding the theory behind the geostatistical methods used can be found in Goovaerts (1997).

4.2.5 Performance assessment

The performance assessment of the rainfall interpolations was based on comparison of the root-mean-square-error (RMSE) and mean bias using a leave-one-out cross-validation. The bias is a measure of accuracy between observed and predicted values. The RMSE is a general measure for the precision of a model. The model with the lowest RMSE and bias was selected for interpolation of the rainfall variable used in the subsequent landslide susceptibility modeling analysis. The forms of the performance measures are,

$$\text{bias} = \frac{1}{n}\sum_{\alpha=1}^{n}[Z^*(u_\alpha) - Z(u_\alpha)], \tag{5}$$

and

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{\alpha=1}^{n}[Z^*(u_\alpha) - Z(u_\alpha)]^2}. \tag{6}$$

4.2.6 Selection of associated rainfall temporal scale

A backward-and-forward stepwise variable selection, based on the Akaike Information Criterion (AIC), was used to select one of the rainfall variables (two weeks, winter or annual) that improves the goodness-of-fit of a GAM model including additional variables representing topography, lithology and land cover (See Section 4.5 for more detail); each rainfall variable was used in the model in a linear and nonlinear form. Thus, the resulting model should provide some empirical evidence regarding which rainfall scale was most associated to the set of landslides and provide possible insights into linear or nonlinear relationship that rainfall may have to landslides. The temporal scale that was selected using this method was used for a subsequent analysis of the relationship of rainfall to other environmental factors.

4.3 Land cover classification

The purpose of the land cover classification was to explore the regional relationship of landslide initiation to different forest cover types. This classification was meant to represent general forest conditions related to disturbances from forest harvesting activities.

4.3.1 Landsat TM data and land cover

Landsat data has been popular for forest classification and other thematic classification for over 30 years (Cohen and Goward, 2004). The spatial resolution (30 m), spectral resolution and the availability of images for all year round make is suitable for classification of land cover for regional mapping of landslide susceptibility.

Six Landsat TM images acquired from 6 July 2006 to 27 July 2006 were used in this study. Before classification, the images were transformed into brightness, greenness, and wetness of the TM TC transformation (Crist et al., 1986). The equations for TC are as follows,

$$\begin{aligned}\text{Brightness} = \; & 0.3037(\text{TM}_1) + 0.2793(\text{TM}_2) + \\ & 0.4743(\text{TM}_3) + 0.5585(\text{TM}_4) + \\ & 0.5082(\text{TM}_5) + 0.1863(\text{TM}_7)\end{aligned} \tag{7}$$

$$\begin{aligned}\text{Greenness} = \; & 0.2848(\text{TM}_1) + 0.2435(\text{TM}_2) + \\ & 0.5436(\text{TM}_3) + 0.7243(\text{TM}_4) + \\ & 0.0840(\text{TM}_5) + 0.1800(\text{TM}_7)\end{aligned} \tag{8}$$

$$\begin{aligned}\text{Wetness} = \; & 0.1509(\text{TM}_1) + 0.1973(\text{TM}_2) + \\ & 0.3279(\text{TM}_3) + 0.3406(\text{TM}_4) + \\ & 0.7112(\text{TM}_5) + 0.4572(\text{TM}_7)\end{aligned} \tag{9}$$

where TM_B refers to the corresponding Landsat TM spectral band.

The Landsat data was downloaded for free from the United States Geological Survey (USGS) Global Visualization Viewer (USGS, 2011). All of the scenes used in this classification were georeferenced by USGS.

4.3.2 Maximum likelihood classifier

The algorithm used for classification of land cover was the maximum likelihood classifier (MLC; Jensen, 2005). MLC is based on a probability density function that is calculated from a training sample; thus, the classes are characterized by mean and (co)variance estimation (Atkinson and Tatnall, 1997). This method assumes that the distribution of class samples is a Gaussian (normal) distribution. A pixel is assigned a class based on which it has the highest probability to belong to. Without prior probability information, an unknown measurement vector *X* is assigned a class *j* if, and only if (Swain and Davis, 1978; Jensen, 2005),

$$p_j \geq p_k \text{ for all } k \text{ and } j \text{ out } 1, 2, \ldots n \text{ possible classes} \tag{10}$$

and,

$$p_j = \frac{1}{2}\ln\left|V_j\right| - \left[\frac{1}{2}\left(X - M_j\right)^T V_j^{-1}\left(X - M_j\right)\right], \tag{11}$$

where M_i is the mean measurement vector for class *j* and V_j is the covariance matrix of class *j* for multiple bands (i.e., TC transformations) of remote sensing data *l* through *m*.

In this study, each Landsat scene was classified using independent training data from other scenes. The training data was comprised of 100 (interpreted) samples per each desired class. Prior to picking the samples, clouds and cloud shadows were masked. The main land cover classes used in this study were associated with a variety of forest stand conditions: exposed ground, open canopy, semi-open canopy forest, and closed-canopy forest. In addition, classes for snow and ice, and water were classified; the non-forest classes represent areas that were masked out of the analysis. The class descriptions are shown in Table 4.1.

Since forest harvesting has and continues to happen on Vancouver Island, the general forest covers were meant to implicitly relate forest activities. The classification was not specific enough to only represent 'logging forests'. However, the general characteristics of these forest types can be used to relate to forestry practices and investigate the relationships to landslides. The decision to base this analysis on different forest class types was based on conclusions made by Schmidt et

al. (2001) that it is necessary to use a refined classification of vegetation to explore the possible associations between landslides and vegetation. The decision of which classes should be used in this study area to represent a variety of forest stand conditions, which may also related to forest activities, was based on the work of Cohen et al. (1995); in that study, different forest stand structures are classified for a Pacific coast study area that has similar forest conditions as Vancouver Island: forests dominated by Western hemlock and Douglas fir.

Table 4.1. Description of land cover classes

Classes	General description	Relationship to forest harvesting
Forest cover		
Exposed ground	Recently disturbed forest area or rock outcrop characterized by exposed surface material/bedrock	Typical condition of a location that has been logged in the time period of the Landsat scene May contain recently planted seedlings
Open forest	A 'young' forest dominated by shrubs and saplings Recent recovery from a disturbance (e.g., logging)	Beginning stages of forest recovery after recent logging Covered by tree saplings and poles
Semi-open forest	Forest with a partially opened canopy and exposed understory	Later stage of forest recovery Covered by tree poles and some sawtimber
Closed forest	Relatively the most developed 'old' forest with closed canopy	Fully recovered forest after logging or forest that has not been logged Majority of forest cover by sawtimber trees
Masked		
Water	All open water	-
Snow and ice	Area characterize by year-long surface cover of snow/ice	-
Agriculture	Area representing all types of agricultural fields	-
Urban	Commercial and residential areas	-

4.3.3 Masking, mosaicking and resampling

A manually digitized snow mask was used to prevent the misclassification of snow and exposed ground land cover (Figure 4.3). Although snow detected in the summer months is typically associated with the snowline at higher elevations, the mask is manually digitized

because the elevation of snow presence varies across mountain peaks. Thus, the MLC was implemented twice on the same training data in masked areas: once with all of the land cover classes, and another without training samples for snow cover. Next, the MLC with snow cover is cropped using the snow mask and used to replace values of the MLC without snow cover. As a result, snow is only classified at high elevations (about >1000 m. a.s.l.) associated with the (semi-) permanent snowline.

Agricultural and urban areas were also masked, which were predominant below 200 m a.s.l. These were masked by extracting agriculture and urban classes from the Circa 2000 Land Cover for Agricultural Regions of Canada, which is land cover classification product produced by Agriculture and Agri-food Canada (AAFC) using Landsat imagery. This data was available for free online from the GeoConnections – Discovery Portal (GeoConnections, 2011).

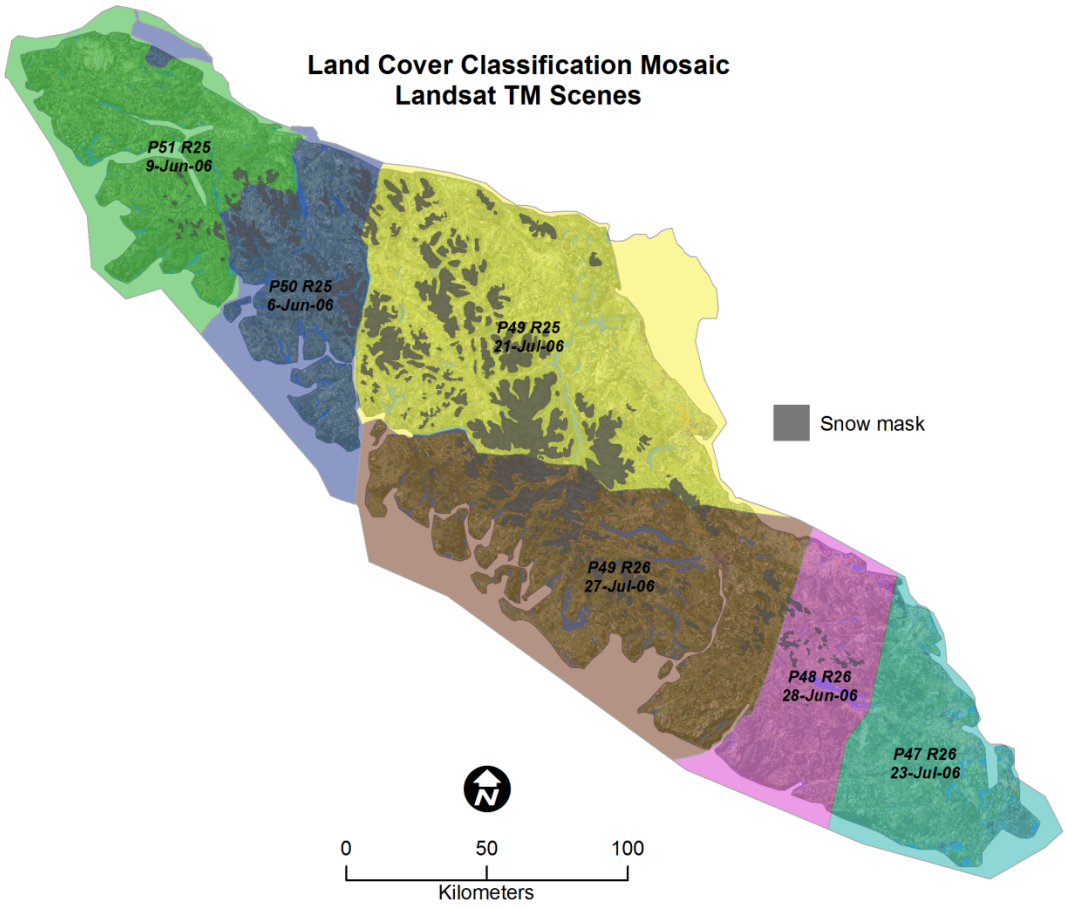

Figure 4.3. Map illustrating the mosaic and snow mask of the land cover classification using Landsat TM scenes. The labels represent the path (P), row (R) and acquisition date for each scene.

Since each Landsat scene was individually classified using MLC, a mosaic of classified scenes was assembled to represent land cover coherently for all of Vancouver Island (Figure 4.3). The decision of where to cut scenes was based on the ability to remove cloud covered areas from the final classification map.

The spatial resolution of the mosaicked scene was resampled before including land cover in the following landslide susceptibility analysis. A nearest-neighborhood resampling method was used to transform the spatial resolution from 30 m to 20 m, the latter being the same spatial resolution of the grids used in the susceptibility analysis.

4.3.4 Accuracy assessment

The accuracy of the land cover classification focuses on highlighting the performance of the forest cover types and the masked out features (urban, agriculture, snow and ice, water) of the resampled-mosaicked land cover map that was used in the landslide susceptibility analysis. A reference data set was assembled by random sampling 100 points across the classified area, which allows for a 95% confidence level with a 10% confidence interval of the performance results. These points were independently interpreted using the same Landsat TM imagery used for the classification. The reference data set was then compared to the classified map values in a confusion matrix: a cross-tabulation of classes observed in the reference data and predicted in the classification. The measures used for estimating classification accuracy were derived from the confusion matrix; these were the overall accuracy and the kappa (κ) coefficient (Foody, 2002).

The overall accuracy is a summary of the total agreement (or disagreement) measured by the proportion of the total number of assigned classes that are correct (Foody, 2002).

Cohen's κ coefficient was used to test the similarity between the reference observations and the classification. The κ coefficient ranges between -1 and +1, where -1 indicates perfect disagreement, +1 indicates perfect agreement and 0 indicates that there is no relationship. The calculation of κ is in the form,

$$\kappa = \frac{p_o - p_c}{1 - p_c} \tag{12}$$

where p_o is proportion of cases that are in agreement and p_c is the proportion of agreement that is expected by chance (Foody, 2004).

4.4 Logging roads

The Euclidean distance from logging roads was calculated to explore the relationship these roads have to landslide initiation. Only distances from roads up to 100 m, which was the maximum distance assumed to have an influence on landslides, were explored. This data was obtained from the British Columbia Digital Road Atlas (BCDRA). Through visual inspection, it was determined that logging roads are best represented as road surface attributes for "loose" and "rough" surfaces in the BCDRA data set. It was important to differentiate the road surface type because paved roads and logging roads have different construction standards.

4.5 Landslide susceptibility modeling

The landslide susceptibility models constructed were based on a total of 9 natural and anthropogenic controls as independent predictor variables (Table 4.2). These included 5 topographic factors (slope, catchment area, plan curvature, profile curvature, and elevation), 2 anthropogenic factors (land cover and distance-to-roads), 1 climatic factor (two weeks, winter or annual rainfall) and 1 geologic factor (lithology). In this study the following GAM and GLM models for predicting landslides were explored; a GAM and GLM using variables for rainfall, land cover and logging, topography and geology (RLTG-GAM and RLTG-GLM); a GAM and GLM using the previous mentioned variables with the exception of rainfall (LTG-GAM and LTG-GLM).

The topographic factors were derived from a 0.75" (~20 m) resolution CDED DEM using the open-source SAGA GIS (Conrad, 2006). The statistical modeling of the landslide susceptibility models were implemented using R, open-source statistical software (R Development Core Team, 2011). SAGA GIS was utilized in the R environment with the RSAGA package (Brenning, 2008). The factors for rainfall, land cover and logging roads (distance-to-road) were based on the methods previously discussed in Sections 4.2, 4.3 and 4.4. The lithology classes (metamorphic, intrusive, volcanic and sedimentary) used for the geology factor were

from the Digital Geology Map of British Columbia (2005). All of the data sets were transformed into raster form with a spatial resolution of 20 m.

Table 4.2. Summary of landslide susceptibility models and variables

Model name	Prediction model	Variables
RLTG-GAM	Generalized additive model	**R**ainfall: annual, winter or two weeks rainfall **L**ogging related: land cover and distance-to-road
RLTG-GLM	Generalized linear model (logistic regression)	**T**opographic: slope, catchment area, plan curvature, profile curvature, and elevation **G**eologic: lithology
LTG-GAM	Generalized additive model	**L**ogging related: land cover and distance-to-road **T**opographic: slope, catchment area, plan curvature,
LTG-GLM	Generalized linear model (logistic regression)	profile curvature, and elevation **G**eologic: lithology

4.5.1 Generalized additive models

A GAM was used to construct a susceptibility model for this region. A GAM is an extension of generalized linear models (GLM) that can represent covariates as linearly or nonlinearly. The common approach for modeling binary response variables in a GLM is logistic regression, which models the logit of a response probability in the form,

$$\ln\left\{\frac{P(X)}{1-P(X)}\right\} = \beta_0 + \beta_1 X_1 + \cdots + \beta_i X_i \tag{13}$$

where $P(X) = \text{Prob}(Y = 1|\, X)$, $X_1, X_2, \ldots, X_i$ are the covariates, $\beta_1, \beta_2, \ldots, \beta_i$ are the regression coefficients and β_0 is the intercept. The logit model ensures that the proportions $P(X)$ will fall between 0 and 1. The main assumption in this model is that the estimation of the response is only linearly dependent on the predictor variables (Hastie and Tibshirani, 1990).

In contrast, a GAM can include non-parametric (or linear) covariates by replacing the general linear term $\beta_0 + \beta_1 X_1 + \cdots + \beta_i X_i$ with the additive relationship,

$$\ln\left\{\frac{P(X)}{1-P(X)}\right\} = \beta_0 + f_1(X_1) + \cdots + f_i(X_i) \quad (14)$$

where $f_i(X_i)$ is an arbitrary function defined by the data. The term additive describes a model as being the sum of its terms (Hastie and Tibshirani, 1990). The advantage of the function term is that the model can be fitted to the data without rigid assumptions regarding the dependence on the response. Thus, a smoothing function $s_i(X_i)$ can be used to estimate $f_i(X_i)$ using flexible specifications of the dependence of the response on the covariates (Wood, 2006). A variety of smoothing functions are explained in detail by Hastie and Tibshirani (1990). For this study, cubic smoothing splines were used to estimate the dependence of the mean response on the predictors for the estimation of GAM (Hastie and Tibshirani, 1990). The degrees of freedom, which describes the flexibility of a smoother to fit to the data, was set to 2. Any statistical comparison of GLMs and GAMs was completed by using ANOVA based on the χ^2 (chi) test statistic.

Akaikes's information criterion (AIC) was used as measure of "goodness of fit" calculated using a model's log-likelihood,,

$$AIC = -2\times\text{log-likelihood} + 2(p+1) \quad (15)$$

where p is the number of predictor variables in the model (Crawley, 2007). AIC penalizes models that have a larger number of predictor variables with $2(p+1)$. When comparing models, a better fit is represented by a lower AIC.

The AIC was applied to automatically determine the best fit of parameters for the models (Crawley, 2007). This was implemented in this study by using a combined back-and-forward variable selection method, where each form of a variable (linear or nonlinear) was implemented in the model and only the form that contributes to a lower AIC were selected for the final model fit.

4.5.2 Model assessment

As proposed by Brenning (2005), spatial cross validation was implemented for the estimation of model performance. The spatial cross validation method used in this study is

similar to *k*-fold cross-validation, but instead of dividing data set into *k* random subsets of equal size, the data set is divided into *k* spatial subsets (Figure 4.4). In this study, the partitioning into subsets was performed by using *k*-means: a simple unsupervised clustering algorithm. Spatial cross validation was repeated 25 times with each of repetition drawing spatial partitions independently from the other replications to test the AUROC and the sensitivity of each model at a high specificity level of 90%.

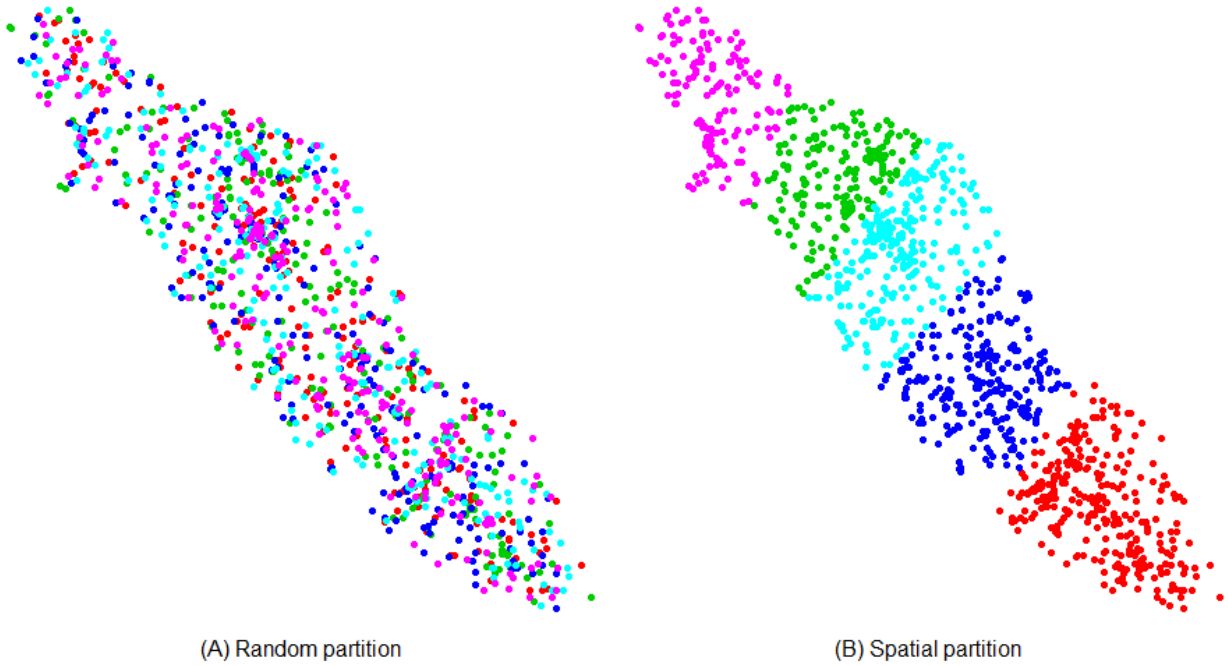

Figure 4.4. Comparison of partitioning of data set using 5-fold (non-spatial) cross-validation (A) to 5-fold spatial cross-validation (B).

4.5.3 Assessing variable importance and nonlinearity

Although all the variables for the susceptibility models were pre-selected, the decision to incorporate a predictor variable as either linear or nonlinear was automated using the stepwise-variable selection. By recording the frequency of nonlinear occurrence in the 25-repeated 5-fold spatial cross-validation, the prevalence of nonlinear variables was observed.

Observing the relative importance of predictor variables was measured in this study by systematically forming a GAM with all predictor variables except one. The associated AUROC

based for each model was tested and trained using the entire study area. This approach was applied for all of the variables. Thus, 8 different models were formed, each with only one predictor variable excluded. The relative importance of a predictor variable was then determined by measuring the difference in model performance from RTLG-GAM (a model with all predictor variables).

4.6 Additional model exploratory analysis

An additional exploratory analysis was included in this study to further investigate relationships between landslide initiation and some of the predictor variables. In particular, the relationships of land cover and lithology to influence the susceptibility of different values of annual rainfall and slope was explored.

These relationships were summarized using conditional density plots that compared the probability characteristics of rainfall/slope to landslide initiation for individual classes of land cover/lithology. In addition, the interaction terms between these variables to predict landslide susceptibility were investigated. Several extensions of the LTG-GLM were modeled to represent combinations of predictor variables as interactions terms. The interactions that were explored include land cover to rainfall, land cover to slope, lithology to rainfall, and lithology to slope. A LTG-GLM, with no interaction terms, was used as the basis for model comparison. The AIC and AUROC (for the entire study area) of these models were measured and compared to observe how the interactions may change model performance.

