# MODELLING THE DISTRIBUTION OF ADVANCE REGENERATION IN LODGEPOLE PINE STANDS IN THE CENTRAL INTERIOR OF BRITISH COLUMBIA

by

**Darin Warren Brooks**

B.Sc. (Hons) University of Saskatchewan, 1996

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
NATURAL RESOURCES AND ENVIRONMENTAL STUDIES
(GEOGRAPHY)

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

December 2012

*Your file Votre référence*
*ISBN: 978-0-494-94154-6*

*Our file Notre référence*
*ISBN: 978-0-494-94154-6*

**Abstract**

The recent mountain-pine beetle outbreak in the Central Interior of British Columbia is leaving unsalvaged stands with minimal silvicultural treatment, raising questions about their ability to regenerate and the implications of this uncertainty to future timber supply and habitat values. No system currently exists to predict, on a landscape level, which pine stands will have adequate stocking of advance regeneration suitable for release upon canopy death. My research takes a ground-truthed, landscape-level approach to modelling, predicting, mapping, and prioritizing stands for salvage or rehabilitation. The resulting model, derived from recursive partitioning of data from 964 sample plots, created a landscape level output with a predictive accuracy of 78%. Across the Sub-Boreal Spruce study area, I estimate that 58% of mature pine-leading stands (approximately 840,000 ha) are likely or very likely to be stocked with at least 600 stems/ha of living understory trees.

**Table of Contents**

## List of Tables

**List of Figures**

**Acknowledgements**

I would like to thank the following people and agencies: Dr. Phil Burton, Dr. Roger Wheate, and Dr. Michael Gillingham (my committee members), who provided me with more latitude than I deserved. This thesis would not have been possible without their patience, guidance, and support. They have been incredible mentors … and consequently I am a better student, teacher, and mentor for having known them. Thank you from the bottom of my heart. I thank my wife, kids, and extended families for their patience as they watched me sacrifice holidays, weekends, family nights and evenings of dance for this thesis; always graciously and unselfishly supporting me in this dream. I can't imagine how I will begin to pay them back. To David Kim from KFM in Prince George, BC: thank you for seeing the importance of this thesis and for providing me with the time, material, equipment, and mentorship that was required. Thank you to everyone who provided funding (whether monetary or in kind): BC Forest Science Program, the Pacific Forestry Centre of the Canadian Forest Service, PFC graduate scholarship, Kim Forest Management Ltd. A special thanks to Heather and Talya for helping me with field research and for providing me with a couple of summers that I will never forget. I thank everyone who provided data to the aggregate dataset: Dave Coates, Craig DeLong, Patience Rakochy, Chris Hawkins, Brad Hawkes, Deb Cichowski, and Phil Burton. And finally, thanks to the College of the North Atlantic for the use of materials and time during the "final push".

## Chapter One: Introduction

British Columbia's (BC) lodgepole pine (*Pinus contorta* var. *latifolia*) forests have recently (~1999 to present) suffered from an epidemic of mountain pine beetles (*Dendroctonus ponderosae*), hereafter MPB. Although such infestations are largely a natural occurrence, the intensity and spatial extent of the present infestation are unprecedented, such that attempts to control the growing pest population have failed (Stockdale et al., 2004; Burton, 2010). In 2004, Pedersen (2004, p. 11) noted "despite the suppression measures, the epidemic as well as the amount of beetle-killed wood continues to increase". The Provincial Aerial Overview Surveys of Forest Health have indicated that the current epidemic reached its peak in 2005, and although mountain pine beetles continue to attack pine stands throughout the province, the outbreak will eventually subside by 2021 (Walton, 2012). In the meantime, because pest-management intervention was not sufficient to combat the current damage, research interest has largely shifted to guide timber salvage and regeneration operations. Social, economic, and operational limitations continue to hamper those attempts to salvage the infested area (Stockdale et al., 2004). Those constraints have contributed to predictions of a future severe timber shortage (Pedersen, 2004; Burton, 2010). Immature pine stands, previously presumed to be immune to pine beetle attack and provide harvestable timber within the next few decades, have become susceptible to pine beetle attack at the peak of the outbreak as the older stands most susceptible to beetle-induced mortality were overrun with beetles and their offspring become increasingly desperate for food (Maclauchlan, 2006). With most of the mature pine forest killed, and with many of the plantations established in the 1970s and 1980s equally compromised or unlikely to reach maturity by the time salvage operations are completed, it is expected that BC will soon be suffering from an unprecedented timber supply fall-down over the midterm of the next 10 to 50 years (Pousette and Hawkins, 2006; Snetsinger, 2011).

Advance regeneration consists of the seedlings, saplings or sprouts that have naturally established in a forest understory before any large-scale disturbance, and can be found naturally established under some mature lodgepole pine stands (Johnson et al., 2003; Burton, 2006). The survival and release of those understory trees has been identified as a valuable mechanism to help avoid timber shortages created by the MPB infestation (Burton, 2006; Coates, 2006; Greisbauer and Green, 2006; Pousette and Hawkins, 2006). The presence of advance regeneration decreases the need for the use of more aggressive intervention and recovery methods to regenerate forest stands by forest managers (Greene et al., 1999).

Forest planners are faced with the quandary of deciding which stands of trees are unlikely to recover and, therefore, should be harvested and planted to speed regeneration and which stands should be left to recover on their own. The lack of knowledge regarding the capacity and growth of advance regeneration to respond to a disturbance will be a major obstacle for the prediction of future harvests and the development of appropriate response efforts (Messier et al., 1999). The ability of advance regeneration to respond to a disturbance can be impacted by factors such as the composition of species within the affected stand, the particular characteristics (e.g., height, age, and diameter) of the trees involved, and the availability of light determined by canopy gaps (Mitchell, 2005).

This potentially high variability among stands has limited advance regeneration modelling using individual tree and stand models such as SORTIE (Hawkins et al., 2012). It is this variability within and among stands that also present a challenge to SORTIE's modelling. SORTIE develops an output that is derived from complex interactions between light, growth, seed dispersal, and mortality—all ecological factors that have a high variability between stands (Sattler, 2009). Coates (2006) suggests that the variability among stands is so great, that stands can only be managed on an individual basis. A landscape-level model for predicting the distribution of advance regeneration is critical for efficient and effective forest planning, particularly in the context of the province's management unit objectives to ensure harvested areas in BC are adequately renewed through stocking standards (McWilliams, 2009).

Despite the widespread occurrence of understory regeneration in lodgepole pine forests, the question remains as to whether those forests will remain adequately stocked upon the death of the overstory. Stocking can be broadly defined through current tree metrics (typically volume based) and early stand conditions that are used to measure the probability of achieving long-term management goals for that stand (Martin et al., 2005; McWilliams, 2009). A fully stocked stand or sample plot is one in which all open space is (or is projected to be) occupied by living trees. To ensure that BC forests continue to return to their pre-harvested conditions, stocking standards have been established to maximize the probability of successful regeneration. It is, therefore, both critical and prudent, for forest planners to assess stand regeneration at a landscape level. The stocking required to regenerate stands, however, is not the same across the landscape (i.e., stands within different biogeoclimatic unit and site series require different stocking standards to achieve renewal). And in some cases harvested areas do not have timber production as their primary objective (set by government and industry planners), and, therefore, have different stocking standards to achieve their particular management objectives.

Dhar and Hawkins (2011) identified three critical research-driven purposes for advance regeneration assessment: forecasting long-term development (yield) of attacked stands, selecting stands for further research, and forecasting impacts on ecological attributes. The data to support estimates of stocking in stands following

MPB are available, but are addressed in several different research projects. There is a need to collapse those data sets into a single usable format (perhaps even differentiating between predictive and operational) that can be used by industry, research entities, and a variety of other stakeholders (Wilford, 2008; Dhar and Hawkins, 2011).

The need to understand the patterns of advance regeneration occurrence has an operational focus. Current reforestation stocking standards – designed primarily to guide the reforestation of clearcuts – may need to be revisited by forest planners. Designation of preferred and acceptable species, as well as their densities, may need to be altered or amended in order to economically utilize post-MPB advance regeneration (Lewis, 2005; Greisbauer and Green, 2006). Commercial logging strategies may also require adjustment on a stand-by-stand basis to provide protection to advance regeneration that may or may not contribute to the midterm timber supply. Forest managers require a better understanding of the cost implications associated with the retention of advance regeneration versus 'starting over' with planting after logging in terms of facilitating important midterm forest harvesting opportunities (Dhar and Hawkins, 2011). Consideration must also be provided to the potential effects of climate change on advance regeneration, and its ability to sequester carbon and thereby mitigate some climate change impacts (Brown et al., 2012). A more complete assessment of the distribution of advance regeneration is required to identify the potential for future challenges such as vulnerability to pests and diseases. As forests continue to respond to climate change, so too do the populations and distributions of numerous insects and fungal pathogens that may pose a threat to the regenerating species (Lewis, 2005).

**Factors Impacting Advance Regeneration and Stand Recovery**

Research on the recovery of stands following a disturbance through advance regeneration is limited because the vast majority of studies have chosen to focus only upon their early development (Messier et al., 1999). Patterns of understory release in forests disrupted by natural disturbance, and the long-term growth, yield and habitat value of such forests are unfortunately poorly documented.

Natural disturbances are important ecological processes that drive observed patterns in ecosystems. The study of disturbance regimes and the interaction of disturbance agents has only recently become a central theme in ecology (Mori, 2011). Aspects of disturbance ecology that require further inquiry include disturbance history, spatiotemporal dynamics, disturbance interactions, regeneration response to disturbance, and the application of resistance and resilience theory to ecosystem management (Wright et al., 2000).

Ecological disturbances are defined as disruptive changes to an ecological system by an external event that makes changes to the resources within the system, but does not necessarily result in the destruction of the ecological system itself (White and Pickett, 1985; Pickett et al., 1989; Johnson and Miyanishi, 2007; Hughes, 2010). Studies indicate that disturbance impacts are defined by more than just the affecting agent (i.e., whether it is insect, fire, wind, or other agents). Pickett et al. (1986) state that the bases for understanding of an ecological disturbance are threefold: 1) identifying the existing ecological system that may be affected by a disturbance; 2) discerning only the changes to the ecological system that are a result of a disturbance; and 3) understanding the consequences of the disturbance.

How effective will advance regeneration be in supporting stand recovery after disturbance? Stand recovery rates vary through advance regeneration. For example, the re-establishment of MPB attacked stands through regeneration can be delayed by five to ten years due to species composition before the beetle attack and the vigour of overstory trees (Bouchard et al., 2005; Coates, 2006). A more complex example of stand recovery after disturbance is the creation of a thinning effect in balsam fir (*Abies balsamea*) stands by spruce budworm (*Choristoneura fumiferana*) in eastern Canada. The natural regeneration is released post-budworm attack, only to be attacked itself 30 or 40 years later – perpetuating a cycle of favourable conditions for subsequent spruce budworm attacks (MacLean and Anderson, 2008). The MPB outbreak in BC is having a profound impact on the makeup of the affected stands. MPB is not unique as a disturbance agent that just kills the overstory, typically serving as a means of releasing the understory regeneration and vegetation; this is true of wind storms, other insect outbreaks, and root rot pockets, etc. (Gautreaux, 1999). Pine-dominated forests have varying amounts of live trees, saplings and seedlings remaining after they have been attacked by MPB, which are collectively referred to as "secondary stand structure" or "secondary structure" (Coates et al., 2006).

Whereas forest fire will kill most seedlings and saplings, windthrow and insect infestations typically kill the overstory trees, but not the regeneration, thereby facilitating its release (Johnson et al., 2003; Roberts, 2004; Burton, 2008b). Natural methods of forest regeneration can be particularly attractive to forest planners when stocking is reliable, or economic and operational factors prohibit the large-scale use of artificial regeneration strategies. Natural regeneration -- whether by seed or through the release of existing seedlings -- offers several advantages to alternative approaches. For example, natural regeneration is cost-effective when compared to more interventionist reforestation strategies and helps to protect the forest's natural diversity (Weetman and Vyse, 1990).

Forest understory species may thrive following bark-beetle attack and this will cause a dominant species shift within the stand. The dwarf shrubs kinnikinnick (*Arctostaphylos uva-ursi*) and twinflower (*Linnaea borealis*) have been documented to increase in cover following MPB-induced canopy opening (Williston et al., 2006). Other studies indicate that plant species richness (particularly grasses) is measurably higher in post-MPB attacked stands and non-tree vegetation can represent one to two-thirds of the $CO_2$ uptake contribution in these stands (Stone and Wolfe, 1996; Bowler et al., 2012). Therefore, the beetle outbreak may be regarded as a stand-releasing event that causes the understory vegetation to assume a more dominant position within the stand as it grows to take the place of the lost canopy trees (Greene et al., 1999; Burton, 2008a; Lindenmayer et al., 2008). Young trees and other vegetation surviving in the understory have a strong potential to thrive because of the new availability of resources created by the lost stand members (Coates et al., 1994). Stands with a healthy understory may recover from MPB attack to yield harvestable timber within a timeframe of 40 to 80 years (Coates and Hall, 2005; Coates et al., 2006).

Understory light availability increases in natural canopy gaps (resulting from the mortality of one or more mature trees) and after forest harvesting (Palik et al., 1997; Burgess and Wetzel, 2000; Oguchi et al., 2006; Boucher et al., 2007), and larger gaps provide more light than smaller gaps (McGuire et al., 2001; Gray et al., 2002; Palik et al., 2003). Large canopy gaps due to dead trees can result in reduced evapotranspiration and, therefore, decrease summer groundwater depletion (Hélie et al., 2005). Light availability is also greater near gap edges than in forest interior positions (Matlack, 1993; Heithecker and Halpern, 2007). High light availability is associated with greater light-saturated photosynthetic rates (Ellsworth and Reich, 1992; Bond et al., 1999) so any death or removal of overstory trees should promote increased growth of advance regeneration (Williams et al., 1999). The new availability of increased light that results from the loss of shading crowns leads to the stimulation of germination and seedling release in many shade-tolerant trees such as *Abies* spp. (McCarthy, 2001). The adaptability of these shade-tolerant species is responsible for their establishment in the understory and further supports their recruitment into the canopy of the stand (Messier et al., 1999). As a result, some tree species are more likely than others to be found within the advance regeneration stratum. For example, shade tolerant species such as the subalpine fir and interior white spruce (a natural hybrid of *Picea engelmannii* and *P. glauca,* common throughout the BC Central Interior) are species commonly found among the advance regeneration found in BC's sub-boreal forests (Coates et al., 1994; Kneeshaw and Burton, 1997). The distribution of other plant species within a region can also impact the success of advance regeneration strategies. For example, the presence of aggressive non-tree vegetation can negatively impact the ability of advance regeneration to respond to overstory mortality (Bassman et al., 1992; Stone and Wolfe, 1996). Knowledge of sapling and seedling growth is also an essential element for the formulation of accurate predictions of future stand conditions (Wright et al., 2000). The species composition within the understory itself may also have an influence on the density of advance regeneration. Advance regeneration is generally more abundant when accompanied by a greater diversity of tree species within the overstory (Arnup, 1996).

Particular characteristics allow some species to respond more positively than others after an insect outbreak. For example, the relative availability of shade and sunlight has been highlighted as particularly influential for advance regeneration. Wright et al. (2000, p. 1528) found "a clear relationship between shade tolerance and the magnitude of the effects of past periods of suppression and release on sapling growth." Species able to tolerate conditions of low light have a propensity to thrive, and possibly adapt to variable light regimes due to changing canopy structure within the understory through advance regeneration (Oliver and Larson, 1996; Messier et al., 1999; McCarthy, 2001). Understory light may remain relatively unchanged for up to five years after a MPB attack because it takes that long for the dead foliage to fall and the residual canopy will continue to shade the vegetation in the understory (Coates and Hall, 2005). Furthermore, pine snags are another enduring source of shade for plants within the understory (Coates and Hall, 2005). The shade from these snags and any surviving trees can help to prevent understory trees from being out-competed by other vegetation (Lieffers and Stadt, 1994).

Although some lodgepole pine seedlings can establish under full canopies, it is generally classified as a shade-intolerant species (Burns and Honkala, 1990; Klinka et al., 2000). Shade-intolerant species tend to be less likely to respond well within the cohort of advance regeneration following a disturbance because they have a more fixed, less adaptable structure and physiology (Messier et al., 1999). Its status as a shade-intolerant species, however, does not preclude lodgepole pine from being found as a natural component of the secondary stand structure surviving an MPB outbreak. The presence of lodgepole pine regeneration within a mature forest can be explained by the presence of uneven canopy closure, which permits sunlight to reach the light-demanding young pines (McCarthy, 2001). Indeed, in cold dry environments where other tree species are at a competitive disadvantage, lodgepole pine may be the dominant species of advance regeneration, and can release to create an uneven-aged stand after MPB attack (Axelson et al. 2009).

Finally, there is evidence that natural regeneration may be assisted through seed rain recruitment (the deposition of seeds spread by bird, wind, humans, and animals) by some species (Moles and Drake, 1999), although Burton (2006) found inconsistencies in the relationship between regeneration densities and proximity to non-pine conifer seed sources. Leadem et al. (1997) state that successful regeneration of stands relies on seed production and dispersal, so there must be some fundamental dependency on the availability of seed from shade-tolerant species, either within the stand or from nearby on the landscape.

## Forest Management Options

Standard industrial forestry in the form of even-aged management and clearcut harvesting promotes the development of simplified stand structures through the application of homogeneous treatments across stands at fixed intervals that are often shorter than the return intervals of natural disturbances in the same region (Coates and Burton, 1997; Palik et al., 2002; Seymour et al., 2002). The loss of structural complexity in managed forests presents concerns for conserving biodiversity, sustaining key ecosystem functions, and maintaining ecological resilience in the face of a changing climate (Franklin et al., 2002; Lindenmayer and Franklin, 2002; Palik et al., 2002; Tews and Jeltsch, 2004; Drever et al., 2006). These concerns over the ecological consequences of simplified forest structures have created interest in developing novel silvicultural systems that promote more natural patterns of stand development by emulating the frequency, scale, and severity of natural disturbances (Coates and Burton, 1997; Franklin et al., 1997, 2002; Palik et al., 2002; Seymour et al., 2002).

## Empirical Models of Species Distribution and Abundance

Ecologists have spent a great deal of time examining the interaction between plant species and their micro- and macro-environments. Through the collection and examination of environmental data, ecologists attempt to discover patterns that assist them in predicting the presence or absence of a particular species, community, or ecosystem (e.g., Franklin, 2009). Unfortunately, the environmental variables used in predictive models are complex and covarying. Their interaction with each other can vary by simply changing where they exist in time or space. For example, even if we understand how variables interact with each other, our understanding can change when the variables are examined at a slightly higher elevation or on a warmer day. This ties a model to a particular geographical context, as the derived model may encounter accuracy errors when the geography changes, even if the combination of variables remains the same (Guisan et al., 1999). Furthermore, environmental variables can be significantly affected by unknown additional lurking variables, (such as anthropogenic interferences) or a sophisticated interaction of multiple variables. That is, our understanding of the variable interaction between $variable_x + variable_y + variable_z$ can change if an additional unknown and unseen $variable_u$ is present. These limitations are especially problematic for ecology, where variability and deviations occur in nature as the norm, not the exception.

General linear models have traditionally been used by ecologists to describe the relationship between causal factors and an observed response (Draper and Smith, 1981; Burnham and Anderson, 2002). Yet despite the well documented drawbacks of relying on techniques such as stepwise multiple regression, the use of these procedures for ecological studies are prevalent (Stephens et al., 2005). Many ecologists sacrifice the potential of making erroneous conclusions in the search for the most parsimonious model. This is important in datasets that have factor interactions too complex for parametric models (Freedman, 1983; Derksen and Keselman, 1992). Recent studies have shown that classification trees, or recursive partitioning, can be more reliable and accurate than traditional parametric linear models (Friedl et al., 1999; Hansen et al., 2000).

**Recursive Partitioning and Classification Trees**

Classification and regression trees are the statistical application of binary trees first introduced by Breiman et al. (1984). Classification trees are an intuitive and easily interpreted type of supervised learning method used in exploring relationships in data. Further, when the explored relationships in the data take the form of a decision tree and are then applied to new data to predict new values, the classification tree becomes a predictive tool (Han et al., 2011). Classification trees are equipped to deal with continuous and categorical data, missing values, and outliers (Moisen, 2008). But unlike linear regression models, which identify and measure the relationship between the response and explanatory variables, classification trees divide the explanatory variables into homogenous groups through a series of recursive partitions. Botanists employ classification-style decision trees in the form of the routinely used dichotomous keys for the correct identification of a plant by following a tree of if-then statements. Classification trees are also prevalent in the fields of medicine and psychology for diagnosis and decision making (Ripley, 1996).

In a standard classification tree, the idea is to split the dataset based on homogeneity of data. The goal is to achieve pure homogeneous groupings of data to 1) describe the systematic structure of the data; and 2) predict unobserved data (De'ath and Fabricius, 2000). For example, consider two variables, tree age and tree vigour, that predict whether a tree is likely to be attacked by mountain pine beetle (1), or not (0). If 90% of the trees that are >80 years old in our training data showed signs of beetle attack, we can split the data here and age becomes a top node in the tree. Further, if it is discovered that any tree with a vigour less than five (ten being the healthiest and one being the unhealthiest), is attacked 80% of the time, then vigour <5 or ≥5 would be the second branch in the tree. Graphically it would look like the sequence of decisions portrayed in Figure 1.

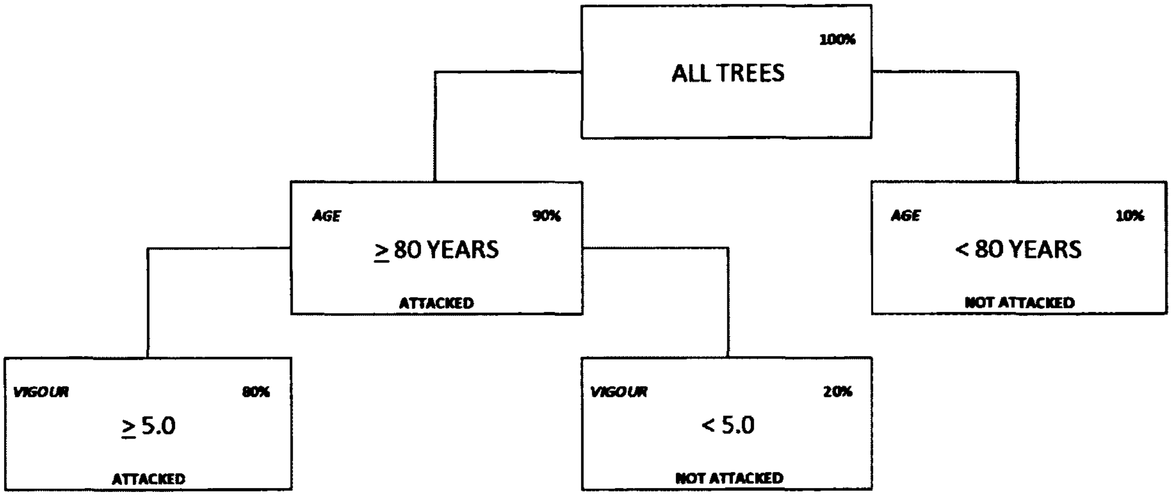



Figure 1. Example of a classification tree illustrating predictor variable splits. Note that the first split at age >80 years old is an internal node, i.e., other nodes can branch from this node. The next split at vigour >5.0 is a terminal node, signalling the end of the branch.

Classification trees are a useful tool for analyzing data because of their visual simplicity. The trees are rules for predicting or explaining the response category using hierarchical binary splits of the explanatory variables. When predicting the category of response, classification trees are used as an algorithm to classify new data. An observation will follow a path in the tree starting in the top or root node and follow its individual splits at the interior or branch nodes, down the path until it reaches a terminal or leaf node where no more splitting occurs (Kim and Yates, 2003). The criteria used to split the branches to achieve the nodes represent the "if-then" model. Classification trees are used to predict membership of cases or objects in the classes of a categorical dependent variable from their measurements on one or more predictor variables. Classification tree analysis is one of the main techniques used in data mining (Hastie et al., 2001).

Each predictor variable is examined to see how well it can divide the node into two groups. If the predictor is continuous, a trial split is made between each category (every unique value is a 'category') of the variable. The process is repeated by moving the split point across all possible division points in the training data until the best improvement is found. This split point is saved as the best possible split for that predictor variable in this node. The process is then repeated for each of the other predictor variables.

A well-recognized advantage of the decision tree representation of a model is that the paths through the decision tree can be interpreted as a collection of rules. The information associated with the textual rules include a node number for reference, a decision of 0 or 1 to indicate (in the application considered here) whether the plot is stocked or not stocked, the number of training observations and the strength or confidence of the decision.

The measurement of predictive accuracy of a variable (i.e., mean decrease in accuracy) is the more meaningful importance indicator (Berk, 2005). The mean decrease in accuracy is defined as the normalized difference between classification accuracy and the accuracy when the variable values have been randomly permuted.

Higher mean decrease in accuracy indicates that a variable is more important to the accuracy of the classification.

The Receive Operating Characteristic (ROC) curve, a diagnostic measure (represented by a graphic curve and a numerical score) that plots false positive against true positive rates, is generally described as one of the most accurate ways to measure the discriminatory power of the resultant classification tree model through comparison of the relationship between sensitivity (in this case, true positives) and specificity (false positives) (Hanley and McNeil, 1982; Beck and Schultz, 1986; Krivec and Matjaz, 2011). In other words, specificity of the tree is the probability that one more not stocked point added to the analysis will be correctly classified as not stocked; and conversely, sensitivity is the probability that one more stocked point added to the analysis will be correctly classified as stocked (Feldman and Gross, 2003). A perfect test (100% sensitive and 100% specific) would score a ROC value of 1.0. The closer the value is to 1.0 (100%), the better the distinguishing capability of the classification tree model. The ROC graphic is a curved line that extends from the 45 degree line to the upper left corner; higher model accuracy is associated with a sharp distinct curve that approaches the top left corner of the graph (Figure 2). The ROC curve is essentially the measurement of the trade-off between sensitivity

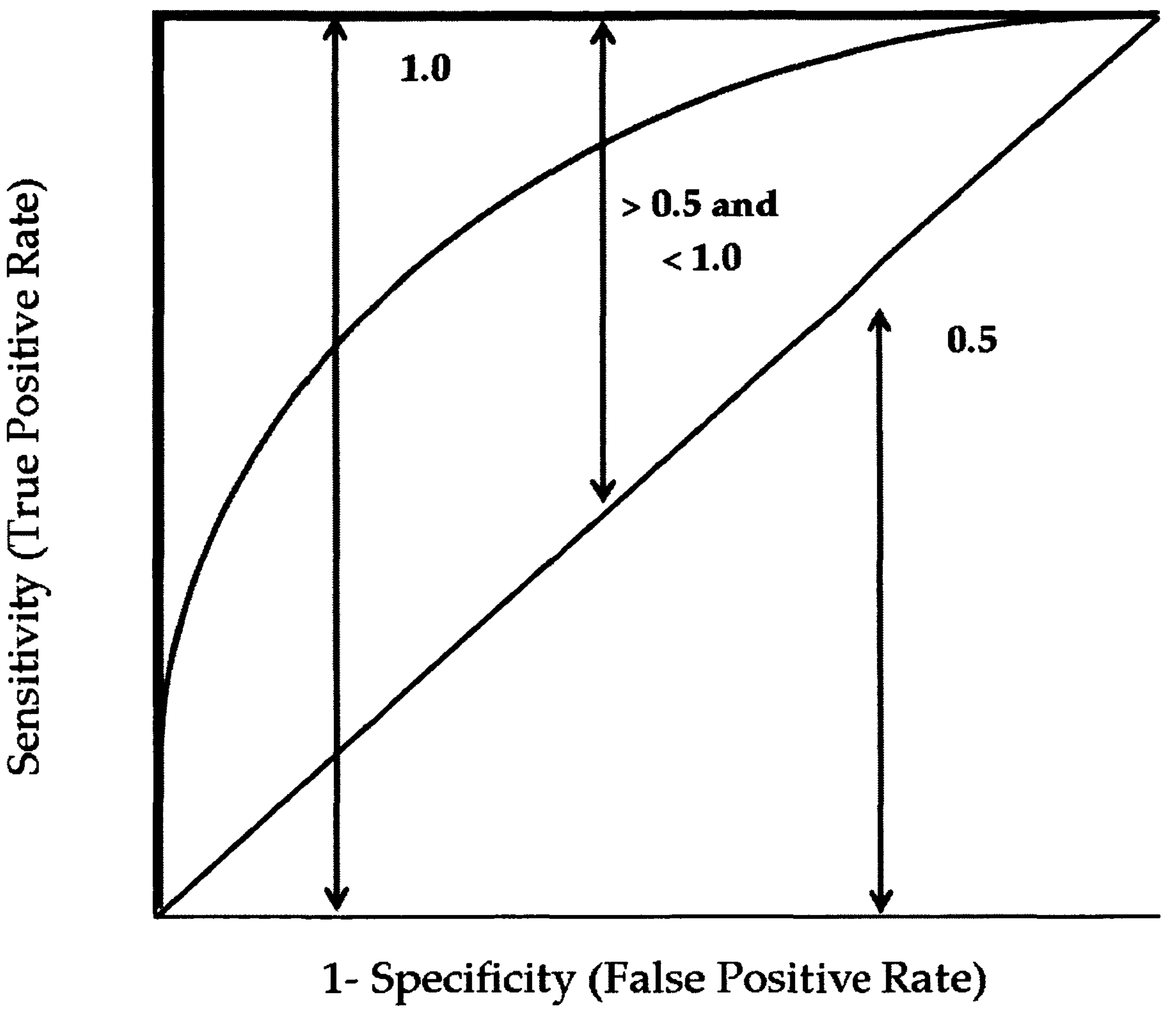



Figure 2. Example of a ROC curve illustrating the trade-off between sensitivity and specificity. A ROC score of 1.0 is a perfect model fit and 0.5 is a purely random model fit.

and specificity. A perfect model would be able to correctly classify all stocked stands as stocked and it would never incorrectly classify a stocked stand as not stocked. A perfect model (represented by the red line on the ROC curve in Figure 2) would follow the y axis to the top of the graph (zero false positives, i.e., zero incorrectly identified) and then follow the x axis along the top of the graph (all true positives, i.e., all correctly identified). A random model (represented by the 45 degree black line in Figure 2) would indicate that every true positive has an equally likely false positive. A ROC value of 0.5 (50% sensitive and 50% specific) demonstrates no discriminative value and is equivalent to assigning true positives with a coin flip (Sherrod, 2006). The strength of the final classified model can be interpreted by assessing the ROC value, also known as AUC (area under the curve).

**Objectives**

The objectives of my thesis are to : 1) devise a predictive model of forest understory stocking, based on publicly available digital data with long shelf-life, that provides forest managers with a decision model to assist in post-beetle stand management planning; and 2) apply the model in a Geographic Information System (GIS) to generate a series of colour-themed maps portraying the probability of understory stocking in stands dominated by mature lodgepole pine in Sub-Boreal Spruce biogeoclimatic zone.

Many factors such as proximity to potential non-pine seed sources, biogeoclimatic subzone, mean annual precipitation, crown closure, and site productivity may impact advance regeneration and stand recovery in the lodgepole pine forests of central British Columbia. Literature suggests that patterns in the abundance of advance regeneration are evident, such that the proper use of stand variables should make a predictive model possible (Kneeshaw and Burton, 1997;

Kayes and Tinker, 2012). Given the far-reaching geographic spread of MPB-affected stands (and the inability of industry to harvest these stands), and the fact that there seem to be some consistent trends in the distribution of advance regeneration, it seems likely that a relatively accurate landscape-level predictive model could be useful (Hawkes et al., 2003). My thesis explores the use of a statistical data-mining technique known as recursive partitioning, or Classification and Regression Tree Analysis, and its integration with GIS geospatial modelling. The results illustrate how the combination of *a priori* information and recursive partitioning can provide an accurate, stable and reliable predictive model of the likely distribution of advance regeneration.

## Chapter Two: Methods

### Study Area

The study area was restricted to the dk, dw2, dw3, mc2, mc3, mk1, mw, and wk1 biogeoclimatic ecosystem classification (BEC) units of the Sub-Boreal Spruce (SBS) biogeoclimatic zone (Klinka et al, 2000) located in central BC (Figure 3). I selected this study area because of the abundance of predominantly even-aged mature pine stands. As a consequence of its composition, the study area selected was one of the hardest hit MPB-attacked regions in BC (Parkins and MacKendrick, 2007), making it a critical area in which to consider the potential for stand recovery. In order to report the results within a forest industry recognized context, the data were clipped to 1:250,000 NTS map tiles 93F, 93G, 93J, 93K, and 93L. The 1:250,000 NTS map tile dataset also allows for logical, efficient data storage and organized dissemination of the resultant data.

The SBS occurs mostly in the central region of BC and dominates the gently rolling Nechako Plateau between the Coast Range and the Cariboo and Rocky Mountains (Meidinger et al. 1991). It ranges from valley bottoms to 1300 m elevation, receives 400-900 mm of precipitation annually, and has a mean temperature range of below 0°C in winter to above 10°C in the summer (Meidinger et al. 1991). The SBS zone is characterized by ten subzones and four common site associations: from driest to wettest, these are referred to as the Lodgepole pine –

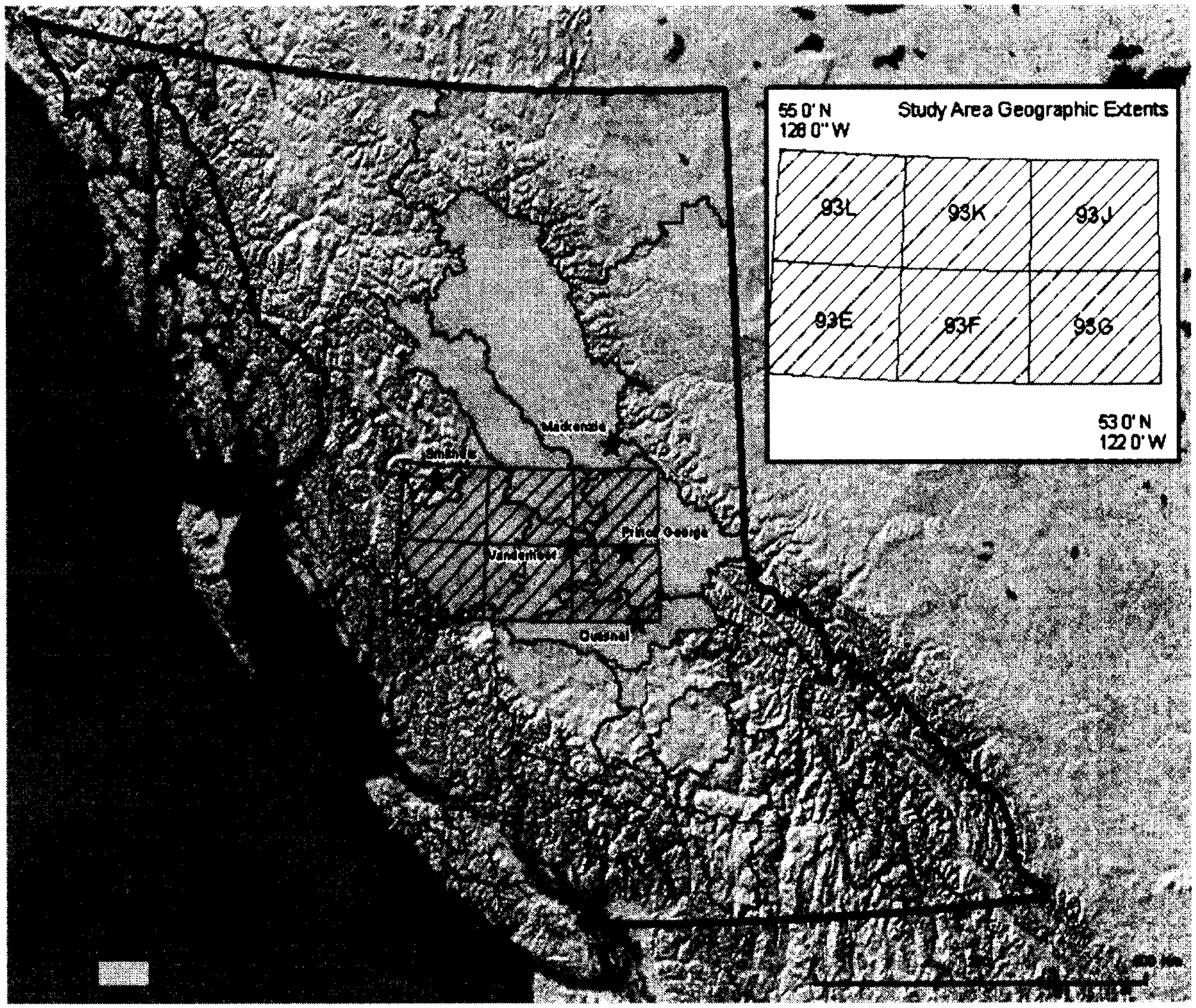



Figure 3. Study area as defined by six National Topographic Survey (NTS) 1:250,000 map sheets (cross-hatched). Grey or green lines denote forest district boundaries.

