Modelling forest composition after a mountain pine beetle epidemic

Colin Robertson[†]* and Trisalyn A. Nelson[††]

[†]Department of Geography, University of Victoria, Victoria, British Columbia, V8W 3P5, Canada. Email: colinr23@gmail.com

[††]Department of Geography, University of Victoria, Victoria, British Columbia, V8W 3P5, Canada. Email: tnelson@mail.geog.uvic.ca

* Corresponding Author: Colin Robertson, University of Victoria, Department of Geography, Victoria, British Columbia, V8W 3P5, Canada
email: colinr23@gmail.com, phone: 250-885-8065

Abstract

The mountain pine beetle (*Dendroctonus ponderosae* Coleoptera: Hopkins) epidemic currently taking place in western Canada is changing the structure of lodgepole pine forests. The altered landscape has created a new challenge for forest managers and requires mapping of the current forest conditions. The goal of this research is to develop a method for mapping the cumulative impact of intense beetle infestation on forest composition. By combining existing data sets on mountain pine beetle infestations, terrain, climate, and forest conditions with regression tree analysis we propose a method for mapping the spatial variability in mountian pine beetle caused pine mortality. To assess the methodology for large-area mapping of post-beetle forest composition, modelling was repeated at multiple spatial scales, and attributed with both fine scale field data and coarse sacle forest inventory polygon data. The most important variables for predicting PPI are percent pine, stem density and the number of nearby infested trees . Accuracies of classifications were within +/- 25% percentage of pine infested between 54-91% of the models, and within +/- 15% percentage of pine infested between 9-64% of the models, depending on the spatial scale and input data used. The methodology is a flexible approach to modelling and mapping rapidly changing forest composition given operational data constraints.

Word count: 208

Keywords : mountain pine beetle, regression tree, infestation severity, forest monitoring, forest mapping

Introduction

The mountain pine beetle, *Dendroctonus ponderosae* (Coleoptera: Hopkins), is a native species of many forest ecosystems in North America (Cole and Amman 1980). Periodically, beetle populations rise to epidemic levels impacting their environment with severe and spatially extensive infestation. Currently, the mountain pine beetle is affecting forests in western Canada on an unprecedented scale: over 9.2 million ha of forest are infested in British Columbia (B.C.) (Westfall 2007). The extent and severity of the current epidemic is due to two key factors. First, lodgepole pine (*Pinus contorta* Dougl. Ex Loud. Var. *latifolia* Engelm.) is the primary host for mountain pine beetle, and the amount of mature lodgepole pine in large contiguous stands has increased due to fire suppression (Taylor and Carroll 2004). Secondly, a warming climate has helped the success of beetle populations by reducing winter brood mortality and allowing the beetle to expand into previously unsuitable habitats (Carroll et al. 2004).

Most large-area maps of mountain pine beetle epidemics delineate the area impacted by the infestation providing no details of the spatial variability of the infestation magnitude or the impact to the forest (Wulder et al. 2006). These maps give the impression that the infestation is uniform in intensity and all trees within the infested region are dead. Such representations do little to assist forest managers tasked with setting priorities for harvesting and treatment. To aid management, updated forest inventory information is required in regions infested by the mountain pine beetle; knowing what trees remain is essential.

While information about the amount of remaining pine is necessary in the short term, the spatial extent of the infestation makes traditional approaches to forest inventory updating impractical due to the time and cost associated with collecting aerial photography or satellite imagery. However, existing forest health monitoring data, which quantify the location and magnitude of infestations, can be combined with forest inventories, which provide data on forest conditions prior to infestation, to model post-mountain pine beetle forest composition.

Modelling the post-mountain pine beetle forest composition requires an understanding of the forest characteristics that lead to spatial variability in the amount of pine that remains in a stand. Susceptibility modelling by Shore and Safranyik (1992) has related stand characteristics to the likelihood of attack and damage by mountain pine beetles. Beetles have historically shown a preference for large diameter trees (Cole and Amman 1980). The mechanism by which beetles select large diameter trees is unknown, but the result is that when large trees are infested, brood production is enhanced the following year. Mass attacked trees having a diameter at breast height (dbh) larger than 25.4 cm are thought to be a source of beetles in the year following their infestation, while smaller trees act as infestation sinks (Safranyik et al 1974). Tree dbh is therefore an important variable in modelling forest damage caused by mountain pine beetle infestations (Shore and Safranyik 1992). Tree size increases with age, so stand age is also related to the susceptibility of a stand to mountain pine beetle attack. However, as lodgepole pine defenses (i.e., resin production) also vary with tree age, the likelihood of infestation starts to increase for trees greater than 60 years of age (Safranyik et al. 1974). Large trees with

thick, healthy phloem offer the most desirable hosts for mountain pine beetles, but typically are only infested when populations reach epidemic levels (Shore et al. 2006). Additional variables often used in modelling a stand's susceptibility to mountain pine beetle caused pine mortality are stand density, with beetles preferring moderately dense stands, and geographic location, which factors in latitude, longitude, and elevation (Shore and Safranyik 1992).