Also, the relationship of distance-to-road to predict landslide initiation for different slope angles was explored. The purpose of this investigation was to illustrate the potential for landslide susceptibility models to provide supportive information to assist in road planning through landslide prone areas. The behaviour of slope and distance-to-road to influence susceptibility was isolated using the RLTG-GAM fitted to a range of values for slope (0-60°) and distance-to-road (0-100 m). The remaining variables were fitted as constant values. Thus, a RLTG-GAM was predicted using the following constant values; elevation (600 m), profile curvature (-0.001), plan curvature (-0.002), catchment area (10000 m), annual rainfall (2000 m), land cover (closed forest), and lithology (intrusive). The RLTG-GAM was used to calculate predicted probabilities of landslide initiation for corresponding slope and distance-to-road values.

4.7 Summary

The analysis of landslide susceptibility was first conducted by preparing and processing data for topographic, anthropogenic, climatic and geologic factors. Topographic factors (slope, catchment area, plan curvature, profile curvature and elevation) are processed from a DEM. Land cover, which partially represents anthropogenic factors, was produced by applying a supervised MLC classification of Landsat TM for scenes from 2006. Another anthropogenic factor, distance-to-road, was calculated using the Euclidean distance from road. A climatic factor for rainfall was formed by applying different geostatistical methods (OK, UK, and OCK) to rainfall weather station data across Vancouver Island and interpolating rainfall for different temporal scales (two weeks, winter, annual). The 'best' performing geostatistical method was used to compare which temporal scale of rainfall pattern was most closely related to the distribution of landslides. This was determined using and stepwise-variable-selection method. Lithology was used as a geologic factor, which was classified by volcanic, sedimentary, intrusive and metamorphic rocks.

After data processing, the environmental factors were used as predictor variables for statistical classification of landslide susceptibility using GAMs and GLMs. The subsequent susceptibility models were analyzed using repeated spatial-cross validation that records AUROC and sensitivity at 90% specificity. For simplification, these methods applied for RLTG-GAM are summarized in a flowchart shown in Figure 4.5.

An additional exploratory analysis was conducted to investigate possible interaction terms between rainfall, slope, lithology and land cover. These were examined by using conditional density plots and model extension of the LTG-GLM that included interaction terms. Also, the relationship of distance-to-road and slope to predict landslide susceptibility was explored by fitting RLTG-GAM with predetermined values for model variables.

Figure 4.5. Flowchart summary of methods for landslide susceptibility modeling of RLTG-GAM.

Chapter 5

Results

5.1 Rainfall interpolation

The interpolation of rainfall at different temporal scales (two weeks, winter and annual) using geostatistical methods (OK, UK and OCK) was based on approximately 53 stations (see section 4.2.1 for removed stations). The observed station values indicate that the winter of 2006-2007 had a maximum rainfall of 5025 mm, which was greater than the maximum annual rainfall for 2006 (4050 mm; Table. 5.1). However, annual rainfall had a higher median (2050 mm) than winter rainfall (1332 mm). Only, stations during the two weeks of rainfall, related to the November 2006 storm, observed values of no rainfall.

The correlation between the smoothed DEM, used as an additional variable for interpolation in OK and OCK, and rainfall amount was greatest in the two weeks rainfall period (0.51), and least for annual rainfall (0.21). The stations were located at relatively low elevations. The median station elevation was 73 m and the maximum was 715 m (Table. 5.1).

Table 5.1. Summary statistics of observed rainfall and elevation associated with weather stations

Rainfall Period	**Median**	**Mean (std. dev.)**	**Minimum**	**Maximum**	**Correlation with elevation**
Two weeks (mm)	333	387 (228)	0	1451	0.51
Winter (mm)	1332	1555 (814)	627	5025	0.32
Annual (mm)	2050	2104 (154)	717	4050	0.21
Elevation*(m)	73	134 (182)	1	715	

*Elevation derived from a smoothed DEM with a ~7 km radius described in Section 4.2.3

5.1.1 Correlation of DEM resolution and rainfall

Examination of the relationship between DEM spatial resolution and rainfall revealed that the correlation of an 800 m DEM was least impacted by the use of different filter sizes to generalize topographic features (Figure 5.1). The values used for this comparison were based on rainfall from 2006-November storm (two weeks) and measured using Spearman's correlation coefficient ρ_{Sp}.

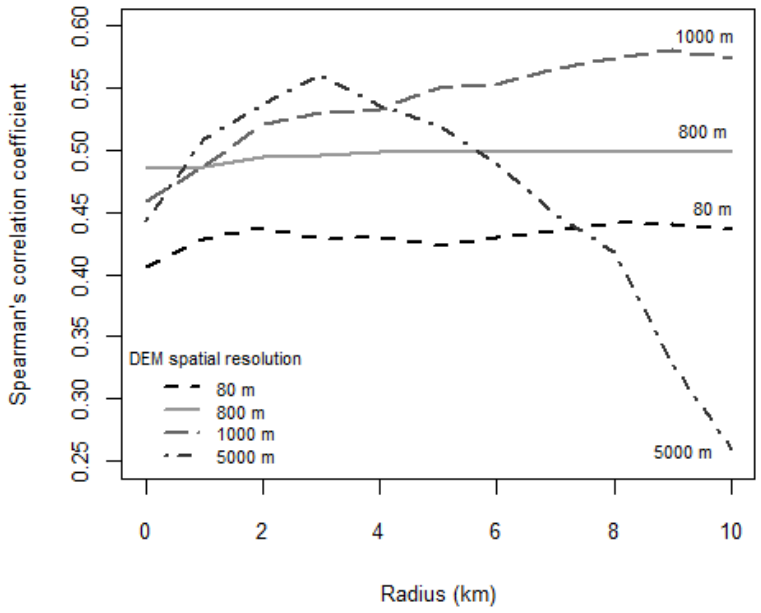

Figure 5.1. Correlations of DEMs with different spatial resolution to rainfall accumulation occuring during the November 2006 storm. Note that a radius of 0 km implies no filter.

The strongest correlation was associated with a DEM having a 1000 m resolution and a filter radius of 9 km ($\rho_{Sp} = 0.58$). The weakest correlation to rainfall was found using a 5000 m DEM with a filter radius of 10 km ($\rho_{Sp} = 0.26$). The 800 m DEMs had the least amount of variation in correlation related to filter size (ρ_{Sp} range: 0.49 to 0.50). Thus, the performance of rainfall interpolation does not rely heavily on the filter radius while using an 800 m DEM.

5.1.2 Semivariogram models of rainfall at different temporal scales

The experimental semivariograms (Figure 5.2 and Figure 5.3) computed for each temporal and geostatistical method (OK, UK and OCK) demonstrated the differences in model structures. There estimated semivariance nuggets were very small for interpolation of rainfall using OK and UK. OCK had a negative nugget for all temporal scales related to the fitting of the co-regionalization semivariogram for elevation and rainfall.

The range of spatial autocorrelation of rainfall increases with length of temporal scale. The longest ranges (157 km to 337 km) were associated to annual rainfall, and followed by winter rainfall (97 km to 124 km). The smallest temporal scale, two weeks, had the shortest ranges (75 km to 85 km).

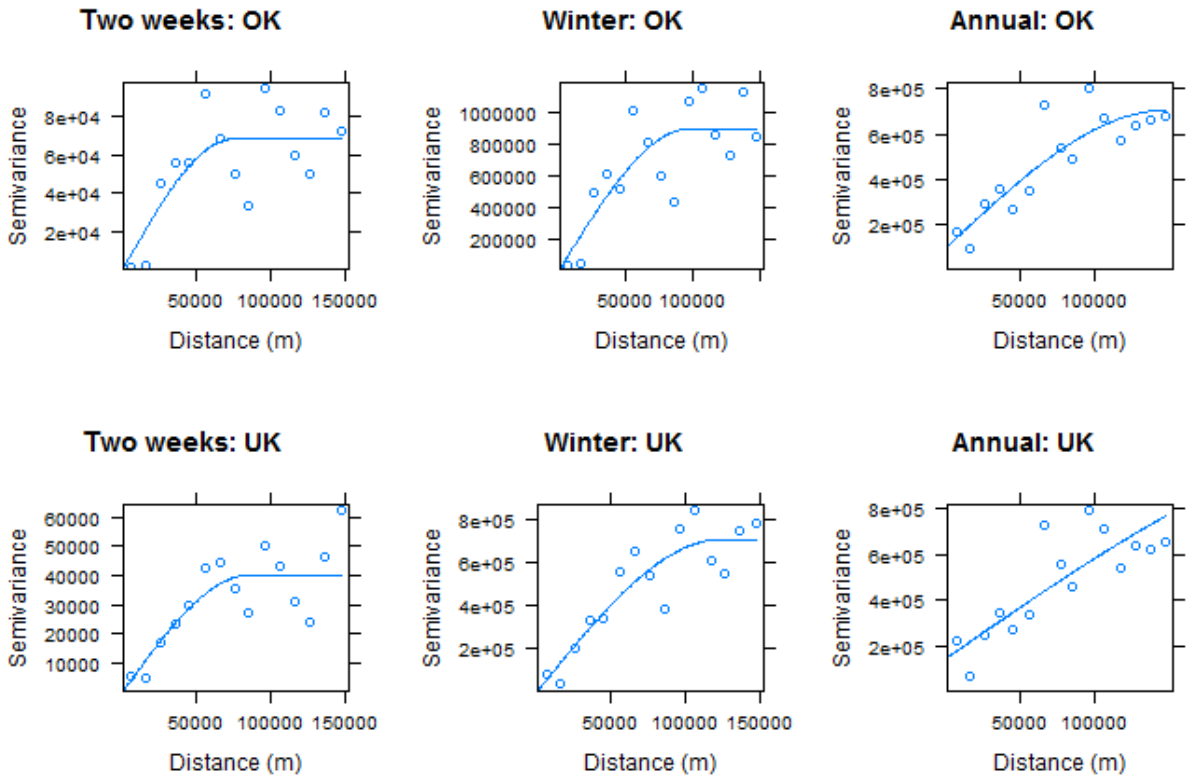

Figure 5.2. Experimental semivariograms and fitted models for geostatistical interpolation using OK and UK.

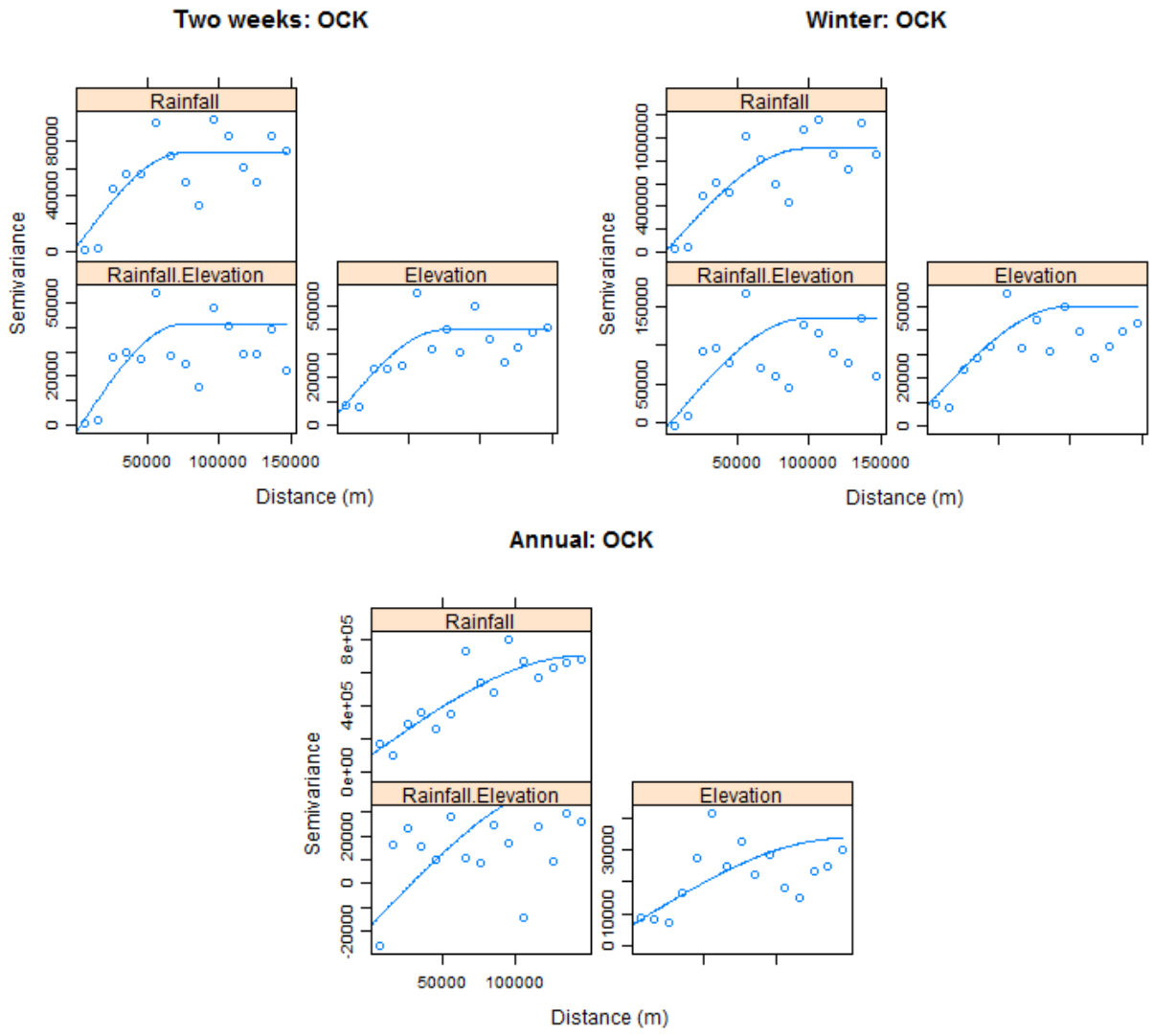

Figure 5.3. Experimental semivariograms for co-regionalization of rainfall and elevation using OCK.

5.1.3 Rainfall distribution

The maps of rainfall interpolation clearly show differences in rainfall distribution related to the geostatistical method used. In particular, interpolation using UK illustrated a strong and direct relationship of rainfall to elevation (Figure 5.6); this was visually apparent by the strong similarity the rainfall pattern to an elevation model. The rainfall patterns of OCK (Figure 5.4) and OK (Figure 5.5) appear to be more generalized, especially OK, which only relies on rainfall values for interpolation.