Huckleberry—Cladonia plant community, the Hybrid spruce – Huckleberry – Highbush-cranberry plant community, the Hybrid spruce – Oak fern plant community, and the Hybrid spruce – Devil's club plant community (Meidinger et al. 1991).

**Field Data Collection**

I selected my sampled stands through a stratified sampling design, using biogeoclimatic subzone as strata (Figure 4) because previous research had indicated strong differences among subzones in the abundance of advance regeneration (Burton, 2006; Coates et al. 2006). Suitable stands, which I referred to as target stands, met the following attribute criteria: stand species composition is lodgepole pine (coded as *PL* or *PLI* in the Province of BC's Vegetation Resource Inventory, VRI) as the leading species with a minimum of 50 percent basal area (as specified by VRI attribute: *SPEC_CD_1* and *SPEC_PCT_1*); the stand has an average age of greater than 60 years, weighted by basal area of the dominant trees for the leading species (specified by VRI attribute: *AGE_1*); and the stand is adjacent to a mature stand dominated by conifer species other than lodgepole pine.

My field sampling was designed to test and calibrate distance-related factors affecting the density of advance regeneration in mature stands dominated by lodgepole pine. In particular, the effects of proximity to potential seed sources from mapped stands or landscape positions dominated by non-pine trees guided the

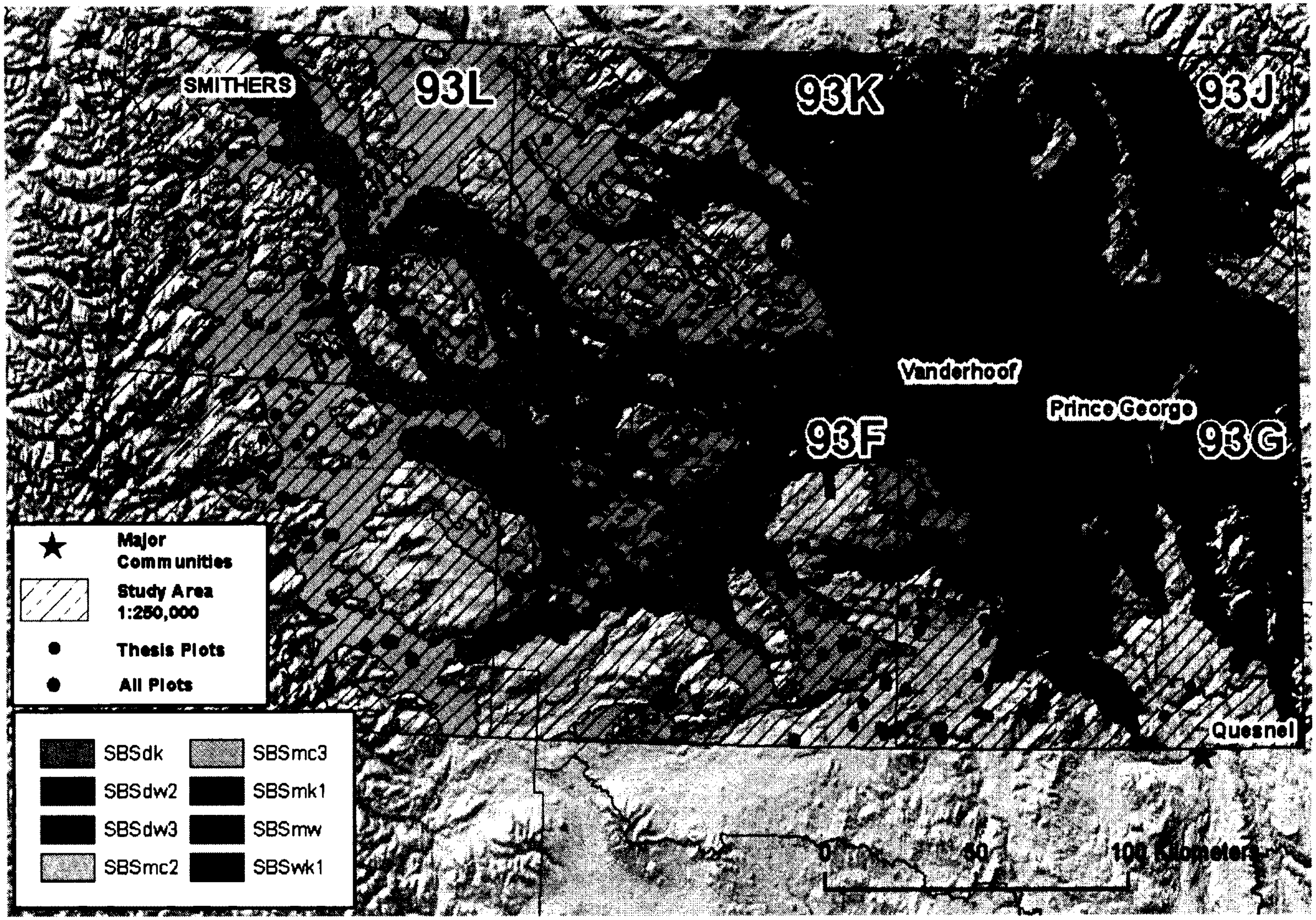



Figure 4. Location of sampled stands (thesis plots) within study area SBS BEC units.

sampling design. Consequently, my sample plots were positioned at 50-m intervals at transects extending from potential seed sources (for hybrid white spruce [*Picea engelmannii* x *glauca*], subalpine fir [*Abies lasiocarpa*], or Douglas-fir [*Pseudotsuga menziesii*]) into pine stands for distances up to approximately 700 m. The number of plots per stand typically varied from 5 to 10, depending on the size of the stand being sampled. Stands dominated by those non-pine species were identified from VRI (forest cover) maps considered current to May 2006 and interpretation of colour 1:15000 digital orthophotos compiled from 2004.

Utilizing a GIS software package, I isolated VRI stand polygons that met suitable target stand criteria. Because a straight transect from the edge of the target stand through the target stand was required, I established a transect point of commencement (POC) for each target stand through computer on-screen digitizing and geographic coordinates (Universal Transverse Mercator Easting (x) and Northing (y)) for the POC and coordinates for each subsequent plot 50 m along the transect line. I transferred the UTM to a handheld global positioning system (GPS) unit. A hardcopy map with UTM coordinates, number of intended plots, transect bearing, and target stand attributes was created for every target stand before I left for field sampling. This was particularly useful when I encountered target stands that had been logged and were no longer useful for data collection. The GPS unit helped me locate the target stand POC in the field and establish a proper bearing line along the intended transect. Both GPS locations and compass bearings were used to ensure the predetermined transect line was being followed. Once the exact location of the POC was located in the field, the predetermined bearing into the candidate stand was established and data collection plots were collected every 50 m until the stand was fully transected.

I established my sample plots by driving a semi-permanent numbered metal pigtail stake into the ground (with a unique number derived from the sample date, transect number, and plot number). A GPS coordinate was collected and recorded for the plot center. An overall site description was recorded, including: the percent slope gradient (measured with a clinometer), slope aspect (measured by compass in degrees), and mesoslope position (Figure 5).

At each plot center, I dug a 50 cm deep soil pit to accurately assess the soil texture (the relative proportions of sand, silt, and clay) and the percentage of coarse fragments. The soil properties and plot position allowed for identification of both soil nutrient regime (amount of essential soil nutrients that are available to the vegetation) and soil moisture regime (the average amount of soil water annually available for evapotranspiration by vegetation).

Using the appropriate BEC unit edatopic grid from field guides site identification and interpretation for the Southwest, Southeast, and North Central portion of the Prince George Forest Region (DeLong et al., 1993; DeLong, 2003; 2004), I was able to assign a field-based site series designation for each plot location.

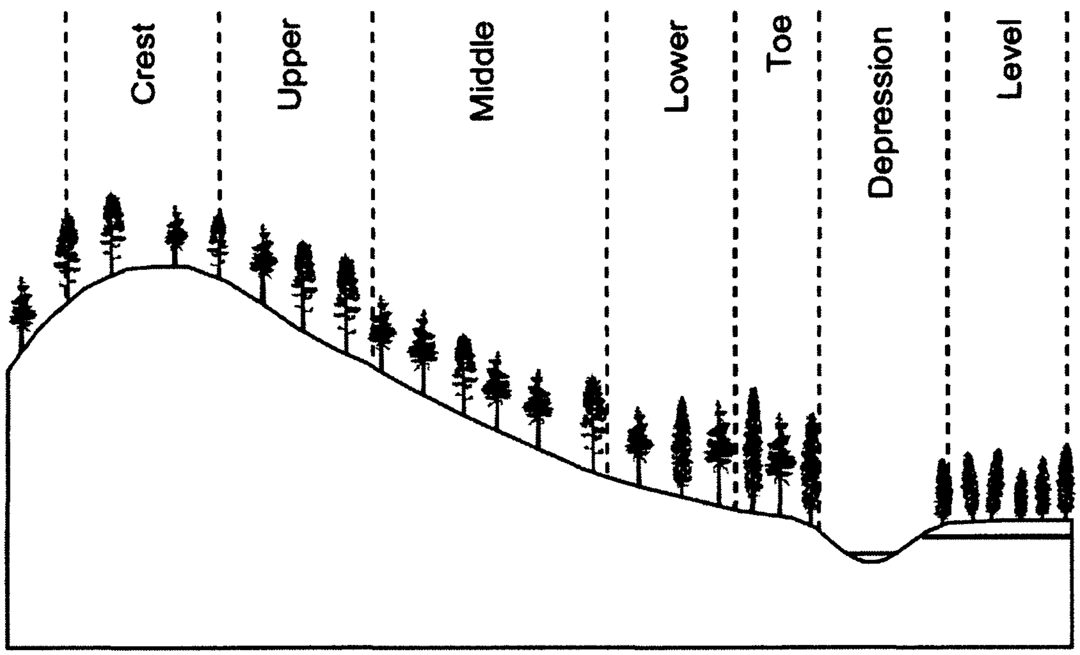



Figure 5. Cross section illustrating site mesoslope positions along a hill.

A cruise plot sweep was conducted at each plot center for basal area estimates using a prism with a basal area factor (BAF) of 4.0. To supplement the cruise data, a plot-level estimate of the number of years since MPB attack was determined by examining the trees for MPB bore holes, pitch tubes, and vigour. I established 3.99-m (50 m$^2$) and 5.64-m (100 m$^2$) radius fixed-area sample plots by attaching a metered tether to the metal pigtail at plot center and flagging the radius. Dominant plant species within the 3.99-m radius were identified. Through ocular estimate, the percent cover and average height (cm) of each species were recorded.

Also within the 3.99 radius sample plot, all seedlings (0 cm – 130 cm tall) were tallied by species, with their vigour, height (cm), and distance to the nearest seedling/sapling recorded. In the 5.64 radius sample plot, all saplings (130 cm tall to 7.5 cm in diameter as measured at 1.3 m height) were tallied by species, with their vigour, diameter at breast height (dbh), and distance to the nearest seedling/sapling recorded. The number of tree seedlings (≥10cm tall) and saplings at each plot represented the amount of advance regeneration. This number was extrapolated to estimate the number of stems/ha of advance regeneration per plot. The stems/ha of all conifer regeneration ≥30 cm in height represented the target or response variable for much of the classification tree analysis. At each of my plot locations, a camera tripod (with built-in level) was placed and levelled at plot center. Digital oblique and upward-directed hemispherical ("fisheye") photographs were taken to assist in site description and subsequent incoming light/canopy openness studies respectively. Finally, three mature trees, that were indicative of the stand's average age, were cored using an increment borer. The cores were protected in soda straws, and tree rings were counted post-field to estimate approximate stand age.

Additional data used for my thesis included an aggregation of raw field data collected from advance regeneration studies in the study area from 1996 to 2007. Although the studies may not have had similar objectives or deliverables as my thesis, I was able to extract portions of the raw data that were suitable inputs for my analysis. In total, I collected data from 162 plots within 37 stands (sampled in 2006 and 2007). These data were combined with data collected in a similar manner (i.e., along transects from non-pine stands into pine-dominated stands) by P.J. Burton and K.D. Coates (Coates et al. 2006), and other analogous plot data designed to representatively sample entire pine-leading stands rather than specifically the effects of plot distance from a non-pine seed source (e.g., Dhar and Hawkins, 2011). In total, 4241 plots were aggregated into a single plot dataset. Although I only collected 162 plots in support of this analysis, the study derives its conclusions from a 964 plot subset of the aggregated plot dataset. Table 1 lists the source, number of plots, and BEC units of all data used as inputs for the thesis classification tree model development. The 964 plots were selected by a GIS query that isolated only those plots within mature pine-leading stands (>50% lodgepole pine by basal area and mapped as >60 years old at the time of sampling) within the SBS dk, dw2, dw3, mc2, mc3, mk1, mw, and wk1 BEC units of 1:250,000 NTS map tiles 93E, 93F, 93G, 93J, 93K, and 93L.

### Post-Field Data Organization

I organized the data from the 964 plot sample data in an ArcGIS (ESRI, 2011) file geodatabase. I subsequently used this dataset as the training data for the classification tree analysis. A common table attribute (advance regeneration stems/ha) was established and populated with field collected data. Classification tree analysis requires the variable being predicted to be categorical.

Table 1. Number of advance regeneration plots sampled per BEC unit by data source.

| Source | Year | BEC Unit (SBS) | | | | | | | | Total Plots |
|---|---|---|---|---|---|---|---|---|---|---|
| | | dk | dw2 | dw3 | mc2 | mc3 | mk1 | mw | wk1 | |
| Brooks | 2007 | - | - | 59 | - | - | 54 | - | 14 | 127 |
| Burton | 2005 | 8 | 21 | 138 | 5 | 36 | - | - | - | 208 |
| CarrotLake | 2006 | - | - | - | - | 100 | - | - | - | 100 |
| Cichowski | 2005 | 5 | - | - | 18 | - | - | - | - | 23 |
| Coates | 2005 | 38 | - | - | 55 | - | - | - | - | 93 |
| DeLong | 2005 | 10 | - | 9 | - | 6 | - | - | - | 25 |
| Fluxnet | 2006 | - | - | - | - | - | 9 | - | - | 9 |
| Hawkes | 2002 | 77 | - | - | - | - | - | - | - | 77 |
| Hawkins | 2005 | - | 23 | 53 | 13 | - | - | 3 | - | 92 |
| NIVMA | 1996 - 2001 | 4 | 1 | - | 8 | 1 | 4 | - | - | 18 |
| Rakochy | 2004 | 150 | - | - | 42 | - | - | - | - | 192 |
| Total Plots | | 292 | 45 | 259 | 141 | 143 | 67 | 3 | 14 | 964 |

As my thesis objective was to build a model that assists in predicting the stocking status of stands, it was necessary to convert the advance regeneration stems/ha to a Boolean attribute of stocked or not stocked. An important qualification should be made with regards to my thesis' use of the attribute "stocked". I applied a conservative estimate of stocking, as the data only refers to the stocking of seedlings (>10cm tall but less than 130cm tall) and saplings (>130cm tall but less than 7.5 cm dbh). My analysis does not take into account germinants (<10 cm tall) or existing trees (>7.5 dbh), even though both could be considered important elements of advance regeneration. Germinants are not infrequent in the understories of the more moist zones of the SBS (Vyse et al., 2009). I also make the assumption that all pine trees >7.5 cm dbh within the plot will not survive post MPB, and therefore are not considered in the stocking calculation. This becomes an important consideration when we consider the growing space actually available to the advance regeneration. The results are consequently a conservative estimate of stocking and could be better identified as "stocked with seedlings and saplings".

I selected a threshold of 600 stems/ha to separate stocked from not stocked stands as several studies referring the percentage of plots meeting or exceeding minimal stocking standards have used 600 stems/ha as the baseline (Bulmer et al., 2002; Burton, 2006; Vyse et. al., 2009). 600 stems/ha is widely recognized as the minimum well-spaced preferred trees/ha in in stocking guidelines for the regeneration of clearcuts (e.g., British Columbia Ministry of Forests 2000), and is threshold beneath which stand rehabilitation measures might be undertaken. Based on that threshold, I created two new attributes in the geodatabase, named alltrees600 and conifers600. The alltrees600 attribute was populated with a 1 if the advance regeneration total stems/ha for the sample plot was equal to or greater than 600 stems/ha and populated with a 0 if the advance regeneration total stems/ha was less than 600 stems/ha. This attribute applied to all tree species within the sample plot. A second 600 stems/ha attribute, conifer600, was also populated with a 1 or 0 following the same criteria as the alltrees600 attribute, but only counting those conifer species in the sample plot. In addition, a third threshold group, MSSpa, was established to assign a stocking value of one or zero according to the appropriate minimum stocking standard of preferred and acceptable species found in the Establishment to Free Growing Guidebook for the Prince George Forest Region (British Columbia Ministry of Forests, 2000). Using the Establishment to Free Growing Guidebook the MSSpa, I identified criteria for each biogeoclimatic unit and site series pairing that contained a study plot. The published MSSpa values and conditions were used to define the level of stocking against which to assess seedling and sapling densities (Appendix 1). MSSpa can be as low as 200 stems/ha to as high as 700 stems/ha in the subzone/site series present in my study area (British Columbia Ministry of Forests, 2000). Plots that had less than the prescribed MSSpa were considered not stocked and plots that met or exceeded the prescribed MSSpa were considered stocked. The "well-spaced" requirement for counting regeneration was ignored, as this value (which typically ranges from 1.0 to 2.5 m between seedlings) is arbitrary and depends on the silvicultural prescription and growth modelling assumptions, while fully mature trees are often observed with stem bases growing <1.0 m apart. This simple but conservative measurement of understory stocking (designed to be applied under open-growing conditions) was critical to the classification tree development and subsequent GIS mapping model.

**Statistical Analyses**

Initially, basic descriptive statistics were applied to the data to examine patterns in the variables. T-tests, R package *t.test* (R Core Team, 2012), were conducted to examine the contribution of non-conifers to the advance regeneration densities. Linear regression models, R package *glm* (R Core Team, 2012), were applied to the data to explore the possibilities of relationships between advance regeneration densities and several key variables. The primary objective of building a predictive geospatial model of the probability distribution of advance regeneration in target pine stands was pursued by developing classification tree models in DTREG (Sherrod, 2006) and R (R Core Team, 2012) , using a recursive partitioning package called *rpart* (Therneau et al., 2012). The classification-tree model required a binary response variable (stocked or not stocked) and publicly available explanatory variables that currently exist in digital format. Therefore, any data I collected in the field that could not be derived from publicly available geospatial data sets (e.g., understory plant cover) were not used as a potential explanatory (predictor) variable. I determined that data derived from: 1) the Vegetation Resource Inventory (VRI, 2006); 2) Predictive Ecosystem Modelling (PEM, 2008 ); and 3) interpolated and elevation adjusted climate data calculated using ClimateBC (Wang et al. 2006) were all publicly accessible and contained key potential predictors. The three datasets together yielded 54 potential predictors (Table 2). The intention was to find the best combination of potential predictors that could be used to create a parsimonious predictive model to accurately classify each 1-ha cell section within target stands across the study area as stocked (1) or not stocked (0). Additionally, each cell would not only be designated as stocked or not stocked, but would also be assigned the probability of being stocked or not stocked. A cross validation argument *cv.tree* function was applied to the full tree to minimize the misclassification error associated with overfitting of the tree. The cross validation pruned back the tree to the optimal number of splits/node pairs. See Appendix 3 for a more detailed description of recursive partitioning, tree pruning using cross-validation, and variable importance calculation.

**Geographical Information System (GIS) Analyses**

The second objective of my thesis was to create a geospatial model that assists in converting the probability of stocking into a geospatial environment. Through the use of classification tree modeling, the rules for determining stocked or not stocked and the probabilities of stocking were determined. One of the passive inputs (included in the model development but not used as a key predictor) into the classification tree analysis was the geographic coordinates of the 100 m cells. The themed polygons portraying the likely distribution and location of stocked and unstocked stands dominated by mature lodgepole pine in Sub-Boreal Spruce

Table 2. Ecological variables from publicly available geospatial data(evaluated as potential inputs) to the probability of stocking classification tree model.

| Variable | Source | Type | Variable | Source | Type |
|---|---|---|---|---|---|
| UTM coordinate (easting) | GIS | 2 | Elevation | GIS | 2 |
| UTM coordinate (northing) | GIS | 2 | Leading Species Live Volume per Hectare at 12.5 cm | VRI | 2 |
| Forest District | VRI | 1 | Leading Species Live Volume per Hectare at 17.5 cm | VRI | 2 |
| Biogeoclimatic Subzone/Variant | VRI | 1 | Second Species Live Volume per Hectare at 12.5 cm | VRI | 2 |
| Site Series | PEM | 1 | Second Species Live Volume per Hectare at 17.5 cm | VRI | 2 |
| Soil Moisture Regime | VRI | 1 | Leading Species Dead Volume per Hectare at 12.5 cm | VRI | 2 |
| Soil Nutrient Regime | VRI | 1 | Leading Species Dead Volume per Hectare at 17.5 cm | VRI | 2 |
| Absolute Moisture Regime | VRI | 1 | Mean Annual Temperature | $C^{bc}$ | 2 |
| Surface Expression | VRI | 1 | Mean Warmest Month Temperature | $C^{bc}$ | 2 |
| Modifying Process | VRI | 1 | Mean Coldest Month Temperature | $C^{bc}$ | 2 |
| Mesoslope Position | VRI | 1 | Temperature Difference Between MWMT and MCMT | $C^{bc}$ | 2 |
| Quadratic Diameter at 12.5 cm | VRI | 2 | Mean Annual Precipitation | $C^{bc}$ | 2 |
| Quadratic Diameter at 17.5 cm | VRI | 2 | Mean Summer (May to Sept.) Precipitation | $C^{bc}$ | 2 |
| Crown Closure | VRI | 2 | Annual Heat: Moisture Index | $C^{bc}$ | 2 |
| Site Index | VRI | 2 | Summer Heat: Moisture Index | $C^{bc}$ | 2 |
| Basal Area | VRI | 2 | Degree-Days Below 0°C | $C^{bc}$ | 2 |
| Tree Cover Pattern | VRI | 1 | Degree-Days Above 5°C | $C^{bc}$ | 2 |
| Vertical Complexity | VRI | 1 | Degree-Days Below 18°C | $C^{bc}$ | 2 |
| Species Composition of Leading Species | VRI | 1 | Degree-Days Above 18°C | $C^{bc}$ | 2 |
| Leading Species Percentage | VRI | 2 | Number of Frost-Free Days | $C^{bc}$ | 2 |
| Species Composition for Second Species | VRI | 1 | Frost-Free Period | $C^{bc}$ | 2 |
| Second Species Percentage | VRI | 2 | Precipitation as Snow | $C^{bc}$ | 2 |
| Percent of Pine in Leading Species | VRI | 2 | Extreme Minimum Temperature over 30 Years | $C^{bc}$ | 2 |
| Percent of NonPine in Stand | VRI | 2 | Hargreaves Reference Evaporation | $C^{bc}$ | 2 |
| Ratio of Pine to NonPine in Stand | VRI | 2 | Hargreaves Climatic Moisture Deficit | $C^{bc}$ | 2 |
| Projected Age for Leading Species | VRI | 2 | Distance (m) to Nearest Nonpine Seed Source (SW direction) | GIS | 2 |
| Projected Height for Leading Species | VRI | 2 | Species Composition of Nearest Nonpine Seed Source | GIS | 1 |

Data Source: VRI (Vegetation Resource Inventory), PEM (Predictive Ecosystem Mapping), $C^{bc}$ (ClimateBC), GIS (derived data using ArcGIS) Data Type: 1 (categorical), 2 (continuous) biogeoclimatic zone. The GIS model incorporates the exported rules from the classification tree to isolate criteria identified in each branch. Each of the terminal nodes in the classification tree could be extracted from the resultant tree individually through a series of single programming statements, with each properly executed statement resulting in a queried output. This, however, would be an inefficient use of the final model, as the result would be a series of if-then strings that would each have to be applied against test data. By creating a classification tree and the splitting rules that make up the tree as a model, all the branches leading to all the terminal nodes could be applied against test data simultaneously. Because the model dataset contained geographical coordinates, the dataset could be imported into a GIS. The resultant is a spatially attributed file that identifies the variables used in the model, the predicted target value for every 100-m raster cell (as stocked or not stocked), the terminal node used for each predicted value, and a probability value for the designation of stocked or not stocked for every 100-m raster cell. The probability value is the critical attribute that is used in creating the predictive colour-theme map. Much like the classification tree output, ArcGIS models are components linked together through connectors such as tools, variables, and iterators (Allen, 2011). The final model output is an extensible geospatial model that can be run within the ArcGIS environment. The model automates the geoprocessing tasks that are required to prepare the data for modelling, geospatial analysis, and colour-themed mapping.

## Chapter Three: Results

### Trends in Advance Regeneration Density

A total of 243 pine-leading stands (containing the 964 plots) were field sampled. A majority of these plots were located in the SBSdk, SBSdw3, SBSmc2, and SBSmc3 biogeoclimatic units and on 01, 03, 04, or 05 site series (Table 3).

Table 3. Number of regeneration plots sampled in pine-dominated stands on different site series in different biogeoclimatic subzones.

| BEC Unit | Site Series | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | Total |
| dk | 142 | 71 | 52 | 0 | 21 | 3 | 2 | 0 | 0 | 1 | 292 |
| dw2 | 16 | 0 | 6 | 13 | 2 | 1 | 2 | 0 | 4 | 1 | 45 |
| dw3 | 38 | 2 | 27 | 45 | 73 | 39 | 14 | 3 | 16 | 2 | 259 |
| mc2 | 97 | 21 | 1 | 0 | 13 | 5 | 0 | 0 | 0 | 4 | 141 |
| mc3 | 20 | 0 | 2 | 114 | 2 | 0 | 5 | 0 | 0 | 0 | 143 |
| mk1 | 12 | 2 | 16 | 4 | 23 | 9 | 0 | 0 | 1 | 0 | 67 |
| mw | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| wk1 | 0 | 0 | 7 | 3 | 3 | 1 | 0 | 0 | 0 | 0 | 14 |
| Total | 326 | 96 | 111 | 181 | 137 | 58 | 23 | 3 | 21 | 8 | 964 |

The 964 plots were located on a range of topographic locations. The percentages on the site meso-slope cross section (Figure 6) quantify the location of the 964 plots according to their topographical locations (as determined by intersection with VRI digital data). It should be noted that although site mesoslope position was one of the potential predictors, it was a more general topographic descriptor named surface expression (VRI, 2006) that became a key predictor variable.

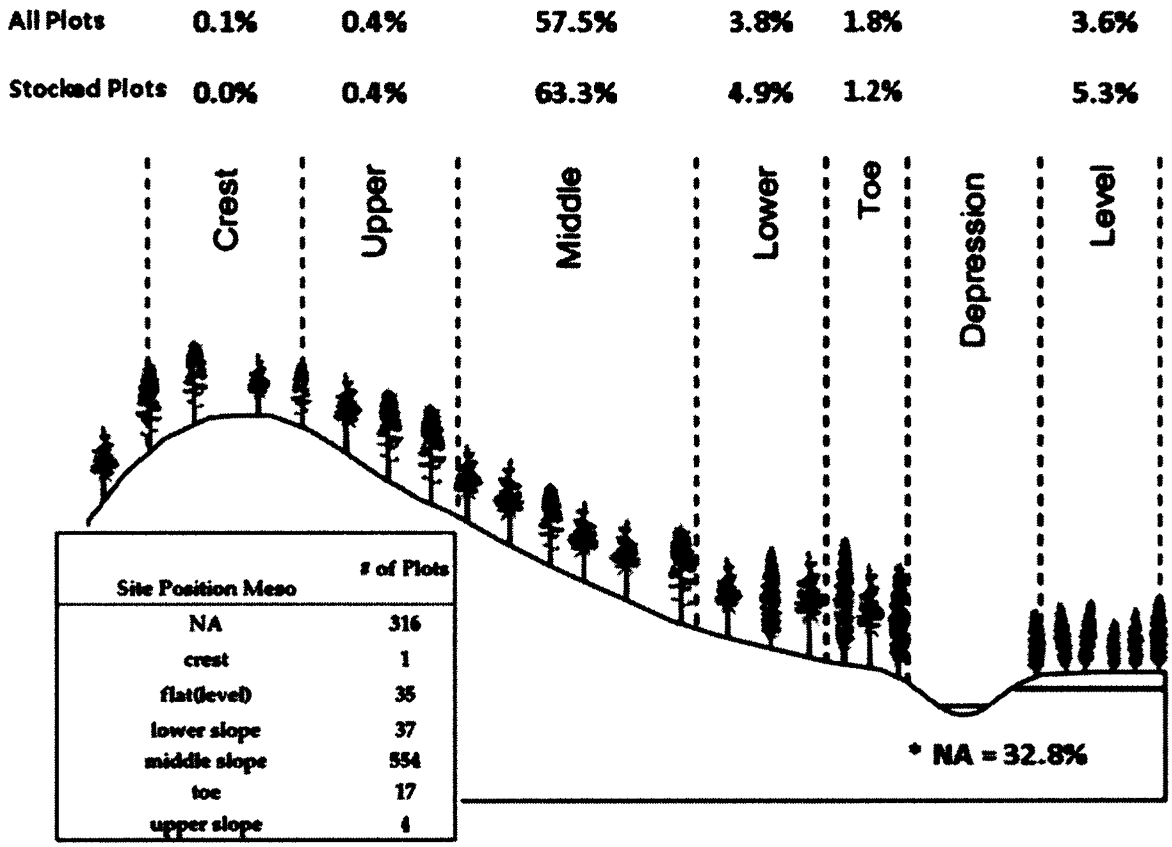



Figure 6. Proportion of the study area plot locations (n=964) among different site meso-slope positions. NA represents plots for which meso-slope value was not available in the VRI

The distribution of plots reflects the current age class structure of pine forests in the Northern Interior, with many mapped as being greater than 100 years of age, but with "old-growth" lodgepole pine and natural stands under 80 years of age much more difficult to find. The distribution of plots by stand age and BEC subzone is shown in Table 4. Cross-tabulating plots by site series and age class, a concentration of sampling in circum-mesic stands 80 to 140 years of age is evident (Table 5). Note that site series 07, 08, 09, 10 have been binned into a single "moist

Table 4. Number of regeneration plots sampled in pine-dominated forest cover polygons by age classes* and biogeoclimatic subzones.

| BEC Unit | Nominal Mapped Age Class * | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 5 | 6 | 7 | 8 | 9 | Total |
| dk | 9 | 63 | 55 | 64 | 101 | 0 | 292 |
| dw2 | 4 | 4 | 3 | 24 | 10 | 0 | 45 |
| dw3 | 2 | 103 | 42 | 22 | 90 | 0 | 259 |
| mc2 | 0 | 2 | 10 | 23 | 104 | 2 | 141 |
| mc3 | 14 | 5 | 0 | 105 | 18 | 1 | 143 |
| mk1 | 0 | 32 | 0 | 4 | 31 | 0 | 67 |
| mw | 0 | 0 | 3 | 0 | 0 | 0 | 3 |
| wk1 | 0 | 14 | 0 | 0 | 0 | 0 | 14 |
| Total | 29 | 223 | 113 | 242 | 354 | 3 | 964 |

* age class 4 is 61 to 80 years, age class 5 is 81 to 100, age class 6 is 101 to 120 , age class 7 is 121 to 140 , age class 8 is 141 to 240 , and age class 9 is 240+ years.

Table 5. Number of regeneration plots sampled in pine-dominated stands by age classes* and sites series, across all biogeoclimatic subzones.

| Site Series | Nominal Mapped Age Class * | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 5 | 6 | 7 | 8 | 9 | Total |
| 01 | 6 | 28 | 48 | 76 | 166 | 2 | 326 |
| 02 | 8 | 27 | 8 | 16 | 36 | 1 | 96 |
| 03 | 3 | 37 | 23 | 17 | 31 | 0 | 111 |
| 04 | 10 | 22 | 10 | 99 | 40 | 0 | 181 |
| 05 | 0 | 64 | 9 | 19 | 45 | 0 | 137 |
| 06 | 1 | 28 | 10 | 5 | 14 | 0 | 58 |
| 07,08,09,10 | 1 | 17 | 5 | 10 | 22 | 0 | 55 |
| Total | 29 | 223 | 113 | 242 | 354 | 3 | 964 |

* age class 4 is 61 to 80 years, age class 5 is 81 to 100, age class 6 is 101 to 120 , age class 7 is 121 to 140 , age class 8 is 141 to 240 years, and age class 9 is 240+ years.

site" category in order to increase sample size. Across all 964 plots, advance regeneration by all species (>10cm tall) averaged a mean density of 1268 stems/ha and a median density of 600 stems/ha. This means that one-half of the plots sampled had more than 600 stems/ha of advance regeneration (my thesis threshold for designating a plot as stocked), and one-half of the plots sampled had less.

A paired-samples t-test was conducted to determine if there was a significant difference between the mean densities of advance regeneration when calculations included and excluded non-conifer species. I found that there was not a significant difference in the mean density between conifers only (mean=1267.87, s.d.=1911.90 stems/ha) and all species (mean=1289.02, s.d.=1909.36 stems/ha) across the study area; $t_{963}$=4.42, P = 0.08. This result suggests that non-conifer species do not significantly influence the mean density of advance regeneration, and by extension, the stocking status of the stand.

An examination of advance regeneration across biogeoclimatic subzones shows that mean densities across all subzones exceed the 600 stems/ha— except in the SBSmw, a subzone with only three samples. Comparison of mean and median advance regeneration densities across the thesis study area proved to be quite valuable, as this comparison indicates; most sub-populations are not normally distributed. Mean regeneration densities for each BEC unit, age class, and forest district respectively are presented in Tables 6 through 8.

Table 6. Descriptive statistics for advance regeneration densities (stems/ha) for all sample plots in each BEC unit. Mean, median, standard deviation, and interquartile ranges are reported.

| BEC Unit | n | Mean | s.d | 25% | Median | 75% |
|---|---|---|---|---|---|---|
| dk | 287 | 743 | 1342 | 0 | 200 | 800 |
| dw2 | 45 | 1504 | 1649 | 200 | 1200 | 1700 |
| dw3 | 262 | 1278 | 1661 | 200 | 750 | 1600 |
| mc2 | 143 | 1086 | 1414 | 200 | 600 | 1400 |
| mc3 | 143 | 2120 | 2992 | 200 | 600 | 3250 |
| mk1 | 61 | 2077 | 2263 | 400 | 1200 | 2700 |
| mw | 3 | 266 | 115 | 200 | 200 | 300 |
| wk1 | 20 | 2035 | 1571 | 850 | 1650 | 3750 |

Table 7. Descriptive statistics for advance regeneration densities (stems/ha) for all sample plots per age class. Mean, median, standard deviation, and interquartile ranges are reported.

| Age Class | n | Mean | s.d | 25% | Median | 75% |
|---|---|---|---|---|---|---|
| 4 | 29 | 231 | 355 | 0 | 100 | 300 |
| 5 | 223 | 1221 | 1597 | 200 | 600 | 1700 |
| 6 | 113 | 1132 | 1435 | 200 | 600 | 1600 |
| 7 | 242 | 1399 | 2082 | 200 | 600 | 1675 |
| 8 | 356 | 1387 | 2136 | 200 | 600 | 1613 |
| 9 | 1 | 2800 | na | 2800 | 2800 | 2800 |

Table 8. Descriptive statistics for advance regeneration densities (stems/ha) for all sample plots per forest district. Mean, median, standard deviation, and interquartile ranges are reported.

| Forest District | n | Mean | s.d | 25% | Median | 75% |
|---|---|---|---|---|---|---|
| Fort St. James | 1 | 1200 | na | 1200 | 1200 | 1200 |
| Nadina | 350 | 746 | 1283 | 0 | 400 | 1000 |
| Prince George | 275 | 1410 | 1663 | 300 | 900 | 1750 |
| Quesnel | 4 | 900 | 621 | 500 | 900 | 1300 |
| Vanderhoof | 334 | 1762 | 2446 | 200 | 750 | 2375 |

Across the entire study area, there was a near equal distribution of stocked and not stocked across the 964 sample plots for all three stocking groups (Table 9).