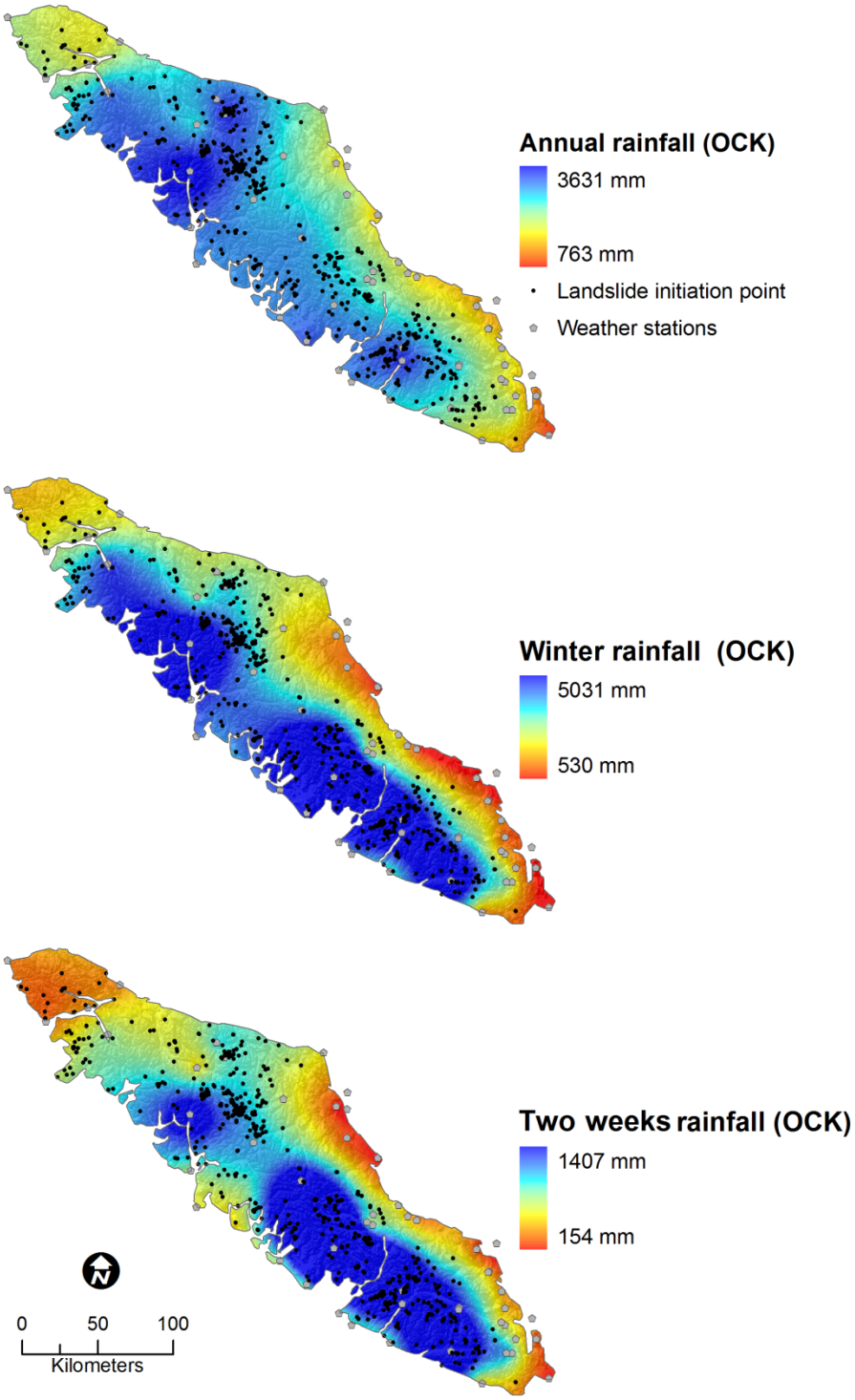

Figure 5.4. Geostatistical interpolation of rainfall using OCK

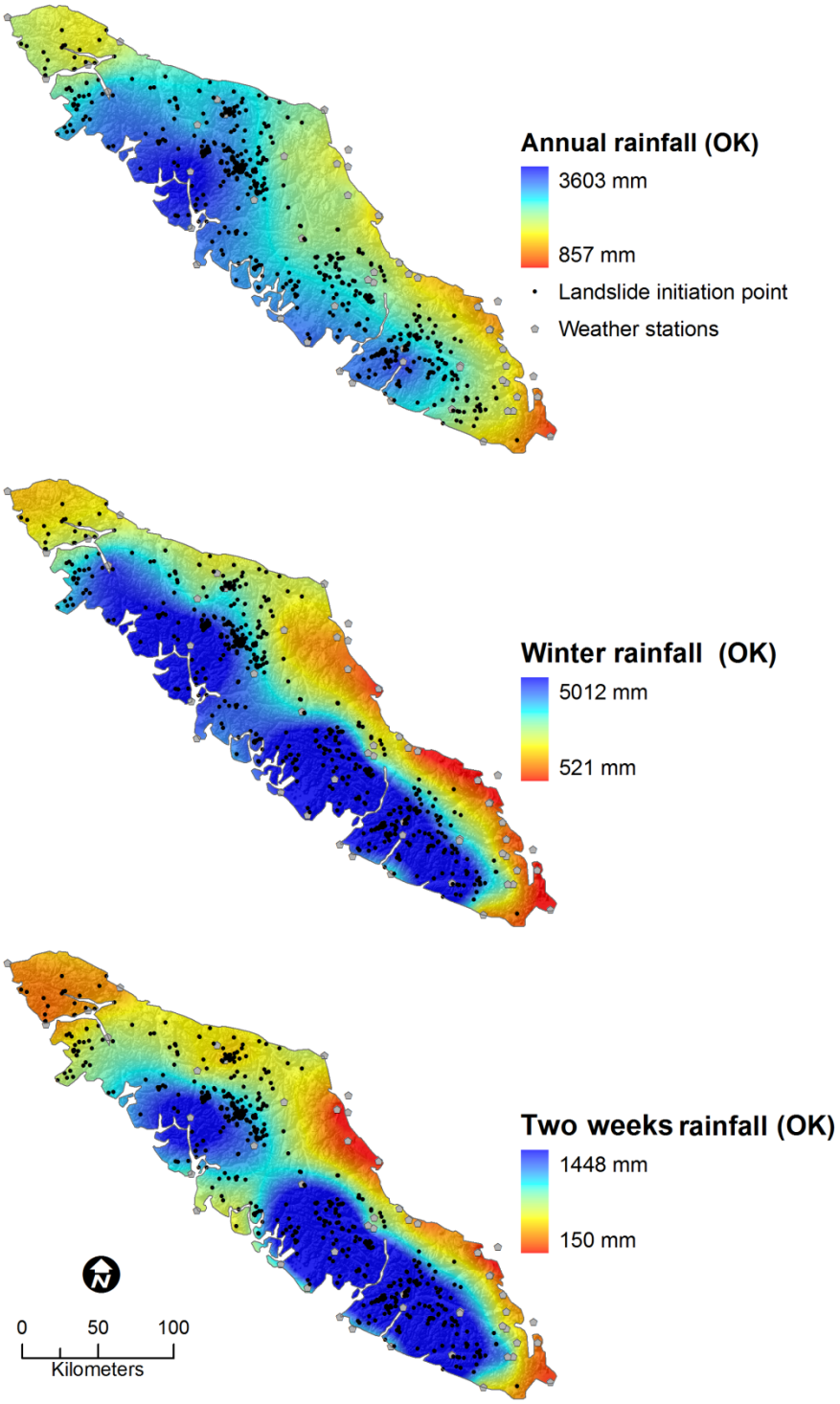

Figure 5.5. Geostatistical interpolation of rainfall with OK

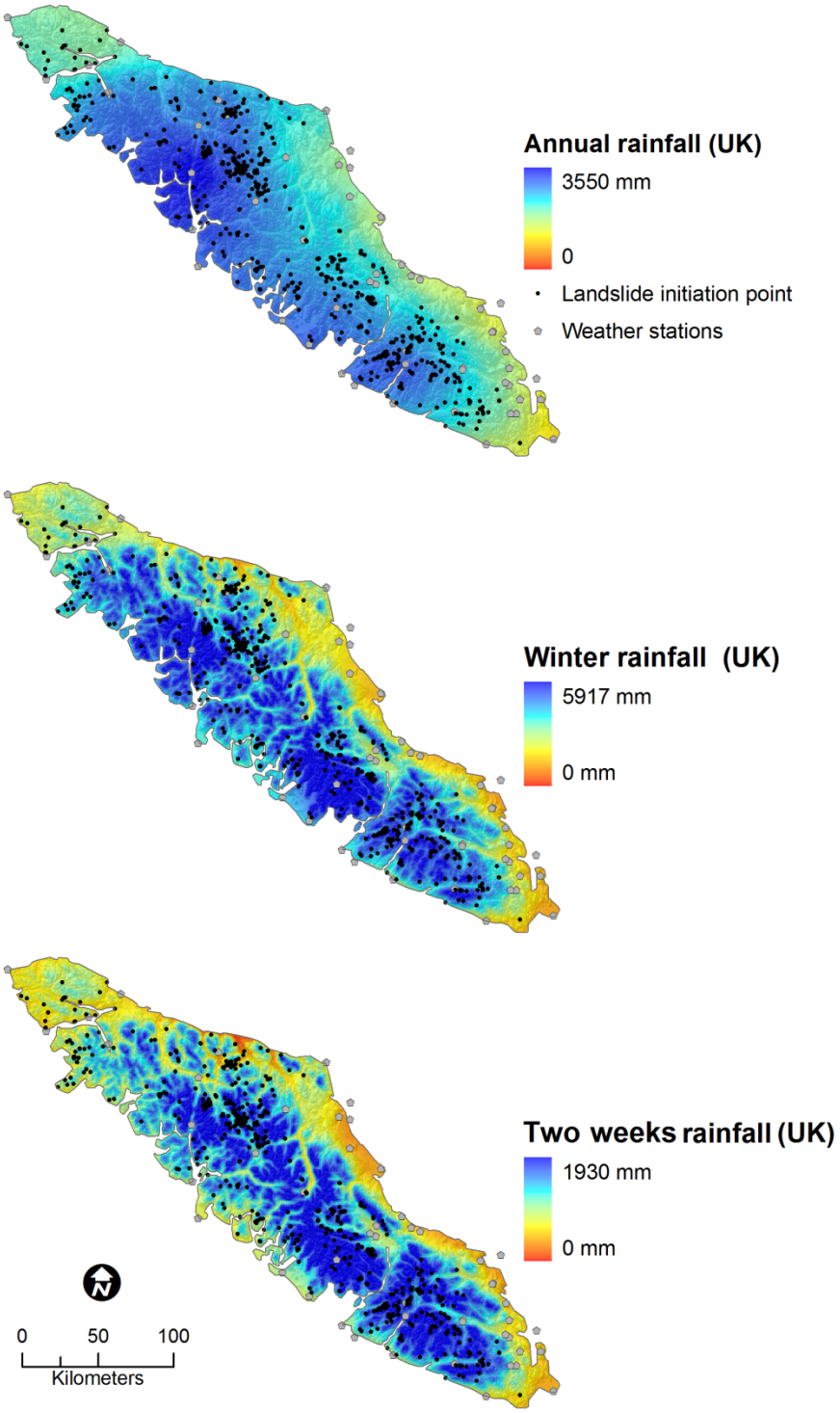

Figure 5.6. Geostatistical interpolation of rainfall with UK

5.1.4 Interpolation performance

Overall, OCK with elevation as co-variable was the strongest performer for all temporal scales (Table 5.1). In terms of two weeks interpolation, OCK had the lowest bias (3.4 mm) and UK (9.6 mm) had the highest. The RMSE of two weeks was lowest for OCK (131.1 mm) and highest for OK (170 mm). In terms of winter rainfall, OK had the lowest bias (16.8 mm) and UK had the highest bias (18.8 mm). The RMSE for winter was lowest for OCK (463 mm) and highest for OK (559.1 mm). In terms of annual rainfall, OCK again had the lowest bias (-4.0 mm) and UK had the highest (-11.6 mm). The RMSE for annual rainfall was lowest for OCK (541.4 mm) and highest for UK (561.4 mm). The positive bias for all values in two weeks and winter interpolation indicate that rainfall amounts were consistently overestimated with each method. In contrast, the negative bias associated with annual rainfall interpolation indicates that rainfall was slightly underestimated.

In general, as the time period for rainfall observations increased, the performance of the geostatistical methods decreased. This was measured by comparing the mean RMSE for each time period prediction (OK, UK and OCK) to the standard deviation of observed rainfall for similarity, which indicates a model's ability to preserve the observed variance (Table 5.1 and Table 5.2). The two weeks interpolation shared the most similar mean RMSE to the standard deviation ($RMSE_{mean}$/Std. dev. = 150/228 mm), followed by winter (523/814 mm) and annual rainfall (550/154 mm).

Table 5.2. Leave-one-out cross-validation results for geostatistical interpolation of rainfall at different temporal scales

Model	Two weeks		Winter		Annual	
	RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)	RMSE (mm)	Bias (mm)
OCK	131.1	3.4	463.0	17.0	541.4	-4.0
UK	149.7	9.6	546.7	18.8	561.4	-11.6
OK	170.0	7.2	559.1	16.8	547.7	-5.4

5.1.5 Temporal relationship of rainfall to landslide initiation

Since OCK had the overall strongest performance for interpolation of rainfall, it was used for the subsequent comparison of which temporal scale provides the 'best' information for prediction of landslide susceptibility.

Based on the automatic variable-selection method, annual rainfall was selected as the temporal scale that was most related to this set of landslides in terms of contribution to model fit in RLTG-GAM.

5.2 Land cover classification

The land cover classification was used to categorize forest into classes that reflect the impacts of logging on Vancouver Island (Figure 5.7). The most common land cover was closed forest (56% of the study area), followed by semi-open forest (19%), open forest (14%) and exposed ground (6%). The masked area, which represented the area of Vancouver Island that was outside of the landslide susceptibility model domain, had the least amount of area covered (4%).

Table 5.3. Confusion matrix for land cover classification. The records that are highlighted are agreements between reference data and the classification prediction.

		Reference data					
Classification prediction	Class	Exposed ground	Open	Semi-open	Closed	Masked	Total
	Exposed ground	3	0	0	0	3	6
	Open	0	3	1	2	1	7
	Semi-open	0	1	12	6	0	19
	Closed	0	0	3	61	1	65
	Masked	0	0	0	0	3	3
	Total	3	4	16	69	8	100

The accuracy of the land cover classification was based on a confusion matrix constructed of 100 randomly sampled points that were separate from the class training set (Table 5.3). The overall accuracy was 82% (±7% at a 95% confidence interval) with a κ coefficient of 0.65.

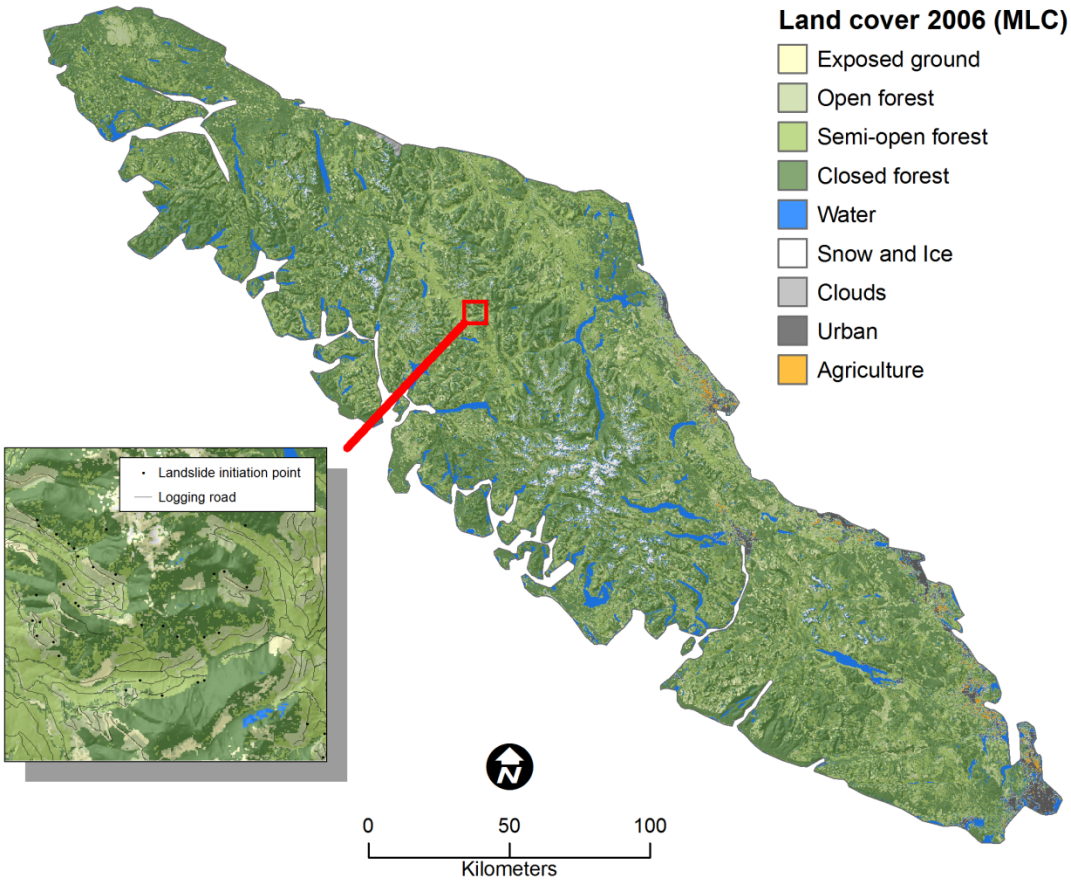

Figure 5.7. 2006 land cover classification map of Vancouver Island from Landsat TM satellite imagery.

The most confusion between reference data and classification prediction was between semi-open forest and closed forest (Table 5.3). Exposed ground had the least confusion; however it should be noted that in some cases a road may be wide enough to be detected from the Landsat imagery (Figure 5.8), thus resulting in an exposed ground classification, which does not necessarily represent an area of recent deforestation.

Visually speaking, the classification appeared to adequately capture the different forest conditions represented as classes (Figure 5.8). A comparison of the classification to a Landsat image in false colour (red: TM band 4, blue: TM band 3, green: TM band 2) show dark red areas represent closed forest, light red represents semi-open forest, bright red to pail red-blue represent open forest, and pale blue-green were areas of exposed ground (Figure 5.8).

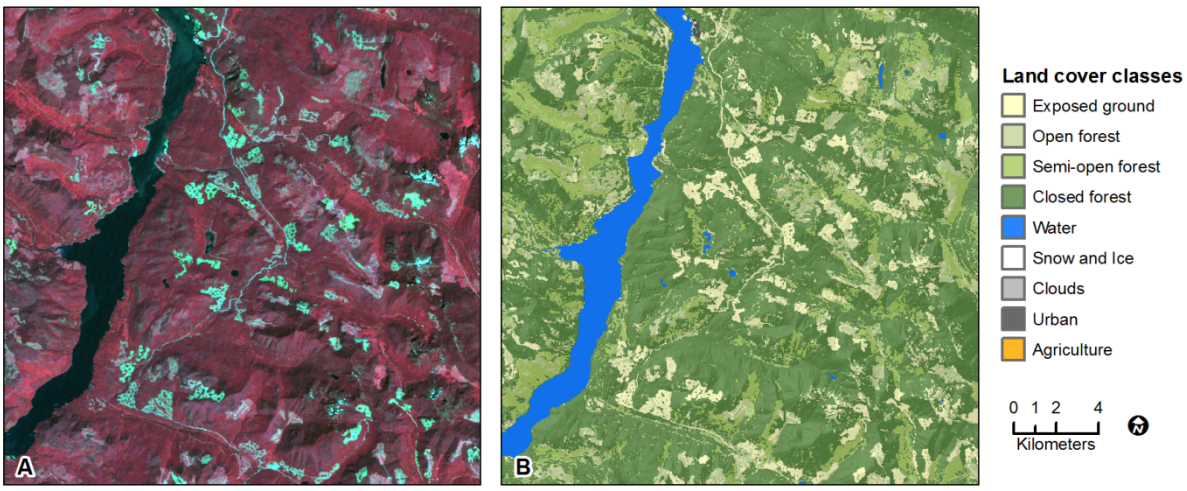

Figure 5.8. Visual comparison of false colour composite Landsat image (A) to land cover classification map (B); the area of interest is of the Hitwatchas Mountain and Nahmit Bay, Vancouver Island.

5.3 Landslide susceptibility

5.3.1 Exploratory analysis of predictor variables

An analysis comparing environmental factors of landslide and non-landslide samples was divided by exploration of continuous and categorical predictor variables. For continuous predictor variables (topography, rainfall and distance to logging roads) descriptive statistics highlighting the difference in continuous-predictor variables of landslide and non-landslide samples are shown in Table 5.4. The Wilcoxon Rank sum test was used to test for difference in AUROC values between samples. All the differences in values of the continuous predictor variables between landslide points and non-landslide points were statistically significant at the 5% level based on Wilcoxon rank sum test, most of the nominal p-values being <0.001 except for distance-to-road with p-value of 0.003.

The AUROC values were calculated using two error estimation techniques, testing on the entire study area and spatial cross-validation. The strongest single predictor of landslide initiation was slope (AUROC > 70%), followed by catchment area, rainfall (AUROC > 65%), plan curvature, and elevation (AUROC > 60%). The weakest single predictors were distance-to-road and profile curvature (AUROC < 60%).

Table 5.4. Descriptive statistics for predictor variables used for modeling landslide susceptibility

Predictor Variable	Median (std. dev.)	Median (std. dev.)	Wilcoxon rank sum test (p-value)	AUROC (%) study area	AUROC (%) SP-CV (std. dev.)
Topography					
Elevation (m)	585 (270)	432 (366)	<0.001	62.7	62.4 (1.1)
Slope (degree)	32 (10)	18 (13)	<0.001	75.5	75.3 (1.3)
Planar curvature	-0.002 (0.013)	0.000 (0.006)	<0.001	64.9	65.1 (0.7)
Profile curvature	-0.001 (0.009)	0.000 (0.005)	<0.001	59.4	59.5 (1.1)
Catchment area (log_{10})	3.91 (0.61)	3.58 (0.55)	<0.001	69.7	69.7 (1.3)
Rainfall					
Two weeks (mm)	482 (185)	418 (191)	<0.001	67.1	68.1 (1.6)
Winter (mm)	2114 (550)	1770 (687)	<0.001	65.9	66.8 (2.5)
Annual (mm)	2775 (340)	2548 (503)	<0.001	65.9	65.4 (2.3)
Logging roads					
Distance-to-road (max. 100 m)	20 (29)	40 (26)	0.003	52.5	52.6 (1.2)

Examination of correlations between predictor variable using Spearman's rank correlation coefficient (ρ_{Sp}) revealed that only a strong inter-correlation exists between winter rainfall and two weeks rainfall (ρ_{Sp} = 0.88; Table 5.5). All other correlations between variables were weaker with $|\rho_{Sp}|<0.59$.

Table 5.5. Correlation matrix of predictor variables using Spearman's correlation coefficient

	Distance-to-road	Annual rainfall	Plan curvature	Profile curvature	Elevation	Slope	Catchment area (log_{10})	Winter rainfall
Annual rainfall	0.089	-	-	-	-	-	-	-
Plan curvature	0.004	-0.091	-	-	-	-	-	-
Profile curvature	0.016	-0.098	0.499	-	-	-	-	-
Elevation	0.075	0.211	-0.181	-0.067	-	-	-	-
Slope	0.155	0.362	-0.126	-0.083	0.493	-	-	-
Catchment area (log_{10})	-0.040	0.169	-0.581	-0.342	0.178	0.134	-	-
Winter rainfall	-0.005	0.58	-0.074	-0.066	0.112	0.360	0.133	-
Two weeks rainfall	-0.083	0.388	-0.114	-0.082	0.271	0.350	0.218	0.877

The sample difference between landslide and non-landslide points related to categorical predictor variables (land cover and lithology) was explored by examining the percentage of points that fall within each predictor variable class and by calculating the odds ratio for individual classes compared to remaining classes (Table 5.6). The highest odds ratio for landslides to occur were open forest (OR = 1.98), followed by semi-open forest (1.18) and exposed ground (1.17). Closed forest canopy had the smallest odds ratio (0.66) compared to the other forest based classes. The higher chance of landslides in open forest was also demonstrated by the higher percentage of landslides that occur in this class (25.7%) compared to samples of non-landslide areas (15.9%).

The lithology showed the highest odds ratio of landslide initiation associated with intrusive rocks (1.34), followed by metamorphic rocks (1.10) and then closely by volcanic rocks types (0.98). The odds ratio of landslide occurring in sedimentary rocks was the smallest (0.46). The difference between landslide and non-landslide point samples related to percentage of occurrence in each lithology class was more difficult to discern then using odds ratio; the percentage difference only varied by a maximum 7%.

Table 5.6. Summary of categorical predictor variables used for modeling landslide susceptibility

Predictor Variable	**Landslide points (%)**	**Non-landslide points (%)**	**Odds ratio**
Land cover class			
Exposed ground	5.6	4.9	1.17
Open	25.7	14.9	1.98
Semi-open	23.6	20.8	1.18
Closed	44.8	54.9	0.66
Masked (Urban, Agri., Clouds)	0.3	4.5	0.07
Lithology class			
Intrusive	61.8	54.8	1.34
Metamorphic	3.4	3.1	1.10
Volcanic	27.9	28.3	0.98
Sedimentary	6.9	13.8	0.46

Insights into the ability of continuous predictor variables to characterize landslide initiation are shown using conditional density plots. These plots are created for variables representing

topography, rainfall and distance-to-road and represent the estimated probability of values of these variables to causes landslides (Figure 5.9).

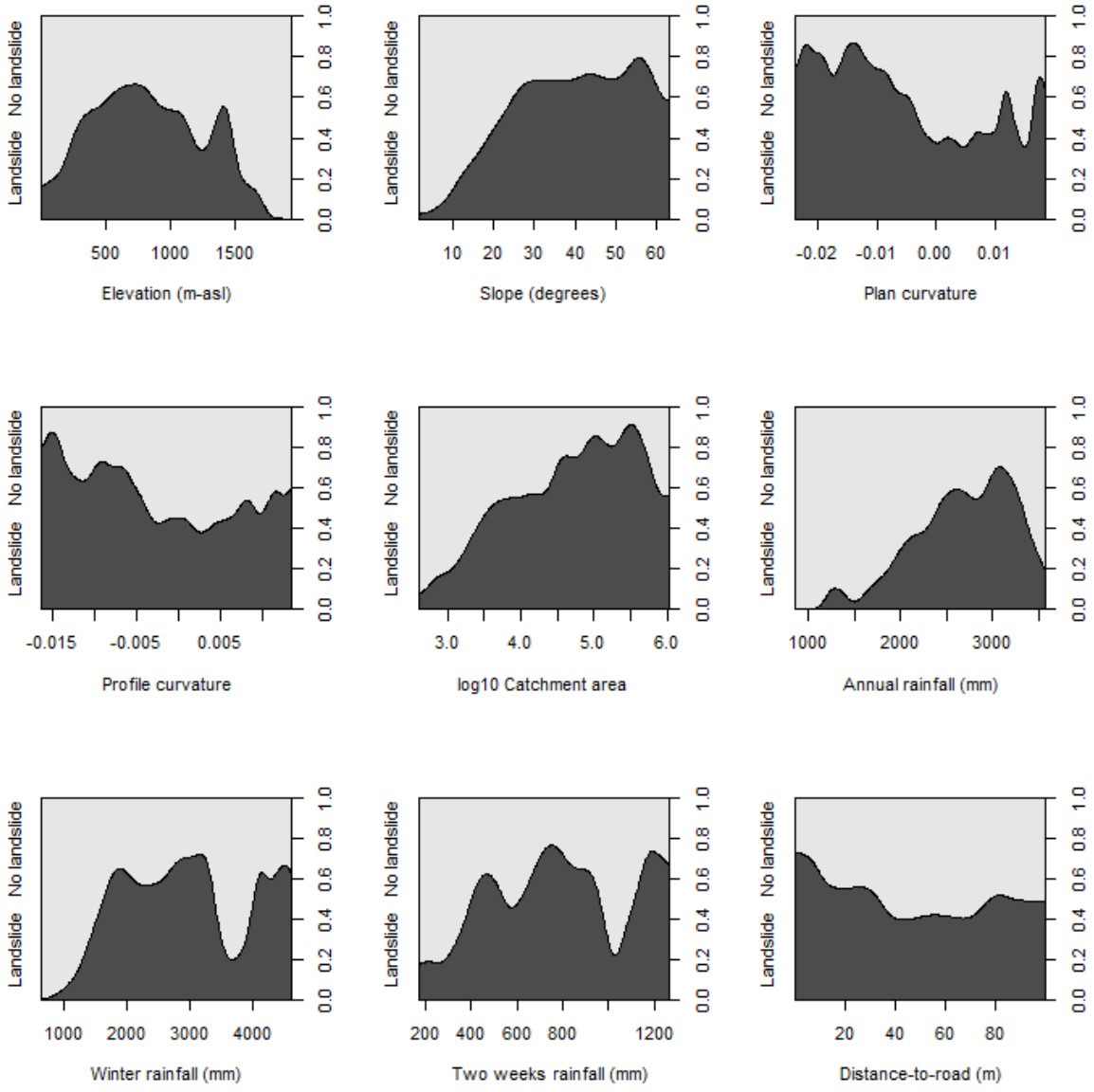

Figure 5.9. Conditional density plots for susceptibility of landslide to occur or not by topographic variables (slope, elevation, plan curvature, profile curvature), rainfall variables and distance-to-road.