Table 9. Total number of stocked and not stocked sample plots across the study area based on three different definitions of acceptable stocking (stocking groups).

| Stocking Group | Stocked | Not Stocked |
|---|---|---|
| All Tree Species (600 stems/ha) | 510 | 454 |
| Conifer Species only (600 stems/ha) | 496 | 468 |
| Minimum Stocking Standards * | 460 | 504 |

Based on the sampling within the study area, all but two BEC units (SBSdk and mw) are greater than 50% stocked for all stocking groups, and as high as 70% stocked in the mk1 and wk1 BEC units (Figures 7 , 8, and 9). The biogeoclimatic unit SBSmw had no stocked plots within the study area. It is worth noting that there were only three plots gathered in the mw subzone variant, all collected from the same stand.

In an attempt to identify any key variables that may have a direct (positive or negative) effect on the advance regeneration densities, several potential key variables were examined through a simple linear regression model. This exercise was conducted for the purposes of data exploration and description only, searching for any potential relationships between the advance regeneration

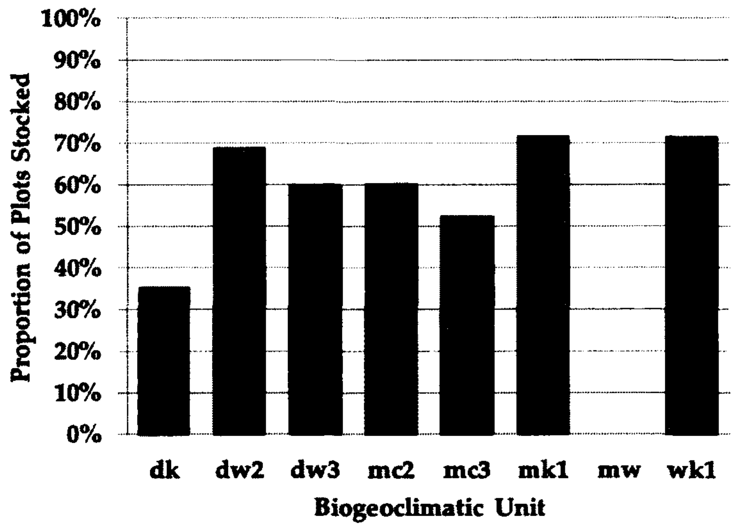



Figure 7. Proportion of 964 plots in pine-leading forest cover polygons in the Northern Interior Forest Region that meet a 600 stems/ha stocking density threshold (for all tree species).

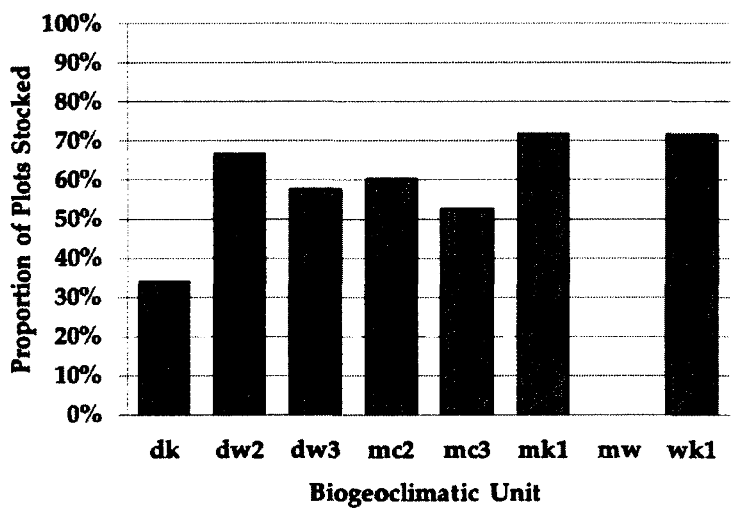



Figure 8. Proportion of 964 plots in pine-leading forest cover polygons in the Northern Interior Forest Region that meet a 600 stems/ha stocking density threshold (for conifer species).

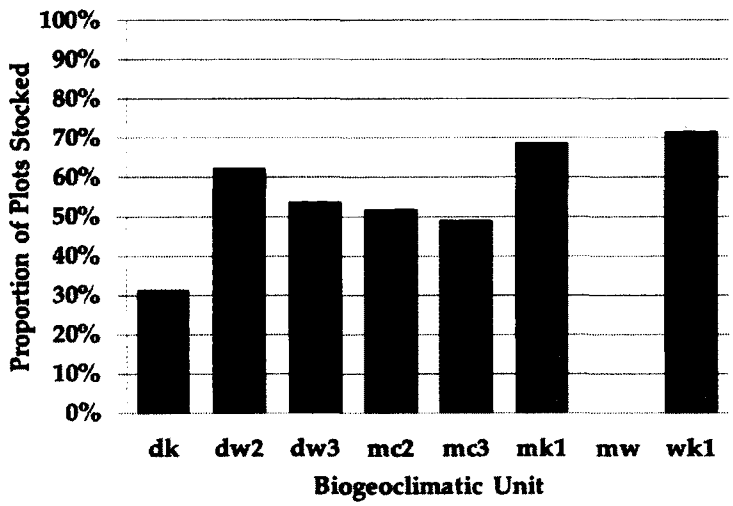



Figure 9. Proportion of 964 plots in pine-leading forest cover polygons in the Northern Interior Forest Region that meet a prescribed stocking density threshold (for minimum stocking standards - preferred and acceptable species).

(positive or negative) effect on the advance regeneration densities, several potential key variables were examined through a simple linear regression model. This exercise was conducted for the purposes of data exploration and description only, searching for any potential relationships between the advance regeneration stems/ha of each stand and basal area (from VRI), distance to nearest seed source (calculated using GIS), live pine volume per ha (VRI), dead pine volume per ha (VRI), mean annual precipitation (ClimateBC), mean annual temperature (ClimateBC), stand height (VRI), and stand age (VRI). The results of the models, presented in Figures 10 through 14, are represented as scatterplots. The exercise did not uncover any variables that had a strong ($r^2$ >0.5) linear relationship with density of advance regeneration. Only distance from nearest seed source yielded a statistically significant result (P=0.007). The $r^2$ (0.008), however, indicates that distance from nearest seed source explains less than 1% of the variation in advance regeneration stems/ha. The resulting linear regression equation for distance from nearest seed source is total stems/ha = 1462.0493 - 0.5851 * m to nearest seed source.

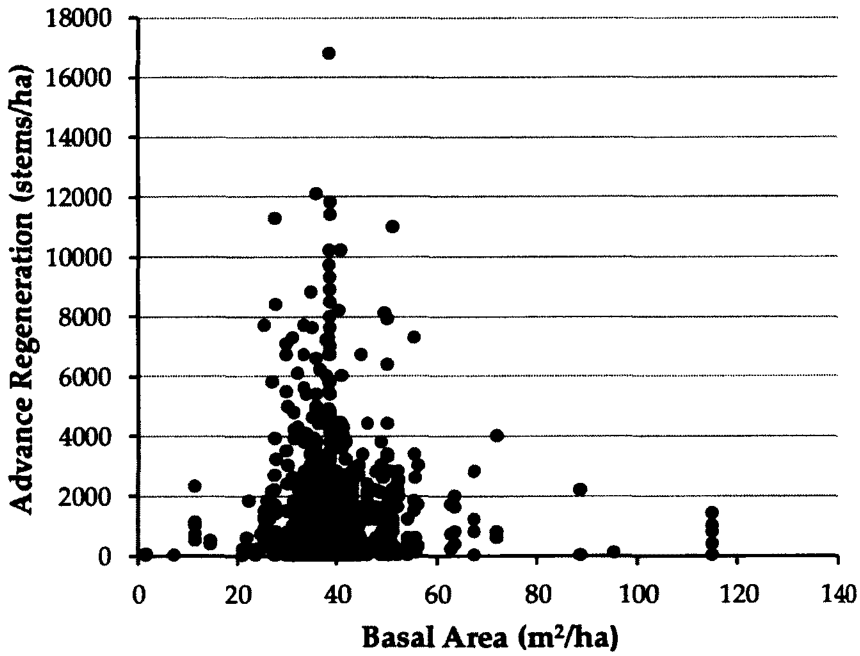



Figure 10. Scatterplot of advance regeneration density vs. stand basal area ($r^2 = 0.0002315$, P = 0.637, F = 0.2228, n = 964).

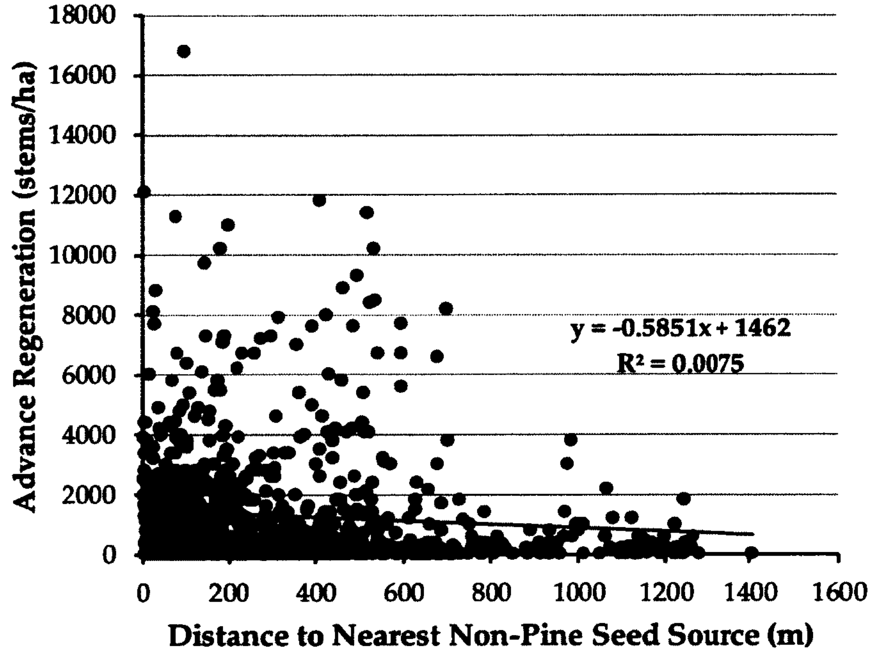



Figure 11. Scatterplot of advance regeneration density vs. distance to the nearest non-pine seed source ($r^2 = 0.008$, $P = 0.007$, $F= 7.239$, $n = 964$).

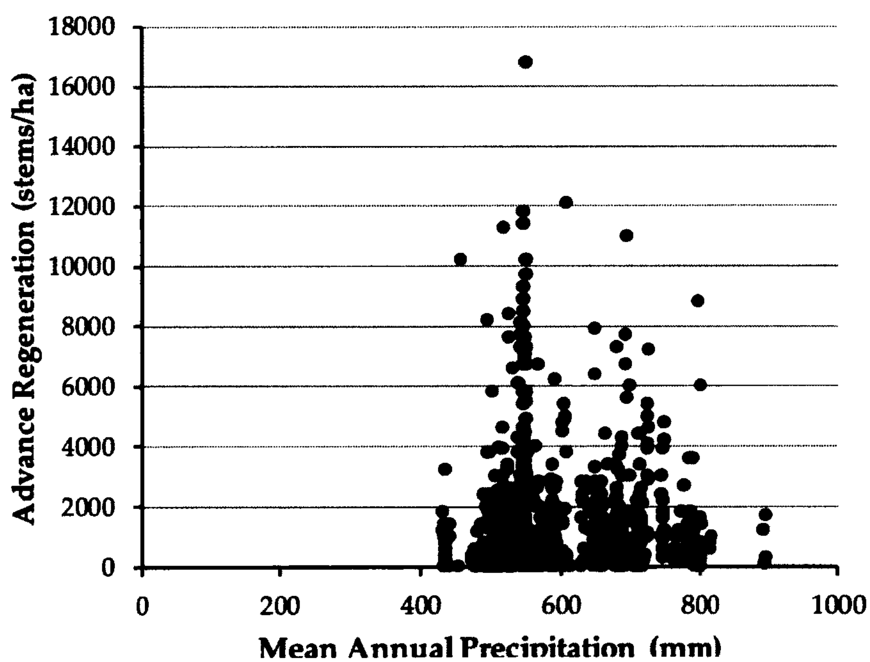



Figure 12. Scatterplot of advance regeneration density vs. distance to the nearest non-pine seed source ($r^2 = 0.003$, $P = 0.058$, $F= 3.596$, $n = 964$.)

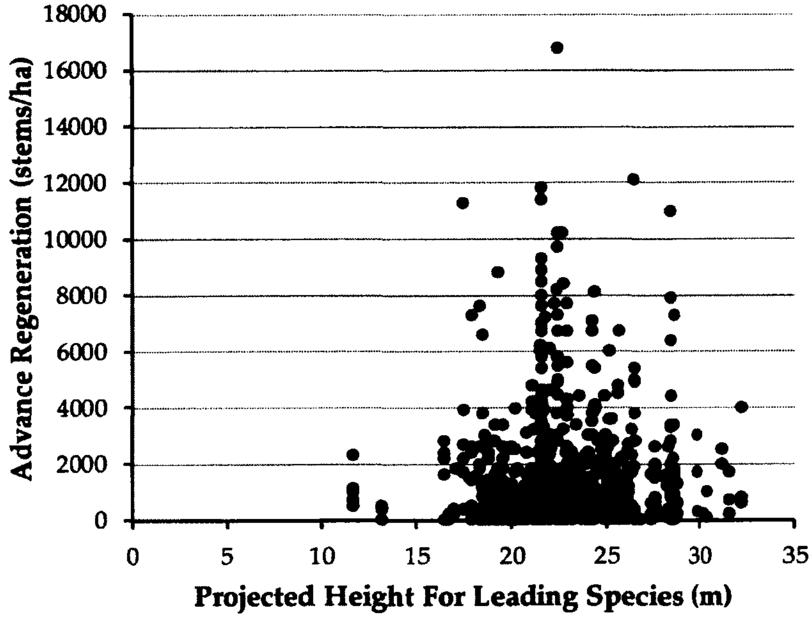



Figure 13. Scatterplot of advance regeneration density vs. stand projected height ($r^2 = 0.001$, $P = 0.436$, $F= 0.6068$, $n = 964$).

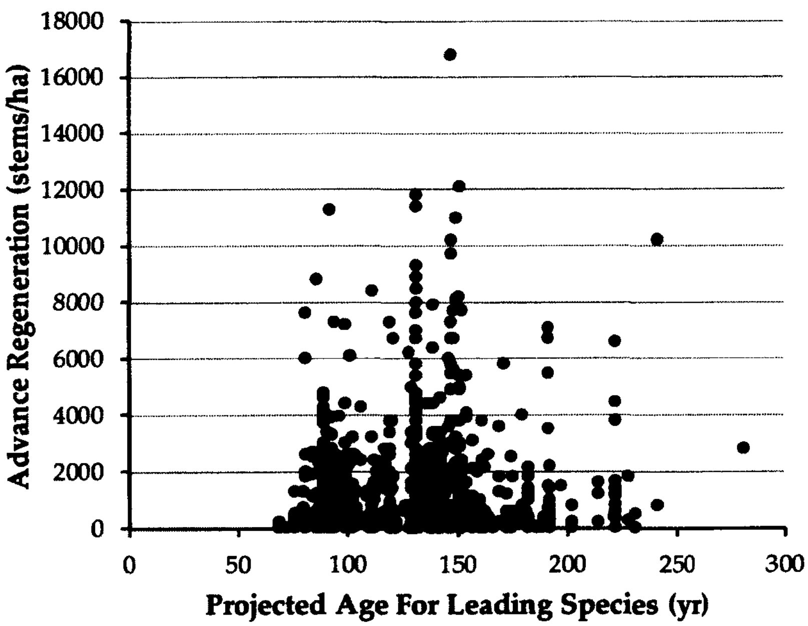



Figure 14. Scatterplot of advance regeneration density vs. stand projected age ($r2 = 0.003$, $P = 0.076$, $F= 3.147$, $n = 964$).

**Classification Tree Analysis**

My model derived from the *rpart* module in R classified the presence/absence of advance regeneration stocking with an accuracy of 77% (11 terminal nodes) for all-tree stocking, 78% (13 terminal nodes) for conifer-only stocking, and 78% (14 terminal nodes) for MSSpa stocking. As expected, the identity and order of predictors deemed important by the models were similar for each of the three stocking groups. The final tree for all three stocking groups had BEC unit, distance to nearest non-pine seed source, projected stand age, and basal area as their top four important variables. It is also interesting that all three stocking groups utilized the categorical variable BEC unit as their initial split, making it the most important variable for all three groups. This would appear to indicate that the density of advance regeneration is influenced by the broad factors associated with biogeoclimatic subzones (i.e., climate, vegetation, soil) and may imply that the model is suitable as an overarching landscape-level predictor.

The critical statistical outputs for the final model (Table 10) indicate that in all three stocking groups, the resultant final tree was larger than it needed to be (i.e., the number of terminal nodes before pruning was overfitting the data) and that a simpler tree with fewer terminal nodes could achieve the same (or better) accuracy. The numbers of terminal nodes after pruning are the final number of nodes associated with the best fitting and most parsimonious classification tree model.

Table 10. Key statistics describing the classification tree accuracy for the advance regeneration prediction model. The classification tree size, accuracy, and evaluation are listed for all three stocking groups. MSSpa is the Minimum stocking standard (various stems/ha) of preferred and acceptable species, particular to each site series in each BEC unit.

| | 600 stems/ha (all tree) | 600 stems/ha conifers) | MSSpa |
|---|---|---|---|
| Number of terminal nodes (before turning) | 24 | 31 | 29 |
| Number of terminal nodes (after pruning) | 11 | 14 | 14 |
| Number of splits (after pruning) | 10 | 13 | 13 |
| Sensitivity (true positive rate) | 82.45% | 80.85% | 71.99% |
| Specificity (false positive rate) | 71.33% | 74.15% | 83.63% |
| Misclassification rate | 0.2282 | 0.2241 | 0.2189 |
| Model accuracy | 77.18% | 77.59% | 78.11% |
| ROC score | 0.8386 | 0.8327 | 0.8528 |

The red arrows in Figure 15 illustrates the path of a rule from initial split to terminal node. The initial split at SUBZVAR (BEC unit) is governed by the rule that the plot must be located in SBSdk or SBSmw to follow this branch.

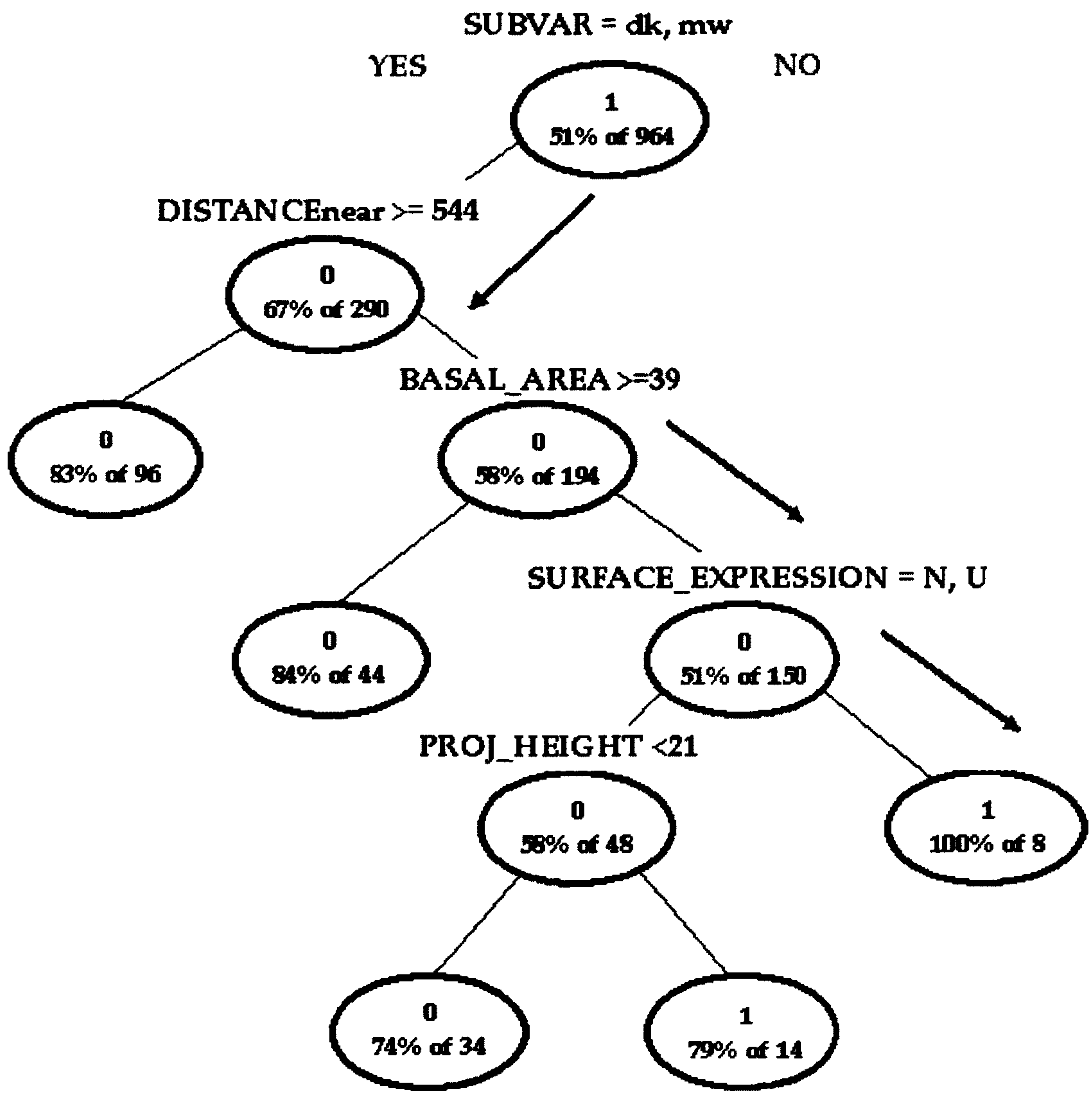



Figure 15. Example of a rule path in the conifer only 600 stems/ha classification tree. The tree begins with an initial split of BEC unit and terminates after a split based on surface expression. This rule can be found in Table 15 as rule number 10.

The second split at DISTANCEnear (distance to nearest non-pine seed source) is governed by a threshold of 544 m to the nearest non-pine seed source. If the nearest distance to a non-pine seed source is ≥544 m, then the rule follows to the left and terminates at the dark green node classified as 0, or not stocked. The rule indicates that 96 plots are present in the split. The rule stipulates that there are only 96 plots that fit both splits (BEC unit and distance to nearest seed source) and that the likelihood of those 96 plots being not stocked is 17%. This provides a low confidence that BEC unit and distance from nearest non-pine seed source alone are good indicators for predicting stocking. If the plot is <544.3 m to the nearest non pine seed source, then the rule follows the branch to the right where it encounters a split at basal area. If the basal area is <39.34 $m^2$/ha, then the rule follows to the right where it encounters its last split of surface expression being either U (undulating) or M (rolling). The plot terminates as a stocked node, where 100% of the 8 plots in the 964 plot dataset that conformed to this set of rules were correctly classified as stocked.

The rule in text form reads as follows:

SUBZVAR=dk, or mw
DISTANCEnear<544.3 m
BASAL_AREA<39.34 $m^2$/ha
SURFACE_EXPRESSION=M (rolling), or P (undulating)

The fully pruned classification trees (for all three stocking groups) provide sets of rules for the prediction of stocked with seedlings and saplings (Figure 16, 17, 18). The trees are interpreted (as per Figure 14) by following the splits in the branches — to the left if the split value in the tree is true and to the right if the split

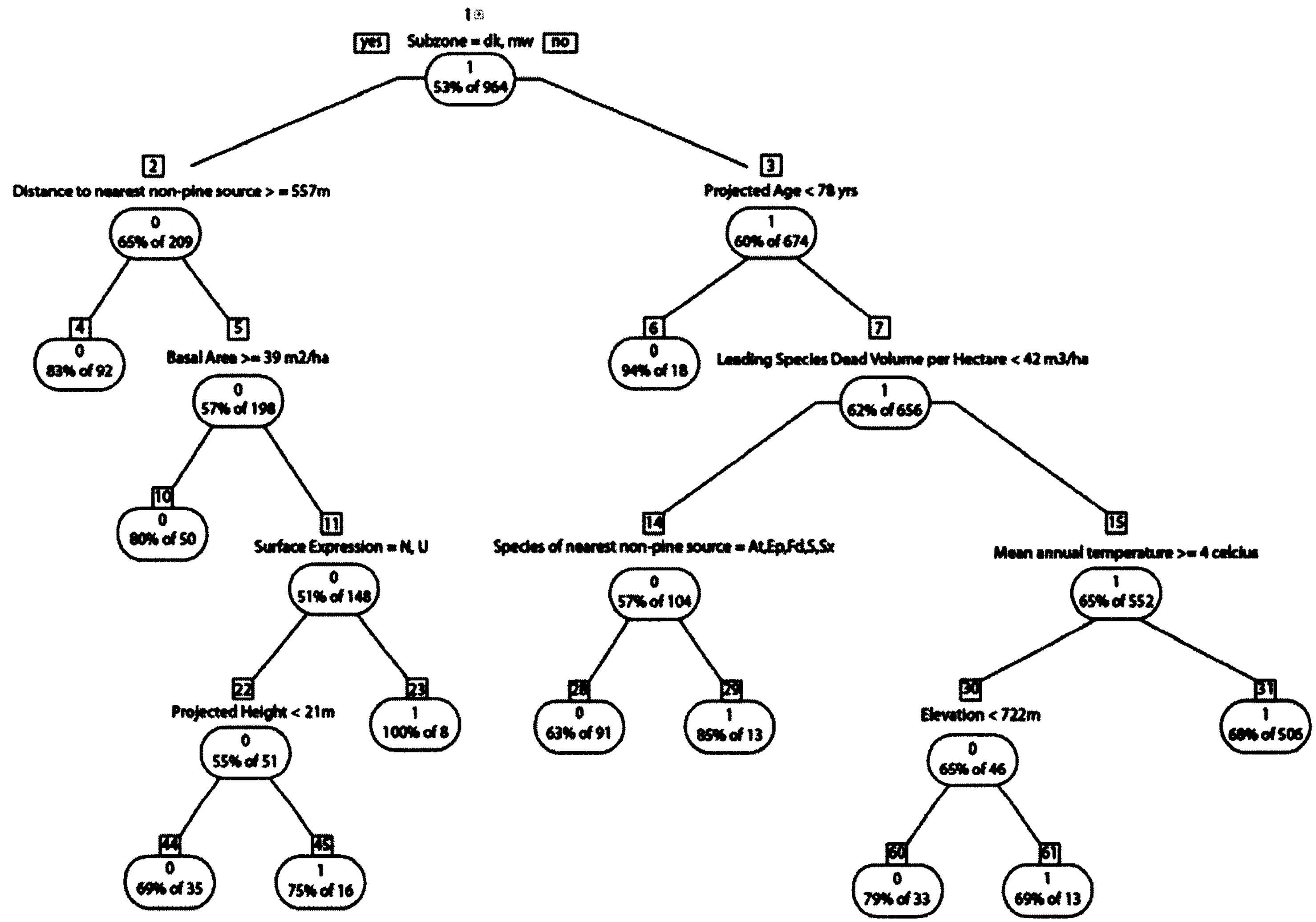



Figure 16. Final pruned (using 10-fold cross-validation) classification tree model for predicting the probability of stocking by all tree species

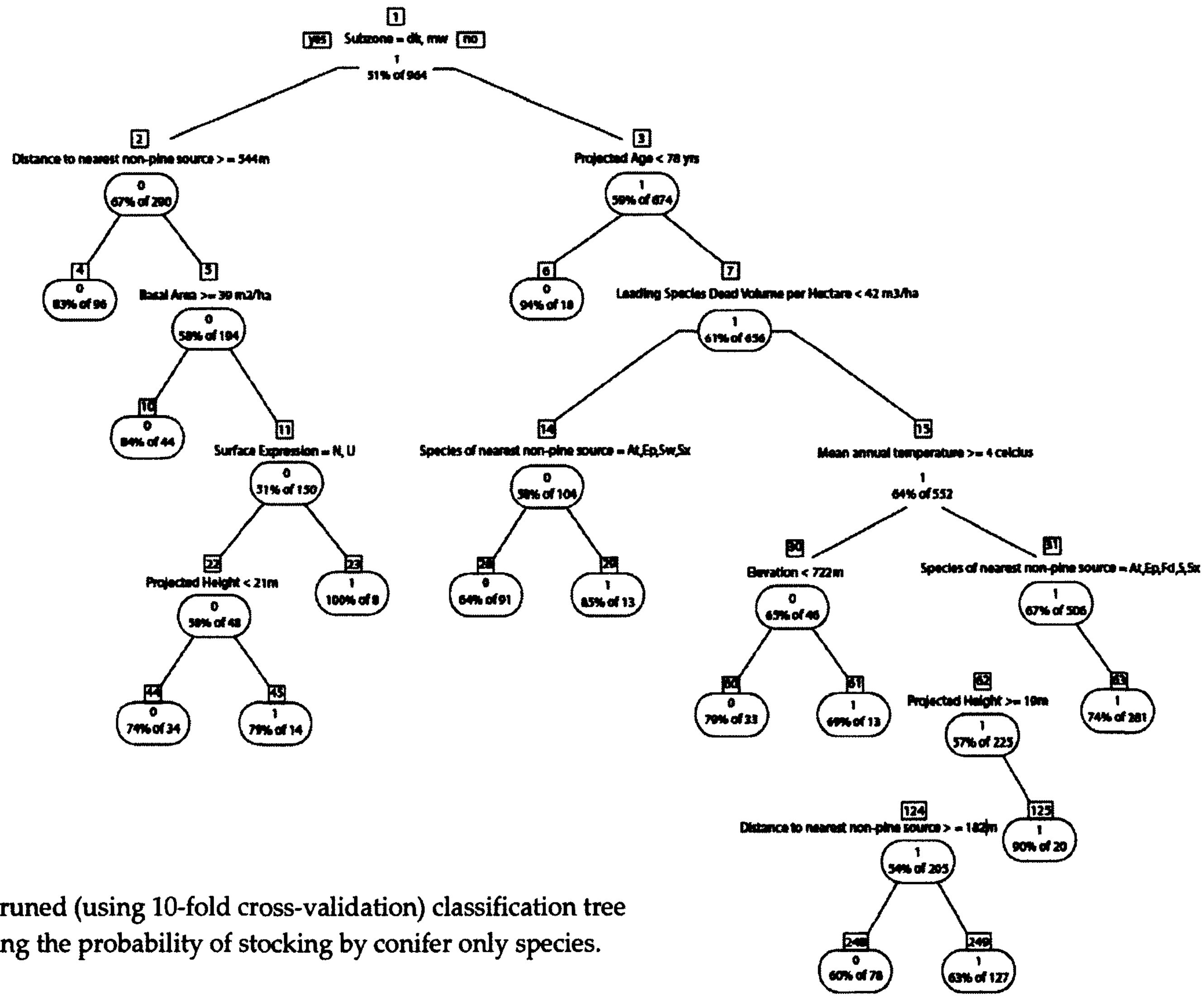



Figure 17. Final pruned (using 10-fold cross-validation) classification tree model for predicting the probability of stocking by conifer only species.

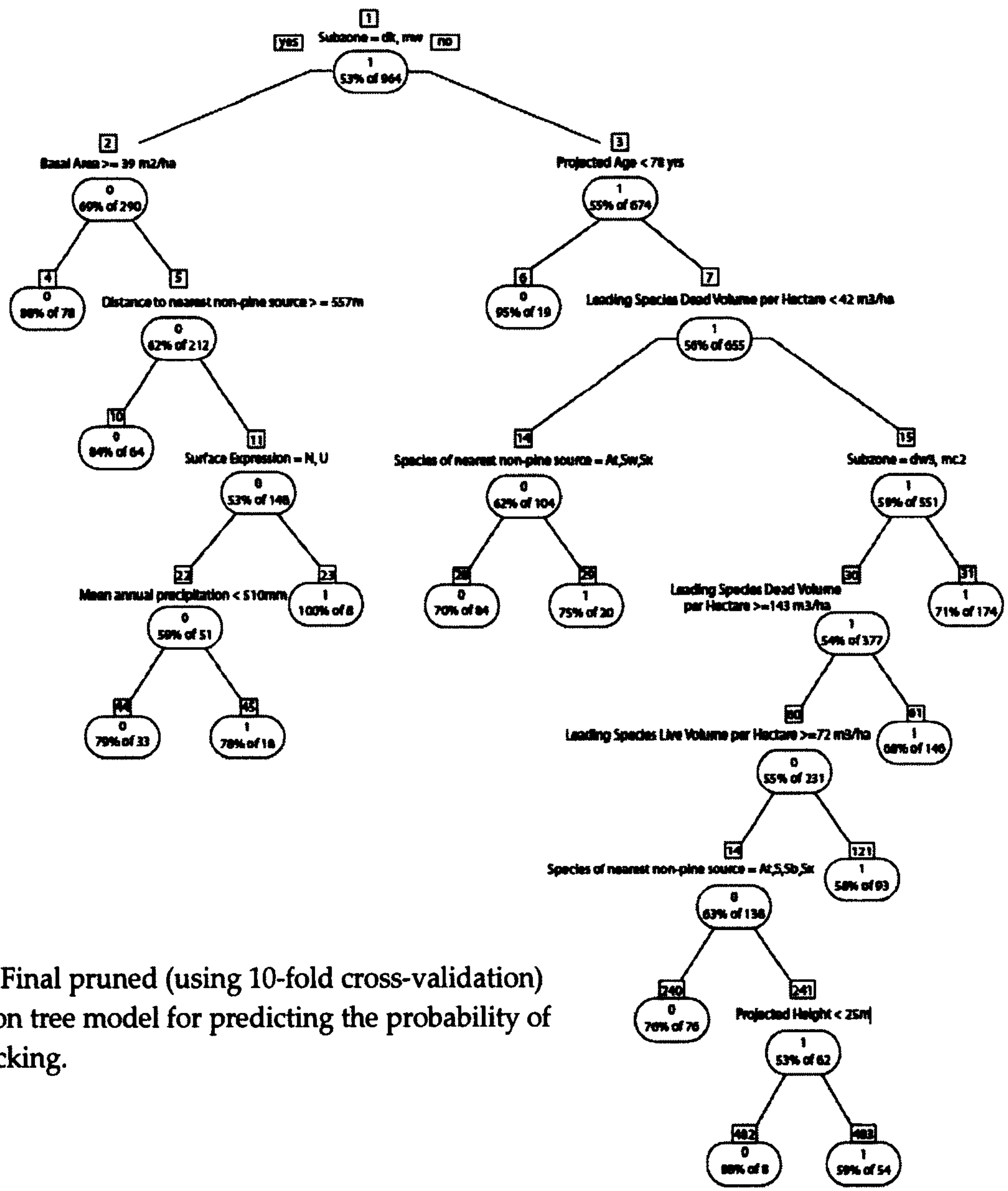



Figure 18. Final pruned (using 10-fold cross-validation) classification tree model for predicting the probability of MSSpa stocking.

value in the tree is false, until a terminal node is reached. The number of root-to-terminal nodes range from a minimum of two splits to a maximum of 14 splits.

The key to building the classification tree is the identification of the important variables used, as they are the only variables used in the resultant rules. The order of appearance in the classification tree does not necessarily indicate the strength or overall importance of a particular variable. What is more important is the variable's overall contribution to the model (i.e., is the model worse without it?)

The important variables for predicting the probability of stocking for all three stocking groups remained consistent. Even though the classification model had access to an input of 54 equally weighted potential predictors, Table 11 details the important variables (and therefore significant ecological factors) selected for each classification tree.

Table 11. List of important variables/significant ecological factors involved in the final model predicting the probability of stocking. MSSpa is the Minimum stocking standard (various stems/ha) of preferred and acceptable species, particular to each site series in each BEC unit.

| Variable/Ecological Factor | Classification Tree | | |
|---|---|---|---|
| | all trees | conifers | MSSpa |
| Biogeoclimatic Unit | ✓ | ✓ | ✓ |
| Projected Height for Leading Species | ✓ | ✓ | ✓ |
| Leading Species Dead Volume/ha at 12.5 cm | ✓ | ✓ | ✓ |
| Elevation | ✓ | ✓ | |
| Basal area | ✓ | ✓ | ✓ |
| Projected Age for Leading Species | ✓ | ✓ | ✓ |
| Species Composition of Nearest Non-pine Seed Source | ✓ | ✓ | ✓ |
| Distance (m) to Nearest Non-pine Seed Source (SW direction) | ✓ | ✓ | ✓ |
| Mean Annual Temperature | ✓ | ✓ | |
| Surface Expression | ✓ | ✓ | ✓ |
| Mean Annual Precipitation | | | ✓ |
| Leading Species Live Volume/ha at 12.5 cm | | | ✓ |

The variable importance for each stocking group is based on the mean decrease in accuracy. The higher the mean decrease in accuracy, the more important the variable is for the overall predictive accuracy of the model (Table 12).