General characteristics of topographic conditions that may lead to landslide initiation were estimated for elevation, slope, plan curvature, profile curvature and catchment area (log_{10}). In terms of elevation, landslides had a higher probability (Prob. ~0.5 to~ 0.6) to occur at heights of 250 m to 1250 m. Slope showed a strong linear increase in probability of landslides from 0° to

25°, where the probability (~0.75) of slope failure levels off. The probability of slope failure for profile and plan curvature was generally highest (~0.8) for concave curvature (represented by negative values), levels off around a curvature of 0 (~0.4), and slightly increased for convex curvatures (~0.6). By considering both plan and profile curvature as concave, it was apparent that gullies or channels have the highest estimated probability to landslide initiation. Catchment area had a general positive linear trend to related to the estimated probability of landslides to occur, with high estimated probabilities (~0.8 to 0.9) peaking at catchment sizes from 10 000 m^2 to 300 000 m^2.

Regarding rainfall, in general wetter conditions result in higher estimated probabilities of landslide initiation, which peak at an estimated probability of 0.6. Landslides related to annual rainfall tend to increase probability gradually as rainfall increases and peaks at around 2500 mm to 3000 mm. Winter and two weeks rainfall both had sharp increases in estimated probabilities at relatively lower rainfall amounts. The influence of winter rainfall on landslides appears to level off at about 1500 mm; two weeks rainfall had a similar plateau at 400 mm, but was not as apparent as winter rainfall. Also notable, winter and two weeks rainfall had a sharp drop in estimated probability at relatively high rainfall amounts, 3500 mm and 1000 mm respectively.

The influence of logging roads to landslide initiation is represented by the variable distance-to-road. The conditional density plot shows that the estimated probability of landslide initiation increases with shorter distances to the roads. The decline in probability of landslide initiation levels off at around 50 m from a road.

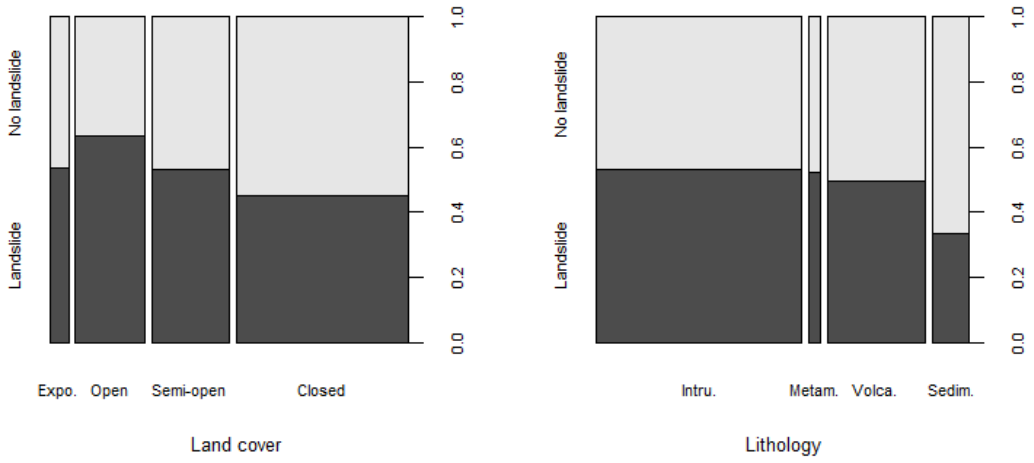

Figure 5.10. Spine plots of land cover and lithology class to landslide initiation. The width of the bar columns represent the proportion of samples available to estimate the probability

The ability of land cover and lithology to characterize landslide susceptibility was illustrated using spine plots (Figure 5.10), which plot the estimate probability of each category/class to initiate landslides. The spine plot for land cover shows that probability of landslide initiation was greatest in open forest canopy (prob.~0.6), followed by exposed ground and semi-open forest canopy (~0.5). Closed forest had the lowest estimated probability (~0.4) of the forest cover types. It should be noted that the small spine plot column width of exposed soil may indicate low confidence in its estimation.

The lithology classes, volcanic, intrusive and metamorphic were very similar (prob.~0.5), while sedimentary rock types had the lowest estimated probability (~0.35). Again, there were a low number of samples for the metamorphic lithology class, which may had led to low confidence in its estimation.

5.3.2 Performance results

The performance results of the spatial cross-validation estimation show that all of the GAMs outperform the GLMs (Figure 5.11, Table 5.7). The difference between the GAMs and GLMs are found to be statistically significant (p-value<0.001) using Wilcoxon signed rank sum test for pairwise model comparison. Also, the incorporation of annual rainfall as a predictor variable only marginally improved spatial cross-validation AUROC estimations compared to models using only the remaining environmental factors (land cover, logging roads, topographic and lithologic). Also, there is no statistical difference between RTLG-GAM with LTG-GAM (AUROC = 83.46% and 83.34%) and RLTG-GLM with LTG-GLM (81.82% and 81.18%).

Table 5.7. Median (and the interquartile range - IQR) of model performance for GAM and GLM models estimated using spatial cross-validation.

	SPCV		Study area	
Model	**AUROC % (IQR)**	**Sensitivity % at 90% specificity (IQR)**	**AUROC %**	**Sensitivity % at 90% specificity**
RLTG-GAM	83.46 (8.59)	52.72 (0.15)	85.61	51.96
LTG-GAM	83.34 (6.64)	52.67 (0.16)	85.16	51.88
RLTG-GLM	81.82 (6.53)	52.65 (0.2.4)	82.82	43.96
LTG-GLM	81.18 (5.34)	52.63 (0.24)	82.56	45.05

The larger range of AUROC values for RLTG-GAM and RLTG-GLM compared to LTG-GAM and LTG-GLM indicated that there was more variation in performance results when rainfall was incorporated as a predictor variable (Figure 5.11A). Therefore, models with rainfall may be overfitting to the training sample. In particular, it appears as if RLTG-GLM had the worst case of overfitting with AUROC values ranging from 66% to 90%. Also, there was a general skewness in the distribution of AUROC results, with the exception of LTG-GLM, towards higher AUROC values.

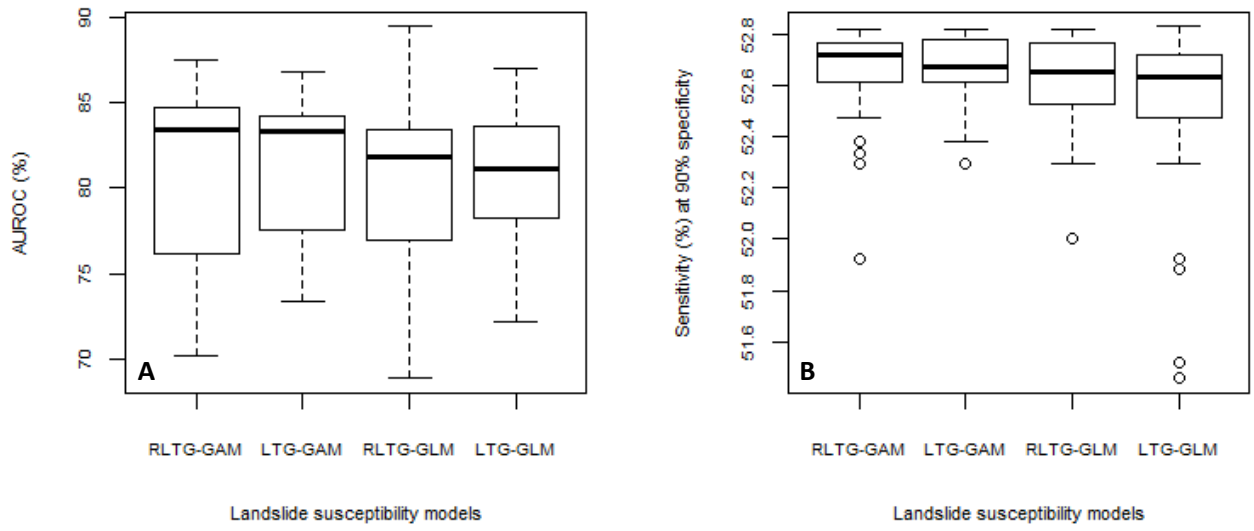

Figure 5.11. Box-and-whisker plots of landslide susceptibility model performances for AUROC (A) and sensitivity at 90% specificity (B).

Model comparison using the entire data for training and testing resulted in a slight overestimation of model performance compared to median performance using spatial cross-validation. This ROC curve illustrated that the general shape of the curve for GAMs and GLMs were respectively similar (Figure 5.12). The GAMs had higher sensitivity performance values when specificity is generally high, compared to the GLMs.

The sensitivity results using spatial cross-valdiation show little variation among GLMs and GAMs with or without a rainfall predictor variable (Table 5.7). However, the box-and-whisker plot of spatiatl cross-valiation results of sensitivity at 90% specificity show that the GAMs slightly outpeform the GLMs (Figure 5.11B); the Wilcoxon signed rank sum tests show that the only statistically significant difference in sensivity performance was between the GAMs and

GLMs (p-value<0.001). Futhermore, the range of performance results were smaller for GAMs than GLMs.

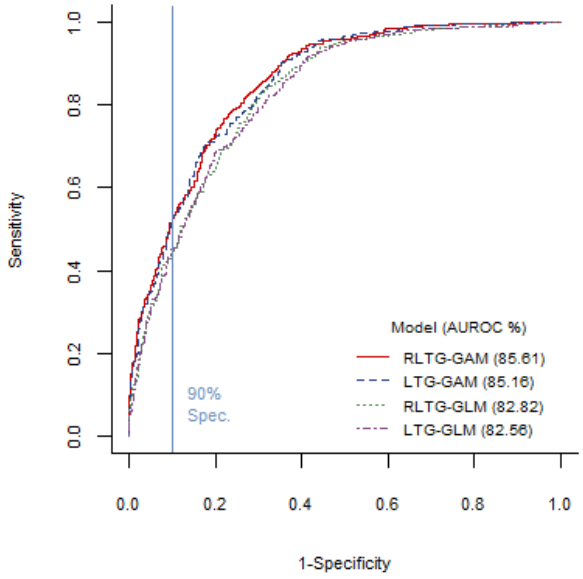

Figure 5.12. ROC curve for landslide susceptibility models (GAM and GLM) trained using a sample of the entire study area.

5.3.3 Susceptibility map

The landslide susceptibility model's capability to discriminate stable and unstable areas can best be appreciated by the examining the percentage of landslide initiation points falling in areas with very high (≥0.8) and low (<0.5) predicted probabilities of landslide initiation (Table 5.8). In terms of levels of susceptibility across the island, 73% of the study area was predicted to have a probability of <0.5 to landslide initiation, which contains 20% of the landslide initiation points in the landslide inventory. The highest probabilities (≥0.8) were predicted for 4% of the study area, which contains 33% of the landslide initiation points. The presence of landslide initiation points per square kilometer was 0.156 and 0.005 for high (≥0.8) and low (<0.5) predicted probabilities.

Visually, the LTG-GAM map illustrates the strong relationship of landslide susceptibility with the topographic variables and logging roads (Figure 5.13). In particular, debris flow channels, which can be characterized as gullies having steep slopes and convex plan curvature, are features that were easily identifiable in the susceptibility map by the highest predicted

probabilities (prob. ≥0.9). Logging roads through intensely forested areas were predicted to have high probabilities, which indicate their importance for landslide initiation prediction (insert map of Figure 5.13). This relationship was most apparent in locations that had a high density of winding logging roads on steep slopes.

In general, locations of low susceptibility (prob. <0.3) were associated with mountain crests and valleys (Figure 5.13).

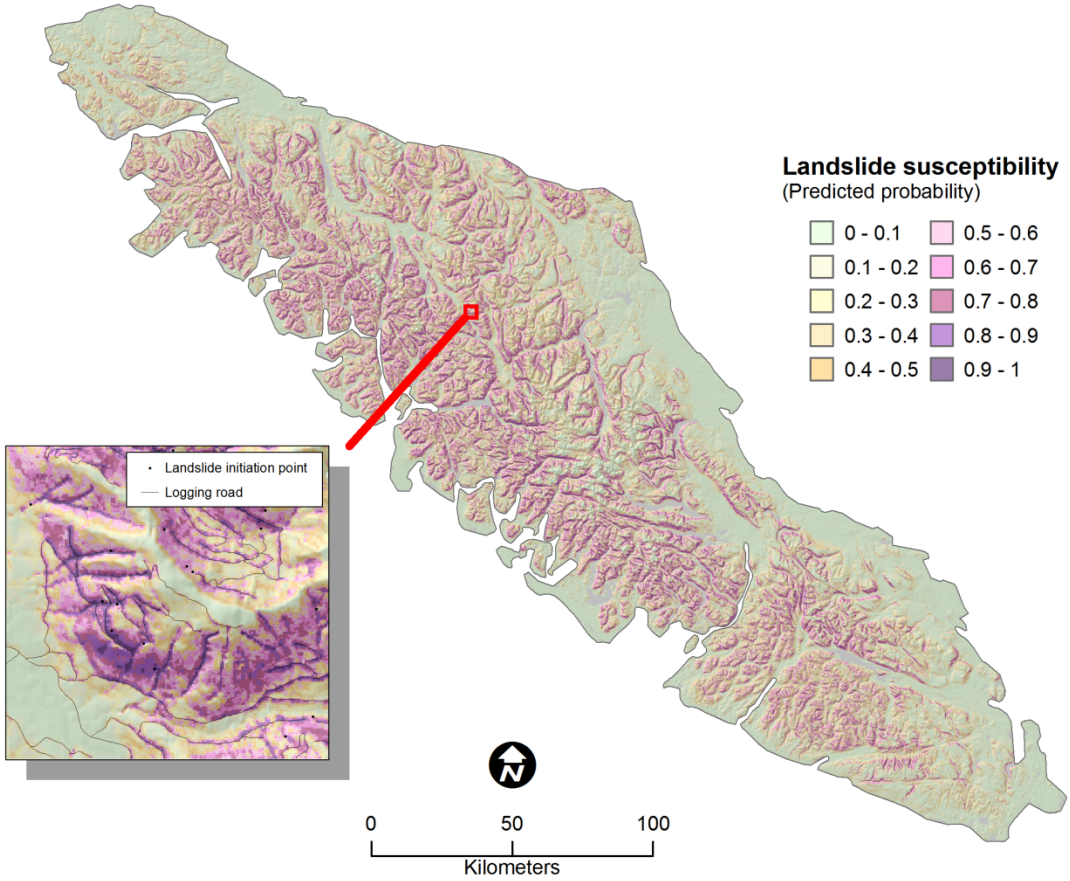

Figure 5.13. Landslide susceptibility map for LTG-GAM. Probability values closer to 0 represent low susceptibility and values closer to 1 represent high susceptibility to landslide initiation.

Table 5.8. Summary of landslide susceptibility classes in terms of percentage of study area and the percentage of landslide initiation points that fall within each class in Figure 5.13 (LTG-GAM).

Predicted probability	Area cover (%)	Cumulative area (%)	Landslides (%)	Cumulative landslide (%)
0.0 - 0.1	34.0	100	0.8	100
0.1 - 0.2	12.5	66.0	1.9	99.2
0.2 - 0.3	9.1	53.6	1.6	97.3
0.3 - 0.4	8.5	44.4	4.7	95.8
0.4 - 0.5	8.8	35.9	10.6	91.1
0.5 - 0.6	9.0	27.2	10.5	80.4
0.6 - 0.7	8.2	18.2	16.9	70.0
0.7 - 0.8	5.8	10	19.9	53.1
0.8 - 0.9	3.1	4.2	19.2	33.2
0.9 – 1.0	1.1	1.1	13.9	13.9

5.3.4 Nonlinearity and variable importance

All of the variables, except catchment area, are included in the majority of model repetitions as nonlinear (Table 5.9). In particular, slope, distance-to-road, elevation and plan curvature were included in the GAMs as nonlinear for all repetitions. Annual rainfall and profile curvature were included in approximately 80% of the repetitions as nonlinear.

The confidence in nonlinear transformations, using a spline function, was related to sampling distribution of landslide points for a given predictor variable. In most cases, the confidence of the spline function for non-parametric smoothing was much weaker at the tail ends of a plot distribution (Figure 5.14). In terms of categorical variables, exposed ground had the least confidence in model transformation relative to other forest based land cover classes. Additionally, metamorphic and sedimentary rocks had the lowest confidence for lithology classes. Exposed ground, metamorphic and intrusive classes had the lowest number of landslide samples compared to their other associated classes (Table 5.5).

The differences in model predictions related to categorical variable was explored in more detail by examining the odds ratios derived from the estimated model coefficients in RLTG-GAM. As a result, it was found that forest land cover classes with lower tree canopy cover had higher levels of susceptibility to landslide initiation (Figure 5.14). The most stable class, closed

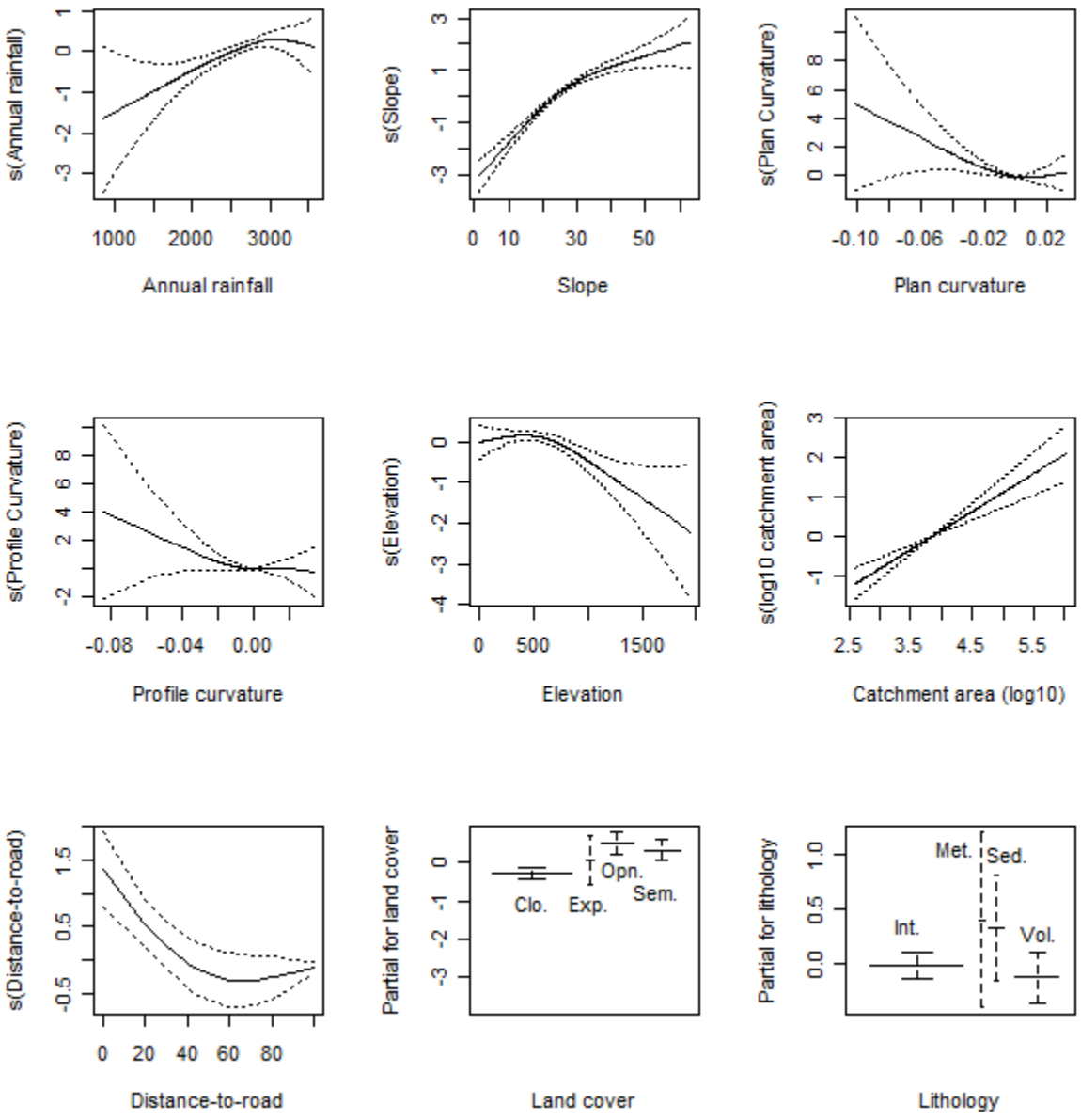

Figure 5.14. Transformation of predictor variables in the generalized additive model for RLTG-GAM that utilize the entire study area as a training sample. A spline function for non-parametric smoothing of the variable, s(variable), indicates a nonlinear transformation. The dotted lines represent confidence bands.

forest, was used as a reference to compare odds ratios (OR) to determine relative levels of susceptibility within land cover. The most susceptible land cover class was open forest with odds of landslide initiation occurring 2.13 times as large as the odds for landslide initiation in a closed forest. Semi-open forest had 1.80 times the odds compared to closed forest, and exposed ground had the lowest odds with 1.39 times the odds of landslide initiation occurring in closed forest.

Table 5.9. Variable selection frequencies and percentage of nonlinear occurrences for GAM models from 25-repeated 5-fold spatial-cross validation

	Nonlinear occurrence (%)	
Variable	**RLTG-GAM**	**LTG-GAM**
Slope	100	100
Distance-to-road	100	100
Elevation	100	100
Plan curvature	100	100
Annual rainfall	83	-
Profile curvature	82	83
Catchment area (log_{10})	33	46

The lowest level of susceptibility in lithology classes was associated with volcanic rocks. Its OR, also derived from RLTG-GAM, was 0.90 compared with intrusive rock (the most abundant lithology class), which implies the landslide initiation was less likely to occur in volcanic rocks than intrusive rocks. Metamorphic and sedimentary rocks were most likely to have occurrence of landslide initiation, with respective odds 1.52 and 1.41 times as large as the odds for landslide initiation in intrusive rocks.

The relative importance of each predictor variable was explored by systematically testing the AUROC performance of RLTG-GAM with one variable removed at a time. Overall, all of the predictor variables contributed to positive increases in AUROC performance (Figure 5.15).

Slope is the largest contributor to AUROC performance (+4.5%), followed by catchment area (log_{10}; +0.97%), land cover (+0.90%), distance-to-road (+0.83%) and elevation (+0.80%). Annual rainfall (+0.45%), plan curvature (+0.20%) and profile curvature (+0.10%) had a much smaller contribution. However, the rock classes for lithological units contributed by far the least to AUROC performance (+0.01%) in an additive model.