The classification tree branches are translated into textual paths that can be interpreted as a collection of rules (Tables 13-15). They are listed in the order of their strength, i.e., highest probability in predicting the target variable (stocked or not stocked). For example: there are 11 textual rules for the stocking group all tree species (listed in Table 13). The two strongest and highest probability rules are:

Predictive Rule (Rank 1) probability = 1.00
biogeoclimatic unit = dk, or mw
distance to the nearest non-pine seed source <557 m
stand basal area <39 $m^2$/ha
surface expression = M (rolling), or P (undulating)

Following the tree in Figure 16, the above predictive rule terminates at node 23. The probability of 1.00 relates to the node text 1:100% of 8. The 1 indicates that the rule results in the prediction of a stocked cell. The 100% of 8 indicate that 8 of the 964 plots remain in the predictive solution at this point of the tree. And of the remaining 8 plots, all 8 plots are both actually stocked and predicted to be stocked, resulting in a 100% prediction or 1.00 probability. This set of rules has the highest confidence for predictive power on test plots. The rule set with the second strongest predictive power terminates at node 29 in Figure 16. The following path through the classification tree (represented by textual rules below) indicates that 13 of the

Table 12. Classification tree variable importance (in order of importance) for each stocking group. Order of importance is determined by the mean decrease accuracy value; a higher mean decrease accuracy values indicate a more important variable for prediction.

| all tree species stocking group | | |
|---|---|---|
| | Variable Importance | MeanDecreaseAcccuracy |
| 1 | Biogeoclimatic Unit | 1.01 |
| 2 | Distance (m) to Nearest Non-pine Seed Source (SW direction) | 0.93 |
| 3 | Projected Age for Leading Species | 0.93 |
| 4 | Leading Species Dead Volume per Hectare at 12.5 cm | 0.91 |
| 5 | Basal area | 0.86 |
| 6 | Projected Height for Leading Species | 0.86 |
| 7 | Mean Annual Temperature | 0.85 |
| 8 | Species Composition of Nearest Non-pine Seed Source | 0.83 |
| 9 | Elevation | 0.81 |
| 10 | Surface Expression | 0.44 |

| conifer only stocking group | | |
|---|---|---|
| | Variable Importance | MeanDecreaseAcccuracy |
| 1 | Biogeoclimatic Unit | 1.02 |
| 2 | Distance (m) to Nearest Non-pine Seed Source (SW direction) | 0.90 |
| 3 | Leading Species Dead Volume per Hectare at 12.5 cm | 0.90 |
| 4 | Projected Age for Leading Species | 0.89 |
| 5 | Basal area | 0.88 |
| 6 | Projected Height for Leading Species | 0.87 |
| 7 | Species Composition of Nearest Non-pine Seed Source | 0.87 |
| 8 | Mean Annual Temperature | 0.86 |
| 9 | Elevation | 0.84 |
| 10 | Surface Expression | 0.54 |

| MSSpa stocking group | | |
|---|---|---|
| | Variable Importance | MeanDecreaseAcccuracy |
| 1 | Biogeoclimatic Unit | 1.04 |
| 2 | Leading Species Dead Volume per Hectare at 12.5 cm | 1.01 |
| 3 | Projected Age for Leading Species | 0.96 |
| 4 | Basal area | 0.95 |
| 5 | Distance (m) to Nearest Non-pine Seed Source (SW direction) | 0.92 |
| 6 | Projected Height for Leading Species | 0.92 |
| 7 | Leading Species Live Volume per Hectare at 12.5 cm | 0.91 |
| 8 | Mean Annual Precipitation | 0.88 |
| 9 | Species Composition of Nearest Non-pine Seed Source | 0.88 |
| 10 | Surface Expression | 0.59 |

Table 13. Translation of classification tree model for predicting the probability of stocking in all tree species 600 stems/ha stocking threshold into textual rules.

| 600 stems/ha (all trees) | | | |
|---|---|---|---|
| | stocked | | not stocked |
| 1 | allsp600=1 prob=1.00<br>SUBZVAR=dk,mw<br>DISTANCEnear< 557.4<br>BASAL_AREA< 38.78<br>SURFACE_EXPRESSION=M,P | 6 | allsp600=0 prob=0.37<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125< 42.33<br>SPECIES_CD_1near=AT,EP,SW,SX |
| 2 | allsp600=1 prob=0.85<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125< 42.33<br>SPECIES_CD_1near=S,SB | 7 | allsp600=0 prob=0.31<br>SUBZVAR=dk,mw<br>DISTANCEnear< 557.4<br>BASAL_AREA< 38.78<br>SURFACE_EXPRESSION=N,U<br>PROJ_HEIGHT_1< 21.35 |
| 3 | allsp600=1 prob=0.75<br>SUBZVAR=dk,mw<br>DISTANCEnear< 557.4<br>BASAL_AREA< 38.78<br>SURFACE_EXPRESSION=N,U<br>PROJ_HEIGHT_1>=21.35 | 8 | allsp600=0 prob=0.21<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>MAT>=4.05<br>Elevation< 722.5 |
| 4 | allsp600=1 prob=0.69<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>MAT>=4.05<br>Elevation>=722.5 | 9 | allsp600=0 prob=0.20<br>SUBZVAR=dk,mw<br>DISTANCEnear< 557.4<br>BASAL_AREA>=38.78 |
| 5 | allsp600=1 prob=0.68<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>MAT< 4.05 | 10 | allsp600=0 prob=0.17<br>SUBZVAR=dk,mw<br>DISTANCEnear>=557.4 |
| | | 11 | allsp600=0 prob=0.06<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1< 77.5 |

Table 14. Translation of classification tree model for predicting the probability of stocking in conifer-only 600 stems/ha stocking threshold into textual rules.

| 600 stems/ha (conifers only) | |
|---|---|
| stocked | not stocked |
| 1 conifer600=1 prob=1.00<br>SUBZVAR=dk,mw<br>DISTANCEnear< 544.3<br>BASAL_AREA< 39.34<br>SURFACE_EXPRESSION=M,P | 8 conifer600=0 prob=0.40<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>MAT< 4.05<br>SPECIES_CD_1near=AT,EP,FD,FDI,S,SX<br>PROJ_HEIGHT_1>=19.25<br>DISTANCEnear>=182.4 |
| 2 conifer600=1 prob=0.90<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>MAT< 4.05<br>SPECIES_CD_1near=AT,EP,FD,FDI,S,SX<br>PROJ_HEIGHT_1< 19.25 | 9 conifer600=0 prob=0.36<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125< 42.33<br>SPECIES_CD_1near=AT,EP,SW,SX |
| 3 conifer600=1 prob=0.85<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125< 42.33<br>SPECIES_CD_1near=S,SB | 10 conifer600=0 prob=0.26<br>SUBZVAR=dk,mw<br>DISTANCEnear< 544.3<br>BASAL_AREA< 39.34<br>SURFACE_EXPRESSION=N,U<br>PROJ_HEIGHT_1< 21.35 |
| 4 conifer600=1 prob=0.79<br>SUBZVAR=dk,mw<br>DISTANCEnear< 544.3<br>BASAL_AREA< 39.34<br>SURFACE_EXPRESSION=N,U<br>PROJ_HEIGHT_1>=21.35 | 11 conifer600=0 prob=0.21<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>MAT>=4.05<br>Elevation< 722.5 |
| 5 conifer600=1 prob=0.74<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>MAT< 4.05<br>SPECIES_CD_1near=BL,SB,SW,SXW | 12 conifer600=0 prob=0.17<br>SUBZVAR=dk,mw<br>DISTANCEnear>=544.3 |
| 6 conifer600=1 prob=0.69<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>MAT>=4.05<br>Elevation>=722.5 | 13 conifer600=0 prob=0.16<br>SUBZVAR=dk,mw<br>DISTANCEnear< 544.3<br>BASAL_AREA>=39.34 |
| 7 conifer600=1 prob=0.63<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=77.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>MAT< 4.05<br>SPECIES_CD_1near=AT,EP,FD,FDI,S,SX<br>PROJ_HEIGHT_1>=19.25<br>DISTANCEnear< 182.4 | 14 conifer600=0 prob=0.06<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1< 77.5 |

Table 15. Translation of classification tree model for predicting the probability of stocking in all minimum stocking standards stocking threshold into textual rules.

| MSSpa (preferred and acceptable) | | | |
|---|---|---|---|
| | stocked | | not stocked |
| 1 | MSSpa=1 prob=1.00<br>SUBZVAR=dk,mw<br>BASAL_AREA< 38.78<br>DISTANCEnear< 557.4<br>SURFACE_EXPRESSION=M,P | 8 | MSSpa=0 prob=0.30<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=79.5<br>DEAD_VOL_PER_HA_SPP1_125< 42.33<br>SPECIES_CD_1near=AT,SW,SX |
| 2 | MSSpa=1 prob=0.78<br>SUBZVAR=dk,mw<br>BASAL_AREA< 38.78<br>DISTANCEnear< 557.4<br>SURFACE_EXPRESSION=N,U<br>MAP>=510 | 9 | MSSpa=0 prob=0.24<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=79.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>SUBZVAR=dw3,mc2<br>DEAD_VOL_PER_HA_SPP1_125>=142.8<br>LIVE_VOL_PER_HA_SPP1_125>=72.22<br>SPECIES_CD_1near=AT,S,SB,SX |
| 3 | MSSpa=1 prob=0.75<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=79.5<br>DEAD_VOL_PER_HA_SPP1_125< 42.33<br>SPECIES_CD_1near=EP,S,SB | 10 | MSSpa=0 prob=0.21<br>SUBZVAR=dk,mw<br>BASAL_AREA< 38.78<br>DISTANCEnear< 557.4<br>SURFACE_EXPRESSION=N,U<br>MAP< 510 |
| 4 | MSSpa=1 prob=0.71<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=79.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>SUBZVAR=dw2,mc3,mk1,wk1 | 11 | MSSpa=0 prob=0.16<br>SUBZVAR=dk,mw<br>BASAL_AREA< 38.78<br>DISTANCEnear>=557.4 |
| 5 | MSSpa=1 prob=0.68<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=79.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>SUBZVAR=dw3,mc2<br>DEAD_VOL_PER_HA_SPP1_125< 142.8 | 12 | MSSpa=0 prob=0.12<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=79.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>SUBZVAR=dw3,mc2<br>DEAD_VOL_PER_HA_SPP1_125>=142.8<br>LIVE_VOL_PER_HA_SPP1_125>=72.22<br>SPECIES_CD_1near=EP,FD,SW<br>PROJ_HEIGHT_1< 24.75 |
| 6 | MSSpa=1 prob=0.59<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=79.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>SUBZVAR=dw3,mc2<br>DEAD_VOL_PER_HA_SPP1_125>=142.8<br>LIVE_VOL_PER_HA_SPP1_125>=72.22<br>SPECIES_CD_1near=EP,FD,SW<br>PROJ_HEIGHT_1>=24.75 | 13 | MSSpa=0 prob=0.12<br>SUBZVAR=dk,mw<br>BASAL_AREA>=38.78 |
| 7 | MSSpa=1 prob=0.58<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1>=79.5<br>DEAD_VOL_PER_HA_SPP1_125>=42.33<br>SUBZVAR=dw3,mc2<br>DEAD_VOL_PER_HA_SPP1_125>=142.8<br>LIVE_VOL_PER_HA_SPP1_125< 72.22 | 14 | MSSpa=0 prob=0.05<br>SUBZVAR=dw2,dw3,mc2,mc3,mk1,wk1<br>PROJ_AGE_1< 79.5 |

964 plots remaining in the predictive solution at this point of the tree are correct with 85% prediction or 0.85 probability.

Predictive Rule (Rank 2) probability = 0.85
biogeoclimatic unit = dw2, dw3, mc2, mc3, mk1, or wk1
stand age >78 yrs
stand estimated dead volume <42 $m^3$/ha
species of the nearest non-pine seed source = Spruce

The ROC evaluations for all three stocking groups indicate that my predictive model is a useful application, as all the ROC scores are above 0.83 (0.5-0.7 are considered low accuracy, 0.7-0.9 are considered useful, and ROC values $\geq$ 0.9 indicate high accuracy) (Manel et al., 2001). The area under the ROC curve (AUC) represents the probability that the model will rank a randomly chosen positive instance (true positive) higher than a randomly chosen negative one (false positive). The AUC scores are 84%, 83%, and 85% for all tree species, conifers only, and MSSpa respectively (Figures 19, 20, and 21).

**Applying the Classification Tree in a GIS Model**

Maps illustrating the probability of stocking were generated for each of the three stocking groups per 1:250,000 map tile. A full series of maps, depicting the probability of stocking themed by five probability classes, for the 1:250,000 map tiles covering the study area are presented in Appendix 2.

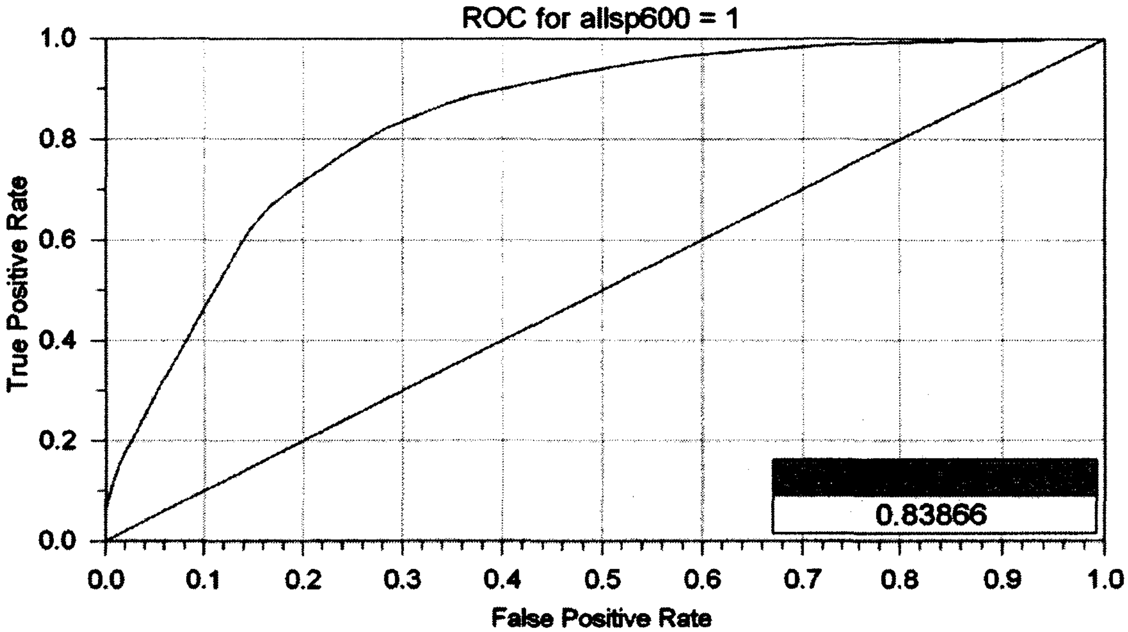



Figure 19. ROC curve chart illustrating the accuracy of the classification tree model (all tree species 600 stems/ha threshold) through area under curve (AUC) statistic.

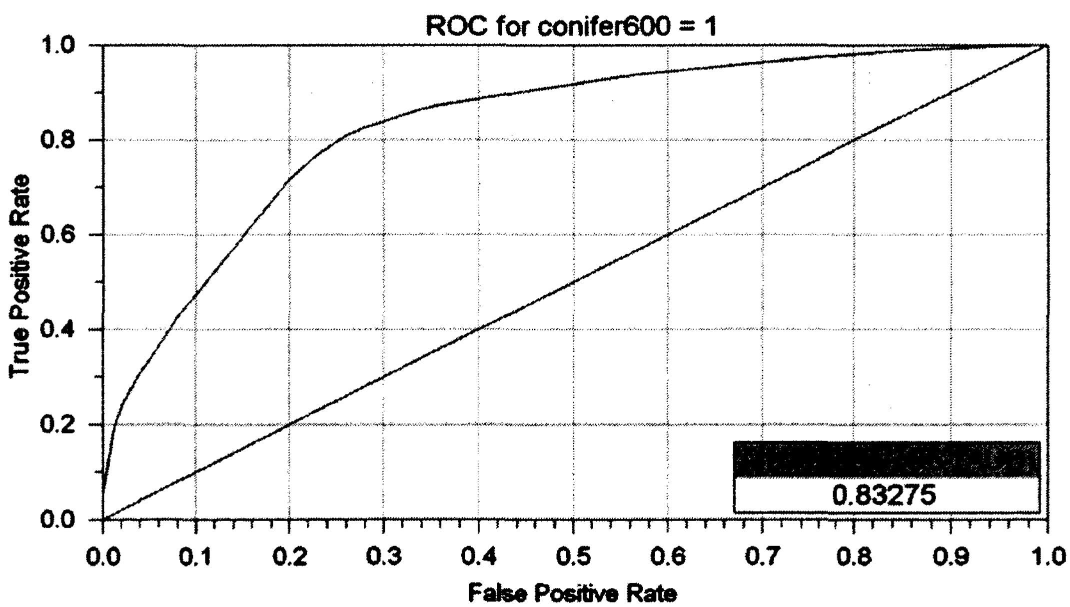



Figure 20. ROC curve chart illustrating the accuracy of the classification tree model (conifer-only threshold) through area under curve (AUC) statistic.

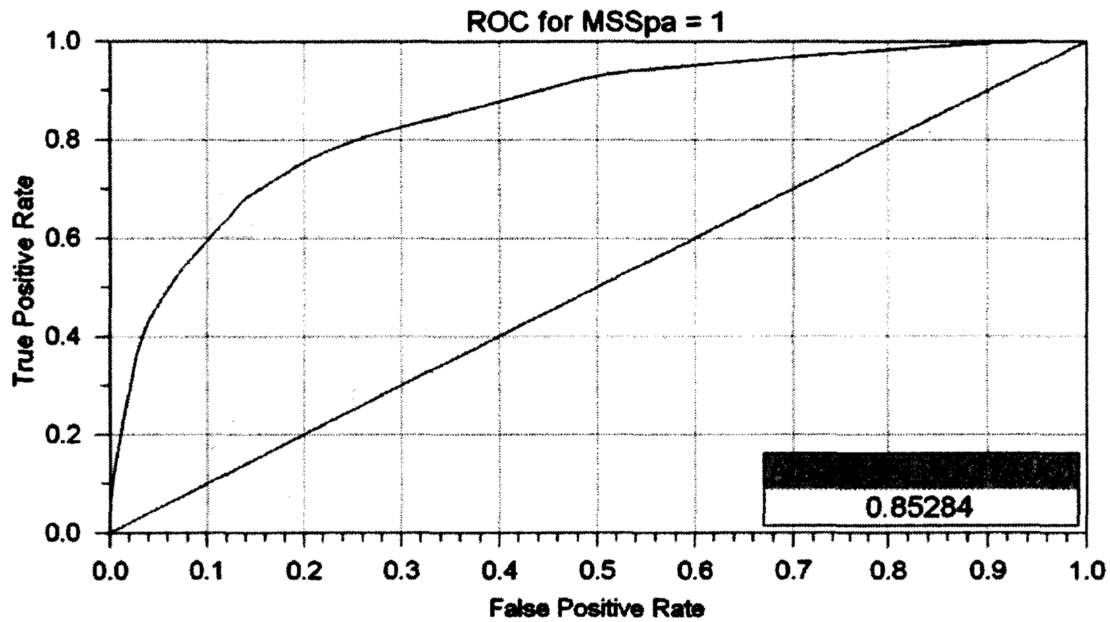


Figure 21. ROC curve chart illustrating the accuracy of the classification tree model (minimum stocking standard threshold) through area under curve (AUC) statistic.

The probability of stocking tables indicate that approximately 63% of the study area (when measured by a stocking threshold of 600 stems/ha for all tree species) is likely to very likely stocked; approximately 60% of the study area (when stocking is measured by a 600 stems/ha threshold for conifers only) is likely to very likely stocked; and approximately 44% of the study area (when stocking is measured by MSSpa) is likely to very likely stocked (Tables 16, 17, and 18). Translated into area, this means that the study area (comprised of 1.4 million ha of mature pine stands) can be expected to have up to 616,000 ha of stocked with seedlings and saplings (measured by MSSpa), 833,000 ha of stocked with seedlings and saplings (measured by 600 stems/ha all tree species), or 878,000 ha of stocked with seedlings and saplings (measured by600 stems/ha conifers only).

Table 16. Probability (as a percentage) of being stocked for each BEC unit derived from the classification tree rule set for conifer only 600 stems/ha stocking threshold.

| BEC unit | Mature Pine (ha) | Probability of Being Stocked (as percentage) 0-20 Very Likely Not Stocked | 20.1-40 Likely Not Stocked | 40.1-60 As Likely Stocked As Not | 60.1-80 Likely Stocked | 80.1-100 Very Likely Stocked |
|---|---|---|---|---|---|---|
| SBSdk all trees | | 40 | 16 | 0 | 5 | 39 |
| SBSdk conifers | 282539 | 40 | 18 | 0 | 5 | 37 |
| SBSdk MSSpa | | 46 | 14 | 6 | 21 | 12 |
| SBSdw2 all trees | | 18 | 15 | 7 | 36 | 23 |
| SBSdw2 conifers | 80876 | 17 | 13 | 0 | 49 | 21 |
| SBSdw2 MSSpa | | 36 | 13 | 10 | 21 | 20 |
| SBSdw3 all trees | | 15 | 15 | 9 | 32 | 29 |
| SBSdw3 conifers | 291425 | 15 | 13 | 0 | 47 | 25 |
| SBSdw3 MSSpa | | 32 | 17 | 7 | 22 | 22 |
| SBSmc2 all trees | | 16 | 15 | 11 | 27 | 31 |
| SBSmc2 conifers | 479996 | 26 | 11 | 0 | 37 | 27 |
| SBSmc2 MSSpa | | 34 | 10 | 9 | 21 | 26 |
| SBSmc3 all trees | | 15 | 7 | 9 | 31 | 38 |
| SBSmc3 conifers | 77284 | 41 | 5 | 0 | 26 | 28 |
| SBSmc3 MSSpa | | 39 | 11 | 12 | 16 | 22 |
| SBSmk1 all trees | | 14 | 6 | 5 | 37 | 38 |
| SBSmk1 conifers | 259730 | 14 | 11 | 0 | 37 | 38 |
| SBSmk1 MSSpa | | 31 | 13 | 3 | 15 | 38 |
| SBSmw all trees | | 71 | 13 | 0 | 3 | 13 |
| SBSmw conifers | 5283 | 70 | 17 | 0 | 3 | 9 |
| SBSmw MSSpa | | 92 | 0 | 0 | 6 | 2 |
| SBSwk1 all trees | | 18 | 3 | 8 | 28 | 44 |
| SBSwk1 conifers | 22014 | 14 | 12 | 0 | 36 | 38 |
| SBSwk1 MSSpa | | 23 | 29 | 1 | 29 | 17 |
| Study Area all trees | | 20 | 13 | 7 | 26 | 33 |
| Study Area conifers | 1499147 | 25 | 12 | 0 | 33 | 30 |
| Study Area MSSpa | | 36 | 13 | 7 | 20 | 24 |

Table 17. Probability (as a percentage) of being stocked for each 1:250,000 NTS maptile derived from the classification tree rule set for conifer only 600 stems/ha stocking threshold.

| NTS 1:250,000 Maptile | Mature Pine (ha) | Probability of Being Stocked (as percentage) 0-20 Very Likely Not Stocked | 20.1-40 Likely Not Stocked | 40.1-60 As Likely Stocked As Not | 60.1-80 Likely Stocked | 80.1-100 Very Likely Stocked |
|---|---|---|---|---|---|---|
| 93E all trees | | 20 | 16 | 11 | 27 | 27 |
| 93E conifers | 161479 | 35 | 11 | 0 | 34 | 20 |
| 93E MSSpa | | 33 | 11 | 9 | 12 | 36 |
| 93F all trees | | 29 | 14 | 4 | 19 | 34 |
| 93F conifers | 360025 | 38 | 13 | 0 | 16 | 33 |
| 93F MSSpa | | 41 | 9 | 10 | 24 | 16 |
| 93G all trees | | 21 | 14 | 7 | 33 | 25 |
| 93G conifers | 220916 | 16 | 15 | 0 | 48 | 20 |
| 93G MSSpa | | 35 | 13 | 8 | 24 | 20 |
| 93J all trees | | 13 | 9 | 6 | 35 | 36 |
| 93J conifers | 278669 | 14 | 12 | 0 | 38 | 36 |
| 93J MSSpa | | 29 | 18 | 4 | 13 | 36 |
| 93K all trees | | 16 | 14 | 9 | 26 | 35 |
| 93K conifers | 305215 | 20 | 10 | 0 | 36 | 34 |
| 93K MSSpa | | 36 | 11 | 6 | 26 | 20 |
| 93L all trees | | 20 | 13 | 8 | 17 | 42 |
| 93L conifers | 172843 | 23 | 15 | 0 | 34 | 29 |
| 93L MSSpa | | 39 | 16 | 5 | 16 | 24 |

Table 18. Probability (as a percentage) of being stocked for each forest district derived from the classification tree rule set for conifer only 600 stems/ha stocking threshold.

| Forest District | Mature Pine (ha) | Probability of being stocked (as percentage) 0-20 very likely not stocked | 20.1-40 likely not stocked | 40.1-60 as likely stocked as not | 60.1-80 likely stocked | 80.1-100 very likely stocked |
|---|---|---|---|---|---|---|
| Skeena Stikine all trees | | 34 | 3 | 14 | 13 | 37 |
| Skeena Stikine conifers | 21505 | 23 | 6 | 0 | 33 | 38 |
| Skeena Stikine MSSpa | | 28 | 18 | 3 | 34 | 18 |
| Mackenzie all trees | | 6 | 23 | 9 | 21 | 41 |
| Mackenzi conifers | 6769 | 9 | 11 | 0 | 48 | 32 |
| Mackenzi MSSpa | | 11 | 12 | 1 | 63 | 14 |
| Fort St. James all trees | | 13 | 5 | 6 | 38 | 38 |
| Fort St. James conifers | 219378 | 16 | 10 | 0 | 40 | 33 |
| Fort St. James MSSpa | | 40 | 8 | 4 | 20 | 29 |
| Nadina all trees | | 26 | 14 | 7 | 20 | 33 |
| Nadina conifers | 554084 | 33 | 13 | 0 | 27 | 27 |
| Nadina MSSpa | | 42 | 12 | 7 | 17 | 22 |
| Prince George all trees | | 19 | 8 | 6 | 36 | 30 |
| Prince George conifers | 285404 | 14 | 13 | 0 | 44 | 29 |
| Prince George MSSpa | | 31 | 16 | 3 | 17 | 33 |
| Vanderhoof all trees | | 17 | 20 | 7 | 21 | 34 |
| Vanderhoof conifers | 376145 | 26 | 13 | 0 | 28 | 33 |
| Vanderhoof MSSpa | | 30 | 13 | 12 | 27 | 18 |
| Quesnel all trees | | 10 | 19 | 18 | 16 | 37 |
| Quesnel conifers | 22746 | 39 | 13 | 0 | 31 | 17 |
| Quesnel MSSpa | | 38 | 19 | 15 | 5 | 23 |
| North Island-Central Coast all trees | | 9 | 18 | 40 | 18 | 14 |
| North Island-Central Coast conifers | 13115 | 47 | 5 | 0 | 30 | 19 |
| North Island-Central Coast MSSpa | | 28 | 1 | 24 | 23 | 23 |

It is important to note that although study plots were located in the Skeena Stakine, Mackenzie, and North Island – Central Coast forest districts, only a fraction of these three forest districts intersected with the study area of interest (Figure 22). Caution should be used when drawing any conclusions involving the probability of stocking within a forest district context.

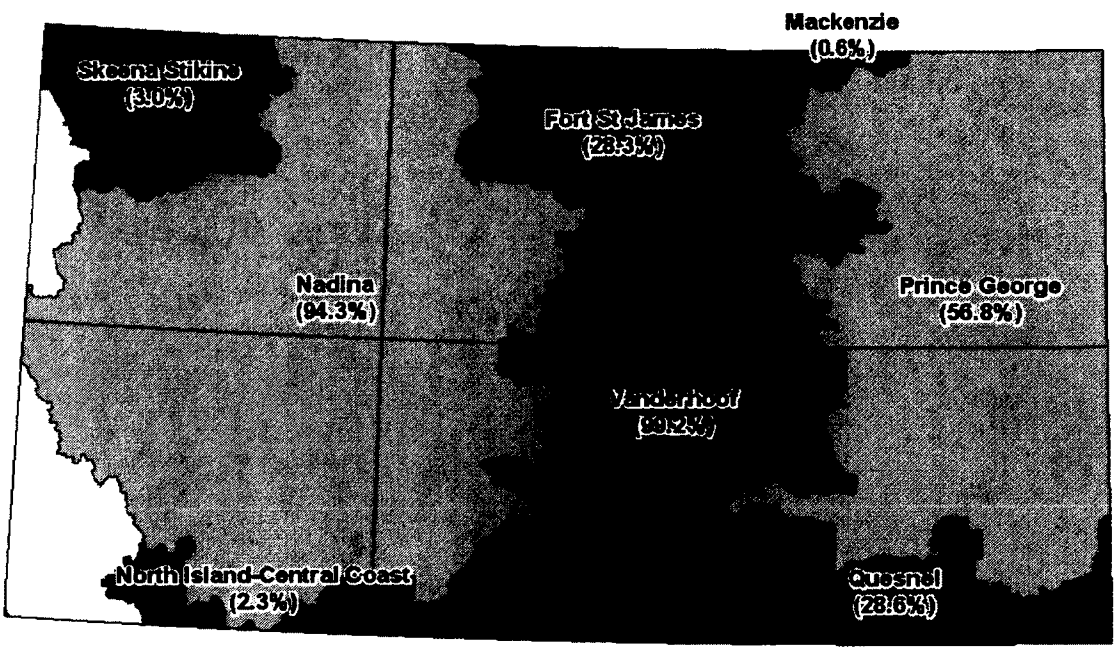



Figure 22. Map illustrating the proportion of forest districts that are within the study area (NTS 1:250,000 93E,F,G,J,K,L).

Figure 23 is an example of a 1:250,000 map predicting the probability of MSSpa stocking in mature pine-leading stands for map tile NTS 93G. The five colours in the legend correspond to a likelihood of stocking (red=0-20%, orange=20-40%, yellow=40-60%, light green=60-80%, and dark green=80-100%). This map and those in Appendix 2 also have a shaded relief in the background for elevation context, major lakes (blue polygons), and major roads (symbolized by black lines).

Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia

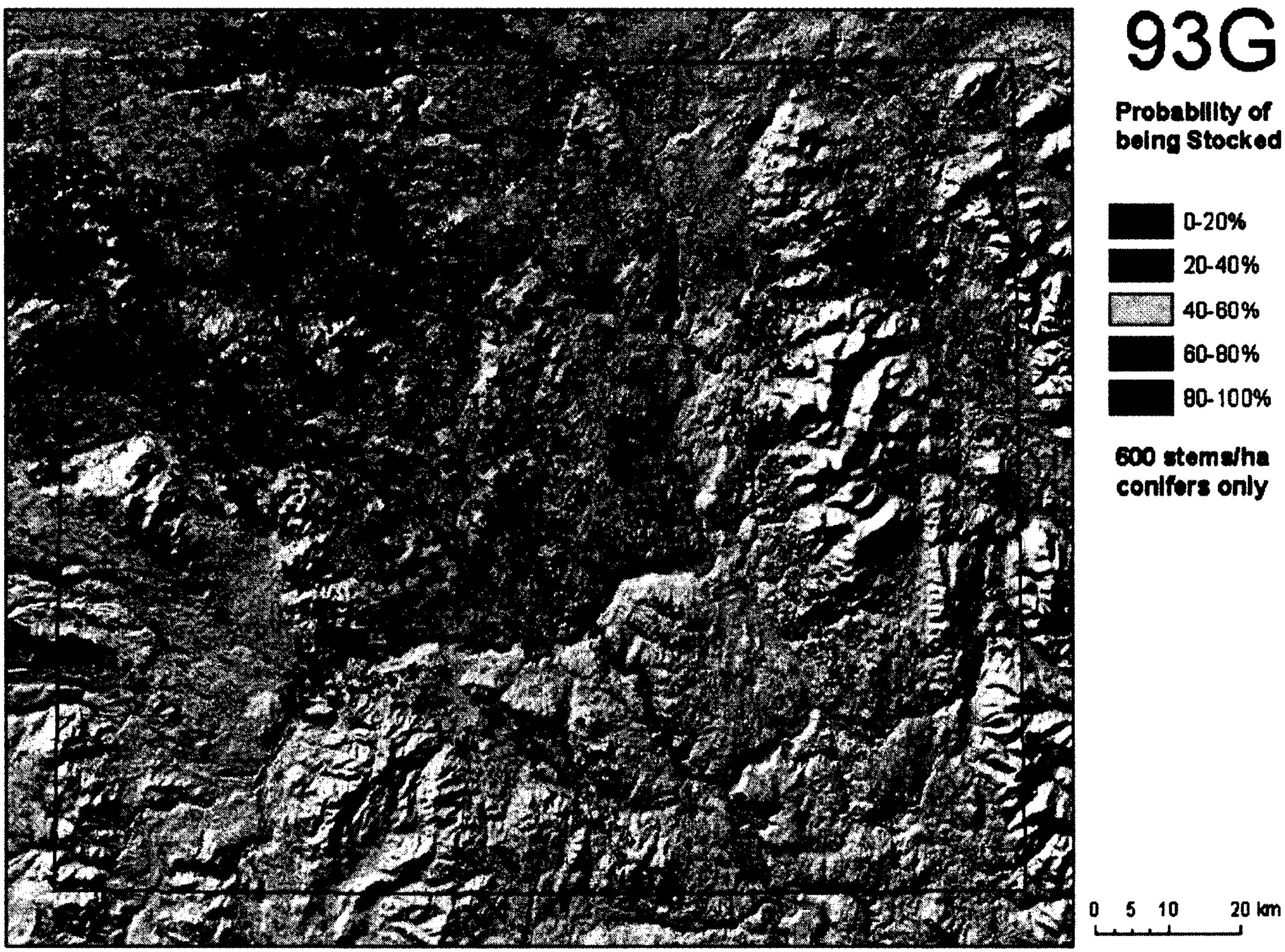



Figure 23. Colour themed map depicting probability of being stocked for conifer only 600 stems/ha stocking threshold within NTS 1:250,000 map tile 93G.

Once the classification tree results are in a geospatial environment, specific probability of stocking percentages can be displayed in isolation of the remaining probability classes and can be provided spatial context. By placing large areas of "very likely" and "very unlikely" stocking in a spatial context, the data can be used as a GIS layer for more sophisticated analysis regarding forest management. For example, Figure 24 shows the isolation of areas that have a >80% probability of stocking. These isolated areas can be easily identified as cells that are very likely to recover naturally without intervention. Conversely, large areas can also be identified as priorities for salvage or rehabilitation operations. Figure 25 shows that by isolating areas that have a ≤20% probability of stocking, areas that may require rehabilitation (including planting) can be targeted. The isolated areas can be combined with datasets such as proximity to mill sites to determine economic viability. Further, they may be targeted for rehabilitation, i.e., treatment to knock down and pile trees, grinding them for pellet fuels or other bioenergy (depending on markets) or burned, and then planted. The large contiguous areas classified as very likely stocked (green pixels in Figure 24) become an important GIS layer with regards to coordinating logging activity. Perhaps these stands could be allocated as no logging or rehabilitation zones. It is recommended that these stands be identified for natural (unaided) recovery. They may be important as a mid-term supply of timber, or for habitat value.