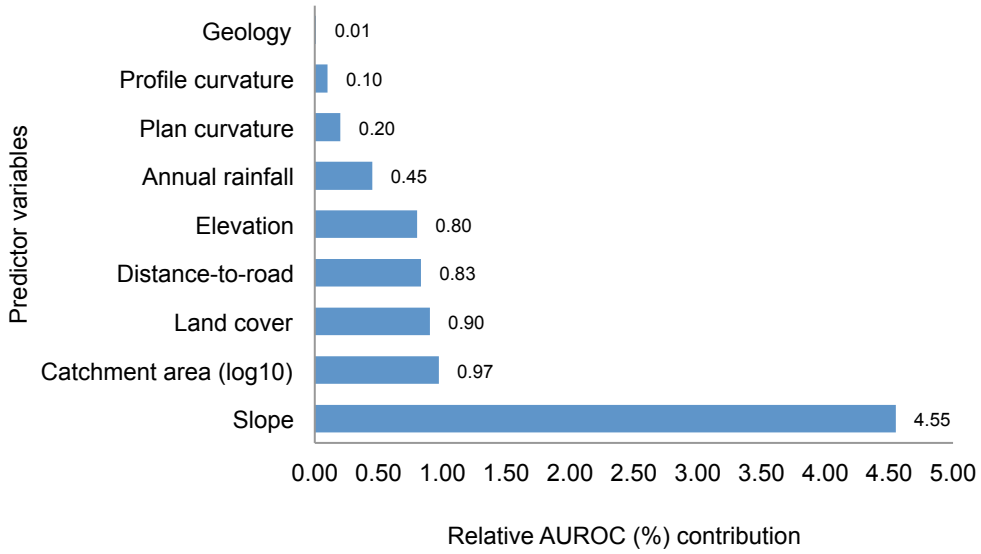

Figure 5.15. The percent of AUROC model improvement is based on a comparison of a model that includes all of the predictor variables (RLTG-GAM).

5.4 Exploring interactions related to landslide initiation

5.4.1 Land cover interactions with rainfall and slope

Some of the relationships between environmental factors and landslide initiation may be better represented as interaction terms in landslide susceptibility models. Examining model extensions of LTG-GLM with interaction terms showed that landslide initiation had different associations to rainfall in each land cover class (Figure 5.16). Land cover with little vegetation (e.g., exposed ground/freshly logged) had its highest estimated probability for failure with lower rainfall amounts (1700 mm to 2200 mm). Areas that have had at least a couple of years to recover after harvesting (open-forest) had higher estimated probability for failure with higher rainfall amounts (2200 mm to 3400 mm). Semi-open forest had generally the same distribution of probability to failure as open forest; however the LTG-GLM model results indicated that the odds of landslide initiation in open forest was about 2 times larger than semi-open forest (Figure 5.16). The estimated probabilities to failure for closed forest were highest near the greatest rainfall amounts (2800 mm to 3400 mm), but were in general lower than exposed ground and open forest.

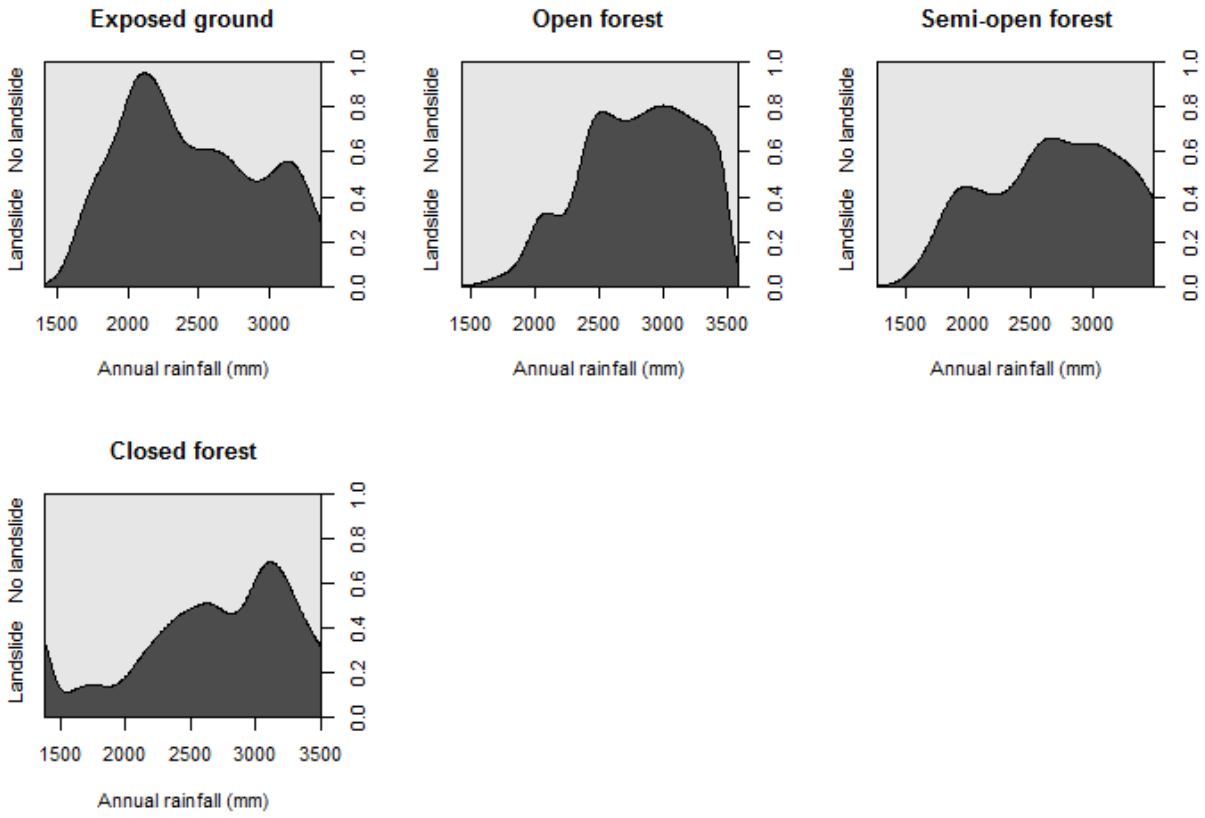

Figure 5.16. Conditional density plots of annual rainfall amount for a given land cover class to initiate landslides

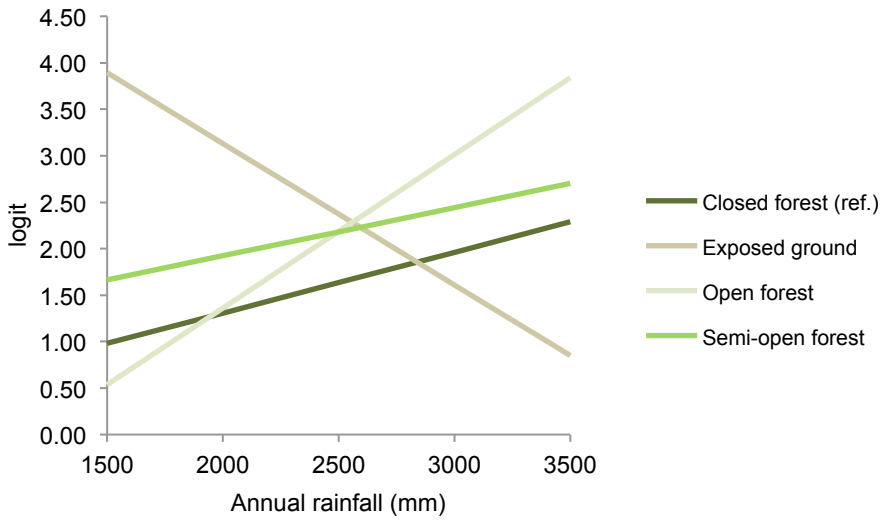

Figure 5.17. Delta logit plot comparing the interactions between annual rainfall and land cover. These values were predicted under otherwise equal conditions using interaction terms in an extension of LTG-GAM.

A comparison of difference in logit values for interaction terms in LTG-GLM showed that open, semi-open and closed forest cover lead to higher susceptibility as the amount of rainfall increased (Figure 5.17). Also, greater rainfall amounts resulted in open forest becoming more susceptible to landslide initiation then semi-open forest and closed forest, in particular around 2500 mm of annual rainfall. The relationship of rainfall with increasing susceptibility in semi-open forest and closed forest was generally the same; however, closed forests had relatively lower susceptibility.

In terms of slope, the conditional density plots illustrated slightly different interactions between land cover and landslide initiation (Figure 5.18). It appears that estimated probabilities of failure were generally highest for steep slope angles in open forest. Failure of slope in exposed ground seems to be more likely at lower slope angles (15° to 35°), and failure in semi-open forest appears to be more likely at steeper slopes (25° to 45°). The estimated probability of failure of closed forest was generally similar to semi-open forest, with the exception that the highest estimated probabilities for failure were at the steepest slopes (50° to 65°).

Exploring interaction terms between slope and land cover (Figure 5.19) show the susceptibility of land cover on slope angle was relatively the same, with odd ratios from 1-1.2, for open forest and semi-open forest at slopes less than 30°. As the slope angle increased (<30°) open and semi-open forest maintained a somewhat similar susceptibility with the odds of landslide initiation being 1.5 times more in open forest than semi-open forest. Landslide initiation in closed forest was consistently less than open forest with odds that ranged from 1.8-2.5 times less than landslide initiation occurring in open forest. Exposed ground had the highest susceptibility of all land cover classes to a slope angle of approximately 25°. The initial odds for exposed ground at a slope of 10° was at least 2.1 times larger than the odds of landslide initiation occurring all other land cover classes.

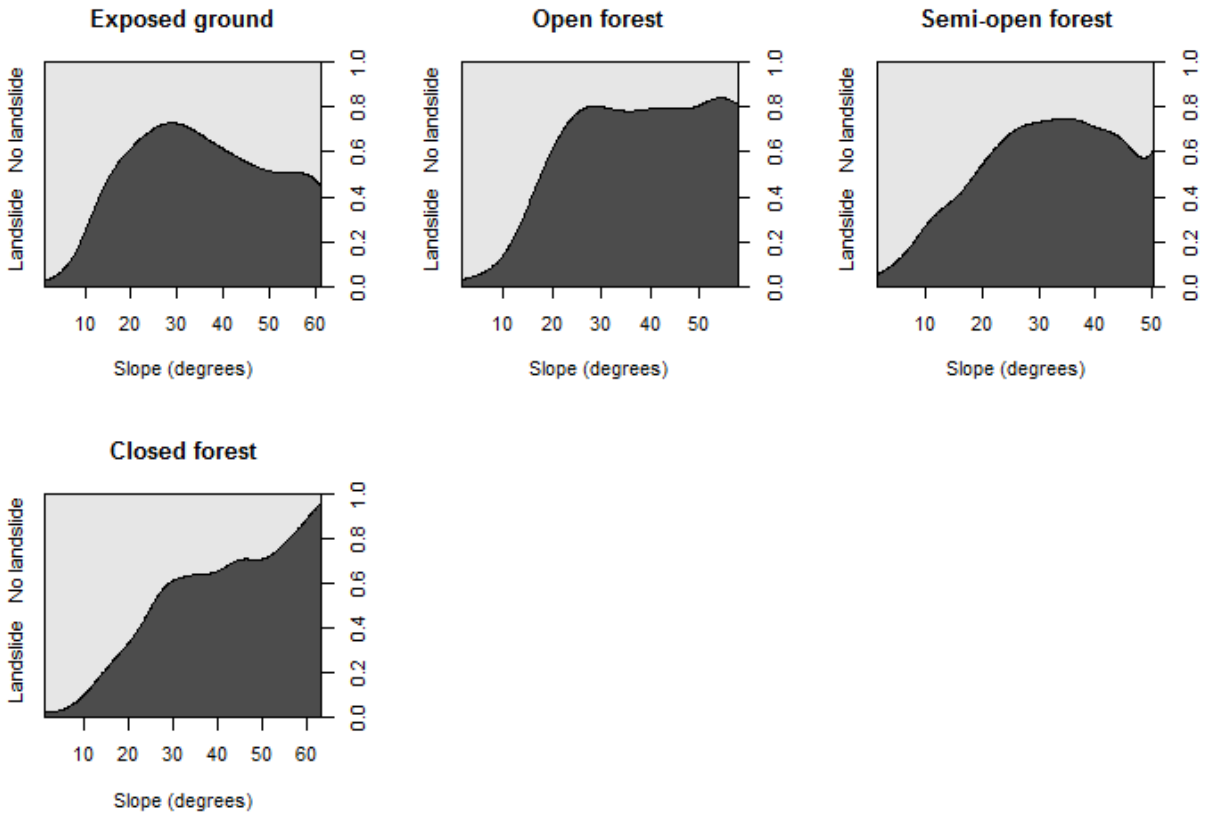

Figure 5.18. Conditional density plots of slope angle for a given land cover class to initiate landslides

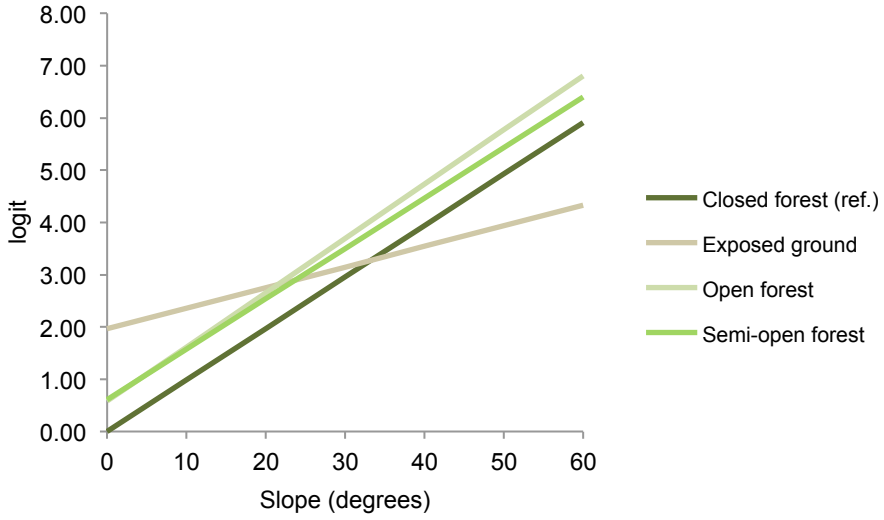

Figure 5.19. Delta logit plot comparing the interactions between slope and land cover. These values were predicted under otherwise equal conditions using interaction terms in an extension of LTG-GAM.

5.4.2 Lithology interactions with rainfall and slope

The conditional density plots in Figure 5.20 indicate some distinct differences in rainfall conditions for given lithology leading to landslide initiation. In particular, there was a threshold for volcanic rocks, where the probabilities of landslides increase sharply around 2000 mm and peaks at about 3000 mm (prob. 0.8). Below this threshold, the probability of landslides was very low (0.1 to 0.0). Sedimentary rocks had generally higher probability of landslides for annual rainfall from 2250 mm to 3250 mm (0.6 to 0.8). Intrusive rocks had a similar relationship to rainfall and landslides as sedimentary rock. Metamorphic rocks had the highest probabilities (0.4 to 0.6) for landslides at the lowest amounts of rainfall from 1500 mm to 2000 mm.

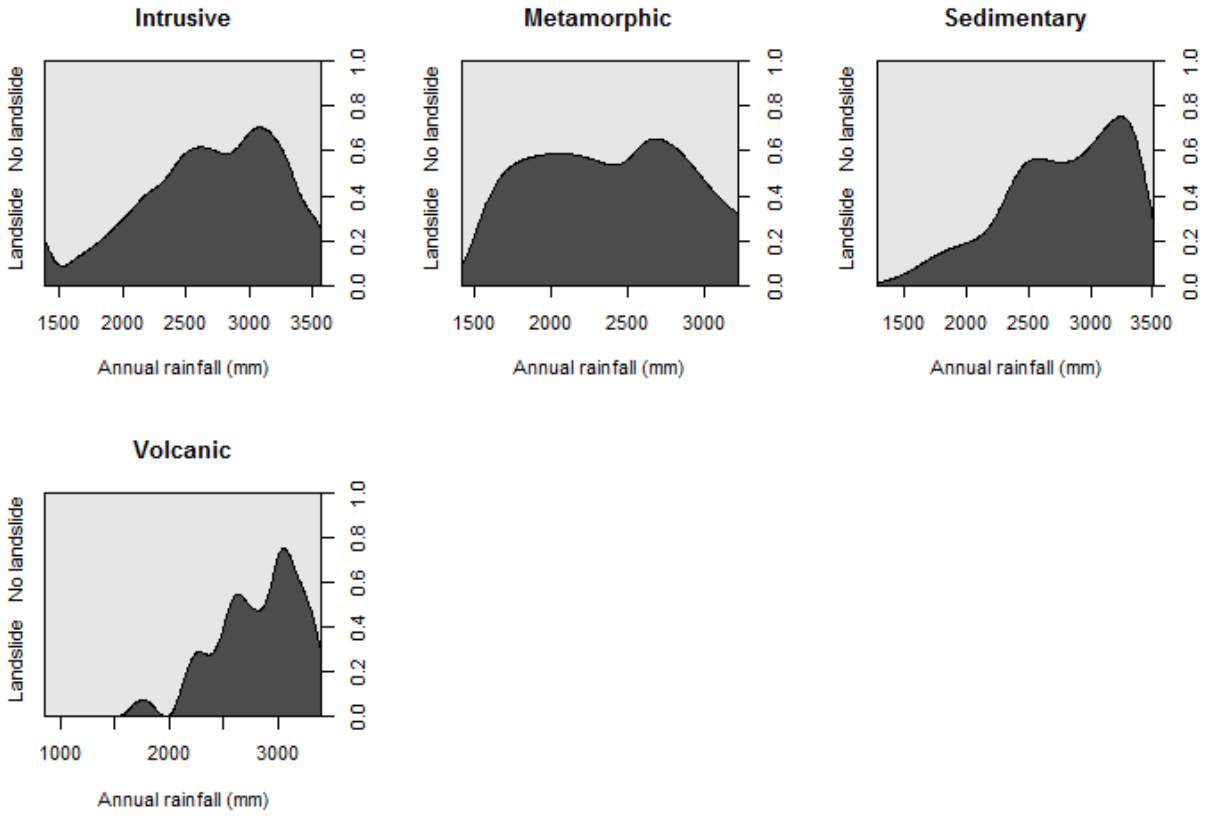

Figure 5.20. Conditional density plots of annual rainfall for a given lithology class to initiate landslides

The interactions between annual rainfall and lithology shown in Figure 5.21 illustrate that intrusive, volcanic and sedimentary had a positive relationship of landslide susceptibility, which increased with higher amounts of rainfall. In contrast, metamorphic rocks illustrated a negative

relationship of landslide susceptibility and annual rainfall. This negative relationship can be related to more landslides occurring in areas of lower annual rainfall. Landslide susceptibility related volcanic rocks only become present after 2250 mm of annual rainfall.

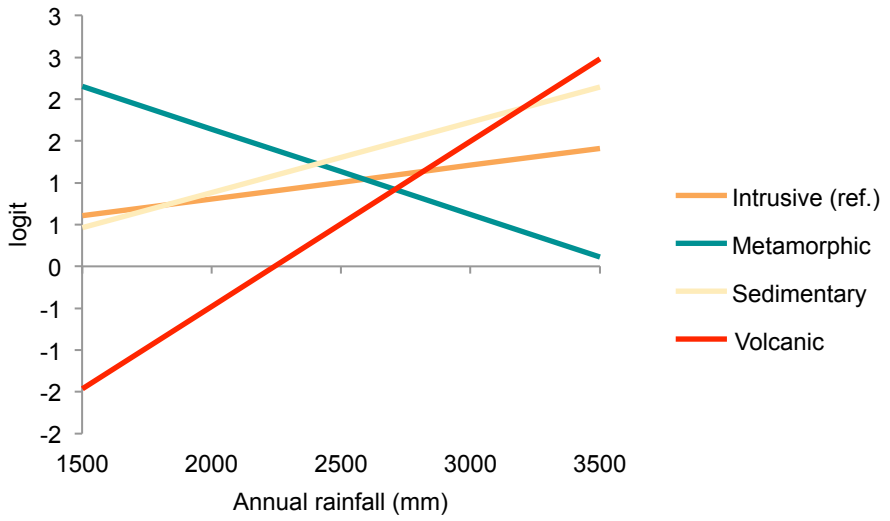

Figure 5.21. Delta logit plot comparing the interactions between annual rainfall and lithology. These values were predicted under otherwise equal conditions using interaction terms in an extension of LTG-GAM.

In terms of slope and lithology, the conditional density plots (Figure 5.22) for intrusive and volcanic rock were similar, which may indicate that hillslope angles leading to landslides were alike. Sedimentary rock had a distinct positive linear relationship of slope and probability of landslide initiation. Metamorphic rock and slope had a threshold for a sharp increase in landslides after approximately 15° slope angle. Also, metamorphic rock had the highest probabilities of failure (> 0.8) for slope angles greater than 40°.

The interaction between slope and landslide susceptibility was positive for all lithology rock classes (Figure 5.23). Intrusive, sedimentary and volcanic rocks demonstrate shared a very similar relationship, illustrating that there was not much difference in the influence of slope for these lithology classes (6.8). However, there was a distinct difference in relationship regarding slope of failure in metamorphic rock. At around 20° slope, metamorphic rock was more susceptible to landslide initiation than the other classes; at a slope of 40° the odds difference increased to being 3.3 times larger than the odds of landslide initiation occurring in the other classes.