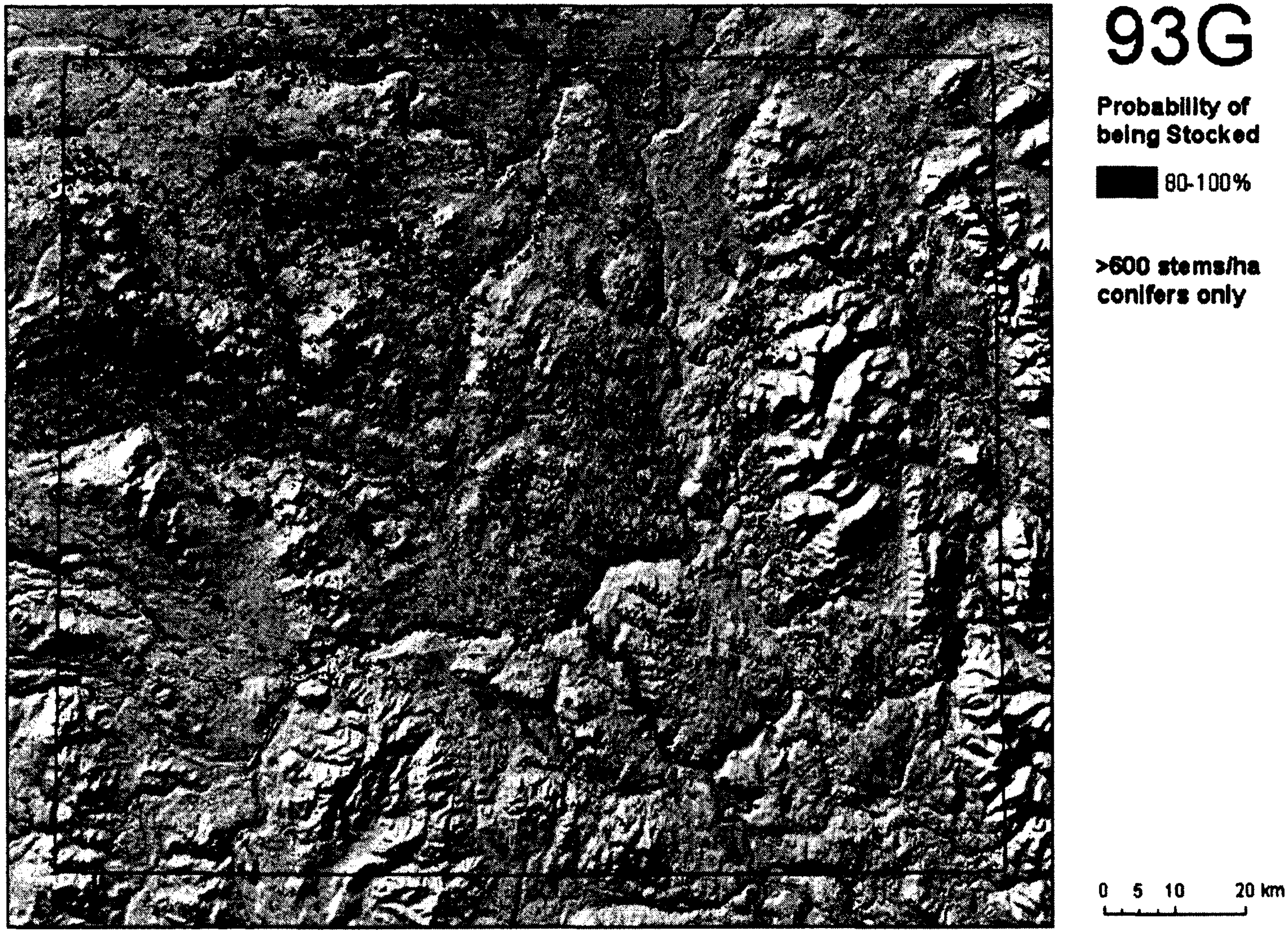



Figure 24. Spatial extent of probability of stocking for map tile NTS 93G. Green areas denote >80% probability of stocking (conifers-only).

Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia

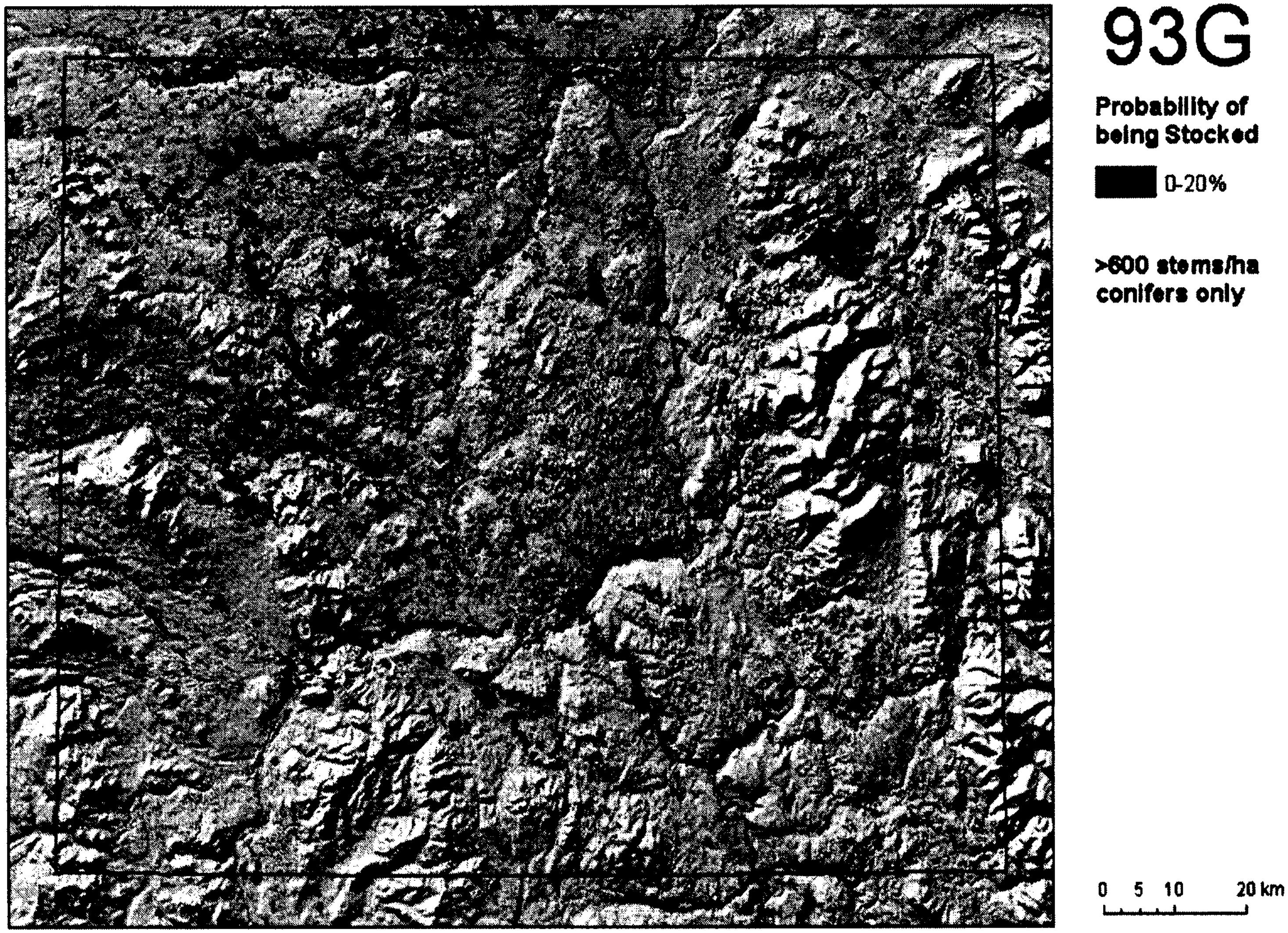



Figure 25. Spatial extent of probability for map tile NTS 93G. Red areas denote <20% probability of stocking (conifer-only).

## Chapter Four: Discussion and Recommendations

The main objective of my thesis was to develop a model predicting the distribution of advance regeneration using publicly available data. Many of the initial inputs to my model have been explored for their predictive strength in previous studies. The resultant model was applied within a study area (NTS 1:250,000 93E, F, G, J, K, L map tiles) and probability maps were constructed (Appendix 2). As per the Establishment to Free Growing Guidebook (BC Ministry of Forests, 2000), MSSpa criteria and thresholds vary with BEC unit and site series. These conclusions are consistent with several current advance regeneration publications such as Vyse et al. (2009) who found that more than half of all plots surveyed in lodgepole pine stands in the Kamloops Timber Supply Area exceeded a threshold of 600 stems/ha. This is echoed in a study by Nigh et al. (2008), which states that over half the stands sampled in the Montane Spruce zone of southern British Columbia had enough advance regeneration (>1000 stems/ha) to form new stands of adequate density. Another study reports 44% to 98% stands contained sufficient stems after MPB attack to be considered stocked (Hawkins et al., 2012). A survey of pre-harvest industrial records has indicated that the percentage of pine stands with a greater than 600 stems/ha are highest in the moist cool (mk) subzone followed by the moist cold (mc) subzone, and then by the dry warm (dw) and dry cool (dk) subzones of the SBS (Burton, 2006). Studies also suggest that SBSdk is unlikely to provide a significant contribution to the mid-term timber supply; in contrast, the SBSdw and SBSmc are thought to be contributors to both the mid- and long-term supplies (Hawkins and Rakochy, 2007). The overall distribution and probability of stocking predicted here aligns well with these conclusions. A closer examination of the resultant data (tabular and spatial) provides evidence to this claim. The SBSmk had the largest likely/very likely stocked probability (as high as 75% for the all trees stocking group) for all BEC units. The results also support the assessment by Hawkins and Rakochy (2007) that SBSdw and SBS mc will likely be the largest contributor to mid- and long-term timber supplies. In the conifer only stocking group, SBSdw has a 71.2% likely/very likely stocked probability (~208,000 ha) and SBSmc2 and SBSmc3 have a 63.3% (~304,000 ha) and 53.8% (~42,000 ha) likely/very likely stocked probability, respectively. The total area associated with these probabilities is ~ 554,000 ha of potential advanced regeneration ≥600 stems/ha or roughly one-third of the existing total ha of the study area.

A closer examination of the colour-themed maps reveals the utility of translating the tree/text based rules into a geospatial output. The maps present the probability of stocking in five coloured classes, and because the scale of the map is known, an ocular estimate of the area of contiguous areas or distance between proximal areas for any of the probability classes can be made. For example, large areas of likely/very likely stocked polygons that are interrupted by smaller areas very unlikely to be stocked (Figure 26) can be identified and subsequently visited and assessed using ground-based advance regeneration grid sampling.

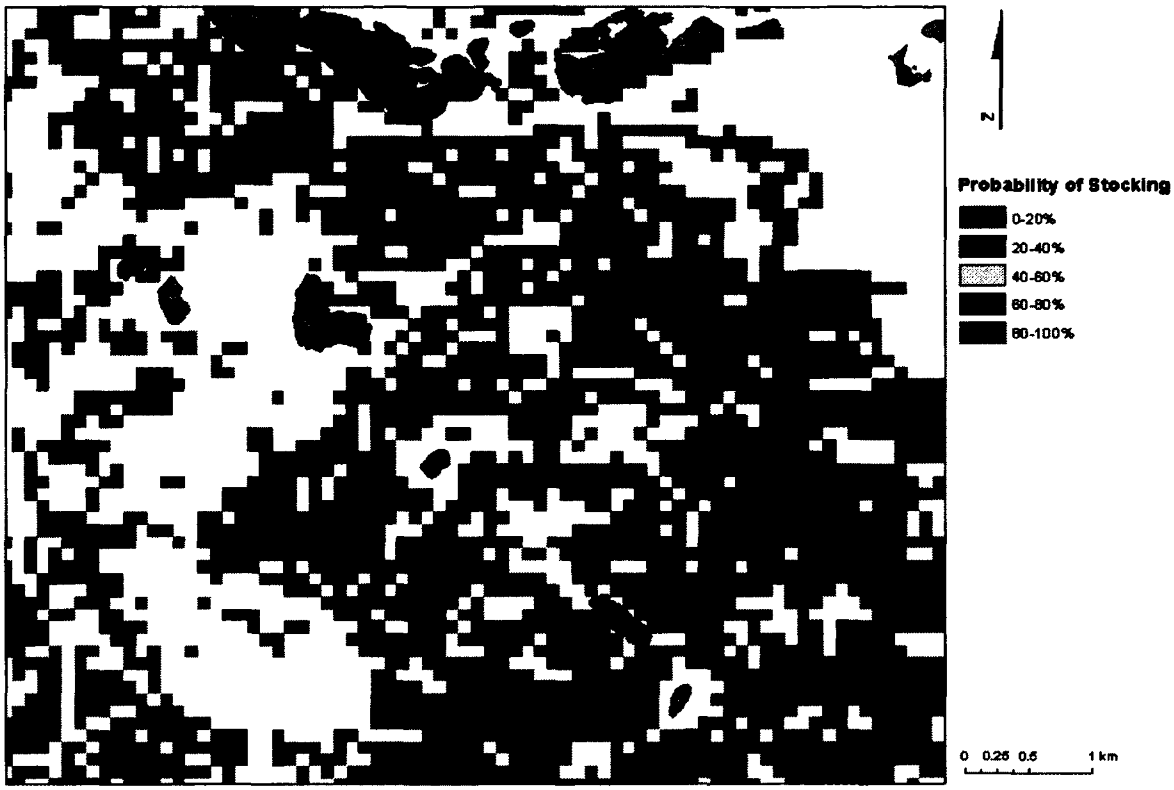



Figure 26. A portion of the mapped final model that illustrates the intersection of large very likely stocked areas (green pixels) with very unlikely stocked cells (red pixels).

The results generated from the classification tree model also support the preliminary results in Hawkins and Rakochy (2007), where it is reported that there was measurably less regeneration in the SBSdk and SBSdw2 than in the SBSdw3 and SBSmc. My classification tree model shows that the predicted probability of the MSSpa group being stocked in SBSdk (12%) is close to half of the probability in

SBSdw (25%) and SBSmc (22%). Vyse et al. (2009) also state that the mean density of stems increased with moisture and elevation, both critical elements in determining the biogeoclimatic subzone of a site. This generalization is further supported by my findings, as mean annual precipitation and BEC unit are defined as important variables in the resulting predictor model. The resultant model indicated that the following variables are key predictors in modelling stocking status: BEC unit, distance to the nearest non-pine seed source in a southwest direction, projected age of the leading species, basal area, leading species of dead volume per hectare at 12.5 cm, surface expression, species composition of the nearest non-pine seeds source, mean annual temperature, mean annual precipitation, projected height of the leading species, and elevation. The key variables selected through recursive partitioning are consistent with other study conclusions, specifically with regards to overstory height (Griesbauer and Green, 2006), overstory mortality (leading species dead volume; Lewis, 2011), basal area (Coates and Sachs, 2012; Nigh et al., 2008), distance to nearest seed sources (LePage et al., 2000; Kaufmann et al., 2008), and moisture (precipitation; Kayes and Tinker, 2012).

**Application to Forest Management**

A landscape-level planning tool is useful for several different aspects of forest planning and management. At the provincial level, this study's model can assist in the implementation of the Forests for Tomorrow Current Reforestation and Timber

Supply Mitigation Strategic Plan 2011-2015 (Ministry of Forests, Lands, and Natural Resource Operations, 2011) by providing critical data to meet several goals.

First and foremost, the resultant of my model (hardcopy maps and textual rule sets) can assist forest managers in the establishment of new and up-to-date assessments of potential timber supply. Tools such as this can play a part in helping estimate the new baseline that will be used for forecasting and planning. The need for current information is echoed in a June 8, 2012, memo from the Inventory Section, Forest Analysis and Inventory Branch: *Approach of the inventory program in 2012-13 to improve inventory information in MPB-affected management units* (Ministry of Forests, 2012). The working memo states that the provincial inventory program is placing a high priority on improving information related MPB-affected areas. The memo specifically states that the advance regeneration in the understory is critical to mid-term timber supply and that the assessment of stocking under these MPB-affected stands is important. The model developed here can play a large role in assisting inventory personnel "short-list" stands that may potentially have this critical mid-term timber supply. The scale of the output is not necessarily fine enough to generate reliable determinations of stand-level stocking, but it is highly useful for directing field validation programs. On a landscape level, the resultant stocking data can be used for more general land-use and resource management plans. By examining the data, areas of sporadic distribution, large anomalies, and "salt and pepper" occurrences can be identified (Figure 27).

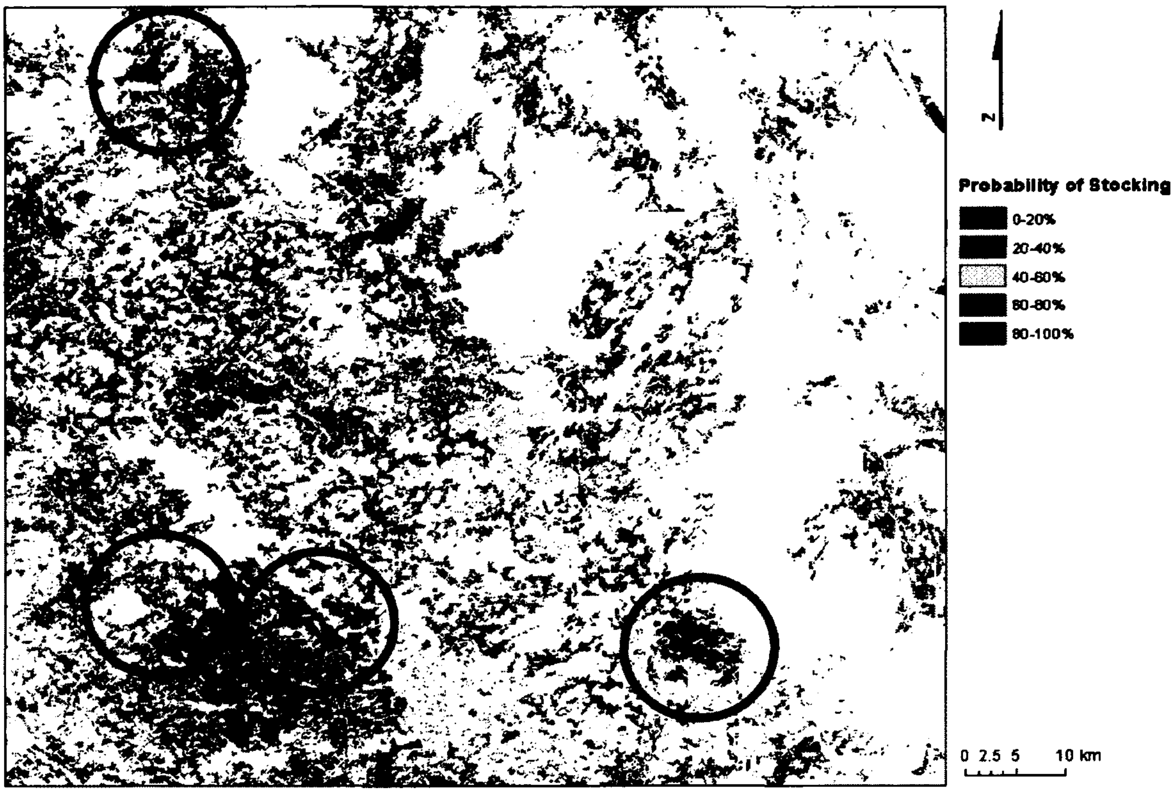



Figure 27. Landscape view of mapped model output. Blue circles indicate large anomalies that are easily identified at a broad scale.

The tabular results from exercising the classification tree have generated important information about how many hectares of likely/very likely to be stocked there are within a defined area. They are, however, limited in their explanatory power, as there is a distinct difference in knowing **how many** hectares of >80% (very likely to be stocked) probability of stocking and **where** these very likely to be stocked areas are on the landscape. Importing the tabular stocking information into a GIS allowed me to spatially locate how the five stocking classes are distributed across the landscape. The ability to spatially locate each of the probability of stocking classes enhances the decision making power of my decision model.

Consider the following example, where a forest manager wants to understand the advance regeneration attributes of a particular management area (Figure 28). Knowing the total hectares of the management area (65180 ha) and total hectares of mature pine within the management area (29815 ha), the forest manager is now in a position to estimate the number of hectares of mature pine that could possibly have advance regeneration. According to the areas reported, the amount of advance regeneration cannot exceed 45% of the management area (or ~ 29,000 ha) as there are only 29,000 ha of mature pine within the area. Assume further that a classification tree model has been used to calculate the probabilities of stocking as per the following: 1) 4858 ha are very unlikely to be stocked; 2) 3210 ha are unlikely to be stocked; 3) 0 ha are as equally likely to be stocked as not stocked; 4) 10280 ha are likely to be stocked; 5) 11467 ha are very likely to be stocked.

The detailed information regarding areas of stocking likelihood provides the forest manager with valuable information regarding the amount of potential

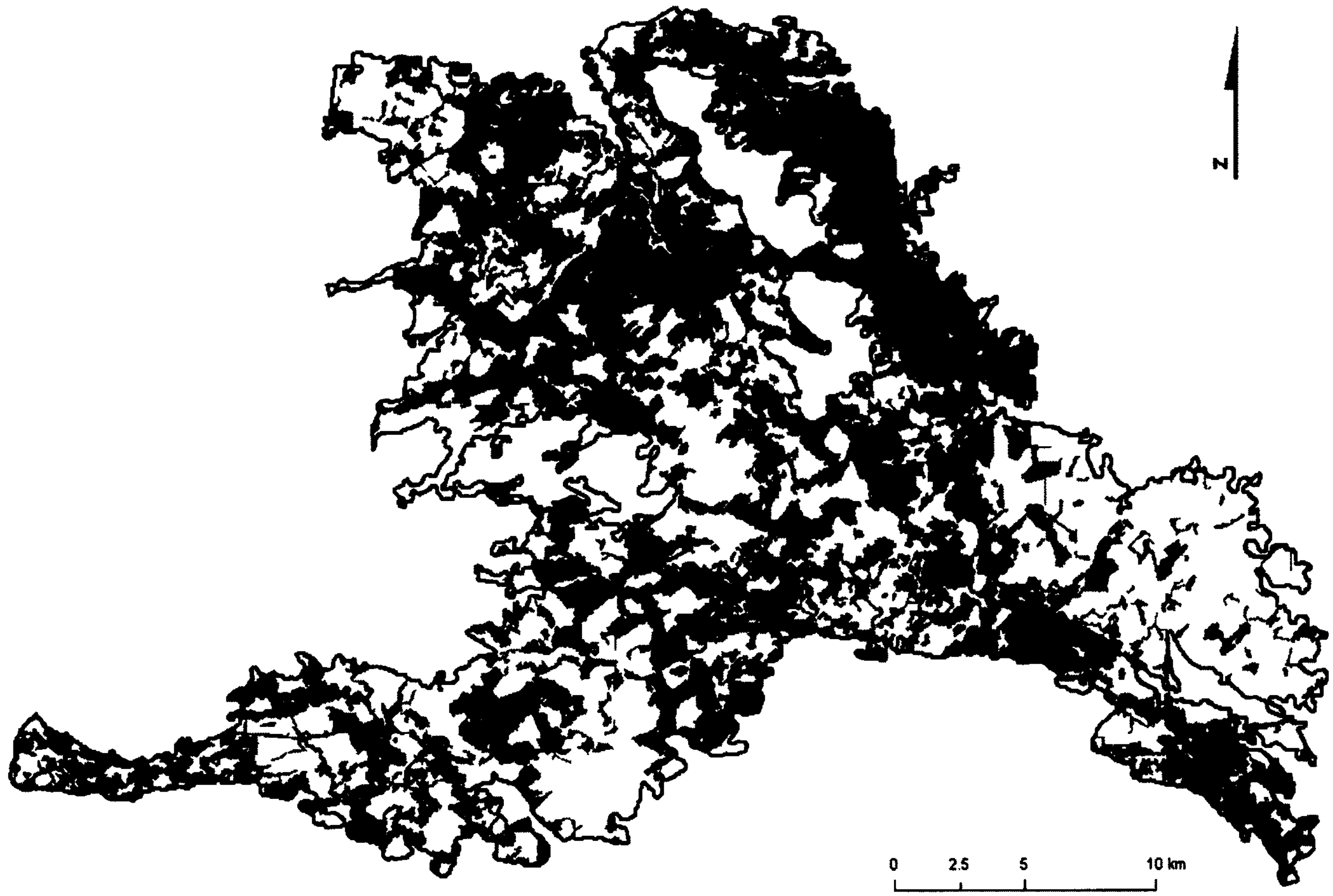



Figure 28. Example of a management area of interest in which a forest manager would like to calculate the amount of advance regeneration under mature pine forest stands by using a probability of stocking model. The grey polygons indicate mature pine forest stands.

advance regeneration in the management area. Without spatial context, however, the planner can only know the probability of stocking with regards to advance regeneration. An assumption regarding the spatial distribution of the likely or unlikely stocked areas cannot be made with the information in hand, i.e., the manager can only know how much stocking is in the management area but not if the stocking is restricted to one particular area or widely dispersed throughout the area. By providing a geospatial context to the probability information, a much clearer and complete picture is provided with regards to stocking (Figure 29). Patterns of large contiguous areas of very unlikely and likely stocking become evident. Large patches can be quickly identified and used to supplement landscape level plans. This information can be used as a landscape level planning tool for activities such as annual allowable cut (AAC) allocation, post-MPB management planning, and even identification of mid- and long-term timber supply stands suitable for future research projects. In a VRI newsletter, Martin (2012) identifies one of the goals for inventory information is to know more about stocking by gathering information regarding small trees under a dead overstory. The primary goal of the BC Government's "Forest for Tomorrow" program, namely to improve the mid- and long-term timber supply and establish resilient forest ecosystems, specifies the need to focus on restoration and reforestation through the identification of sites which have had the greatest negative impact on the mid- and

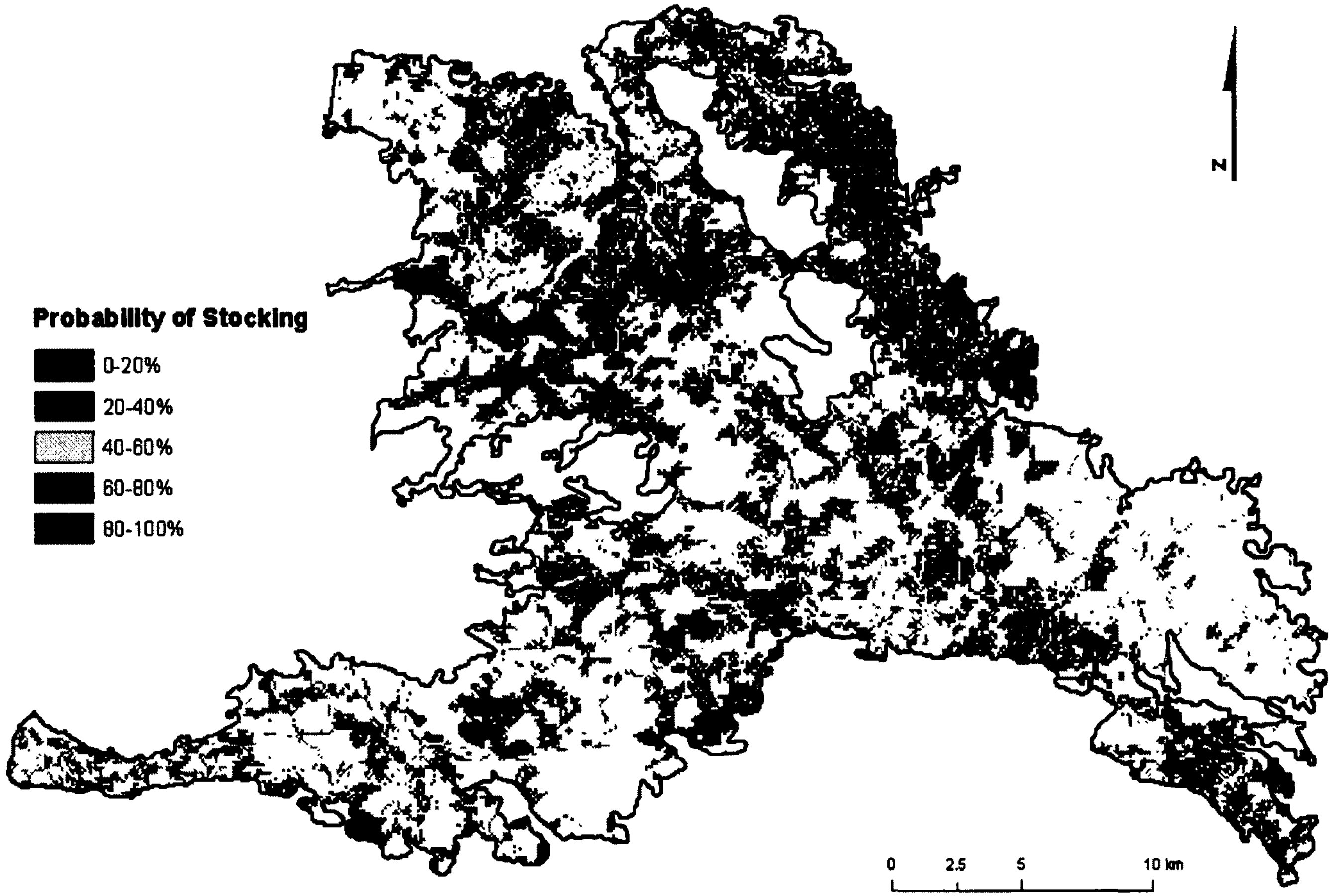



Figure 29. Example of a BEC area of interest in which a forest manager would like to calculate the probability of stocking. Probability of stocking is depicted in the five coloured classes. Spatial patterns of stocking are evident when viewing data in a geospatial context

long-term timber supplies. To effectively mitigate mid-term timber supply shortfalls, it is imperative to have an overarching plan that helps differentiate between stands that have sufficient advance regeneration (and therefore represent the mid-term supply) and stands that will not be contributing to the mid- to long-term timber supply. Stands that do not have sufficient advance regeneration can be identified through the use of a landscape level model and then targeted for harvesting and subsequent replanting. Alternatively, a working knowledge of the percent of mature pine without adequate advance regeneration may be useful for old forest retention in the quest for biodiversity conservation and preservation of critical wildlife habitat. Large areas of mature pine that are classified as very unlikely to be stocked pose no benefit to the mid- and long-term timber supply. These areas can be prioritized for logging efforts; effectively removing post-MPB stands from the inventory without removing any potential mid- and long-term timber represented as advance regeneration, especially large areas classified as very unlikely to be stocked (Figure 30). Roads (red lines with black borders in Figure 30) and existing cutblocks (thick black lines) are added to the maps to provide context for accessibility. The red pixels represent areas very unlikely to be stocked areas.

Goal three in "Forest for Tomorrow's" strategic plan is to develop and implement innovative approaches to reforesting forests damaged by catastrophic disturbance. One of the strategies to achieve this is to engage stakeholders in the development of a strategic and tactical plan. Whether the stakeholders are the Province, the licensees / stewards of the forest, non-traditional users of the forest, or the public, strategic plans must begin with the most current and accurate knowledge possible. The predictive model can help provide a region-wide overview of where the highest potential for mid- and long-term timber supply is located. A GIS-ready dataset of this scale can be used as an informative base layer for refining existing land use plans.

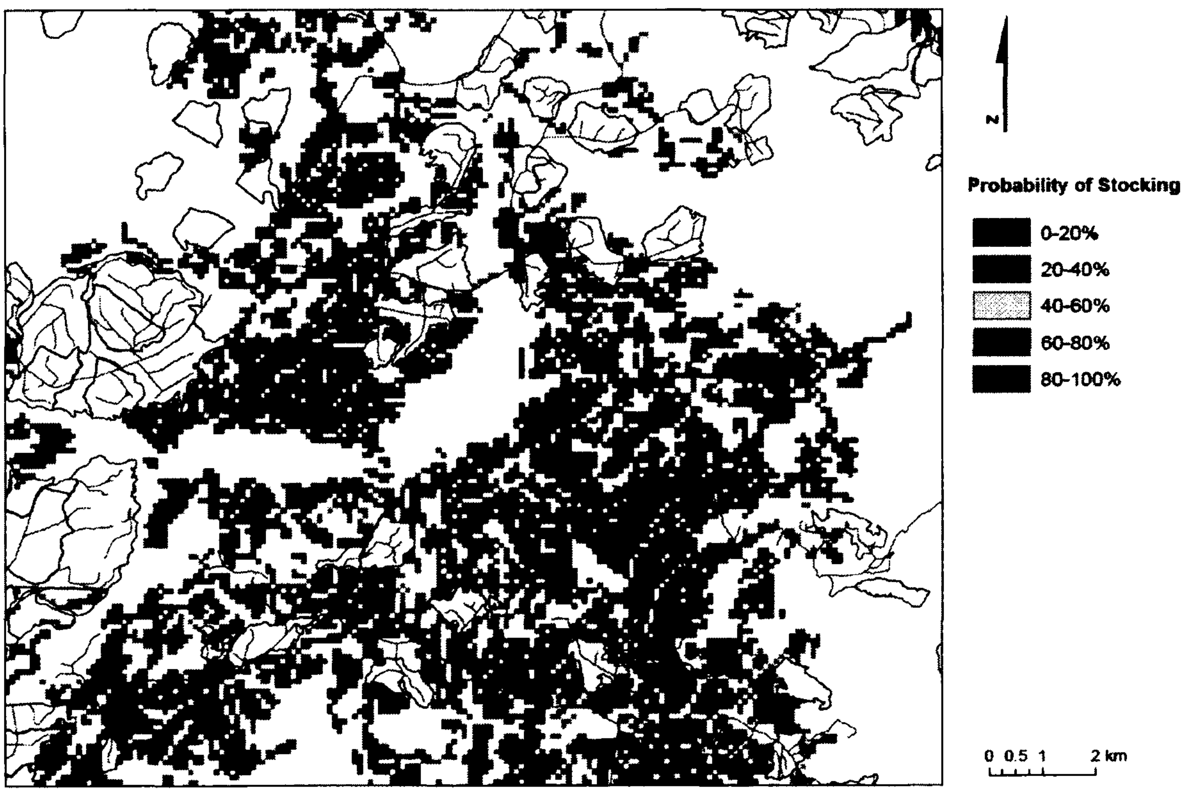



Figure 30. A portion of the final model map that illustrates the location of large, very unlikely to be stocked mature pine areas (red pixels). These areas can be targeted for large scale logging and silviculture programs.

Griesbauer and Green (2006) identify two potentially dangerous and expensive scenarios associated with the perceived need to tend advance regeneration: supplementing advance regeneration through planting of unstocked patches and the requirement for thinning of overstocked understories. Both present a potential danger in the form of falling MPB-killed snags and would require the removal of snags for safety reasons. For the planting scenario, Griesbauer and Green (2006) suggest restricting planting only to highly productive sites that have little advance regeneration. The predictive model outputs can be used as a GIS overlay to help forest managers find the intersection between highly productive sites (site index layer) and probability of stocking (generated by the classification tree model). This model may help to economically identify areas that could have regeneration augmented through planting. Conversely, stands predicted to be stocked on highly productive sites may also be identified and flagged for field inspection.

**Assessment of the Modelling Approach**

R (R Core Team, 2012), DTREG (Sherrod, 2006), and ArcGIS (ESRI, 2011) were used to establish an intuitive workflow to create fully attributed feature classes using publicly available digital data input. The resulting feature classes have correct projection and coordinate systems, are topologically clean, fully attributed, and are GIS-ready for input as a working dataset. The working parts of the model are extensive yet invisible to the user. The users cannot add more tools or explanatory variables to the model without building a new model, as this would jeopardize the integrity of the classification tree solution. To create a stocked / not stocked map for another study area or for pine stands that do not fit the mature pine criteria, the models would have to be recreated and re-run in both R and ArcGIS.

This working model is interpolative, in that its training data were sparsely distributed and it cannot make predictions with regards to advance regeneration stocking beyond the limits of data used to create the model. In the case of my model, the limitations lie more in the organization and definition of variables, rather than quantitative maximum and minimum values.

It would be misleading to claim that the model developed in this study is unique. Other software packages have been used to predict natural regeneration under mountain pine beetle attacked stands, most notably SORTIE and PROGNOSIS$^{BC}$ (Smith, 1990; Ferguson and Carlson, 1993; Ribbens, 1994; Sattler, 2009). The distinction, however, is not in the intended application of the model but rather the scale and ownership of the model input variables and the ultimate utility of the model output. Although those are mechanistic simulation models, and the work reported here resulted in an empirical statistical model, perhaps the key distinction is the emphasis on the ability to map predicted stocking across a broad region. My study input data are derived solely from publicly available geospatial data that can be downloaded from provincial online repositories or requested from BC government agencies such as British Columbia's Land and Resource Data Warehouse (LRDW) and its Integrated Land Management Bureau (ILMB). Most forest planners and managers with access to the internet or an internal data warehouse have access to my model's inputs. The model does not require additional field data collection on the part of the analyst.