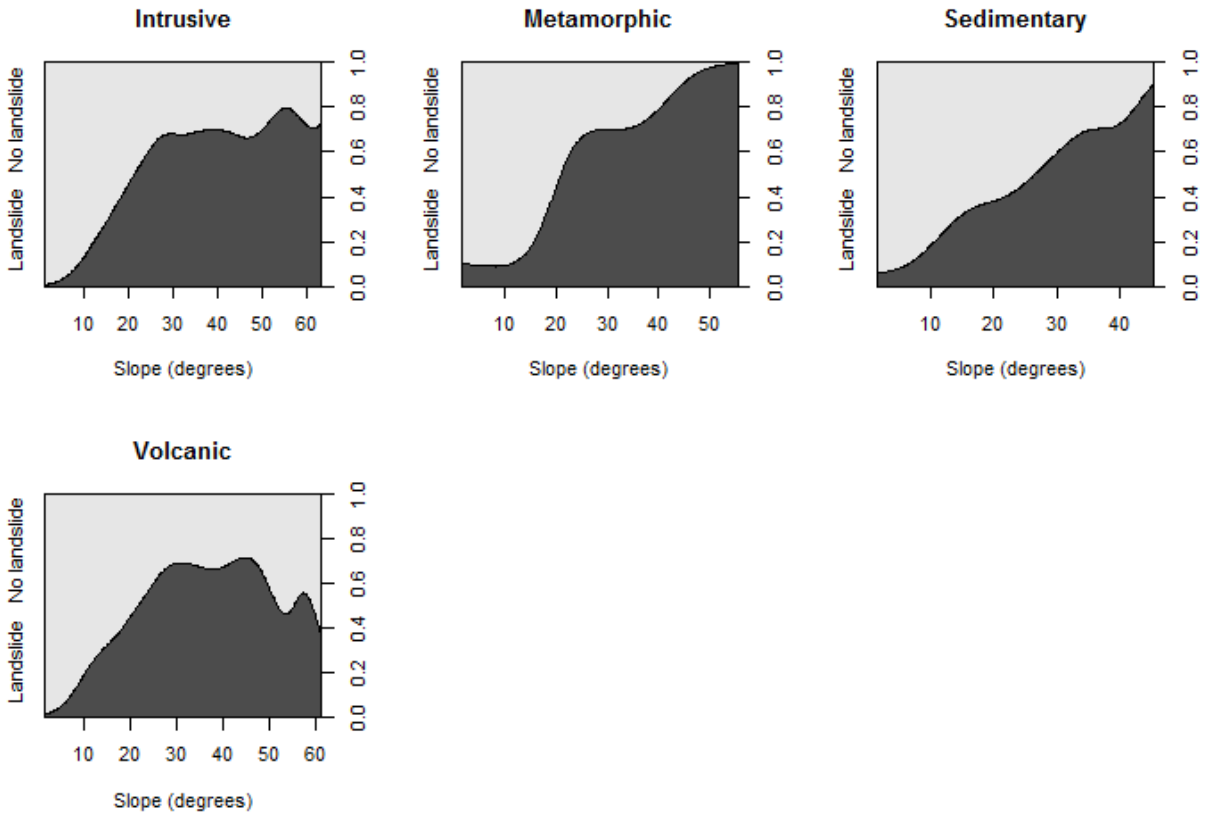

Figure 5.22. Conditional density plot of slope for a given lithology class to initiate landslides

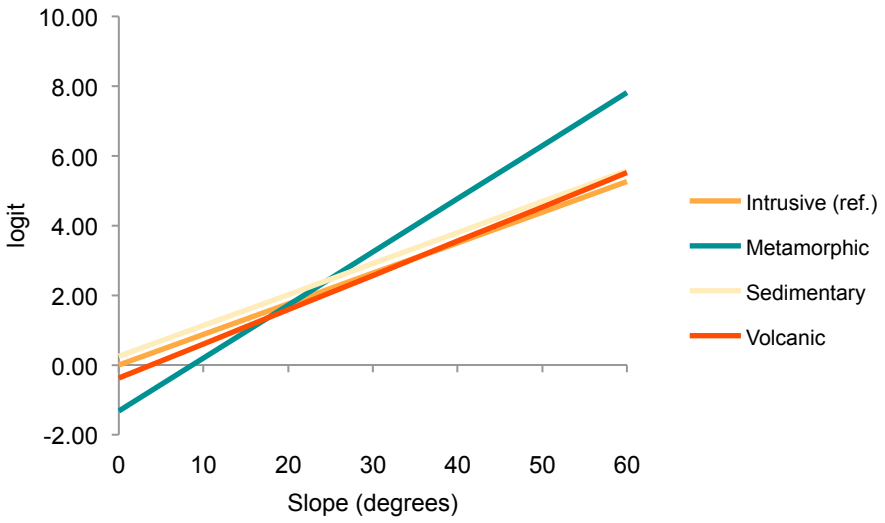

Figure 5.23. Delta logit plot comparing the interactions between slope and lithology. These values were predicted under otherwise equal conditions using interaction terms in an extension of LTG-GAM.

5.4.3 Interactions and model performance

In all cases, the inclusion of an interaction term in a GLM slightly improved model fit and performance over the null model (LTG-GLM; Table 5.10). The interaction of rainfall and land cover provided the most model improvement, followed by the interaction of rainfall and lithology. Interactions between slope and land cover/lithology also improved the landslide susceptibility model, but not to the same extent as rainfall. The interaction between slope and land cover resulted in better model performance than the interaction between slope and lithology.

Table 5.10. Comparison of rainfall and slope interactions to land cover and lithology.

GLM Model	AIC	AUROC
LTG-GLM	1291.21	83.44
Rain: land cover	1283.44	83.94
Rain: rock class	1283.53	83.89
Slope: land cover	1290.47	83.73
Slope: lithology	1294.07	83.51

5.4.4 Distance-to-road and slope

A plot of predicted probability of landslide initiation to slope and distance-to-road demonstrates several non-linear relationships (Figure 5.24). The probability of landslide susceptibility related to distance-to-road leveled off at distance of approximately 60 m. The odds of landslide initiation at 60 m distance were 4 times larger than the odds of landslide initiation occurring at a distance of 0 m from a logging road. Furthermore, slope appears to have had a stronger impact on increasing susceptibility as the distance-to-road decreased.

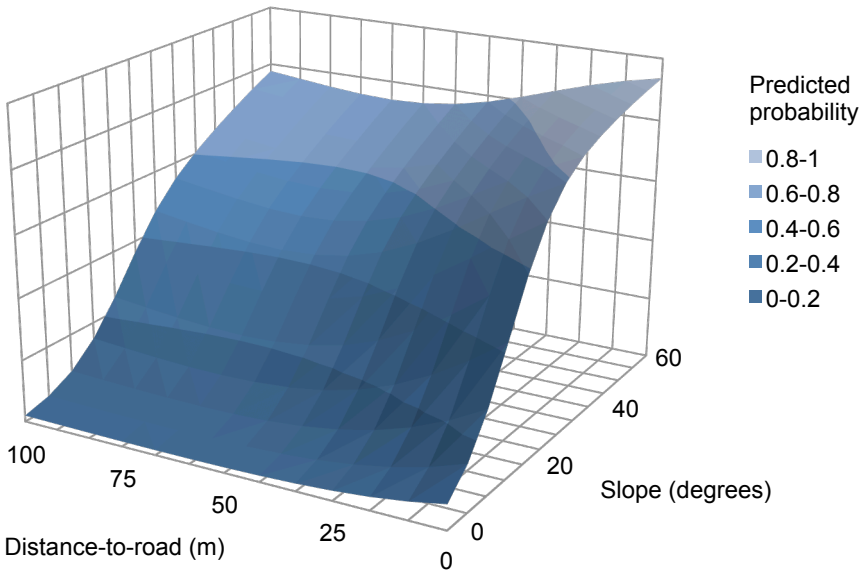

Figure 5.24. The probability of slope failure for a given distance-to-road and slope angle

5.5 Summary

Rainfall interpolation using OCK was found as the best geostatistical method for all temporal scales in this study. The automatic stepwise-variable-selection procedure determined annual rainfall to best reflect the distribution of landslides in this study. The accuracy of the land cover classification was 82%. Closed forest was the most abundant forest class.

The inclusion of rainfall as a predictor variable only marginally enhanced the performance of landslide susceptibility models. There was no statistical difference between a GAM using rainfall and not. However, overall the GAMs performed significantly better than the GLMs in terms of AUROC.

Examining the relative contribution of each predictor variable showed that slope, catchment area, land cover, distance-to-road, and elevation contributed the most to improve landslide susceptibility model performance. The land cover class that was most susceptible to landslide initiation was open forest followed by exposed ground, semi-open forest and closed forest. A threshold for landslide susceptibility from distance-to-road was found to be 60 m; at this distance the odds of landslide initiation occurring were 4 times less than the odds of landslide initiation at 0 m distance from logging roads.

Modeled interactions illustrated the impact of annual rainfall conditions on landslide initiation for a given land cover classes. Exposed ground was the most susceptible land cover class to landslide initiation at lower annual rainfall amounts; it was also the most susceptible land cover class to landslide initiation at gentle slopes. At rainfall amounts greater than 2500 mm open forest was the most susceptible to landslide initiation. Closed forest remained the least susceptible land cover class for all annual rainfall conditions.

The interactions between annual rainfall and lithology illustrated that metamorphic rock was the most susceptible lithology class to landslide initiation at lower annual rainfall amounts. Metamorphic rock was also much more susceptible to landslide initiation for lower slope angles.

Based on the model predictions, the most susceptible 4% of the study area had 29 times higher density of landslide initiation points than the least susceptible 73% of the study area (0.156 versus 0.005 landslides/km^2).

Chapter 6

Discussion

6.1 Interpolating mountain rainfall

The general pattern of rainfall across Vancouver Island was best represented by the geostatistical interpolation methods using OCK and OK. The general pattern of annual rainfall amount was higher in the western portion of the island, which represents the windward side of the Vancouver Island Ranges. In contrast, interpolated rainfall amounts on the leeward side were much lower. Rainfall patterns in all maps also show drier conditions in lowlands located north and south on the island.

Strong variations in spatial-temporal patterns of rainfall were found in OK, UK and OCK maps. The pattern of annual rainfall illustrates higher amounts of rainfall occurring in the northern area of the Vancouver Island Ranges, particularly adjacent to the northeast portion of the mountain ranges. The winter rainfall pattern had the wettest areas divided in the northwest and southwest portions of the ranges, with a zone of less wetness in the centre of the island. Rainfall patterns related to the 2006 November storm were concentrated in the southern part of the mountain ranges, with a pocket of high rainfall around the Tahsis village weather station located in the northwest.

The general pattern of 2006 annual rainfall interpolated with OCK is fairly similar to average precipitation on Vancouver Island. The OCK rainfall on the windward side of the insular mountains was 2500-3500 mm; the mean annual precipitation for this west is >3000 mm (McKenney et al., 2006). On the leeward side, OCK rainfall was 750-2000 mm; the mean annual precipitation for the east is 800-1200 mm (McKenney et al., 2006).

For this study, cokriging yields the best rainfall interpolation for all temporal scales: two weeks of rainfall associated with an extreme storm in 2006, winter rainfall for 2006-2007, and

2006 annual rainfall. OCK with elevation as a co-variable usually obtained the smallest prediction errors, which was also the case in a study of rainfall interpolation by Goovaerts (1999). The OCK maps of rainfall show less detail than then UK. The UK maps are more detailed because the interpolation of rainfall is greatly influenced by the pattern of the DEM (Goovaerts, 2000). Perhaps, the smoothing of the interpolation in OCK, related to the combination of the cross-semivariograms, provided better results in that rainfall is more generalized than UK, but less generalized than OK; the UK values may have been too heterogeneous to capture the regional pattern of rainfall in this area. Furthermore, the interpolation of cokriging, which uses a linear combination of rainfall observations elevation, allows for flexible fitting (Goovaerts, 1999). In contrast, UK relies on an elevation trend for interpolation of rainfall.

The most difficult problem with the interpolation of rainfall in the mountainous terrain of Vancouver Island is the distribution and number of point observations. The small number of stations (53) covering the ~32 000 km^2, is a low number to adequately represent the scale and variability of rainfall occurring in a mountain area; about 62% of the weather stations were located near a coast. Additionally in terms of characterizing landslide susceptibility with rainfall, it is important to have a rain gauge network that captures the climatic conditions where landslides are occurring. Thus, the elevation range and spatial distribution of the rain gauge network should be somewhat similar to that of the landslides under analysis. Furthermore, the removal of some stations in the analysis may have had a strong impact on rainfall patterns. Thus, the relationship of landslides to rainfall may also be affected by the weather stations available for interpolation, especially in areas where rainfall is particularly heterogeneous.

6.2 Land cover accuracy assessment

The overall accuracy of 82% can be considered ‘very satisfactory’ for classification of land cover. The test sample will always affect the measure of accuracy performance of a classifier. The quality of test samples is controlled by availability of adequate ground reference data, such as images with higher spatial resolution or on site observations that correspond with the time of the images used for classification. In addition, an rigorous sampling strategy is required to produce a statistically valid analysis (Congalton and Green, 1999). Due to time constraints, this

study used a simple random sample of 100 points across the study area. Although, this approach is generally unbiased, other techniques (e.g. stratified random sampling) can be implemented to ensure that the proportions of each class in the final classification map are properly represented in the test sampling (Congalton and Green, 1999). Since the training and test samples are both collected through interpretation of the Landsat TM imagery, it would be more effective to use an assessment approach that utilizes all training samples, such as repeated *k*-fold cross-validation or even better, repeated spatial cross-validation (Brenning et al., n.d.).

6.3 Geomorphological interpretation of models

In this study, rainfall as a predictor variable did not contribute significantly to predictive improvement for landslide susceptibility modelling. Different rainfall temporal scales were investigated to determine if the relationship to landslide initiation is stronger at a particular temporal scale. By analysing this relationship for an extreme storm occurring in November 2006 (two weeks), the winter 2006-2007 rainfall and the 2006 annual rainfall, it was determined that annual rainfall had the strongest relationship to this set of landslides, which had little impact on model performance. Annual rainfall amount only increased the likelihood of landslide initiation after 2275 mm.

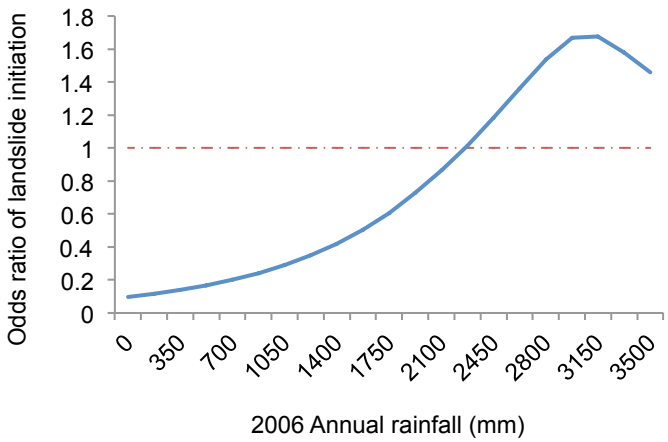

Figure 6.1. Odds of 2006 annual rainfall (mm) amounts for landslide initiation to occur for RLTG-GAM under otherwise equal conditions.

Also, the odds only increased to a maximum of 1.7 at 3150 mm (Figure 6.1). This seems to be related the general regional pattern of annual rainfall and the distribution of landslides. Essentially, the pattern of 2006 annual rainfall characterized the regional topographic pattern of rugged mountains in the west and smoother flatlands in the southeast and north locations of Vancouver Island (Figure 6.2).

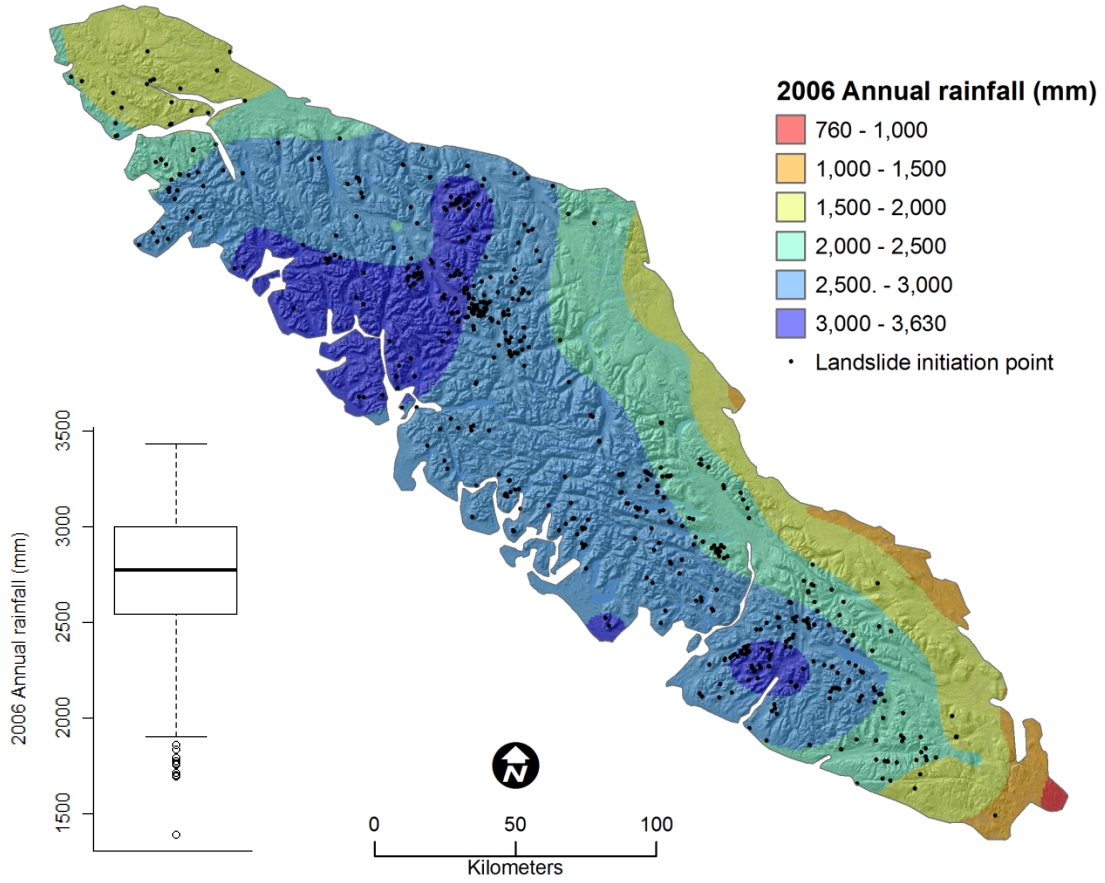

Figure 6.2. The amount of 2006 annual rainfall associate to the landslide initiation points. The rainfall values were interpolated using OCK. The box-and-whisker plot shows the range of annual rainfall values for associated with the landslide initiation points.

Furthermore, the inclusion of rainfall as a predictor variable may increase uncertainty in model predictions. The IQR of AUROC values calculated using repeated spatial cross-validation for a GAM (or GLM) is much wider in a model including rainfall in its predictions (Figure

5.11A). These results are an indication that annual rainfall may contribute random error into a susceptibility model that leads to overfitting.

The majority of the landslides in this analysis occurred during the winter rainy season of 2006-2007. By classifying the forest land cover before the rainy season using scenes from the summer of 2006, this study was able to capture the forest preconditions of landslides. As a result, it was possible to gain insights into the probability of landslides to occur in a given forest condition. In general, forest classes with less forest canopy cover were more likely of have the occurrence of landslide initiation. However, exposed ground, which represents recently logged forest, did not have the highest odds for slope failure. Open forest land cover, which represents forest areas with recent growth of vegetation after logging, was found as having the highest odds ratio compared to closed forest (OR = 2.13) for landslide initiation. This empirically supports previous findings that there is a lag period between logging areas and an increased proneness to landslides (Swanston and Swanson, 1976; Wu and Mckinnell, 1979; Sidle et al., 2006).

Lithology classes for sedimentary, volcanic, intrusive and metamorphic have been observed to have different characteristics of slope stability in British Columbia (Guthrie, 2005; Sterling and Slaymaker, 2007; Pike et al., 2010). However, the relationship of landslide initiation to these lithology classes in this study was not very strong. Previous landslide research of areas along the coast of British Columbia had found similar results (Rollerson et al., 1998). In this study, lithology provided the least model improvement in terms of AUROC when individual variables were compared. Also, the odds ratio only slightly varied among different lithology classes. It has been found on Vancouver Island that rock groups within lithology classes have different relationships to landslide activity (Guthrie, 2005). Therefore, the relationship of landslide to lithology for more detailed classes should be examined to see if specific rock types may improve model predictions of landslide initiation.

The modeled interactions showed that different land cover conditions can influence the amount of rainfall and the friction angle required for slope failure. Hydrological and mechanical conditions of soil are strongly influenced by vegetation cover (Sidle and Ochiai, 2006). In this regional study, the amount of annual rainfall related to landslides was found to vary depending on the general density of vegetation cover. In general, vegetation is the main control for the amount and timing of rainfall reaching the soil (Sidle and Ochiai, 2006). In coniferous forests, such as the ones predominant across Vancouver Island, interception of rainfall is of particular

significance. For dense coniferous canopies, interception has been observed to capture 30-50% of annual precipitation (Dingman, 1994). Furthermore, evapotranspiration in temperate climates is typically lowest for exposed soil and increases with rates that are 5-10 times higher in forests (Jones, 1997).

For this study, the susceptibility of slope angle to landslide initiation was related also related to the density of vegetation cover. In general, landslide initiation occurred on steeper slopes for more densely cover forests. Vegetation roots contribute to the overall soil shear strength and have been observed to be more important for slope stability than evapotranspiration (Greenway, 1987; Sidle and Ochiai, 2006). In general, roots can provide mechanical strength by anchoring the lower soil mantle to a more stable substrate (Greenway, 1987). Forest harvesting that leads to destruction of forest understory drastically increases landslide occurrence (Dhakal and Sidle, 2003). Also, tree roots may be most important for slope stability during high intensity rainstorms or snowmelt; thus, the deterioration of tree roots increases the susceptibility to landslide initiation during such events (Sidle, 1992). Therefore, land cover representing forest density can act as an important surrogate for unknown soil conditions associated to landslide initiation for large regional studies.

Overall, the model improving quality of the interactions highlights the interconnectivity between rainfall, slope, land cover and lithology. This interconnectivity can be expected in a complex geomorphic process, such as landslides, where combinations of environmental factors are involved. Furthermore, the unique rainfall and slope conditions for land cover and lithology indicate that the relationship to landslide initiation is not completely an additive effect. Thus, the impact of rainfall and slope angle on landslides has some dependence on land cover and/or lithology.

A visual comparison of the GAM using land cover, distance to logging roads, lithology, and topographic attributes for predictor variables (LTG-GAM) to previous studies, which applied susceptibility models to watersheds on Vancouver Island, show landslide susceptibility is much higher in gullies in this study. The models by Chung et al. (2002) and Goetz et al. (2011) predicted the susceptibility to debris slide initiation for Tsitika and Klanawa river watershed, respectively. In these previous models and the models used in this study, slope was the strongest factor for determining areas of high susceptibility. The best performing models in Goetz et al. (2011) show logged areas having higher levels of landslide susceptibility. Goetz et al. (2011)

also included distance to logging roads as a predictor variable. In comparison, areas in this current study adjacent to logging roads appeared to have higher levels of susceptibility, as indicated by areas with higher predicted probabilities (Figure 5.13).