SORTIE, a spatially explicit, mixed species forest dynamics simulator, relies heavily on the complex field-based measurements of light transmitted through forest canopies and gaps (Canham et al., 1999). PROGNOSIS$^{BC}$, a growth and yield model, is fueled by tree counts within a stand and dbh measurements for each tree and requires one record of data per tree in the input .txt files.

Although PROGNOSIS$^{BC}$ and SORTIE-ND modelling may be superior in their ecological portrayal of the mechanisms of stand development, they require considerable detail for input data that limits their application to case study simulations in a few representative stand types. LeMay et al. (2002) used PROGNOSIS $^{BC}$ to examine the natural regeneration beneath complex stands in the Interior Douglas-Fir biogeoclimatic zone in the Nelson, Kamloops, and Cariboo Forest Regions. Their final report states that PROGNOSIS$^{BC}$ is best suited for models with a ground-based inventory. Similarly, the data needed to feed a SORTIE-ND and PROGNOSIS$^{BC}$ hybrid model for predicting natural regeneration in MPB-attacked stands in central and southeastern BC necessitated the collection of ground-based measurements such as individual tree dbh, total tree height, height to live crown, maximum crown diameter, and ratio of live crown to tree height (Sattler, 2009).

In contrast, the landscape-level geospatial approach developed here can be applied across a large range of geographic variability and forest stand types with a simple probability of occurrence output. The model uses coarsely collected variables for its input (a majority derived from Vegetation Resource Inventory GIS files), and consequently sacrifices individual stand anomalies for a broad-brush overview of all stands within a study area of interest. This empirical model does not simulate or otherwise address specific understory dynamics (e.g., in terms of growth release, competition among trees or with brush). The ultimate intention of the model is to create the most parsimonious algorithm derived from the most publicly available data. The more complicated the model becomes, the more limited its function.

The process of creating predictive models for the field of ecology is highly contentious. At times, models are built with good intentions only to fall into disuse, or worse, misuse (Bunnell, 1989). There are multiple ways to look at modelling ecological processes, and therefore there are multiple ways to build a model. The inherent strength of my decision tree model is that it accommodates the possibility of multiple models by creating multi-branches of rules that all represent a possible predictive algorithm, with each branch assigned a probability of accuracy. A single regression model could provide a single parsimonious equation that draws from the most significant variables as they apply to the dataset used to create the model. This, however, "locks" the model into a coefficient + $variable_1$ + $variable_2$ + $variable_n$ format, resulting in a single equation to fit the snapshot of data used to build it. With a static list of key predictor variables (and coefficients associated with these variables) the model weakens when key variables are missing from subsequent datasets. Through concepts such as surrogate splitters, the decision tree model can adapt to missing input data and can result in many rule sets with many combinations of contingency that result in a stand, or indeed parts of a forest stand, being stocked with advance regeneration or not. The forest response to MPB or other canopy disturbances is complex, and therefore cannot be easily modeled with a simple catch-all equation. Many stands sharing the same characteristics may respond differently to disturbance. A model that allows for combinations of contingency provides a more multi-faceted approach to predicting the probability of stocking.

With each statistical method used to create models, come multiple assumptions, limitations, and biases. The temptation to explore more complex interactions between variables, and therefore create more complex models is fueled by the ease of data accessibility and increased computing power. The more complex the model, the more difficult it becomes to evaluate (Bunnell, 1989; Kimmins, 2005). Just because a complex model can be built, doesn't necessarily mean it should be built.

It is important to recognize that even the most inappropriate and unrelated variables thrown together into a modelling process may result in a predictive algorithm. Like parametric models, the addition of variables (overfitting) tends to increase the model fit. Variables, however, that have little connection to each other or the ecological process that they are trying to model tend to create a "snapshot" model. A model that fits only the dataset from which it was derived therefore defeats the predictive purpose of the model. The reliance on predictor variables that were derived from static snapshot datasets run into problems when validation is required and in most cases the model misfires (Thompson, 1995; Graham, 2001; Knapp and Sawilowsky, 2001). The variables used in the development of my model were based on known or hypothesized ecological mechanisms and results from preliminary studies.

A second consideration in developing predictive models is the temptation to prioritize all the variables determined from a purely *a priori* approach. This is best described by Thompson's basketball team paradox. Thompson (1995) suggests that modelling is similar to a basketball team picking the best player first, the second-best player second (in the context of the first player's strengths and weaknesses), etc.. This is in direct contrast to the "all possible players'" approach used in constructing a second team, where the five players that play together best as a team are selected. This leaves the distinct possibility that the second team may have a roster that does not include one of the first team players. He goes further to state that the "best team," although possibly comprised of weaker players, may still be stronger as a team, than one made of all-stars selected through a purely linear process (Thompson, 1995). It was important in the development of this study's model that all combinations of variables were tested to ensure that the "best team" was selected. This can be illustrated in my thesis by examining the potential key variable distance from nearest non-pine seed source. The $r^2$ value of density of advance regeneration when plotted against distance to seed source was only 0.0064, ($P = 0.007$, $n = 964$). This means that as a single variable considered alone, distance to seed source explains less than 1% of the variance in regeneration density. Yet it was one of the top five important variables, i.e., factor where a split in the decision tree occurred, for all three stocking groups (and one of the top two for conifer only and all trees, second only to BEC unit in both cases). When taken into account with other predictors, the strength of distance to seed source increases and its importance as a primary splitter in the decision tree becomes evident.

The goal of predictive ecological models should be intuitiveness, parsimony and extensibility, namely the ability to add on or modify the model without breaking what is already there. The final key important variables included in my study's model are ecologically sound and publicly available data that form an intuitive parsimonious model that has both explanatory and predictive value.

This study began with two main objectives: to create an accurate predictive model of understory stocking probability in mature pine-leading stands for forest planners to use in mid- to long-term planning initiatives, and to create cartographic and tabular output portraying the results of the predictive model. The first objective was accomplished by collecting advance regeneration data in the summers of 2006 and 2007 and combining it with collaborator data to produce an extensive, geo-referenced data library. The second objective was achieved by creating a recursively partitioned classification tree model using R *rpart*. The model input variables were limited to readily and publicly available data sources. The final model was composed of data from the Vegetation Resource Inventory, ClimateBC, and the biogeoclimatic ecosystem classification system The resulting model achieved a 78% accuracy in correctly predicting stocked and non-stocked locations in mature pine-leading SBS forests. For the final objective, the predictive model generated using the R package *rpart* was applied against a study area by creating a geospatial model in ArcGIS 10. The geospatial model partitions the data into the following probability of being stocked classes: 0-20%, 20-40%, 40-60%, 60-80%, and 80-100%. Area statistics were gathered for each membership class and a large-format colour themed map was generated to display the results for each NTS 1:250,000 map tile. The variables used in the most parsimonious predictive model are consistent with those reported as ecologically important in a number of similar studies (e.g., Ferguson, 1984; Murphy et al., 1998; Sattler, 2009; Coates et al., 2009).

With BC's interior timber supply compromised by one of the largest disturbances in history, forest managers and planners are implementing strategies to help understand the mid- and long-term implications. These implications reach further than timber supply and will directly impact wildlife, carbon storage, hydrology, and tourism (Coates et al., 2009). Most of these strategies require current and accurate knowledge of the forest land base in the wake of the mountain pine beetle outbreak. The need for inventory data, not only what is in the canopy but also what is beneath it, is obvious. Hopefully, the provincial government is paying close attention to new inventories, or their proxy in the form of modelling output, before it proceeds with new harvesting, rehabilitation, and silviculture regulations.

## Chapter Five: Conclusions

The results of my thesis show that one way to predict the presence or absence of advance regeneration beneath mature pine stands in BC's northern interior is to create a model that works with the complex scenarios supportive of understory development, rather than trying to simplify them. The response of mature pine to the MPB events of the last decade is complicated, and therefore requires a multiple model approach to prediction. The classification decision tree developed in this thesis provides a multiple branch model that proved to be 78% accurate in predicting whether a stand was stocked or not with seedlings and saplings (>600 stems/ha). Further, a geospatial link between the predictive model and GIS provides an enhanced understanding of the prediction, assisting forest managers to discover spatial patterns and the clustering or dispersion of seedling and sapling stocking at a landscape level.

The work presented in my thesis contributes to the ongoing effort to better understand British Columbia's forests post-MPB. Several studies have examined the key variables involved in predicting advance regeneration at a stand level. The model developed in this thesis fills a gap in knowledge with regards to understanding the likely distribution of understory stocking at a landscape level. An accurate broad-brush tool is necessary in assisting with broad-brush management strategies. My predictive model is built to help forest managers make decisions at the landscape level. Although my model output generates a resultant probability of stocked pixels at a 100m size, it is the patterns of aggregated pixels that contain the information and not the single pixels themselves. The results of this thesis also contribute to the scientific community with the introduction of recursive partition analysis for stocking presence evaluation. Although not a new technique, it is presented here as a viable alternative to the more traditional step-wise regression statistical techniques. A second contribution, and an area for future research, is the integration of recursive partitioning and GIS. By providing geospatial awareness to predicted stocking probabilities across a landscape-wide area of interest, patterns previously unseen at a local scale may now emerge and potentially drive new research.

By using the outputs of my model, forest planners may also have the information needed to augment silviculture strategies such as variable-retention harvesting. Variable-retention harvest systems have been developed as one method for creating more natural stand structures, particularly in forest types that rarely experience stand-replacing natural disturbances such as severe forest fires. Variable retention harvest systems retain living and dead structural elements of the pre-harvest stand to restore structural complexity in managed forestlands (Franklin et al., 1997). The relative novelty of formal variable-retention harvest systems, however, leaves many questions about their impacts on stand development unanswered. Simple questions such as how different patterns of overstory retention influence the development of regeneration and growth in residual trees cannot be readily answered for most forest types (Gordon, 1973). More importantly, we have only anecdotal evidence gathered from studies of tree responses to natural disturbances or more traditional management practices to link regeneration and residual tree dynamics following variable retention harvesting to the underlying physiological mechanisms that drive these responses. Overstory density reductions and canopy gap formation alter resource availability, which should impact tree physiological performance and growth. The planning of salvage logging operations after MPB, whether by clearcut or variable-retention harvesting, is further complicated by the presence of tree species other than lodgepole pine within the impacted stands. A significant percentage of the pine stands contain other coniferous tree species in both the overstory and the understory (Coates et al., 2009). Incomplete forest inventories, coupled with the widespread use of unsuitable growth and yield models, appear insufficient to account for these stand and understory variations. Therefore, there exists a need to develop strategies for information gathering to augment these models and to subsequently aid in the strategic planning of salvage and regeneration responses. My model's ability to predict the probability of stocking can aid in the adoption of forestry techniques such as variable-retention.

The model's efficiency and efficacy, however, could be significantly supplemented with a broader geographic (i.e., biogeoclimatic) range of input data and the addition of several unused variables, such as prevailing wind speed and direction. The predictive model is interpolative and therefore can only be used to predict within the limits of the input data. In this study's case, only plot data from the SBS dk, dw2, dw3, mc2, mc3, mk1, mw, and wk1 were used to develop the model. Further study incorporating a more diverse sample of biogeoclimatic zones (or perhaps using temperature and precipitation relationships to extrapolate the equivalence of subzones not sampled) is suggested to enhance the model from a central British Columbia based model to a province-wide model. Although climate data were selected by the model construction algorithm, wind direction and speed data were not considered in this model iteration. Wind data can be generated in a GIS environment; I did not, however, have access to a landscape level wind dataset at the time of this study. Research has indicated that prevailing wind direction and speeds clearly play a role in the recruitment of seeds that contribute to the advance regeneration (Smidt and Blinn, 1995; Pardy, 1997; Wieland et al., 2011). In addition to the availability and favourability of substrate, abundance and proximity to seed sources (such as parent trees) are considered to be key variables in advance regeneration (Astrup et al., 2008). Given that a majority of the mature pine understories sampled in my study area were populated with species other than pine, advance regeneration establishment must have been aided by some form of seed recruitment mechanism such as wind. Incorporating the wind data into the existing model as an extensible component seems to be a logical extension to this project.

This study's predictive model provides forest planners with a unique perspective on advance regeneration inventory. By looking at the landscape as a series of predicted probabilities, detailed inventory work and the continuation of stand-level advance regeneration studies can be focused. Finally, by combining what is known about understories from stand-level advance regeneration studies with a landscape level probability model, forest planners can provide legislators and the public with varying scales of management plans and proactive forest health mitigation strategies.

**References**

Acuna, E., and C.A. Rodriguez. 2004. Meta analysis study of outlier detection methods in classification. Technical paper, Department of Mathematics, University of Puerto Rico at Mayaguez. Available from academic.uprm.edu/eacuna/ paperout.pdf. *In* proceedings IPSI 2004, Venice.

Allen, D. W. 2011. Getting to Know ArcGIS ModelBuilder. ESRI Press. Redlands, CA.

Arnup, R.W. 1996. Site and stand factors influencing the abundance and distribution of coniferous advance growth in northeastern Ontario Northern Ontario Development Agreement (NODA) Note No. 26. Natural Resources Canada, Canadian Forest Service, Great Lakes Forestry Centre, Sault Ste. Marie, ON.

Astrup, R., K.D. Coates, E. Hall.2008. Recruitment limitation in forests: lessons from an unprecedented mountain pine beetle epidemic. Forest Ecology and Management 256:1743–1750.

Axelson, J.N., R.I. Alfaro, and B.C. Hawkes. 2009. Influence of fire and mountain pine beetle on the dynamics of lodgepole pine stands in British Columbia. Forest Ecology and Management 257:1874–882.

Banerjee, O., L.E. Ghaoui, and A. d'Aspremont. 2008. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. The Journal of Machine Learning Research 9:485–516.

Bassman, J.H., J.C. Zwier, J.R. Olson, and J.D. Newberry. 1992. Growth of advance regeneration in response to residual overstory treatment in northern Idaho. Western Journal of Applied Forestry 7:78–81.

Beck, J.R., and E.K. Schultz. 1986. The use of ROC curves in test performance evaluation. Archives of Pathology and Laboratory Medicine 110: 13–20.

Berk, R.A. *2005*. Data mining within a regression framework. *In:* Maimon O., Rokach L., editors. Data mining and knowledge discovery handbook. Springer, New York, NY.

Bond, B.J., B.T. Farnsworth, R.A. Coulombe, and W.E. Winner. 1999. Foliage physiology and biochemistry in response to light gradients in conifers with varying shade tolerance. Oecologia 120: 183–192.

Bouchard, M., D. Kneeshaw, and Y. Bergeron. 2005. Mortality and stand renewal following the last spruce budworm outbreak in mixed forests of western Québec. Forest Ecology and Management 204: 297–313.

Boucher, J.F., P.Y. Bernier, H.A. Margolis, and A.D. Munson. 2007. Growth and physiological response of eastern white pine seedlings to partial cutting and site preparation. Forest Ecology and Management 240: 151–164.

Bowler, R., A.L. Fredeen, M.G. Brown, and T.A. Black. 2012. Residual vegetation importance to net $CO_2$ update in pine-dominated stands following mountain pine beetle attack in central British Columbia, Canada. Forest Ecology and Management 269: 82–91.

Breiman, L. 2001. Statistical modeling: The two cultures. Statistical Science 16:199–215.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. Classification and Regression Trees. Chapman and Hall (Wadsworth Inc), New York, NY.

Brown, M.G., T.A. Black, Z. Nesic, A.L. Fredeen, V.N. Foord, D.L. Spittlehouse, R. Bowler, P.J. Burton, J.A. Trofymow, N.J. Grant, and D. Lessard. 2012. The carbon balance of two lodgepole pine stands recovering from mountain pine beetle attack in British Columbia. Agricultural and Forest Meteorology 153: 82–93.

British Columbia Ministry of Forests. 2000. Establishment to free growing guidebook. Prince George Forest Region. Revised edition, Version 2.2. Forest Practices Code of British Columbia Guidebook. Forest Practices Branch, BC Ministry of Forests, Victoria, BC.

Bulmer C., M. Krzic, and K. Green. 2003. Soil productivity and forest regeneration success on reclaimed oil and gas sites. Final report to Oil and Gas Environmental Fund April 2003. Available at: https://circle.ubc.ca/handle/2429/9044.

Bunnell, F.L. 1989. Alchemy and uncertainty: what good are models? Gen. Tech. Rep. PNW-GTR-232. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.

Burgess, D., and S. Wetzel. 2000. Nutrient availability and regeneration response after partial cutting and site preparation in eastern white pine. Forest Ecology and Management 138:249–261.

Burnham, K.P., and D.R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd Edition. Springer-Verlag, New York, NY.

Burns, R.M., and B.H. Honkala (Technical coordinators). 1990. Silvics of North America. Volume 1 and 2. U.S. Department of Agriculture. Agriculture Handbook 654. Forest Service, United States Department of Agriculture, Washington, DC.

Burton, P.J. 2006. Restoration of forests attacked by mountain pine beetle: Misnomer, misdirected, or must-do? BC Journal of Ecosystems and Management 7:1–10.

Burton, P.J. 2008a. The mountain pine beetle as an agent of forest disturbance. BC Journal of Ecosystems and Management 9:9-13.

Burton, P.J. 2008b. The potential role of secondary structure in forest renewal after mountain pine beetle. Canadian Silviculture May 2008:26–29.

Burton, P.J. 2010. Striving for sustainability and resilience in the face of unprecedented change: the case of the mountain pine beetle outbreak in British Columbia. Sustainability 2:2403-2423.

Burton, P.J., D.C. Sutherland, N.M. Daintith, M.J. Waterhouse, and T.A. Newsome. 2000. Factors influencing the density of natural regeneration in uniform shelterwoods dominated by Douglas-fir in the Sub-Boreal Spruce Zone. Working Paper No. 47. BC Ministry of Forests, Research Branch, Victoria, BC.

Canham, C.D., D.K. Coates, P. Bartemucci, and S. Quaglia. 1999. Measurement and modelling of spatially explicit variation in light transmission through interior cedar–hemlock forests of British Columbia. Canadian Journal of Forest Research 29: 1775–1783.

Coates, K.D. 2006. Silvicultural approaches to managing MPB damaged stands: Regeneration and midterm timber supply. Presentation at the Northern Silviculture Committee Winter Workshop: January 16–18, 2006. Silviculture tactics to lessen the downfall. Prince George, BC.

Coates, K.D., and P.J. Burton. 1997. A gap-based approach for the development of silvicultural systems to address ecosystem management objectives. Forest Ecology and Management 99:339–356.

Coates, K.D., C. DeLong, P.J. Burton, and D.L. Sachs. 2006. Abundance of Secondary Structure in Lodgepole Pine Stands Affected by Mountain Pine Beetle. Report for the Chief Forester. Bulkley Valley Centre for Natural Resources Research and Management, Smithers, BC.

Coates, K.D., and E.C. Hall. 2005. Implications of alternative silvicultural strategies in mountain pine beetle damaged stands. Technical Report for Forest Science Project Y051161. Bulkley Valley Centre for Natural Resources Research and Management, Smithers, BC.

Coates, K.D., S. Haeussler, S. Lindeburgh, R. Pojar, and A.J. Stock. 1994. Ecology and silviculture of interior spruce in British Columbia. BC FRDA Report No. 220. BC Ministry of Forests, Victoria, BC.

Coates, K.D., T. Glover, B. Henderson. 2009. Abundance of secondary structure in lodgepole pine stands affected by mountain pine beetle in the Cariboo-Chilcotin; mountain pine beetle working paper 2009–20. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC.

Coates, K.D., and D.L. Sachs. 2012. MPB Impacted Stands Assessment Project: Current State of Knowledge Regarding Secondary Structure in Mountain Pine Beetle Impacted Landscapes; unpublished MFLNRO document. Smithers, B.C.

De'Ath, G., and K.E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81: 3178–92.

DeLong C, D. Tanner, and M.J. Jull. 1993. A field guide for site identification and interpretation for the southwest portion of the Prince George Forest Region. Land Management Handbook 24. British Columbia Ministry of Forests, Victoria, BC.

DeLong C. 2003. A field guide for site identification and interpretation for the southeast portion of the Prince George Forest Region. Land Management Handbook 51. British Columbia Ministry of Forests, Victoria, BC.

DeLong C. 2004. A field guide for site identification and interpretation for the north central portion of the Prince George Forest Region. Land Management Handbook 54. British Columbia Ministry of Forests, Victoria, BC.

Derksen, S., and H. J. Keselman. 1992. Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. British Journal of Mathematical and Statistical Psychology 45:265–282.

Dhar, A., and C.D.B. Hawkins. 2011. Regeneration and growth following mountain pine beetle attack: A synthesis of knowledge. BC Journal of Ecosystems and Management 12:1–16.

Dhurandhar, A., and A. Dobra. 2008. Probabilistic Characterization of Random Decision Trees. Journal of Machine Learning Research 9: 2287–2314.

Draper, N., and H. Smith. 1981. Applied Regression Analysis. (2nd edition). John Wiley and Sons, New York, NY.

Drever, C.R., G. Peterson, C. Messier, Y. Bergeron, and M.D. Flannigan. 2006. Can forests management based on natural disturbances maintain ecological resilience? Canadian Journal of Forest Research 36: 2285–2299.

Elith J., J.R., Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77:802–13.

Ellsworth, D.S., and P.B. Reich. 1992. Leaf mass per area, nitrogen content and photosynthetic carbon gain in *Acer saccharum* seedlings in contrasting forest light environments. Functional Ecology 6:423–435.

ESRI. 2011. ArcGIS Desktop: Release 10. Environmental Systems Research Institute. Redlands, CA.

Feldman, D., and S. Gross. 2003. Mortgage Default: Classification Trees Analysis, Discussion Paper No. 3. The Pinhas Sapir Center For Development Tel-Aviv University, Israel.

Ferguson, D.E. 1984. Needed: guidelines for defining acceptable advance regeneration. USDA Forest Serv Res Note INT-341, 5 p. Intermountain Forest and Range Experiment Station, Ogden, UT.

Ferguson, D. E., and C.E. Carlson. 1993. Predicting Regeneration Establishment with the Prognosis Model. USDA Forest Service, Intermountain Research Station, Ogden, Utah. Research Paper INT- 467.

Franklin, J.F., D.R. Berg, D.A. Thornburgh, J.C. Tappeiner. 1997. Alternative silvicultural approaches to timber harvesting: variable retention harvest systems. *In*: Kohm, K.A., Franklin, J.F. (Eds.), Creating a Forestry for the 21st Century. Island Press, Washington, DC.

Franklin, J.F., T.A. Spies, R. Van Pelt, A.B. Carey, D.A. Thornburgh, D.R. Berg, D.B. Lindenmayer, M.E. Harmon, W.S. Keeton, D.C. Shaw, K. Bible, and J. Chen. 2002. Disturbances and structural development of natural forest ecosystems with silvicultural implications, using Douglas-fir forests as an example. Forest Ecology and Management 155: 399–423.

Franklin, J. 2009. Mapping Species Distributions. 1st edition. Cambridge University Press. New York, NY.

Friedl, M. A., C. Brodley, and A. Strahler. 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. IEEE Transactions on Geoscience and Remote Sensing 37: 969–977.

Friedman, J.H., and J.J. Meulman. 2003. Multiple additive regression trees with application in epidemiology. Statistics in Medicine 22:1365–1381.

Freedman, D. A. 1983. A note on screening regression equations. The American Statistician 37: 152–155.

Gautreaux, R. 1999. Vegetation Response Unit Characterizations and Target Landscape Prescriptions, Kootenai National Forest, 1999. USDA Forest Service, Northern Region, Kootenai National Forest, Libby, MT.

Gordon, D.T. 1973. Released advance reproduction of white fir and red fir: Effects on growth, damage, and mortality. USDA Forest Service, Research Paper. Pacific Southwest Forest and Range Experiment Station, Berkeley, CA.

Graham, J. M. 2001. The ethical use of statistical analyses in psychological research. Paper Presented at the annual meeting of Division 17 (Counseling Psychology) of the American Psychological Association, Houston, TX.

Gray, J.R., G.D. Glysson, and D.S. Mueller. 2002. Comparability and accuracy of fluvial-sediment data: A view from the U.S. Geological Survey. Proceedings of the American Society of Civil Engineers, Hydraulic Measurements and Methods Symposium, July-August, 2002, Estes Park, CO.

Greene, D.F., J.C. Zasada, L. Sirois, D. Kneeshaw, H. Morin, I. Charron, and M.J. Simard. 1999. A review of the regeneration dynamics of North American boreal forest tree species. Canadian Journal of Forest Research 29:824-839.

Griesbauer, H., and S. Green. 2006. Examining the utility of advance regeneration for reforestation and timber production in unsalvaged stands killed by the mountain pine beetle: Controlling factors and management implications. BC Journal of Ecosystems and Management 7:81–92.

Guisan, A., S.B. Weiss, A.D. Weiss. 1999. GLM versus CCA spatial modeling of plant species distribution. Plant Ecology 143:107–122.

Han, J., M. Kamber, and J. Pei. 2011. Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann. San Francisco, CA.

Hanley, J.A., and B.J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143: 29–36.

Hansen, M.C., R.S. DeFries, J.R.G. Townshend, and R. Sohlberg. 2000. Global land cover classification at 1km spatial resolution using a classification tree approach. International Journal of Remote Sensing 21:1331-1364.

Hastie, T., R. Tibshirani, and J. Friedman. 2001. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, New York, NY.

Hawkes, BC, S.W. Taylor, C. Stockdale, T.L. Shore, R.I. Alfaro, R.A. Campbell, and P. Vera. 2003. Impact of mountain pine beetle on stand dynamics in British Columbia. *In* Mountain pine beetle symposium: Challenges and solutions, October 30–31, 2003, Kelowna, BC.

Hawkins, C.D.B., A. Dhar, N.A. Balliet, and K.D. Runzer. 2012. Residual mature trees and secondary stand structure after mountain pine beetle attack in central British Columbia. Forest Ecology and Management 277:107–115.

Hawkins, C., and P. Rakochy. 2007. Stand-level effects of the mountain pine beetle outbreak in the central British Columbia interior. Mountain pine beetle initiative working paper 2007-2006, mountain pine beetle initiative project #8.23. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC.

Heithecker, T.D., and C.B Halpern. 2007. Edge-related gradients in microclimate in forest aggregates following structural retention harvests in western Washington. Forest Ecology and Management 248: 163–173.

Hélie, J.F., D.L. Peters, K.R. Tattrie, and J.J. Gibson. 2005. Review and synthesis of potential hydrologic impacts of mountain pine beetle and related harvesting activities in British Columbia. Mountain Pine Beetle Initiative Working Paper 2005-23. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC.

Hughes, A.R., 2010. Disturbance and diversity: an ecological chicken and egg problem. Nature Education Knowledge 1: 26–33.

Johnson, E.A., H. Morin, K. Miyanishis, R. Gagnon, and D.F. Greene. 2003. A process approach to understanding disturbance and forest dynamics for sustainable forestry. *In* P.J. Burton, C. Messier, D.W. Smith, and W.L. Adamowicz, editors. Towards Sustainable Management of the Boreal Forest. NRC Research Press, Ottawa, ON.

Johnson, E.A., and K. Miyanishi (eds.). 2007. Plant Disturbance Ecology: The Process and the Response. Academic Press, San Diego, CA.

Kaufmann, M.R., G.H. Aplet, M. Babler, W.L. Baker, B. Bentz, M. Harrington, B.C. Hawkes, L.S. Huckaby, M.J. Jenkins, D.M. Kashian, R.E. Keane, D. Kulakowski, C. McHugh, J. Negron, J. Popp, W.H. Romme, T. Schoennagel, W. Shepperd, F.W. Smith, E.K. Sutherland, D. Tinker, T.T. Veblen. 2008. The status of our scientific understanding of lodgepole pine and mountain pine beetles – a focus on forest ecology and fire behavior. Global Fire Initiative technical report 2008-2, Arlington, VA.

Kayes, L.J., and D.B. Tinker. 2012. Forest structure and regeneration following a mountain pine beetle epidemic in southeastern Wyoming. Forest Ecology and Management 263: 57–66.

Kim, H., and S. Yates. 2003. Missing value algorithms in decision trees. *In* Bozdogan, H. editor, Statistical Data Mining and Knowledge Discovery. Chapman and Hall/CRC, Boca Raton, FL.

Kimmins, J.P., C. Welham, B. Seely, M. Meitne, R. Rempel, and T. Sullivan. 2005. Science in forestry: why does it sometimes disappoint or even fail us. Forestry Chronicle 81: 723–734.

Klinka, K., J. Worral, L. Skoda, and P. Varga. 2000. The distribution and synopsis of ecological and silvical characteristics of tree species of British Columbia's forests. Canadian Cartographics Ltd., Coquitlam, BC.

Knapp, T.R., and S.S. Sawilowsky. 2001. Constructive criticisms of methodological and editorial practices. The Journal of Experimental Education 71:65–79.

Kneeshaw, D.D., Y. Bergeron, and L. DeGrandpré. 1998. Early response of *Abies balsamea* seedlings to artificially created openings. Journal of Vegetation Science 9:543–550.

Kneeshaw, D.D. and P.J. Burton. 1997. Canopy and age structures of some old sub-boreal Picea stands in British Columbia. Journal of Vegetation Science 8:615–626.

Khoshgoftaar, T., and E.B. Allen. 2001. Controlling overfitting in classification-tree models of software quality. Empirical Software Engineering 6:59–79

Klock, R., and J. Mullock. 2008. The Weather of British Columbia: Graphic Area Forecast 31. The Meteorological Service of Canada and NavCanada, Kelowna, BC.

Kreakie BJ, Y. Fan, and T.H. Keitt. 2012. Enhanced migratory waterfowl distribution modeling by inclusion of depth to water table data. PLoS ONE 7: e30142.

Krivec, J., and G. Matjaz. 2011. Data Mining Techniques for Explaining Social Events, Knowledge-Oriented Applications in Data Mining, Kimito Funatsu (Ed.). Jožef Stefan Institute, Slovenia.

Leadem, C., S.L. Gillies, K.H. Yearsley, V. Sit, D.L. Spittlehouse, and P.J. Burton. 1997. Field studies of seed biology. Land Management Handbook No. 40. British Columbia Ministry of Forests, Victoria, BC.

LeMay, V., P. Marshall, H. Temesgen, A.A. Zumrawi, B. Hassini, K. Froese, C. Lencar, and R. Martin. 2002. Development of natural regeneration and juvenile height growth models for complex stands of Southeastern and Central British Columbia. Report prepared for: Forest Renewal BC, Victoria, BC.

LePage, P.T., C.D. Canham, K.D. Coates, and P. Bartemucci. 2000. Seed sources versus substrate limitations of seedling recruitment in interior cedar– hemlock forest of British Columbia. Canadian Journal of Forest Research 30:415–427.

Lewis, R.J. 2000. An Introduction to Classification and Regression Tree (CART) Analysis. Presented at Annual Meeting of the Society for Academic Emergency Medicine. Available at: http://www.saem.org/download/lewis1.pdf. Accessed June 2012.

Lewis, K.J., and I. Hartley. 2005. Rate of deterioration, degrade and fall of trees killed by mountain pine beetle: A synthesis of the literature and experiential knowledge. Mountain Pine Beetle Initiative Working Paper 2005-14. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC.

Lewis, K.J. 2011. Forest health and mortality of advance regeneration following canopy tree mortality caused by the mountain pine beetle. MPB Working Paper 2010-03. Canadian Forest Service, Pacific Forestry Centre, Victoria, BC.

Lieffers, V.J., and K.J. Stadt. 1994. Growth of understory Picea glauca, *Calamagrostis canadensis*, and *Epilobium angustifolium* in relation to overstory light transmission. Canadian Journal of Forest Research 24:1193–1198.

Lindenmayer, D.B., P.J. Burton, and J.F. Franklin. 2008. Salvage Logging and Its Ecological Consequences. Island Press, Washington, DC.

Lindenmayer, D.B., and J.F.Franklin. 2002. Conserving Forest Biodiversity: A Comprehensive Multiscaled Approach. Island Press, Washington, DC.
McCarthy, J. 2001. Gap dynamics of forest trees: A review with particular attention to boreal forests. Environmental Reviews 9:1–59.

McGuire, J.P., R.J. Mitchell, E.B. Moser, S.D. Pecot, D.H. Gjerstad, and C.W. Hedman.2001. Gaps in a gappy forest: plant resources, longleaf pine regeneration, and understory response to tree removal in longleaf pine savannas. Canadian Journal of Forest Research 31: 765–778.

McWilliams, J. 2009. A Review and Analysis of the Effect of BC's Current Stocking Standards on Forest Stewardship. Association of BC Forest Professionals. Vancouver, BC.

Maclauchlan, L. 2006. Status of mountain pine beetle attack in young lodgepole pine stands in central British Columbia. Report to the Chief Forester, Jim Snetsinger, at the 2006 Forest Health Review Committee Meeting, Victoria, BC.

MacLean, D.A., and A.R. Andersen. 2008. Impact of a spruce budworm outbreak in balsam fir and subsequent stand development over a 40-year period. Forestry Chronicle 84:60–69.

Manel, S., H.C., Williams, and S.J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38:921–931.

Martin, P.J., B. Bancroft, K. Day, and K. Peel. 2005. A new basis for understory stocking standards for partially harvested stands in the British Columbia interior. Western Journal of Applied Forestry 20:5–12.

Martin, P. 2012. Improving inventory information in MPB-affected management units. Forest Analysis and Inventory Branch. Available from http://www.for.gov.bc.ca/hts/vri/newsletter/improving_inventory_in%20mpb_units_2012.pdf

Matlack, G. R. 1994. Vegetation dynamics of the forest edge—trends in space and successional time. Journal of Ecology 82:113–123.

Meidinger, D., J. Pojar, and W.L. Harper. 1991. Sub-Boreal Spruce zone. *In* Ecosystems of British Columbia. Special Report Series No. 6. BC Ministry of Forests, Victoria, BC. Available from: http://www.for.gov.bc.ca/hfd/pubs/Docs/Srs/Srseries.htm

Messier, C., R. Doucet, J. Ruel, Y. Claveau, C. Kelly, and M.J. Lechowicz. 1999. Functional ecology of advance regeneration in relation to light in boreal forests. Canadian Journal of Forest Research 29:812–823.

Ministry of Forests, Lands, and Natural Resource Operations. 2011. Forests for Tomorrow Current Reforestation and Timber Supply Mitigation Strategic Plan 2011 to 2015.

Ministry of Forests, Lands and Natural Resources Operations, Inventory Section, Forest Analysis and Inventory Branch. 2012. Approach of the inventory program in 2012-13 to improve inventory information in MPB-affected management units.
Mitchell, J. 2005. Review and synthesis of regeneration methods in beetle-killed stands following mountain pine beetle (*Dendroctonus ponderosae*) attack: A literature review. Mountain Pine Beetle Initiative Working Paper 2005-16. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC.

Moisen, G. G. 2008. Classification and regression trees. *In*: Jørgensen, S.E. and B.D. Fath (Editor-in-Chief). Encyclopedia of Ecology, volume 1. Oxford, U.K.

Moles, A.T., and D.R. Drake. 1999. Potential contributions of the seed rain and seed bank to regeneration of native forest under plantation pine in New Zealand. New Zealand Journal of Botany 37: 83–93.

Mori, A.S. 2011. Ecosystem management based on natural disturbances: hierarchical context and non-equilibrium paradigm. Journal of Applied Ecology 48: 280–292.

Murphy, T.E.L., D.L. Adams, and D.E. Ferguson. 1999. Response of advance lodgepole pine regeneration to overstory removal in eastern Idaho. Forest Ecology Management 120:235–244.

Nigh, G.D., J.A. Antos, and R. Parish. 2008. Density and distribution of advance regeneration in mountain pine beetle killed lodgepole pine stands of the Montane Spruce zone of southern British Columbia. Canadian Journal of Forest Research 38:2826–2836.

Oguchi, R., K. Hikosaka, T. Hiura, and T. Hirose. 2006. Leaf anatomy and light acclimation in woody seedlings after gap formation in a cool-temperate deciduous forest. Oecologia 149: 571–582.

Oliver, C., and B. Larson. 1996. Forest stand dynamics (update edition). McGraw-Hill Inc., New York, NY.

Palik B.J., R.J Mitchell, and J.K Hiers. 2002. Modeling silviculture after natural disturbance to sustain biodiversity in the longleaf pine *Pinus palustris* ecosystem: balancing complexity and implementation. Forest Ecology and Management 155: 347–356.