The models used in this current study predicted debris slides as well as debris flows, which suggest why the gullies are shown as having a high susceptibility compared to the previous studies. Since it appears that gullies represent areas of higher susceptibility, the models created in this study may be favoring the prediction of debris flows more than debris slides. Therefore, it would be recommended to explore landslide susceptibility models individually by landslide type to determine if the differences in predictions are significant for this study area.

Generally speaking, a geomorphic process such as a landslide can have nonlinear relationships to environmental factors (Phillips, 2003; Goetz et al., 2011). The predominant selection of nonlinear forms of predictor variables in the landslide susceptibility models indicates the strong prevalence of nonlinear relationships between environmental factors and landslide initiation. In particular, the relationship to landslide initiation was selected as nonlinear for slope, elevation, plan curvature and distance-to-road in all model repetitions. Also, other variables, such as annual rainfall and profile curvature were for the majority selected as nonlinear. Only catchment area ($\log_{10}$) demonstrated a consistent linear relationship to landslides (Table 5.8). This predominant selection of predictor variables as nonlinear provides further empirical evidence for the presence of nonlinearities in the relationship of environmental factors to landslides (Goetz et al., 2011).These results strengthen the argument for the use of the GAM to capture complex geomorphic processes that are difficult to represent in a linear form. Especially, since the GAM can have a flexible variable selection process that allows for the selection of linear and nonlinear relationships. Overall, the present results suggest that the GAM cannot perform significantly worse than a GLM, which is consistent with similar finding of pervious geomorphological studies (Brenning, 2009; Goetz et al., 2011).

Nonlinearity may become more prevalent in variables when performing larger regional analysis. Goetz et al., (2011) completed a study of landslide susceptibility modeling in the Klanawa River watershed on the southwestern coast of Vancouver Island. Part of this study explored the ability to enhance susceptibility prediction using a GAM. It was determined that there was no statistically significant difference between using a GAM or a GLM; the GAM performed only marginally better than the GLM. In this (current) study, differing results are

found. The Vancouver Island results indicate that the GAM performs significantly better than the GLM. One of the key differences between this regional study of Vancouver Island and the study of the Klanawa River watershed is the scale of analysis. The total size of the Klanawa watershed is 610 km^2, which is a small fraction of the ~32000 km^2 area of Vancouver Island. With larger scale comes greater heterogeneity in values of environmental factors that affect landslides. For example, rainfall, vegetation cover/type, lithology, will differ across the island. However, to be sure of this relationship, it would be recommended to do a scale comparison using the same landslide inventory and environmental factors.

6.6 Limitations of statistical approach

Statistical models are now the most common method for landslide susceptibility modeling. However, these empirical methods may have a cloud of uncertainty in terms of process-related physical meaning. Models based on a stepwise-variable-selection process may further increase the difficultly to interpret or provide physical meaning to model results (Guzzetti et al., 1999). In general, statistical models only provide an implicit explanation of the physical processes involved (Carrara, 1993). Meaning that the variables may be highly correlated, which is typical of geomorphological studies, and that only a general prior decision is made on what variables may be important. The consequent model may lack physical meaning because the statistical model attempts to broaden the factors and generalize the group of landslide types (Guzzetti et al., 1999). Therefore, it is important to conduct separate analyse for known factors that influence a different spatial distribution of landslide events.

Chapter 7

Summary and conclusions

The purpose of this study was to explore natural and anthropogenic controls influencing landslide initiation at a regional scale for Vancouver Island. The relationship of environmental factors, such as topography, lithology, rainfall and land cover, to landslide initiation was explored using logistic regression and generalized additive models.

This study was the first of its kind to analyze the relationship of land cover and annual rainfall to landslide initiation at a large regional scale on Vancouver Island. Geostatistical techniques (OK, UK, and OCK) were used to interpolate mountain rainfall for temporal dates corresponding to the landslide inventory. Land cover was classified from Landsat TM images to represent different forest canopy densities, which were used as a proxy vegetation condition related to forestry activities.

A landslide susceptibility map based on data for land cover, logging roads, lithology and topography predicted the probability of landslide initiation to occur. In this map, the presences of landslide initiation points per square kilometer were 0.156 and 0.005 for high ($\geq$0.8) and low (<0.5) predicted probabilities. The most susceptible 4% of the study area predicted 33% of the landslide initiation points. Similarly, the least susceptible 73% of study the area predicted 20% of the landslide initiation points. Thus, there is a trade-off between having smaller areas classified as highly susceptible to landslide initiation, and the general predictive ability to detect more landslides.

Annual rainfall may not be necessary for building landslide susceptibility models in this study area. This research has shown that annual rainfall did not significantly improve model prediction performance.

The ability to model non-linear relationships for landslide susceptibility using a GAM provided significant improvements over the common method of using a GLM. This finding is

different from a comparison of another study conducted in a much smaller area on Vancouver Island, where there was no significant difference in predictive performance. It is hypothesized that non-linear relationships between predictor variables and landslide initiation are more apparent; however, a detailed analysis comparing GLM and GAMs with the same set of landslides at different regional scales should be conducted to confidently confirm this relationship.

Through exploring statistical interaction terms using GLMs, gainful insights into the relationships between rainfall, slope, land cover and lithology have been made. In terms of land cover, it has been empirically shown that logging areas (exposed ground) have higher landslide susceptibility at lower slope angle than other forest cover types. In addition, these logged areas are also more prone to landslides occurring with relatively lower annual rainfall amount than the other forest classes. For example, a closed forest canopy, representing old-growth forest, is more prone to landslides when annual rainfall amount is much higher

The successful modeling of the relationship between logging roads and landslide susceptibility opens a door to new methods for road planning on hillslopes in landslide prone areas. After an adequate model has been produced, such as the LTG-GAM, the inputs can be modified to form scenarios of landslide susceptibility. Thus, the location of a planned road can be drawn into to the road inventory and updated in the susceptibility model. The subsequent result will show predicted impacts of a planned road to increase the susceptibility of landslides.

A significant outcome of this analysis is the development of an adequate method for regional landslide susceptibility modeling for predicting areas of debris flow and debris slides on Vancouver Island. In particular, the methods used in this study may be able to assist forest development in British Columbia for areas of unstable terrain by providing maps of landslide susceptibility that predict smaller percentage of highly hazardous areas. The flexibility of the models presented in this research allow for application of these methods for other areas prone to landslide initiation. In particular, the open availability of national coverage of DEMs (for topographic controls), Landsat imagery (for land cover classification), road inventories and geologic information across Canada suggest that production of a national scale susceptibility map is highly feasible.

Appendix A

List of Acronyms

AUROC – Area under the receiver operating characteristic curve

CDED – Canadian Digital Elevation Data

DEM – Digital elevation model

GAM – Generalized additive model

GLM – Generalized linear model

GIS – geographical information system

LTG – Land cover and logging roads, topographic and geologic (lithology)

MLC – Maximum likelihood classifier

OCK – Ordinary co-kriging

OK – Ordinary kriging

UK – Universal kriging

RLTG – Rainfall, land cover and logging roads, topographic and geologic (lithology)

Appendix B

List of rock types associated to lithology classes

Table B.1. Summary of lithology on Vancouver Island

Lithology class	Rock Type
Metamorphic	
	Greenstone, greenschist
	Lower amphibolite/kyanite grade
	Othrogneiss
Sedimentary	
	Chert, siliceous agrillite, siliciclastic
	Coarse clastic
	Conglomoerate, coarse clastic
	Limestone bioherm/reef
	Limestone, marble, calcareous
	Limestone, slate, siltstone, argillite
	Mudstone, siltstone, shale fine clastic
	Undivided
Intrusive	
	Dioritic
	Diabase, basaltic
	Feldspar porphyritic
	Gabboic to dicritic
	Grandoioritic
	Quartz dioritic
	Undivided
Volcanic	
	Basaltic
	Bimodal
	Calc-alkaline
	Volcaniclastic
	Undivided

Appendix C

Geostatistical parameters

Table C.1. Summary of parameters for geostatistical models interpolating rainfall

Period	Model	Nugget	Sill	Range	Fit
Two weeks	OK	0	68472	74622	Spherical
	UK	0	40026	85519	Spherical
	OCK				
	Rain-elevation	-3305	44640	74623	Spherical
	Rain	2522	68472	74623	Spherical
	Elevation	4332	35581	74623	Spherical
Winter	OK	0	899635	97423	Spherical
	UK	0	710943	124084	Spherical
	OCK				
	Rain-elevation	-6917	141385	97423	Spherical
	Rain	5838	899634	97423	Spherical
	Elevation	8195	41770	97423	Spherical
Annual	OK	100086	604490	147430	Spherical
	UK	148866	1013751	337417	Spherical
	OCK				
	Rain-elevation	-17884	62274	147430	Spherical
	Rain	100086	604490	147430	Spherical
	Elevation	6455	27202	147430	Spherical

Appendix D

GLM model fit summary

D.1 RLTG-GLM

Table D.1. Summary of model fit for RLTG-GLM (null)

Predictor variables	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-7.421	0.785	-9.45	0.0000	***
Slope	0.093	0.008	12.33	0.0000	***
Profile curvature	-13.070	11.354	-1.15	0.2497	
Plan curvature	-21.308	10.278	-2.07	0.0382	*
Elevation	-0.001	0.000	-2.46	0.0137	*
Catchment area (log_{10})	1.071	0.157	6.81	0.0000	***
Distance-to-road	-0.011	0.002	-4.56	0.0000	***
Land cover-exp.	0.162	0.326	0.50	0.6187	
Land cover-open	0.720	0.188	3.83	0.0001	***
Land cover-urb./agr.	-2.123	0.916	-2.32	0.0205	*
Land cover-sem-open	0.531	0.179	2.97	0.0030	**
Annual rainfall	0.001	0.000	3.59	0.0003	***
Geology-metamorphic	0.330	0.403	0.82	0.4124	
Geology-sedimentary	0.303	0.263	1.15	0.2490	
Geology-volcanic	-0.081	0.163	-0.50	0.6201	

D.2 RLTG-GLM with rainfall and land cover interaction

Table D.2. Summary of model fit for rain: land cover GLM

Predictor variables	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-7.432	0.927	-8.02	0.0000	***
Slope	0.095	0.008	12.35	0.0000	***
Profile curvature	-13.520	11.398	-1.19	0.2355	
Plan curvature	-19.400	10.384	-1.87	0.0617	.
Elevation	-0.001	0.000	-2.35	0.0190	*
Catchment area (log_{10})	1.081	0.159	6.81	0.0000	***
Distance-to-road	-0.011	0.002	-4.57	0.0000	***
Land cover-exp.	6.181	1.927	3.21	0.0013	**
Land cover-open	-1.942	1.391	-1.40	0.1626	
Land cover-urb./agr.	1.727	4.057	0.43	0.6703	
Land cover-sem-open	0.881	1.104	0.80	0.4245	
Annual rainfall	0.001	0.000	2.49	0.0126	*
Geology-metamorphic	0.293	0.407	0.72	0.4720	
Geology-sedimentary	0.266	0.265	1.00	0.3159	
Geology-volcanic	-0.093	0.165	-0.57	0.5718	
Land cover-exp.: rainfall	-0.002	0.001	-3.12	0.0018	**
Land cover-open: rainfall	0.001	0.001	1.93	0.0532	.
Land cover-urb./agr.: rainfall	-0.002	0.002	-0.92	0.3559	
Land cover-sem-open: rainfall	0.000	0.000	-0.32	0.7492	

D.3 RLTG-GLM with slope and land cover interaction

Table D.3. Summary of model fit for slope: land cover GLM

Predictor variables	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-7.559	0.806	-9.38	0.0000	***
Slope	0.098	0.010	10.02	0.0000	***
Profile curvature	-13.600	11.243	-1.21	0.2264	
Plan curvature	-21.657	10.303	-2.10	0.0356	*
Elevation	-0.001	0.000	-2.45	0.0141	*
Catchment area (log_{10})	1.068	0.157	6.80	0.0000	***
Distance-to-road	-0.011	0.002	-4.56	0.0000	***
Land cover-exp.	1.966	0.723	2.72	0.0066	**
Land cover-open	0.577	0.522	1.11	0.2684	
Land cover-urb./agr.	-0.203	1.575	-0.13	0.8973	
Land cover-sem-open	0.606	0.459	1.32	0.1870	
Annual rainfall	0.001	0.000	3.58	0.0003	***
Geology-metamorphic	0.321	0.407	0.79	0.4299	
Geology-sedimentary	0.296	0.264	1.12	0.2628	
Geology-volcanic	-0.081	0.164	-0.49	0.6207	
Land cover-exp.: slope	-0.059	0.022	-2.73	0.0063	**
Land cover-open: slope	0.005	0.017	0.31	0.7581	
Land cover-urb./agr.: slope	-0.059	0.046	-1.28	0.2013	
Land cover-sem-open: slope	-0.002	0.017	-0.12	0.9071	

D.4 RLTG-GLM with rain and lithology interaction

Table D.4. Summary of model fit for rain: lithology GLM

Predictor variables	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-6.748	0.847	-7.97	0.0000	***
Slope	0.094	0.008	12.35	0.0000	***
Profile curvature	-12.595	11.346	-1.11	0.2670	
Plan curvature	-21.793	10.255	-2.13	0.0336	*
Elevation	-0.001	0.000	-2.33	0.0199	*
Catchment area (log_{10})	1.072	0.158	6.78	0.0000	***
Distance-to-road	-0.011	0.002	-4.45	0.0000	***
Land cover-exp.	0.137	0.324	0.42	0.6722	
Land cover-open	0.724	0.189	3.82	0.0001	***
Land cover-urb./agr.	-2.159	0.919	-2.35	0.0189	*
Land cover-sem-open	0.528	0.181	2.92	0.0035	*
Annual rainfall	0.000	0.000	1.73	0.0841	
Geology-metamorphic	3.678	2.217	1.66	0.0971	.
Geology-sedimentary	-0.800	1.394	-0.57	0.5658	.
Geology-volcanic	-4.424	1.507	-2.94	0.0033	**
Geology-metamorphic: rainfall	-0.001	0.001	-1.57	0.1172	
Geology-sedimentary: rainfall	0.000	0.001	0.78	0.4359	
Geology-volcanic: rainfall	0.002	0.001	2.91	0.0037	**

D.5 RLTG-GLM with slope and lithology interaction

Table D.5. Summary of model fit for slope: lithology GLM

Predictor variables	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-7.363	0.800	-9.21	0.0000	***
Slope	0.088	0.009	9.57	0.0000	***
Profile curvature	-13.630	11.402	-1.20	0.2319	
Plan curvature	-20.715	10.259	-2.02	0.0435	*
Elevation	-0.001	0.000	-2.39	0.0170	*
Catchment area (log_{10})	1.077	0.158	6.82	0.0000	***
Distance-to-road	-0.011	0.002	-4.55	0.0000	***
Land cover-exp.	0.169	0.326	0.52	0.6047	
Land cover-open	0.720	0.189	3.81	0.0001	***
Land cover-urb./agr.	-2.103	0.902	-2.33	0.0198	*
Land cover-sem-open	0.529	0.179	2.95	0.0031	**
Annual rainfall	0.001	0.000	3.69	0.0002	***
Geology-metamorphic	-1.321	1.176	-1.12	0.2612	
Geology-sedimentary	0.247	0.608	0.41	0.6852	
Geology-volcanic	-0.376	0.426	-0.88	0.3775	
Geology-metamorphic: slope	0.065	0.042	1.53	0.1272	
Geology-sedimentary: slope	0.001	0.025	0.05	0.9637	
Geology-volcanic: slope	0.011	0.014	0.74	0.4619	

Appendix E

Landslide susceptibility map example

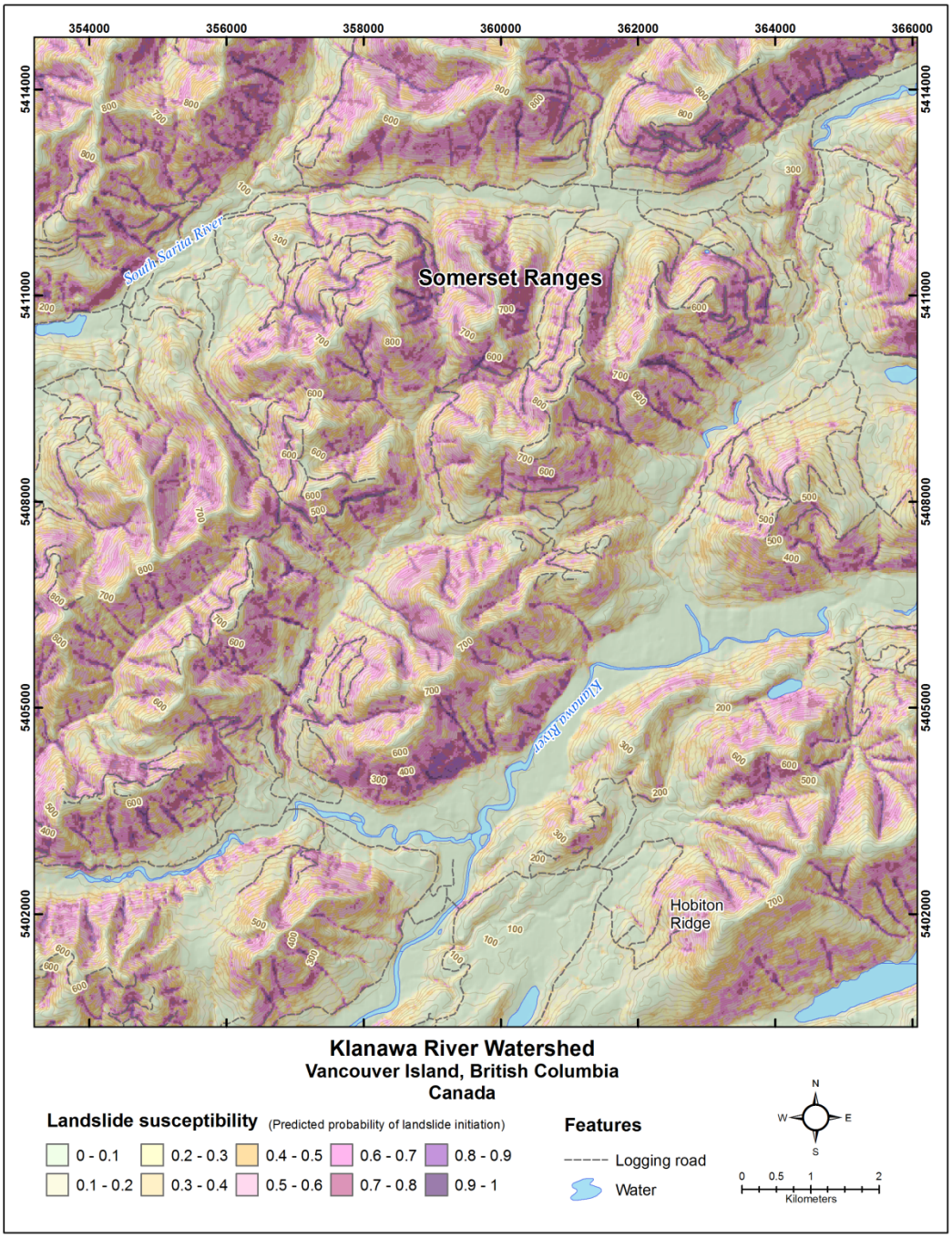

Figure E.0.1. Example landslide susceptibility map created using LTG-GAM for the Klanawa River watershed, Vancouver Island, British Columbia, Canada.

References

Ahmed, S., and G. De Marsily (1987), Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity, Water Resources Research, 23(9), 1717-1737.

Aleotti, P. (2004), A warning system for rainfall-induced shallow failures, Engineering Geology, 73(3-4), 247-265.

Arbia, G., R. Benedetti, and G. Espa (1999), Contextual classification in image analysis : an assessment of accuracy of ICM, Computational Statistics & Data Analysis, 30, 443-445.

Atkinson, P. M., and A. R. L. Tatnall (1997), Introduction Neural networks in remote sensing, International Journal of Remote Sensing, 18(4), 699-709.

BC Ministry of Forests (2002). Forest road engineering guidebook. For. Prac. Br., B.C. Min. For., Victoria, B.C. Forest Practices Code of British Columbia Guide book. http://www.for.gov.bc.ca/tasb/legsregs/fpc/fpcguide/guidetoc.htm.

BC Ministry of Transportation and Infrastructure (2011), Avalanche and weather programs, [Online] Accessed March 2011 from https://pub-apps.th.gov.bc.ca/saw-paws/weatherstation.

BC Wildfire Management Branch (2011), Climate, [Online] Accessed March 2011 from http://bcwildfire.ca/aboutus/organization/cariboo/climate.htm.

Barnett, P. J. (2004), Methods for Remote Engineering Geology Terrain Analysis in Boreal Forest Regions of Ontario, Canada, Environmental and Engineering Geoscience, 10(3), 229-241.

Barry, R. (2008), Mountain Weather and Climate, 3rd ed., Cambridge University Press, Cambridge.

Basist, A., G. D. Bell, and V. Meentemeyer (1994), Statistical Relationships between Topography and Precipitation Patterns, Journal of Climate, 7(9), 1305-1315.

Biggs, G. R. (2003), The oceans and climate, Cambridge University Press, Cambridge, U.K.

Blahut, J., C. J. van Westen, and S. Sterlacchini (2010), Analysis of landslide inventories for accurate prediction of debris-flow source areas, Geomorphology, 119(1-2), 36-51.

Brayshaw, D., and M. a. Hassan (2009), Debris flow initiation and sediment recharge in gullies, Geomorphology, 109(3-4), 122-131.

Brenning, a. (2005), Spatial prediction models for landslide hazards: review, comparison and evaluation, Natural Hazards and Earth System Science, 5(6), 853-862.

Brenning, a., M. Grasser, and D. a. Friend (2007), Statistical estimation and generalized additive modeling of rock glacier distribution in the San Juan Mountains, Colorado, United States, Journal of Geophysical Research, 112(F2), 1-10.

Brenning, A. (2008), Statistical geocomputing combining R and SAGA: the example of landslide susceptibility analysis with generalized additive models, in: Böhner J., Balschke T., and Montanarella L. (Eds.), Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie, 19, pp. 23-32.

Brenning, A. (2009), Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection, Remote Sensing of Environment, 113(1), 239-247.

Brenning, A., S. Long, and P. Fieguth (n.d.), Detecting rock glacier flow structures using Gabor filters and IKONOS imagery, Remote Sensing of Environment.