Palik, B.J., R.J. Mitchell, S. Pecot, M. Battaglia, and M. Pu. 2003. Spatial distribution of overstory retention influences resources and growth of longleaf pine seedlings. Ecological Applications 13: 674–686.

Palik, B.J., R. Mitchell, G. Houseal, and N. Pederson. 1997. Effects of canopy structure on resource availability and seedling responses in a longleaf pine ecosystem. Canadian Journal of Forest Research 27:1458–1464.

Pardy, A.B. 1997. Forest succession following a severe spruce budworm outbreak at Cape Breton Highlands National Park. M.Sc.F. thesis, University of New Brunswick, Fredericton, N.B.

Parkins, J., and N. MacKendrick. 2007. Assessing community vulnerability: A study of the mountain pine beetle outbreak in British Columbia, Canada. Global Environmental Change 17:460–471.

Pedersen, L. 2004. Expedited timber supply review for the Lakes, Prince George, and Quesnel Timber Supply Areas. Public Discussion Paper. BC Ministry of Forests and Range, Victoria, BC.

PEM - Ecosystem and Terrain Mapping Data Inventory [GIS shapefile]. 2008. Land and Resource Data Warehouse. Available http: https://apps.gov.bc.ca/pub/dwds/addProducts.do. [Accessed: May, 2012].
Pickett, S. T. A., and P. S. White, editors. 1985. The ecology of natural disturbance and patch dynamics. Academic Press, NewYork, NY.

Pousette, J., and C. Hawkins. 2006. An assessment of critical assumptions supporting the timber supply modelling for mountains-pine-beetle-induced allowable annual cut uplift in the Prince George Timber Supply Area. BC Journal of Ecosystems and Management 7:93–104.

R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Refaeilzadeh, P., L. Tang, and H. Liu. 2009. Cross Validation. Encyclopedia of Database Systems. Springer, New York, NY.

Ribbens, E., Jr., J.A. Silander, and S.W. Pacala. 1994. Seedling recruitment in forests: Calibrating models to predict patterns of tree seedling dispersion. Ecology 75:1794-1806.

Ripley, B.D. 1996. Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge, UK.

Roberts, M.R. 2004. Response of the herbaceous layer to natural disturbance in North American forests. Canadian Journal of Botany 82:58–64.

Sattler, F.D. 2009. A hybrid model to estimate natural recruitment and growth in stands following mountain pine beetle disturbance. M.Sc. thesis, University of British Columbia, Vancouver, BC.

Seymour R.S., A.S. White, and P.G. deMaynadier. 2002. Natural disturbance regimes in northeastern North America – evaluating silvicultural systems using natural scales and frequencies. Forest Ecology and Management 155: 357–367.

Sherrod, P.H. 2006. DTREG Software for Predictive Modelling and Forecasting User Manual. Available from www.dtreg.com.

Smidt M., and C.R. Blinn. 1995. Logging for the 21st Century: Forest Ecology and Regeneration, FO-06517. University of Minnesota, Minneapolis, MN.

Smith, S.M. 1990. The Prognosis Model Adapted for Dry-Belt Douglas-Fir in British Columbia. Report prepared by Stephen Smith & Associates, Victoria, B.C. for BC Ministry of Forests.

Snetsinger, J. 2011. Prince George Timber Supply Area: Rationale for Allowable Annual Cut (AAC) Determination. British Columbia Ministry of Forests, Mines and Lands. Available from http://www.for.gov.bc.ca/hts/tsa/tsa24/tsr4/24ts11ra.pdf.

Stephens, P.A., S.W. Buskirk, G.D. Hayward, and C. Martinez del Rio. 2005. Information theory and hypothesis testing: a call for pluralism. Journal of Applied Ecology 42: 4–12.

Strobl C., A.L. Boulesteix, and T. Augustin. Unbiased Split Selection for Classification Trees Based on the Gini Index. 2007. Computational Statistics and Data Analysis 52: 483–501.

Strobl, C., J. Malley, and G. Tutz. 2009. An Introduction to Recursive Partitioning: Rational, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. Psychological Methods 14:323–348.

Stockdale, C., S. Taylor, and B. Hawkes. 2004. Incorporating mountain pine beetle impacts on stand dynamics in stand and landscape models: A problem analysis. *In:* Mountain Pine Beetle Symposium: Challenges and Solutions, October 30–31, 2003, Kelowna, BC T.L. Shore, J.E. Brooks, and J.E. Stone (editors). Information Report BC-X-399. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC.

Stone, W.E., and M.L. Wolfe. 1996. Response of understory vegetation to variable tree mortality following a mountain pine beetle epidemic in lodgepole pine stands in northern Utah. Plant Ecology 122:1–12.

Sutherland, J. 2012. Guidance for assessing FSP stocking standards alignment with addressing immediate and long-term forest health issues. Memorandum. File: 280-30/FSP. Ministry of Forests, Lands, and Natural Resource Operations. Available from https://www.for.gov.bc.ca/hfp/silviculture/Guidance%20for%20assessing%20FSP%20stocking%20standards%20June%2021%202012.pdf

Tews, J. and F. Jeltsch. 2004. Modelling the impact of climate change on woody plant population dynamics in South African savannah. BMC Ecology 4:17.

Therneau, T.M, B. Atkinson, and B. Ripley. 2012. Recursive Partitioning. R package version 3.1-52. http://CRAN.R-project.org/package=rpart.

Thompson, Bruce. 1995. Stepwise Regression and Stepwise Discriminant Analysis Need Not Apply here: A Guidelines Editorial. Educational and Psychological Measurement 55:525-534.

Vyse, A., C. Ferguson, D.J. Huggard, J. Roach, and B. Zimonick. 2009. Regeneration beneath lodgepole pine dominated stands attacked or threatened by the mountain pine beetle in the south central Interior, British Columbia. Forest Ecology and Management 258: S36–S43.

VRI - Forest Vegetation Composite Polygons and Rank 1 Layer [GIS shapefile]. 2006. Land and Resource Data Warehouse. DataBC. Available http: http://apps.gov.bc.ca/pub/dwds/addProducts.do?orderId=1206753&packagedProductId=-1&indirect=true. [Accessed: May, 2007].

Walton, A. 2012. Provincial-Level Projection of the Current Mountain Pine Beetle Outbreak: Update of the infestation projection based on the Provincial Aerial Overview Surveys of Forest Health conducted from 1999 through 2011 and the BCMPB model (year 9). BC Forest Service: Victoria, B.C. Available from http://www.for.gov.bc.ca/hre/bcmpb/Year9.htm.

Wang, T., A. Hamann, D.L. Spittlehouse, and S.N. Aitken. 2006. Development of scalefree climate data for western Canada for use in resource management. International Journal of Climatology 26:383–397.

Weetman, G., and A.H. Vyse. 1990. Natural regeneration. *In:* Regenerating British Columbia's forests. D.P. Lavender, R. Parish, C.M. Johnson, G. Montgomery, A. Vyse, R.A. Willis, and D. Winston (editors). University of British Columbia Press, Vancouver, BC.

White, P.S., and S.T.A. Pickett. 1985. Natural disturbance and patch dynamics: an introduction. *In*: Pickett, S.T.A.; White, P.S., eds. The ecology of natural disturbance and patch dynamics. Academic Press, New York, NY.

Wieland, L.M, R.C.G, Mesquita, P.E.D. Bobrowiec, T.V. Bentos, and G.B. Williamson. 2011. Seed rain and advance regeneration in secondary succession in the Brazilian Amazon. Tropical Conservation Science 4(3):300–316.

Williams, H., C. Messier, and D.D. Kneeshaw. 1999. Effects of light availability and sapling size on the growth and crown morphology of understory Douglas-fir and Lodgepole pine. Canadian Journal of Forest Research 29:222–231.

Williston, P., D. Cichowski, and S. Haeussler. 2006. The response of caribou terrestrial matforming lichens to mountain pine beetles and forest harvesting in the East Ootsa and Entiako areas: Final Report -2005 - Years 1 to 5. A report to Morice-Lakes Innovative Forest Practices Agreement, Prince George, BC, the Bulkley Valley Centre for Natural Resources Research and Management, and BC Parks, Smithers, BC.

Whittingham, M, P. Stephens, R.B. Bradbury, and R.P. Freckleton. 2006. Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology 75:1182–1189.

Wilford, D. 2008. Timber Growth and Value Conference Proceedings closing remarks. Bulkley Valley Centre for Natural Resources Research and Management. Smithers, BC. Available from http://bvcentre.ca/files/research_reports/07-07ProceedingsTimberGrowth.pdf

Wright, E.F., C.D. Canham, and K.D. Coates. 2000. Effects of suppression and release on sapling growth for 11 tree species of northern, interior British Columbia. Canadian Journal of Forest Research 30:1571–1580.

## Appendix 1

The attached table are the published MSSpa values and conditions that were used to define the level of stocking against which to assess seedling and sampling densities. Minimum Stocking Standards (MSSpa) for preferred and acceptable species (BC Ministry of Forests, 2000)

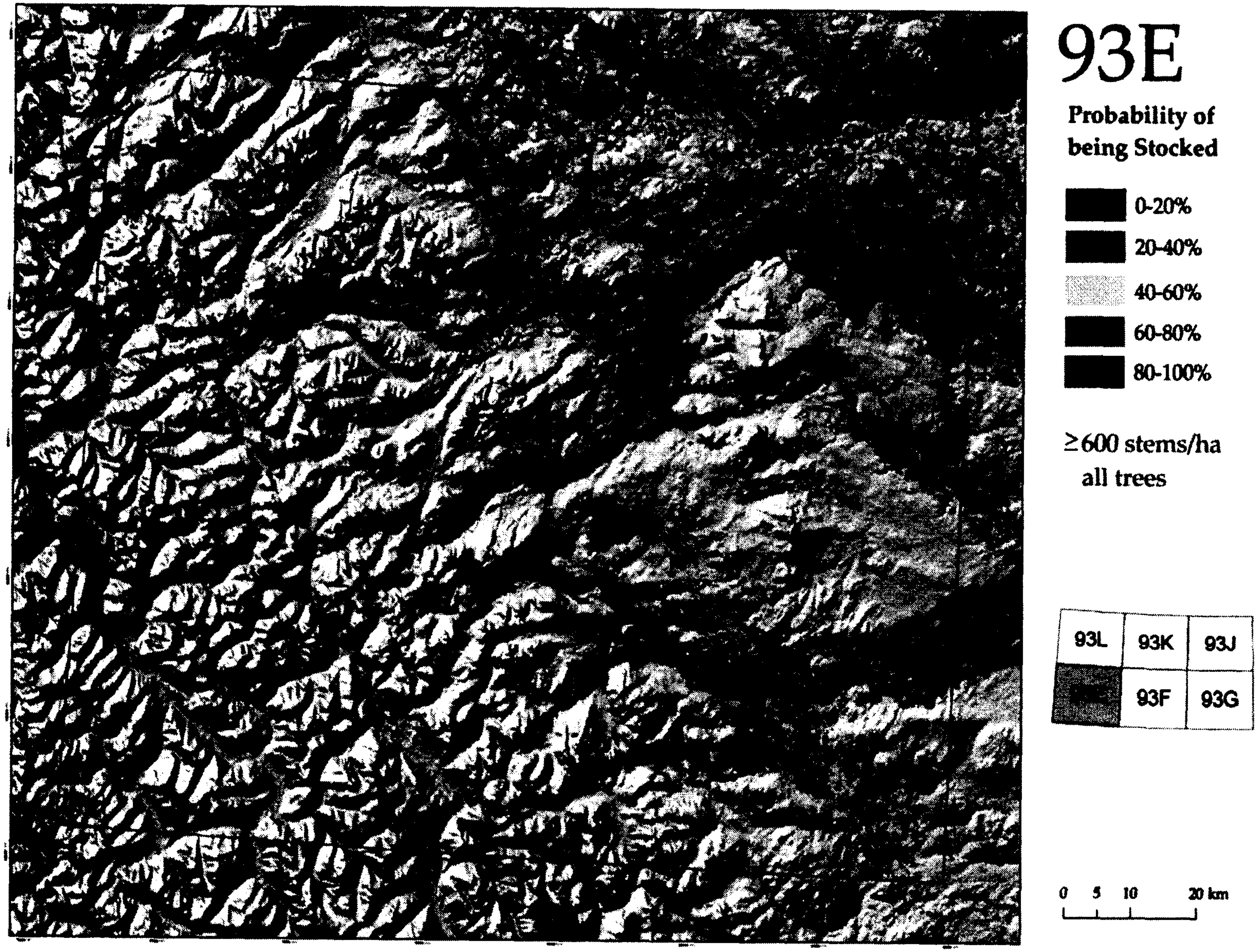
| BEC Unit | SiteSeries | TSSpa | MSSpa | MSSp |
|---|---|---|---|---|
| SBSdk | 1 | 1200 | 700 | 600 |
| SBSdk | 2 | 1000 | 500 | 400 |
| SBSdk | 3 | 1200 | 700 | 600 |
| SBSdk | 5 | 1200 | 700 | 600 |
| SBSdk | 6 | 1200 | 700 | 600 |
| SBSdk | 7 | 1000 | 500 | 400 |
| SBSdw2 | 1 | 1200 | 700 | 600 |
| SBSdw2 | 6 | 1200 | 700 | 600 |
| SBSdw2 | 8 | 1200 | 700 | 600 |
| SBSdw2 | 9 | 1200 | 700 | 600 |
| SBSdw3 | 1 | 1200 | 700 | 600 |
| SBSdw3 | 3 | 1200 | 700 | 600 |
| SBSdw3 | 4 | 1200 | 700 | 600 |
| SBSdw3 | 5 | 1200 | 700 | 600 |
| SBSdw3 | 6 | 1200 | 700 | 600 |
| SBSdw3 | 7 | 1200 | 700 | 600 |
| SBSdw3 | 8 | 1200 | 700 | 600 |
| SBSdw3 | 9 | 1000 | 500 | 400 |
| SBSdw3 | 10 | 400 | 200 | 200 |
| SBSmc2 | 1 | 1200 | 700 | 600 |
| SBSmc2 | 3 | 1200 | 700 | 600 |
| SBSmc2 | 5 | 1200 | 700 | 600 |
| SBSmc2 | 6 | 1200 | 700 | 600 |
| SBSmc2 | 10 | 1000 | 500 | 400 |
| SBSmc3 | 1 | 1200 | 700 | 600 |
| SBSmc3 | 4 | 1200 | 700 | 600 |
| SBSmc3 | 5 | 1200 | 700 | 600 |
| SBSmc3 | 7 | 1200 | 700 | 600 |
| SBSmk1 | 3 | 1200 | 700 | 600 |
| SBSmk1 | 5 | 1200 | 700 | 600 |
| SBSmw | 1 | 1200 | 700 | 600 |
| SBSwk1 | 5 | 1200 | 700 | 600 |

# Appendix 2

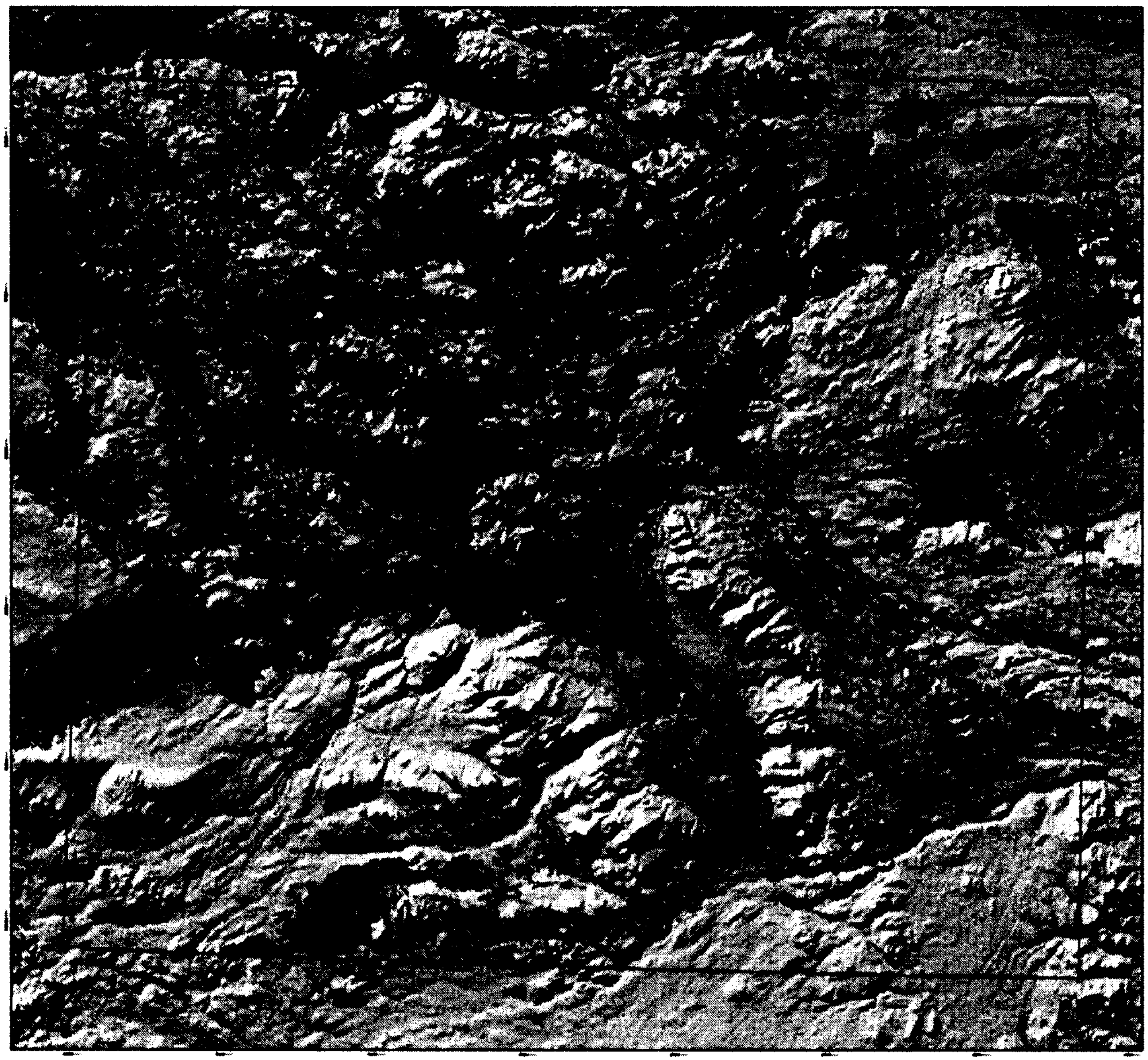
**Colour themed maps depicting probability of being stocked within NTS 1:250K map tiles 93E,F,G,J,K,L**

The following maps are colour-themed maps built from the classification tree model rules presented in this thesis. This is a full series of probability of stocking maps for all three stocking groups: 1) ≥600 stems/ha conifer only; 2) ≥600 stems/ha all tree species; and 3) MSSpa (minimum stocking standards for preferred and acceptable species).

Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia



Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia

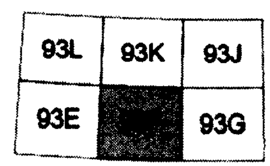


# 93F

**Probability of being Stocked**

- 0-20%
- 20-40%
- 40-60%
- 60-80%
- 80-100%

≥600 stems/ha
all trees




0 5 10 20 km

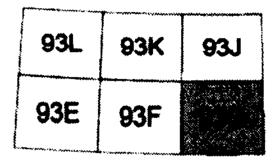
Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia

# 93G

**Probability of being Stocked**

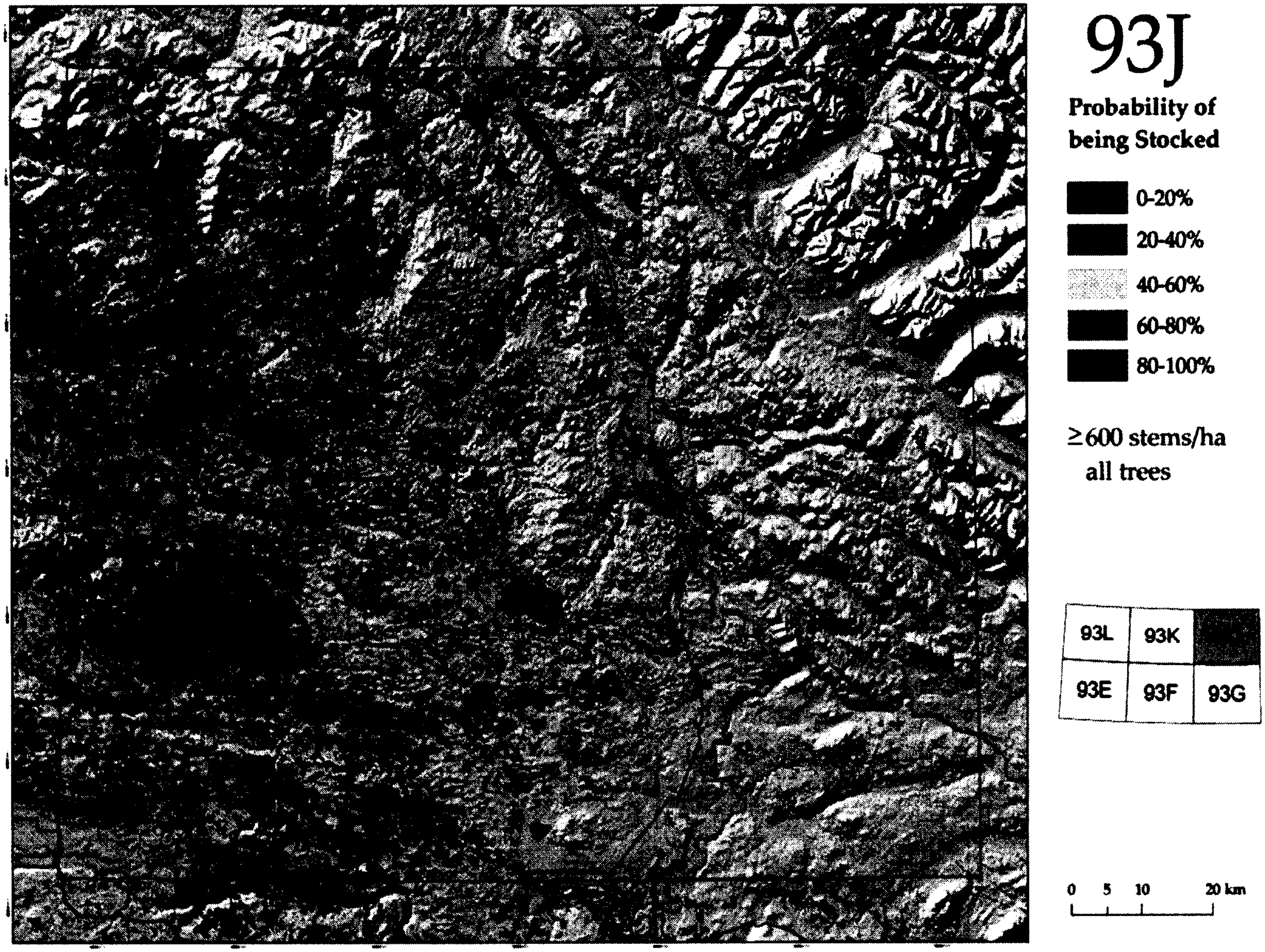
- 0-20%
- 20-40%
- 40-60%
- 60-80%
- 80-100%

≥600 stems/ha
all trees




0 5 10 20 km

Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia



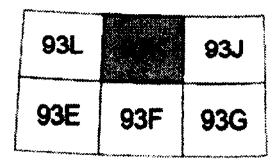
Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia

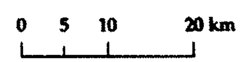
# 93K

**Probability of being Stocked**

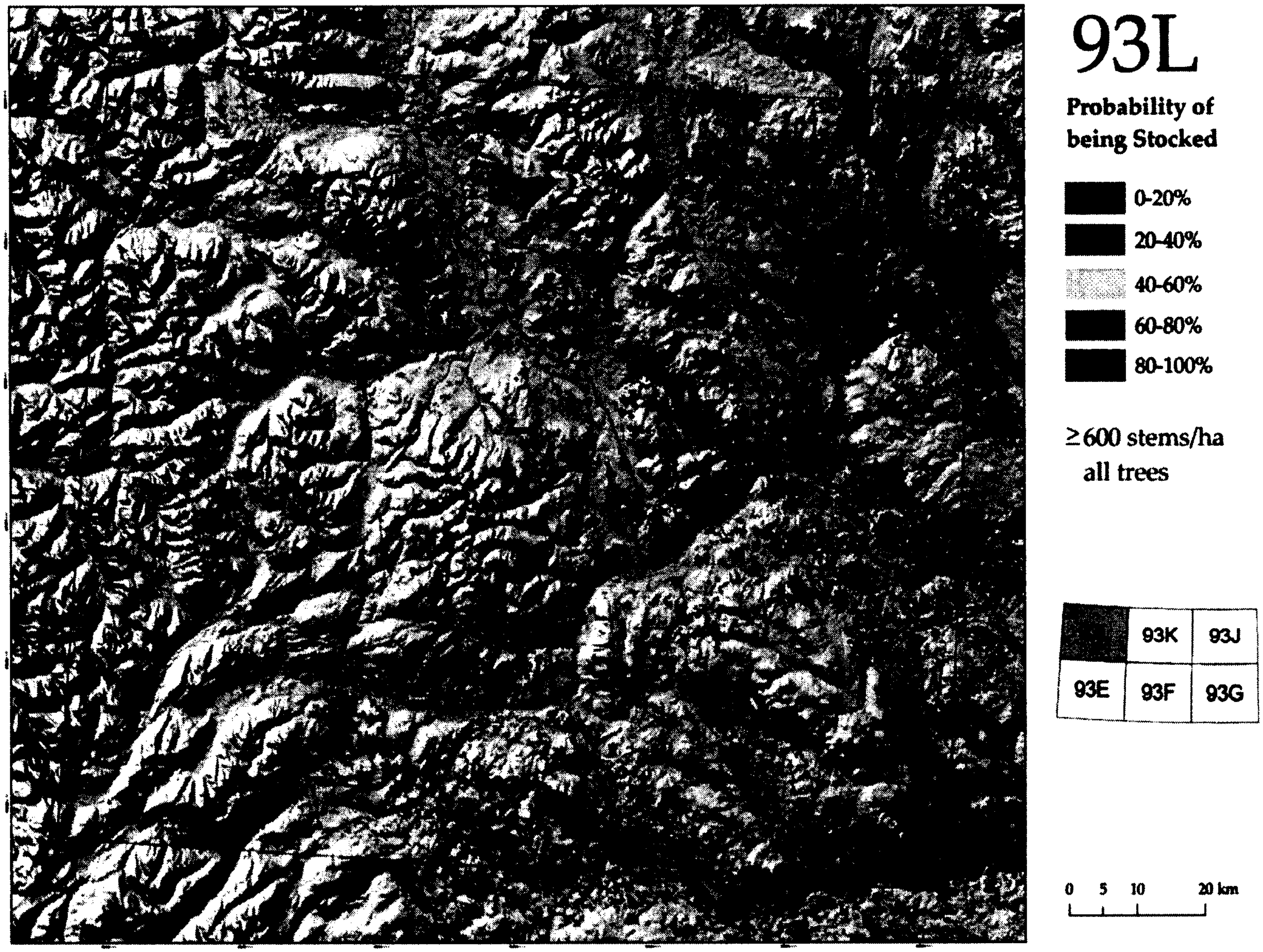
- 0-20%
- 20-40%
- 40-60%
- 60-80%
- 80-100%

≥600 stems/ha
all trees

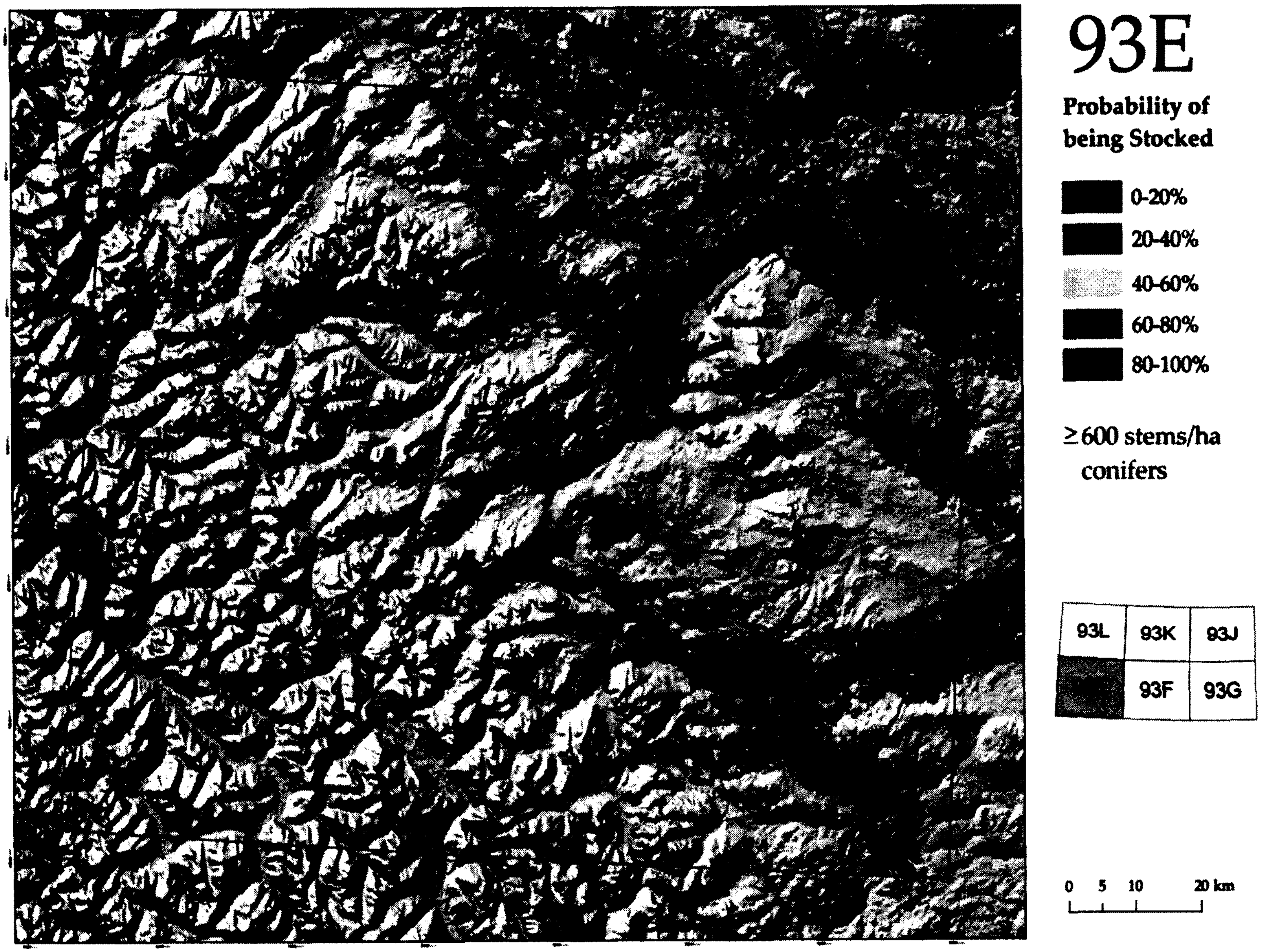


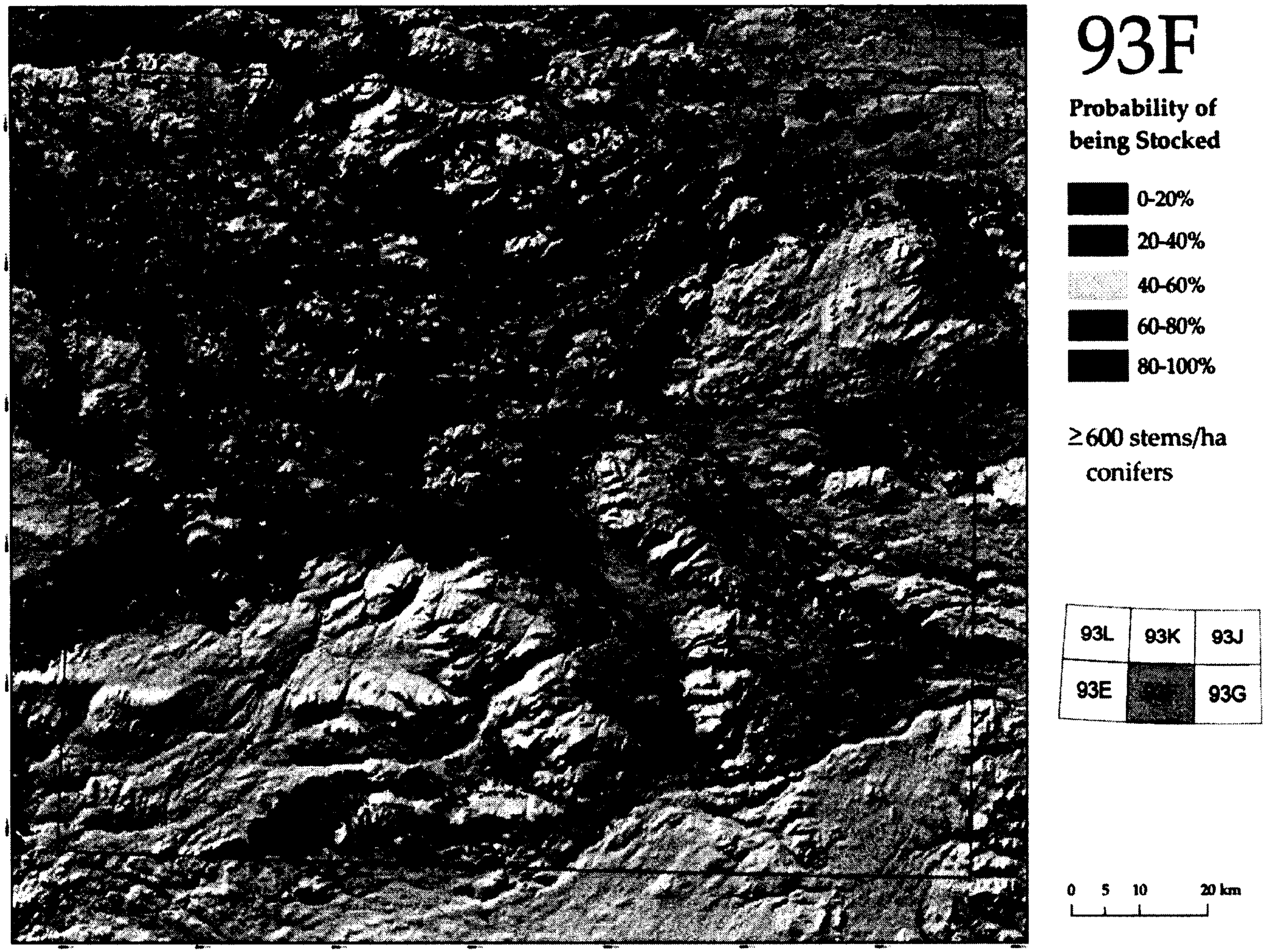
Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia



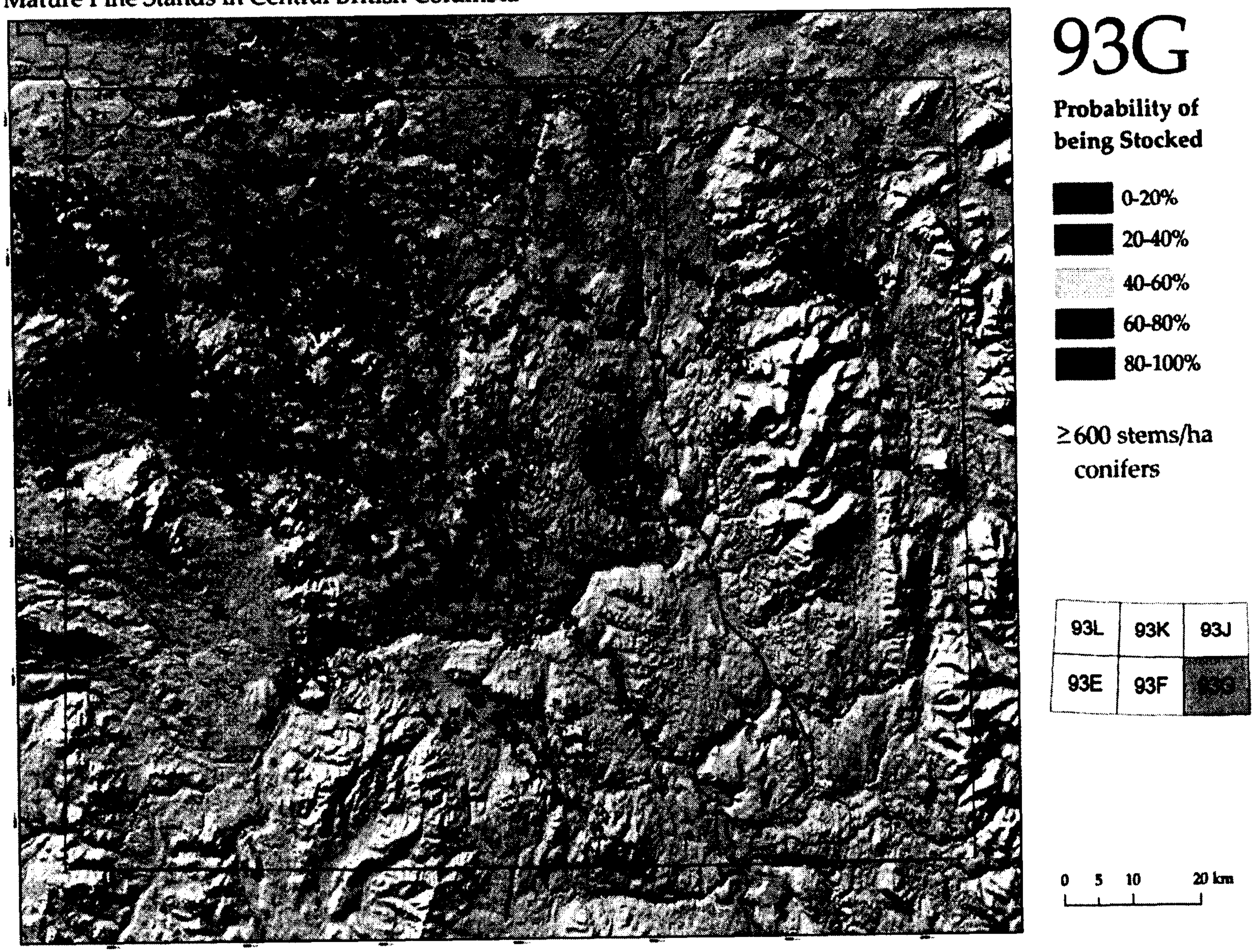
Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia



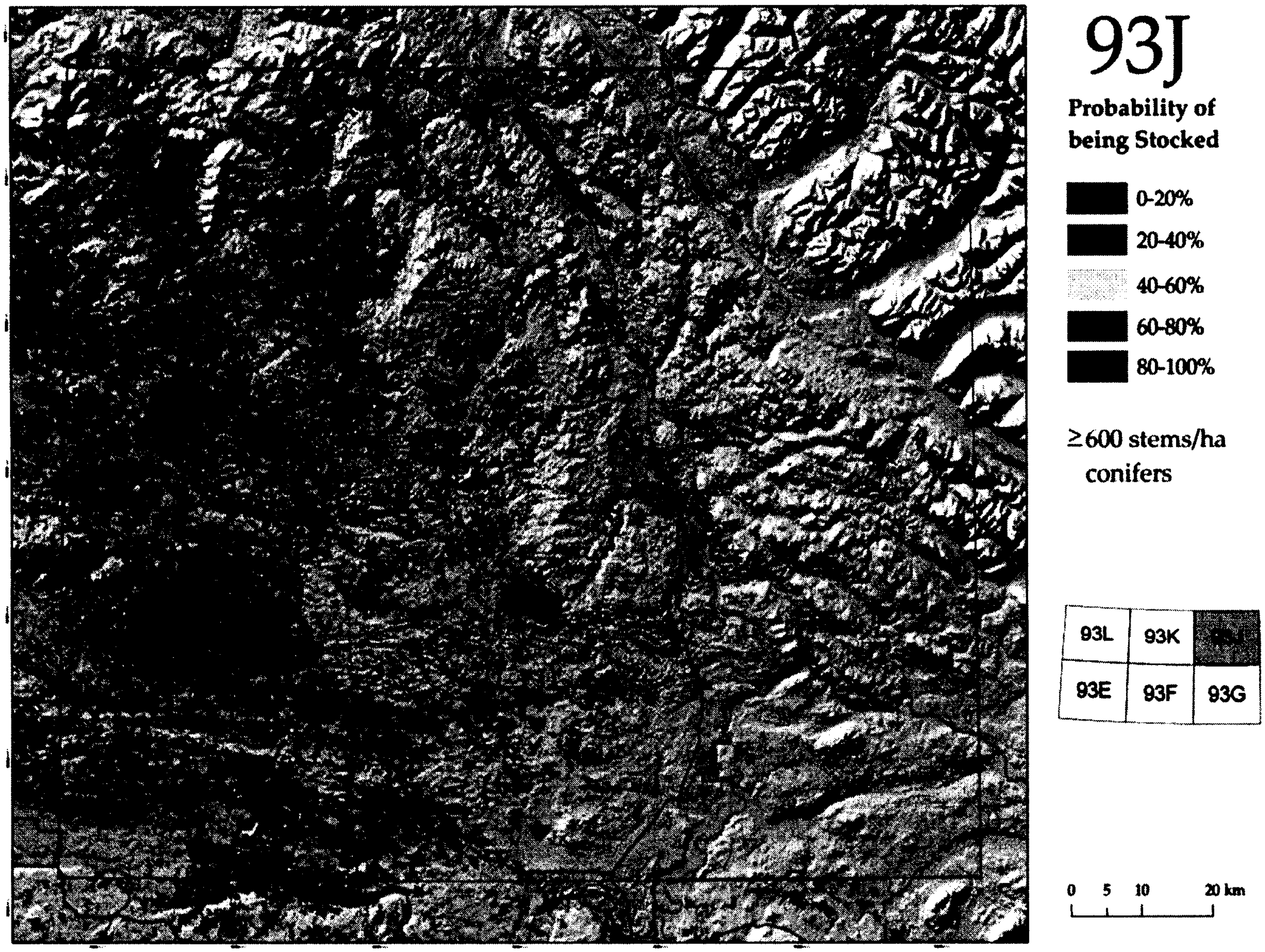
Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia



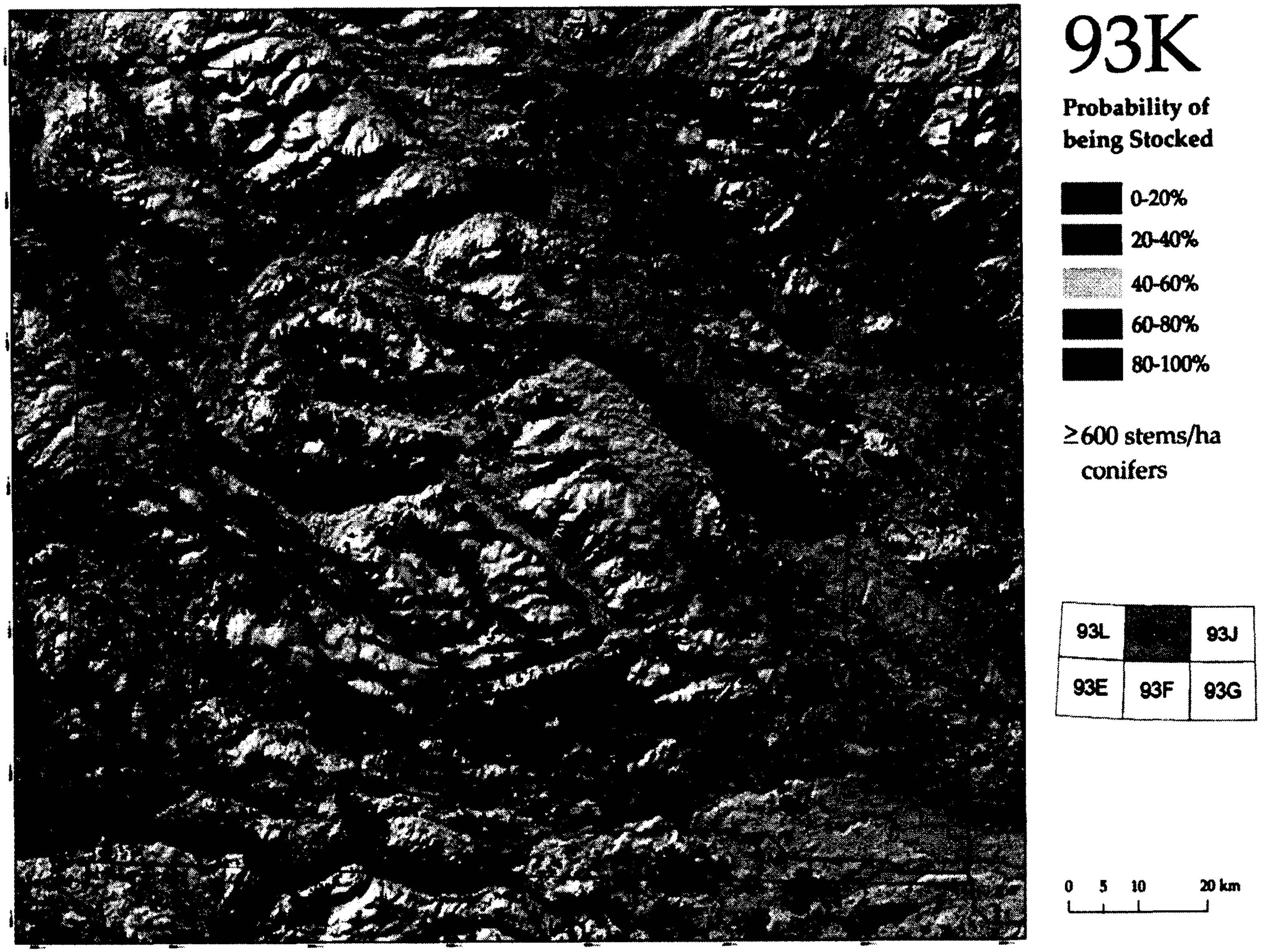
## Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia



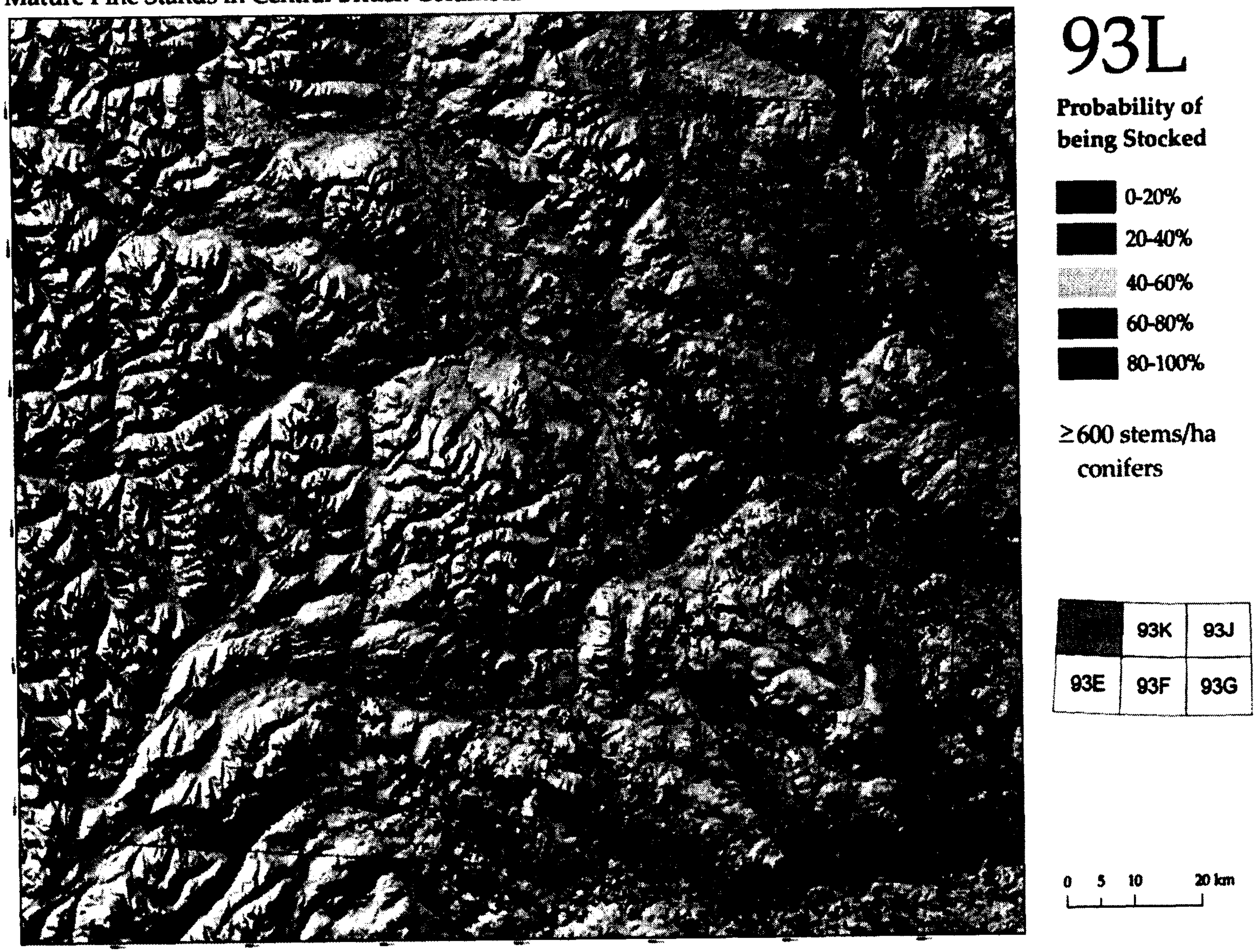
Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia



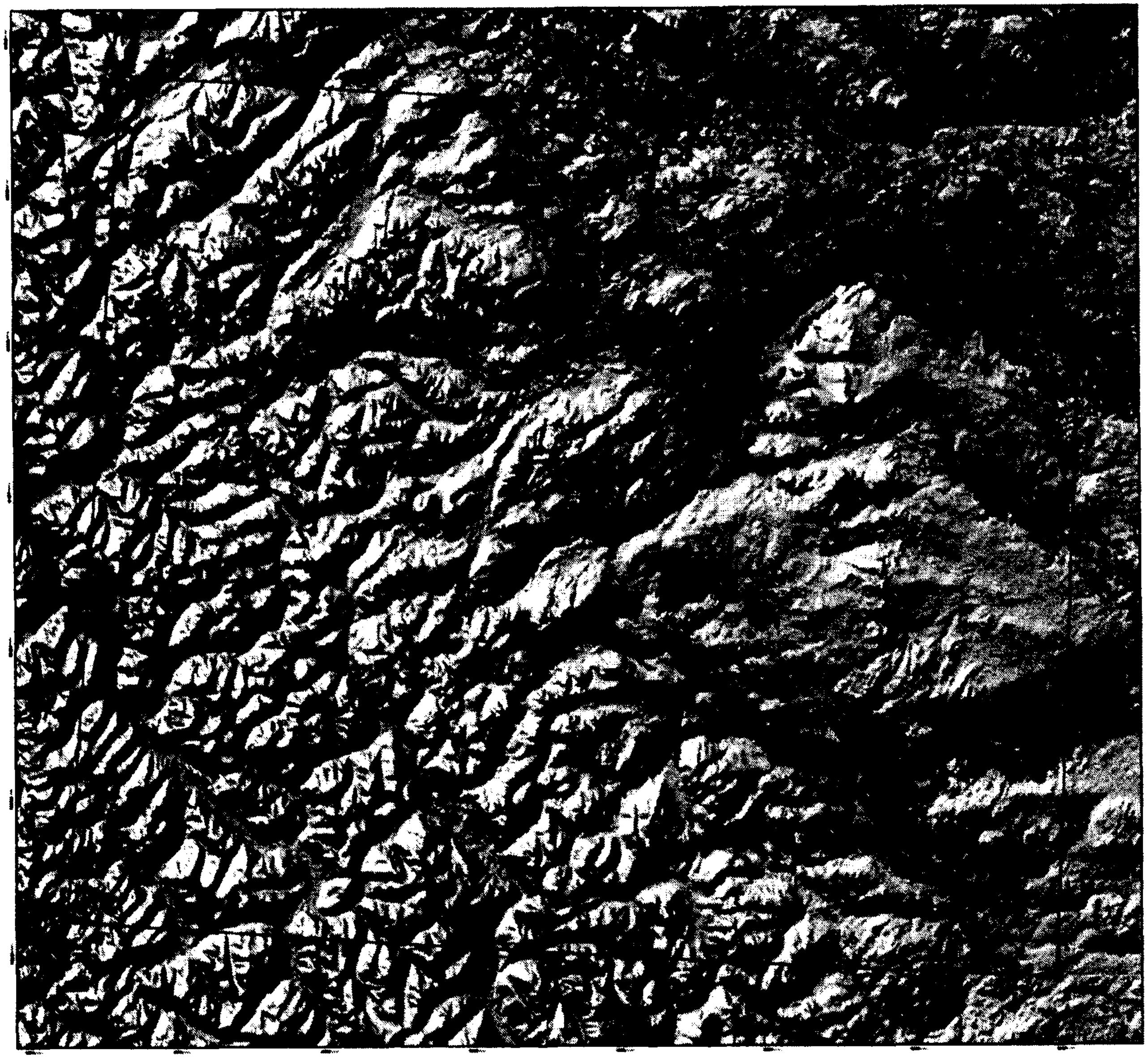
Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia



Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia



Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia

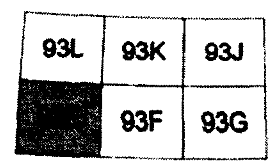


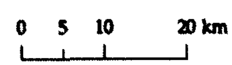
# 93E

**Probability of being Stocked**

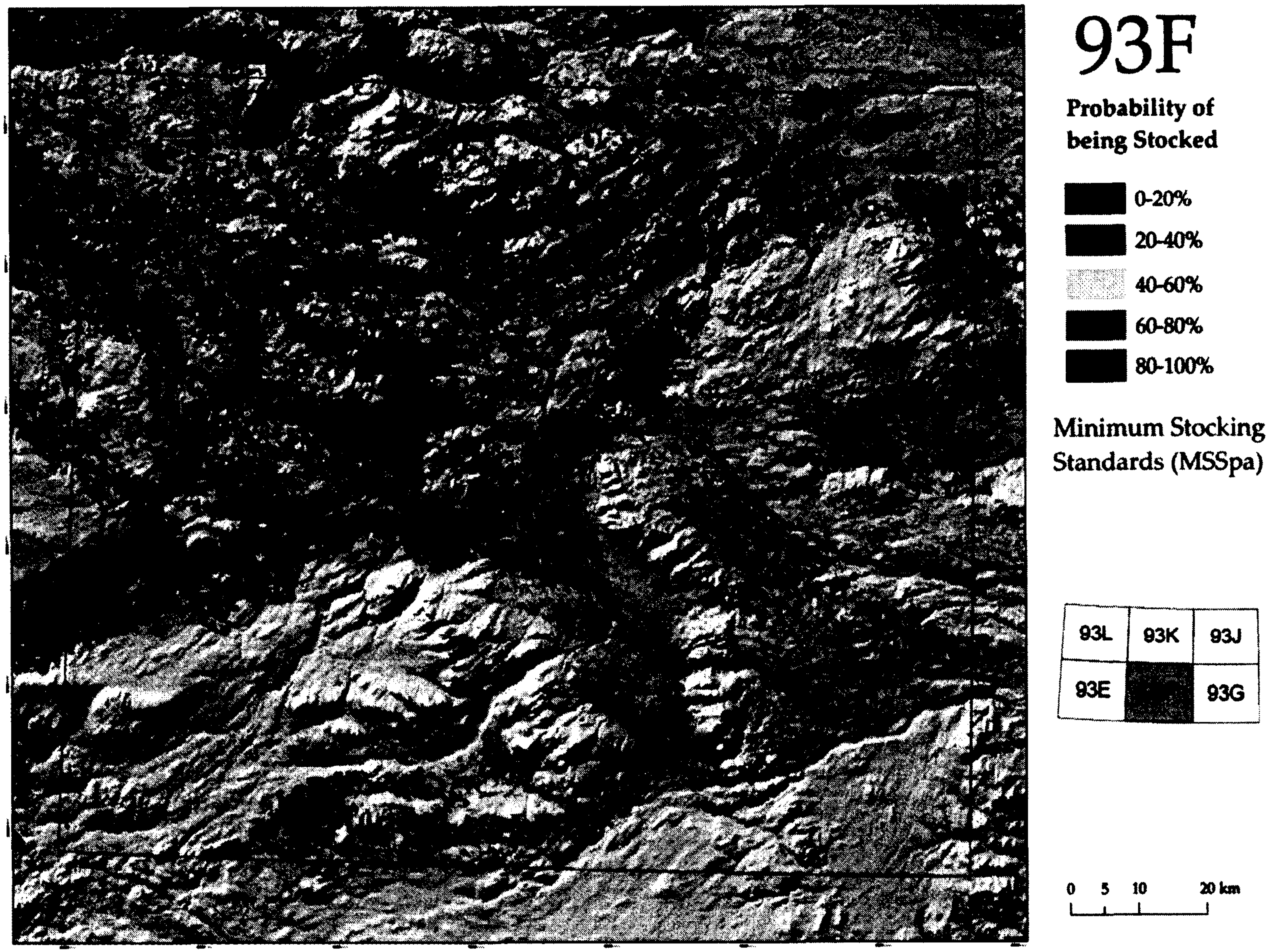
- 0-20%
- 20-40%
- 40-60%
- 60-80%
- 80-100%

Minimum Stocking Standards (MSSpa)



Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia



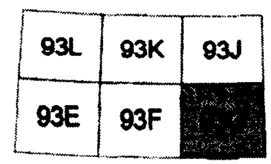
Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia

# 93G

**Probability of being Stocked**

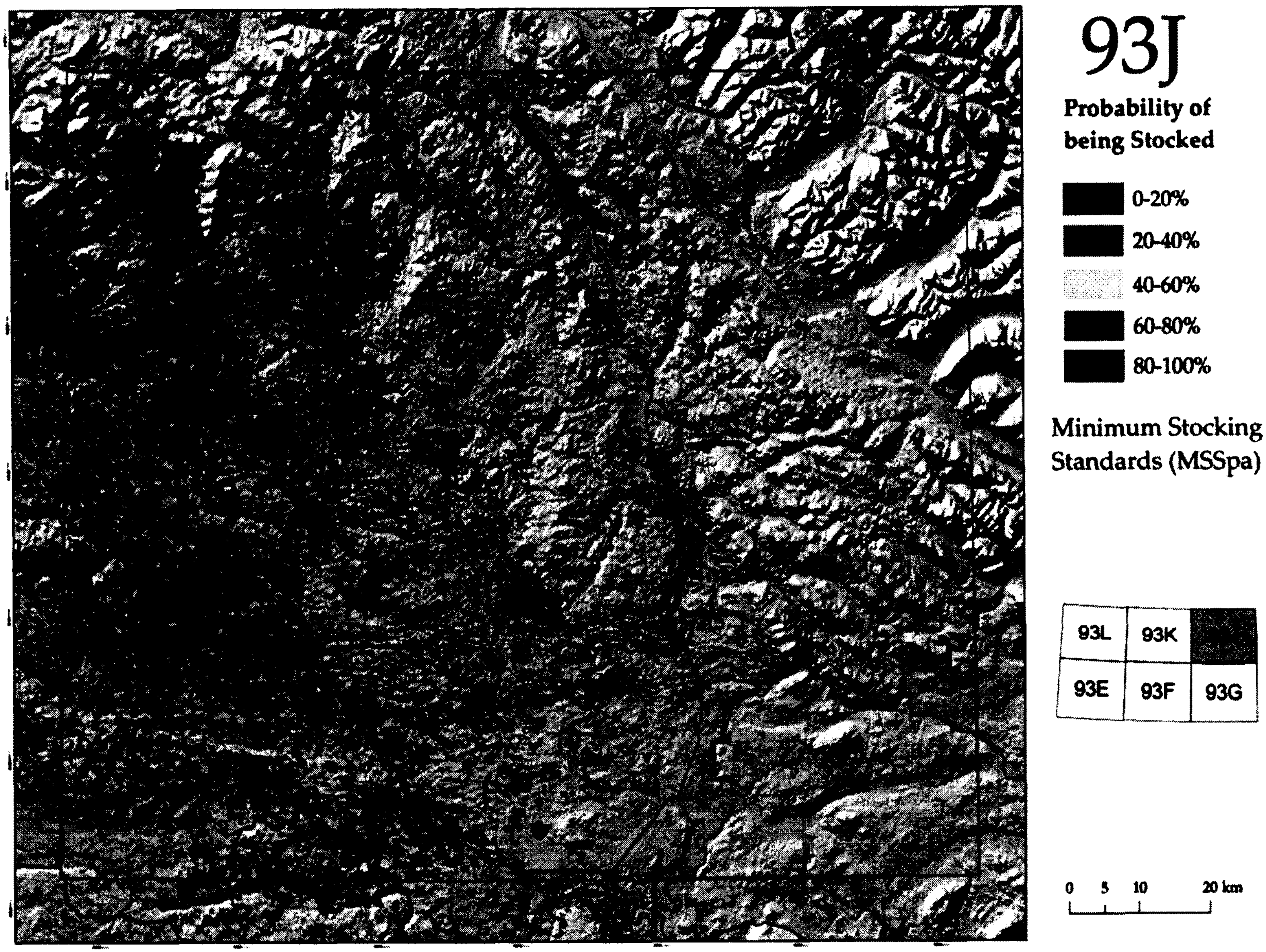
- 0-20%
- 20-40%
- 40-60%
- 60-80%
- 80-100%

Minimum Stocking Standards (MSSpa)

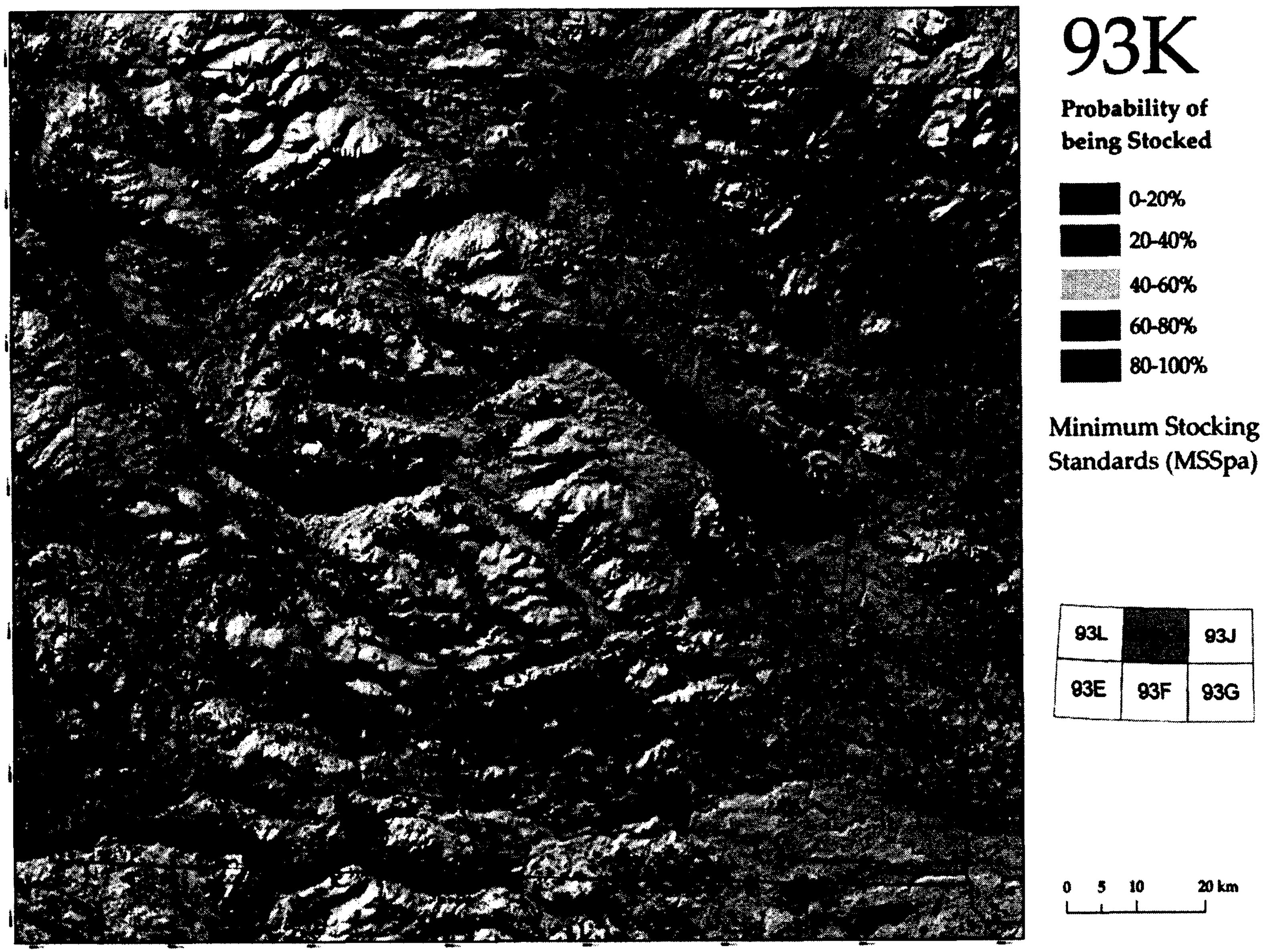



0 5 10 20 km

## Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia



Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia



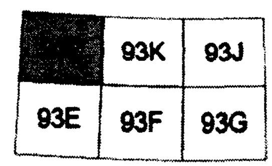
Predicted Locations and Stocking Probability of Advance Regeneration Under Mature Pine Stands in Central British Columbia

# 93L

**Probability of being Stocked**

- 0-20%
- 20-40%
- 40-60%
- 60-80%
- 80-100%

Minimum Stocking Standards (MSSpa)




0 5 10 20 km

## Appendix 3

### Classification Tree Analysis: Recursive Partitioning, Cross Validation, and Variable Importance

For classification trees built with a categorical target variable, the determination of what category to assign a node is more complex: it is the category that minimizes the misclassification cost for the observations in the node. In the simplest case, every row that is misclassified has a cost of 1 and every row that is correctly classified has a cost of 0.

A problematic issue in recursive partitioning is the decision of how large to build the tree (Breiman et al., 1994; Khoshgoftaar and Allen, 2001). Too large a tree means excessive branches, which in turn, mean excessive nodes. This represents an over-fitting of the model. If two trees provide equivalent predictive accuracy, the simpler tree is preferred because it is less sensitive to outliers and spurious observations, easier to understand, and faster to use for making predictions. Limits must be placed on the size of the resultant tree. Without limits, a tree could be built so large that there is terminal node for every case in the original dataset. In addition to it being computationally expensive, this situation would represent a solution that would be too difficult to interpret and it would have no applicability to new cases, i.e., the model would be over-fit. Pre-pruning of the tree can occur by simply providing limits to the classification routine. The analyst can either delimit the number of observations necessary for a node split or program the maximum number of branch levels allowed to be calculated. Both methods will artificially stop the classification splitting before the tree becomes too large.

Generally, the classification is provided a generous berth and stops when there are no more statistically significant splits to be made, (i.e., maximum purity of nodes is achieved). This strategy usually results in a larger tree than necessary with over-fitting. A tree with maximum nodes will result in the best fit model for the dataset. This, however, is not necessarily the best model for all datasets. There is a decision cost associated with producing an overfitted model, necessitating the identification of the optimal sized tree that will best fit subsequent datasets. Parametric models use penalization strategies such as Akaike Information Criterion (AIC) model fit measurements to ensure parsimony is achieved. In the case of nonparametric modelling, a widely accepted method of model fit is tree pruning (Strobl, 2009). Through the use of v-fold cross validation – one pruning method — the resultant can be pruned back to an optimal tree size (Dhurandhar and Dobra, 2008). Cross validation is widely used statistical strategy to evaluate or compare learning algorithms or decision tree models through a generalization error. The role of the v-fold cross validation is to separate the data into a test group and a validation group (folds). These groups, however, are created in such a manner that each point is 'crossed-over' between being a test point and validation point, ensuring that every data point is evaluated (Refaeilzadeh et al., 2009). This method differs from the hold-out method, where data isolated as a validation set are never evaluated. In v-fold cross validation, the working dataset is separated into v equal parts (10 equal parts in the case of 10 v-fold). A test classification tree is built with v – 1 held back for validation. The training data are run and the test data are run as an independent check against the training data for accuracy. The results of both are compared and then stored as the initial test. The process of splitting off the first v-1 test data against training data is automatically conducted nine more times (in the case of 10 v-fold), each time with a new independent validation dataset. Once the process has been run 10 times, the classification error rate calculated for each of the ten test runs are averaged together to provide a generalization error or cross-validation cost (Dhurandhar and Dobra, 2008). The tree size that produces the minimum cross validation cost is pruned to the number of nodes matching the size that produces the minimum cross validation cost. The literature does not indicate that using more than ten folds improves the accuracy of the generalized error Backward pruning requires significantly more calculations than forward pruning, but the tree sizes are much more optimally calculated (Sherrod, 2006).

When the target variable and the predictor variables are categorical (and both are multivariate, i.e., more than two categories), this creates a more mathematically complex process. To perform an exhaustive search, the classifier must evaluate a potential split for every possible combination of categories of the predictor variable. The number of splits is equal to $2^{(k-1)}-1$ where k is the number of categories of the predictor variable. For example, if there are 5 categories, 15 splits are tried; if there are 10 categories, 511 splits are tried; if there are 16 categories, 32,767 splits are tried; if there are 32 categories, 2,147,483,647 splits are tried. Because of this exponential growth, the computation time to do an exhaustive search becomes prohibitive when there are more than about 12 predictor categories (Sherrod, P.H. DTREG Predictive Modeling Software, personal communication, March 10, 2011).

If the target variable is binary and has only two possible categories, as is the case for a stocked or non-stocked condition as posed in my thesis, the exhaustive search is conducted with efficiency. The ideal split would divide a group into two nodes in such a way that all of the observations in the left node are the same (have the same value as the target variable) and all of the observations in the right node are the same – but different from the left node. This is referred to as purity. If such a split can be found, then you can exactly and perfectly classify all of the observations by using just that split, and no further splits are necessary or useful. Such a perfect split is possible only if the observations in the node being split have only two possible values of the target variable.

Unfortunately, perfect splits do not occur often in nature, so it is necessary to evaluate and compare the quality of imperfect splits. Various criteria have been proposed for evaluating splits, but they all have the same basic goal, which is to favour homogeneity within each right/left node and heterogeneity between the right/left nodes. The heterogeneity, or dispersion, of target categories within a node is called the "node impurity". The goal of splitting is to produce nodes with minimum impurity.

The impurity of every node is calculated by examining the distribution of categories of the target variable for the rows in the group. A "pure" node, where all rows have the same value of the target variable, has an impurity value of 0 (zero). When a potential split is evaluated, the weighted average of the impurities of the two nodes is subtracted from the impurity of the node from which they were split. This reduction in impurity is called the improvement of the split. The split with the greatest improvement is the one used. Improvement values for splits are shown in the node information that is part of the generated report.

Variable importance is the relative importance that each variable plays in the splitting of the tree into nodes, both as primary or surrogate splitters. A surrogate splitter is an imputation technique that is employed when rows of the dataset have missing values. If a variable is called on as a primary splitter in the building of a tree and it has missing data, the developed surrogate will take its place and conduct the split as it was developed in the initial tree building (Acuna and Rodriquez, 2004). It is important to note that a variable's importance is not measured solely by how early it enters the tree to act as a splitter. A strong surrogate splitter may be more "important" to the classification model even though it enters the tree later than a weaker primary splitter (Lewis, 2000; Sherrod, 2006). The loss of an important variable in the decision model will likely weaken the model as a whole. Variable importance can also act as a recruiter for subsequent analysis, as it is clear that a variable with a higher variable importance score is likely a significant predictor of the response variable (Breiman, 2001). The variable that contributes the highest improvement measure (i.e., the variable that has the greatest effect on error rate increase) achieves a score of 100 (Banerjee et al., 2008). The measures are based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Friedman and Meulman, 2003; Strobl et al., 2007). The relative influence (or contribution) of each variable is scaled so that the sum adds to 100, with higher numbers indicating stronger influence on the response (Elith et al., 2008). The remaining contributing predictor variables are scored relative to the most important variable. Importance, however, may refer to how important a variable is to the overall goodness of model fit, or it may refer to how important the variable is to the model's predictive ability.