British Columbia Ministry of Forests (2012), Biogeoclimatic zones of British Columbia, [Online] Accessed March 2012 from http://www.for.gov.bc.ca/hre/becweb/resources/maps/map_download.html.

Brown, J. D. (2004), Knowledge, uncertainty and physical geography: towards the development of methodologies for questioning belief, Transactions of the Institute of British Geographers, 29(3), 367-381.

Carrara, A. (1993), Uncertainty in evaluating landslide hazard and risk, in: Nemec J., Nigg J.M., and Siccardi F. (Eds.), Prediction and perception of natural hazards, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 101-109.

Carrara, A. (2008), Comparing models of debris-flow susceptibility in the alpine environment, Geomorphology, 94(3-4), 353-378.

Carrara, A., G. Crosta, and P. Frattini (2003), Geomorphological and historical data in assessing landslide hazard, Earth Surface Processes and Landforms, 28(10), 1125-1142.

Chatwin, S. (2005). Managing landslide risk from forest practices in British Columbia. British Columbia For. Prac. Board Spec. Investig. Rep. 14, BC For. Prac. Board, Victoria, BC.

Chang, K.-T., and S.-H. Chiang (2009), An integrated model for predicting rainfall-induced landslides, Geomorphology, 105(3-4), 366-373.

Chung, C.-J., and A. G. Fabbri (1999), Probabilistic Prediction Models for Landslide Hazard Mapping, Photogrammetric Engineering & Remote Sensing, 65(12), 1389-1399.

Chung, C-J., Bobrowsky, P., and Guthrie, R.H. (2001), Quantitative prediction model for landslide hazard mapping, Tsitika and Schmidt Creek Watersheds, Northern Vancouver Island, British Columbia, Canada. In: Bobrowsky, P., 2001. Geoenvironmental mapping – Method, theory and practice. A.A. Balkema Publishers, Lisse, Netherlands, pp. 697-716.

Chung, C.-J., and A. G. Fabbri (2003), Validation of Spatial Prediction Models for Landslide Hazard Mapping, Natural Hazards, 30(3), 451-472.

Cihlar, J. (2000), Land cover mapping of large areas from satellites : Status and research priorities, International Journal of Remote Sensing, 21(6-7), 1093-1114.

Cingolani, A., D. Renison, M. R. Zak, and M. R. Cabido (2004), Mapping vegetation in a heterogeneous mountain rangeland using Landsat data: an alternative method to define and classify land-cover units, Remote Sensing of Environment, 92(1), 84-97.

Circa 2000 Land Cover for Agricultural Regions of Canada [computer file] (2009), Ottawa, Ont.: Agriculture and Agri-food Canada. Available from ftp://ftp.agr.gc.ca/pub/outgoing/aesb-eosgg/LCV_CA_AAFC_30M_2000_V12.

Cohen, W. B., M. Fiorella, J. Gray, E. Helmer, and K. Anderson (1998), An Efficient and Accurate Method for Mapping Forest Clearcuts in the Pacific Northwest Using Landsat Imagery, Photogrammetric Engineering & Remote Sensing, 64(4), 293-300.

Cohen, W. B., T. A. Spies, and M. Fiorella (1995), Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, U.S.A, International Journal of Remote Sensing, 16(4), 721-746.

Cohen, W. B., and S. N. Goward (2004), Landsat's Role in Ecological Applications of Remote Sensing, BioScience, 54(6), 535-545.

Cohen, W. B., and T. a. Spies (1992), Estimating structural attributes of Douglas-fir/western hemlock forest stands from Landsat and SPOT imagery, Remote Sensing of Environment, 41(1), 1-17.

Congalton, R. G., and K. Green (1999), Assessing the accuracy of remotely sesnsed data, CRC press, Boca Raton, Florida.

Conrad, O. (2006), SAGA — program structure and current state of implementation, in: Böhner J., McCloy K., and Strobl J. (Eds.), Göttinger Geographische Abhandlungen, 115, pp. 39-52.

Crawley, M. J. (2007), The R Book, John Wiley & Sons, London, UK.

Creutin, J., and C. Obled (1987), Objective analyses and mapping techniques for rainfall fields: an objective comparison, Water Resources Research, 18(2), 413-431.

Crist, E. P., R. Laurin, and R. C. Cicone (1986), Vegetation and soils information contained in transformed Thematic Mapper data, in: , IGRASS '86 Symposium, ESA Publ. Division, SP-254, Zürich, Switzerland, pp. 1465-1470.

Crosta, G. (1998), Regionalization of rainfall thresholds: an aid to landslide hazard evaluation, Environmental Geology, 35(2-3), 131-145.

Cruden, D. M. (1991), A simple definition of a landslide., Bulletin of the International Association of Engineering Geology, 43, 27-29.

Dai, F. C., C. F. Lee, and Y. Y. Ngai (2002), Landslide risk assessment and management: an overview, Engineering Geology, 64(1), 65-87.

Dai, F. C., and C. F. Lee (2002), Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong, Geomorphology, 42(3-4), 213-228.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, and P. P. Pasteris (2008), Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, International Journal of Climatology, 2064(March), 2031-2064.

Daly, C., R. P. Neilson, and D. L. Phillips (1994), A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain, Journal of Applied Meteorology, 33(2), 140-158.

Defries, R. S., and J. R. G. Townshend (1994), International Journal of Remote Sensing NDVI-derived land cover classifications at a global scale, International Journal of Remote Sensing, 15(17), 37-41.

Dhakal, A. S., and R. C. Sidle (2003), Long-term modelling of landslides for different forest management practices, Earth Surface Processes and Landforms, 28(8), 853-868.

Dikau, R., D. Brunsden, L. Schrott, and M. L. Ibsen (1996), Landslide recognition: identification, movement and causes, John Wiley & Sons, Toronto.

Dingman, S. L. (1994), Physical hydrology, Macmillan, New York.

Environment Canada (2011), National Climate Data Information Archive, [Online] Accessed March 2011 from http://climate.weatheroffice.gc.ca/climateData/canada_e.html.

Foody, G. M. (2002), Status of land cover classification accuracy assessment, Remote Sensing of Environment, 80(1), 185-201.

Foody, G. M. (2004), Thematic Map Comparison : Evaluating the Statistical Significance of Differences in Classification Accuracy, Photogrammetric Engineering & Remote Sensing, 70(5), 627-633.

Forest Practices Board (2009), Landslide occurrence following major rain storms on Vancouver Island, [Online] Accessed March 2012 from http://www.fpb.gov.bc.ca/WorkArea/DownloadAsset.aspx?id=5064.

Franco-Lopez, H., A. R. Ek, and M. E. Bauer (2001), Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method, Remote Sensing of Environment, 77(3), 251-274.

Frattini, P., G. Crosta, and A. Carrara (2010), Techniques for evaluating the performance of landslide susceptibility models, Engineering Geology, 111(1-4), 62-72.

Friedl, M. A., and C. E. Brodley (1997), Decision Tree Classification of Land Cover from Remotely Sensed Data, Remote Sensing of Environment, 61, 399-409.

Funk, C., and J. Michaelsen (2004), A Simplified Diagnostic Model of Orographic Rainfall for Enhancing Satellite-Based Rainfall Estimates in Data-Poor Regions, Journal of Applied Meteorology, 43, 1366-1378.

GeoBase (2011), GeoBase: A portal for no fee access to quality geospatial data, [Online] Accessed March 2011 from http://www.geobase.ca.

GeoConnections (2011), GeoConnections: Discovery portal, [Online] Accessed March 2011 from http://geodiscover.cgdi.ca/.

Giannecchini, R. (2006), Relationship between rainfall and shallow landslides in the southern Apuan Alps (Italy), Natural Hazards and Earth System Science, 6(3), 357-364.

Gjertsen, A. K. (2007), Accuracy of forest mapping based on Landsat TM data and a kNN-based method, Remote Sensing of Environment, 110, 420-430.

Glade, T. (2000), Applying Probability Determination to Refine Landslide-triggering Rainfall Thresholds Using an Empirical “Antecedent Daily Rainfall Model”, Pure and Applied Geophysics, 157(6-8), 1059-1079.

Goetz, J. N., R. H. Guthrie, and A. Brenning (2011), Integrating physical and empirical landslide susceptibility models using generalized additive models, Geomorphology, 129(3-4), 376-386.

Goovaerts, P. (1997), Geostatistics for Natural Resources Evaluation, Oxford University Press, New York.

Goovaerts, P. (1999), Using elevation to aid the geostatistical mapping of rainfall erosivity, Catena, 34(3-4), 227-242.

Goovaerts, P. (2000), Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, Journal of Hydrology, 228(1-2), 113-129.

Greenway, D. R. (1987), Vegetation and slope stability, in: Anderson M.G. and Richards K.S. (Eds.), Slope Stability, John Wiley & Sons, Chichester, UK, pp. 187-230.

Grimes, D. I. F., and E. Pardo-Igúzquiza (2010), Geostatistical Analysis of Rainfall, Geographical Analysis, 42(2), 136-160.

Gritzner, M. L., W. A. Marcus, R. Aspinall, and S. G. Custer (2001), Assessing landslide potential using GIS, soil wetness modeling and topographic attributes, Payette River, Idaho, Geomorphology, 37(1-2), 149-165.

Guan, H., J. L. Wilson, and O. Makhnin (2005), Geostatistical Mapping of Mountain Precipitation Incorporating Autosearched Effects of Terrain and Climatic Characteristics, Journal of Hydrometeorology, 6(6), 1018-1031.

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

Guthrie, R., and S. G. Evans (2004), Analysis of landslide frequencies and characteristics in a natural system, coastal British Columbia. Earth Surface Processes and Landforms, 29, 1321-1339.

Guthrie, R. (2005), Geomorphology of Vancouver Island: mass wasting potential. MOE Research Report , RR01 (includes maps). http://www.for.gov.bc.ca/hfd/library/documents/bib96153.pdf

Guthrie, R. H., S. J. Mitchell, N. Lanquaye-Opoku, and S. G. Evans (2010a), Extreme weather and landslide initiation in coastal British Columbia, Quarterly Journal of Engineering Geology and Hydrogeology, 43(4), 417-428.

Guthrie, R. H., S. J. Mitchell, and S. G. Evans (2010b), Extreme weather and landslide initiation in coastal British Columbia, Journal of Engineering Geology, (November), 417-428.

Guzzetti, F., A. Carrara, M. Cardinali, and P. Reichenbach (1999), Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy, Geomorphology, 31(1-4), 181-216.

Guzzetti, F., S. Peruccacci, M. Rossi, and C. P. Stark (2007), The rainfall intensity–duration control of shallow landslides and debris flows: an update, Landslides, 5(1), 3-17.

Guzzetti, F., P. Reichenbach, F. Ardizzone, M. Cardinali, and M. Galli (2006), Estimating the quality of landslide susceptibility models, Geomorphology, 81(1-2), 166-184.

Haapanen, R., A. R. Ek, M. E. Bauer, and A. O. Finley (2004), Delineation of forest/nonforest land use classes using nearest neighbor methods, Remote Sensing of Environment, 89, 265-271.

Haberlandt, U. (2007), Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event, Journal of Hydrology, 332(1-2), 144-157.

Hand, D. J. (1997), Construction and assessment of classification rules, John Wiley & Sons, Chichester.

Hastie, T. J., and R. J. Tibshirani (1990), Generalized Additive Models, Chapman & Hall, London, UK.

Hastie, T., and R. Tibshirani (1986), Generalized additive models, Statistical Science, 1(2), 297-318.

Hodgson, E. A. (1946), British Columbia Earthquake, The Royal Astronomical Society of Canada, 40(8), 285-319.

Howes, D. E., and E. Kenk (1997), Terrain classification system for British Columbia (Version 2). Ministry of Environment, Ministry of Crown Lands, Victoria, B.C., Canada.

Huang, J., S. T. Lacey, and P. J. Ryan (1996), Impact of forest harvesting on the hydraulic properties of subsurface soil, Soil Science, 161(2), 79-86.

Jakob, M. (2000), The impacts of logging on landslide activity at Clayoquot Sound, British Columbia, Catena, 38(4), 279-300.

Jensen, J. R. (2005), Introductory digital image processing: a remote sensing perspective, 3rd ed., Prentice-Hall, Upper Saddle River, NJ.

Jones, J. A. A. (1997), Global hydrology: processes, resources and environmental management, Longman, Essex, UK.

Kaldova, J., and C. Rosenfeld (1998), Geomorphological hazards in high mountain areas, Kluwer Academic Publishers, Netherlands.

Kidd, C. (2001), Satellite rainfall climatology: a review, International Journal of Climatology, 21(9), 1041-1066.

Krajewski, W. F. (1987), Cokriging radar-rainfall and rain gage data, Journal of Geophysical Research, 92(D8), 9571-9580.

Krajewski, W. F., and J. a. Smith (2002), Radar hydrology: rainfall estimation, Advances in Water Resources, 25(8-12), 1387-1394.

Krajina, V. J. (1969), Ecology of Western North America: Ecology of forest trees in British Columbia, Vancouver: Dept. of Botany, 2(1), 1-146.

Lee, S., and K. Min (2001), Statistical analysis of landslide susceptibility at Yongin, Korea, Environmental Geology, 40(9), 1095-1113.

Mathews, W. H. (1979), Landslides of central Vancouver Island and the 1946 earthquake, Bulletin of the Seismological Society of America, 69(2), 445-450.

Massey, N.W.D., MacIntyre, D.G., Desjardins, P.J. and Cooney, R.T. (2005) Digital Map of British Columbia: Whole Province, B.C. Ministry of Energy and Mines, GeoFile 2005-1.

McKenney, D., P. Papadopol, K. Campbell, K. Lawrence, and M. Hutchinson (2006), Spatial models of Canadian and North American-wide 1971/2000 minimum and maximum temperature, total precipitation and derived bioclimatic variables., Sault Ste. Marie, Ontario.

Meidinger, D., and J. Pojar (1991), Ecosystems of British Columbia, Crown Publications Inc., Victoria, B.C.

Meisina, C., and S. Scarabelli (2007), A comparative analysis of terrain stability models for predicting shallow landslides in colluvial soils, Geomorphology, 87(3), 207-223.

Merino, A., J. M. Edeso, M. J. González, and P. Marauri (1998), Soil properties in a hilly area following different harvesting management practices, Forest Ecology and Management, 103(2-3), 235-246.

Montgomery, D. R., K. M. Schmidt, H. M. Greenberg, and W. E. Dietrich (2000), Forest clearing and regional landsliding, Geology.

Montgomery, D. R., and W. E. Dietrich (1994), A physically based model for the topographic control on shallow landsliding, Water Resources Research, 30(4), 1153.

Muller, J. E. (1977), Geology of Vancouver Island: Guidebook, Vancouver, B.C.

Natural Resources Canada (2012), Mean total precipitation, The Atlas of Canada. [Online] Accesses March 2012 at http://atlas.nrcan.gc.ca/site/english/maps/environment/climate/precipitation/precip.

Pack, R. T., D. G. Tarboton, and C. N. Goodwin (1998), The SINMAP Approach to Terrain Stability Mapping, in: Moore D. and Hungr O. (Eds.), 8th IAEG Congress, Vancouver, A.A. Balkema, Rotterdam, pp. 1157-1165.

Pal, M., and P. M. Mather (2003), An assessment of the effectiveness of decision tree methods for land cover classification, Remote Sensing of Environment, 86(4), 554-565.

Park, N. W., and K. H. Chi (2008), Quantitative assessment of landslide susceptibility using high resolution remote sensing data and a generalized additive model, International Journal of Remote Sensing, 29(1), 247-264.

Phillips, J. D. (2003), Sources of nonlinearity and complexity in geomorphic systems, Progress in Physical Geography, 27(1), 1-23.

Pike, R. G., T. E. Redding, D. R. Moore, R. D. Winkler, and K. D. Bladon (2010), Compendium of forest hydrology and geomorphology in British Columbia, Kamloops. B.C. Min. For. Range, For. Sci. Prog., Victoria B.C. and FORREX Forum for Research and Extension in Natural Resources, Kamloops, B.C. Land Manag. Handb. 66. www.for.gov.bc.ca/hfd/pubs/Docs/Lmh/Lmh66.htm.

R Development Core Team (2011), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org.

Richards, J. A., and J. Xiuping (2006), Remote sensing digital image analysis, 4th ed., Springer, New York.

Rickli, C., and F. Graf (2009), Effects of forests on shallow landslides – case studies in Switzerland, Landscape, 44, 33-44.

Ritter, D. F., R. C. Kochel, and J. R. Miller (2002), Process geomorphology, fourth., McGraw-Hill, New York.

Rollerson, T. P. (1992), Relationships between landscape attributes and landslide frequencies after logging: Skidegate Plateau, Queen Charlotte Islands. B.C. Minist. For., Victoria, BC. Land Manage. Rep. 76.

Rollerson, T. P., C. Jones, K. Trainor, and B. Thomson. (1998), Linking post-logging landslides to terrain variables: Coast Mountains, British Columbia–preliminary analyses. Pages 1973–1979 in Proc. 8th Int. Congr., Int. Assoc. Eng. Geol. and Environ. Balkema, Rotterdam, The Netherlands. Issue 3.

Rollerson, T. P., T. Millard, and B. Thomson (2002), Using terrain attributes to predict post-logging landslide likelihood on southwestern Vancouver Island. B.C. Minist. For., Res. Sect., Vanc. For. Reg., Nanaimo, BC. For. Res. Tech. Rep. TR-015.

Rogan, J., J. Franklin, and D. A. Roberts (2002), A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery, Remote Sensing of Environment, 80, 143-156.

Rogers, G. C. (1980), A documentation of soil failure during the British Columbia earthquake of 23 June, 1946, Canadian Geotechnical Journal, 17, 122-127.

Rowe, W. D. (1994), Understanding Uncertainty, Risk Analysis, 14(5), 743-750.

Sader, S. a., and J. C. Winne (1992), RGB-NDVI colour composites for visualizing forest change dynamics, International Journal of Remote Sensing, 13(16), 3055-3067.

Sammori, T., Y. Okura, and Y. Horie (1993), Numerical experimental investigation of the effects of parameters on shallow landslides, Shin Sabo J. (Japan Society of Erosion Control Engineering), 46(1), 3-12.

Schmidt, K. M., J. J. Roering, J. D. Stock, W. E. Dietrich, D. R. Montgomery, and T. Schaub (2001), The variability of root cohesion as an influence on shallow landslide susceptibility in the Oregon Coast Range, Canadian Geotechnical Journal, 38(5), 995-1024.

Schwab, J. W. (1983), Mass wasting: October-November 1978 storm Rennell Sound, Queen Charlotte Islands, British Columbia.

Schwab, J. W. and M. Geertsema (2008). Terrain stability mapping on British Columbia forest lands: a historical perspective, In: J. Locat, D. Perret, D. Tunnel, D. Demers et S. Leroueil. Proceedings of the 4th Canadian Conference on Geohazards : From Causes to Management. Presse de l'Université Laval, Québec, 594 p.

Sharples, J. J., M. F. Hutchinson, and D. R. Jellett (2005), On the Horizontal Scale of Elevation Dependence of Australian Monthly Precipitation, Journal of Applied Meteorology, 44, 1850-1865.

Sidle, R. C. (1992), A theoretical model of the effects of timber harvesting on slope stability, Water Resources Research, 28(7), 1897-1910.

Sidle, R. C., and H. Ochiai (2006), Landslides: Processes, prediction, and land use, American Geophysical Union, Washington, D.C.

Sidle, R., a Ziegler, J. Negishi, a Nik, R. Siew, and F. Turkelboom (2006), Erosion processes in steep terrain—Truths, myths, and uncertainties related to forest management in Southeast Asia, Forest Ecology and Management, 224(1-2), 199-225.

Stefanov, W. L., M. S. Ramsey, and P. R. Christensen (2001), Monitoring urban land cover change : An expert system approach to land cover classification of semiarid to arid urban centers, Remote Sensing of Environment, 77, 173-185.

Sterlacchini, S., C. Ballabio, J. Blahut, M. Masetti, and a. Sorichetta (2011), Spatial agreement of predicted patterns in landslide susceptibility maps, Geomorphology, 125(1), 51-61.

Sterling, S., and O. Slaymaker (2007), Lithologic control of debris torrent occurrence, Geomorphology, 86, 307-319.

Swain, P. H. and S. M. Davis (1978), Remote sensing: the quantitative approach, McGraw-Hill, New York.

Swanston, D. N., and F. J. Swanson (1976), Timber harvesting, mass erosion, and steepland forest geomorphology in the pacific northwest, North.

Terlien, M. T. J. (1998), The determination of statistical and deterministic hydrological landslide-triggering thresholds, Environmental Geology, 35(2-3), 124-130.

Thompson, W. (1964), How and why to distinguish between mountain and hills, The Professional Geographer, 16, 6-8.

Tribe, S., and M. Lier (2004), The role of aerial photograph interpretation in natural hazard and risk assessment, in: , 2004 International Pipeline Conference (IPC2004), ASME, Calgary, Alberta, Canada, pp. 1-6.

USGS (2011), USGS Global Visualization Viewer, [Online] Accessed March 2011 from http://glovis.usgs.gov.

Varnes, D. J. (1978), Slope movement type and processes, in: Schuster R.L. and Krizek R.J. (Eds.), Landslides: Analysis and control, TRB, National Research Council, Washington, D.C., pp. 11-33.

VanDine, D. F., and S. G. Evans. (1992), Large landslides on Vancouver Island, British Columbia. Pages 193–201 in Proc. of Symp. on Geotechnique and Natural Hazards. Vanc. Geotech. Soc., Vancouver, BC.

van Westen, C. J., T. W. J. Asch, and R. Soeters (2005), Landslide hazard and risk zonation—why is it still so difficult?, Bulletin of Engineering Geology and the Environment, 65(2), 167-184.

van Westen, C., E. Castellanos, and S. Kuriakose (2008), Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview, Engineering Geology, 102(3-4), 112-131.

Wilson, E. H., and S. A. Sader (2002), Detection of forest harvest type using multiple dates of Landsat TM imagery, Remote Sensing of Environment, 80(3), 385-396.

Wood, S. N. (2006), Generalized additive models: an introduction with R, Chapman & Hall, New York.

Wu, T. H., and P. Mckinnell (1979), Strength of tree roots and landslides on Prince of Wales Island , Alaska, , 1966.

Zweig, M. H., and G. Campbell (1993), Reciever-operating characteristic (ROC) plots, Clinical Chemistry, 39, 561-577.