The impact of infestation intensity and duration is also an important factor when modelling the amount of pine that will remain post-mountain pine beetle infestation, yet quantitative research in this area is more limited. Generally, the number of trees with dbh greater than 10 cm will be proportional to the total pine mortality. The size of the beetle population, and potential for brood production, are also related to the amount of pine mortality (Safranyik and Carroll 2006), although other factors complicate this relationship, especially at a sub-stand scale. Such factors include topographic variables (Amman 1973; Safranyik et al. 1973), the susceptibility of adjacent areas (Safranyik et al. 1992), and long range dispersal (Furniss and Furniss 1972). Previous research has also suggested that beetle dispersal patterns may vary in different environmental conditions (Robertson et al. in press), so climatic and site characteristics may affect the infestation severity at different locations. The temporal aspect of the relationships between these factors, infestation severity and pine mortality is largely unknown.

The goal of this paper is to develop an approach to map forest composition, or the amount of remaining pine after mountain pine beetle infestation, using existing data

sources. Regression tree classifiers are used to model percent pine infested (PPI) based on forest health monitoring data and variables known to influence the spatial variability of forest susceptibility to infestation (e.g., forest age, percentage, pine elevation). Given that ecological relationships are scale dependent (Levin 1992), we anticipate the input variables that will generate the most accurate regression tree model will vary depending on the spatial scale of model output (Woodward 1987; Nelson et al. 2006a). Therefore, the impact of spatial resolution of regression tree model is also explored. To ensure the method developed is operational, the effect of data on classification accuracy is also investigated by comparing the output of regression tree models that are generated using different data sets. The first data sets covers a large area but lacks attribute detail (the forest inventory), while the second data set is available for only a small area but has more detailed attribution.

Study Area

The research study area is the Morice Timber Supply Area (TSA), in west central (54º 24' N 126º 38' W) British Columbia (Figure 1). The current epidemic in the Morice area dates back to the early 1990s (Nelson et al. 2006b). Lodgepole pine cover approximately 50% of the forested landbase in the TSA, and the northern region approximates the northwestern range of mountain pine beetle suitable habitat. Beetle populations are known to be univoltine throughout the TSA, meaning beetles emerge once per year in late summer. Other dominant species include hybrid white spruce (*Picea egelmannii* (Parry) *x glauca* (Moench) Voss) and subalpine fir (*Abies lasiocarpa* (Hook.) Nutt). The TSA lies primarily in the Sub-Boreal Spruce (SBS) Biogeoclimatic Ecosystem Classification (BEC) zone (although high elevations are in the Englemann Spruce – Subalpine Fir zone). The BEC system classifies landscapes based on site and climatic features. Zones are areas influenced by the same regional climate. Subzones, divided into zonal (typical), and variant (atypical), are geographically related ecosystems (Eng and Meidinger 1999). Within the SBS, there are two subzones: Moist Cold (SBSmc) at moderate elevations, and the warmer Dry Cool (SBSdk) subzone at lower elevations (Meidinger et al. 1991).

Data

Aerial Surveys

Mountain pine beetle monitoring in the Morice TSA has been extensive. Heli-GPS surveys, undertaken annually since 1995, provide yearly datasets of outbreak progression. Through helicopter based surveys the locations of infested trees are recorded with GPS points, and the number of infested trees estimated (Nelson et al. 2006b). Crown foliage discolouration is used as a proxy for infested trees. Once a tree is infested, it takes approximately one year for the foliage to turn from green to red. Two years after infestation an attacked tree typically appears grey as the needles have dropped off (Safranyik et al 1974). The area of infestation represented by each GPS point recorded in a helicopter survey varies, but is estimated to have a maximum area of a circle with a 100 m radius (Nelson et al. 2006b). Field verification of heli-GPS data has found that estimates at each point location are accurate to within 10 trees, 92.6% of the time (Nelson et al. 2006b). A limitation of heli-GPS surveys is that all locations are mapped as clusters of trees (i.e., one point with an attribute for number of trees), when in reality, the spatial distribution of infested trees may be clustered or dispersed (Wulder et al. 2006). This type of representation is used because mountain pine beetle attacked trees are typically found as ‘spot infestations’, due to the pheromone dynamics associated with mass attacking large trees (see Geiszler et al. 1980), however at epidemic population levels, infestations occur as large contiguous patches (Safranyik and Carroll 2006). One approach to dealing with this source of uncertainty is to aggregate the counts of the damaged trees over a larger spatial area to form a measure of infestation density (Nelson et al. 2006b).

Field Surveys

Detailed data on mountain pine beetle infested forest composition were collected through plot-based field surveys that were undertaken in the Morice TSA during the summer months of 2003 and 2004. Plots were established at locations identified as infested during the heli-GPS surveys. Thus, all plots were located where mountain pine beetles had been present for at least one year. Plots were circular with an area of 200 m^2 (or .02 ha). In each plot, all trees with dbh > 10 cm were surveyed, and trees were categorized as healthy pine, newly attacked pine (green foliage), previously attacked pine (red foliage), dead pine (grey foliage), or non-pine. Previously attacked and dead pine were summed and divided by the total number of pine trees in the plot, giving the percentage of pine infested (PPI). Recently attacked trees with green foliage were excluded as they can not be surveyed aerially. Within the Morice TSA, 231 plots were established in 2003, and 144 were established in 2004, totaling 375 locations where the percentage of pine infested by mountain pine beetles is known.

Forest Inventory Database

British Columbia Ministry of Forests' inventory is the primary data set available for characterizing forest conditions across the province. The forest inventory polygon (FIP) attributes used in this study were forest age, percentage pine, and dbh. Forest inventory data are primarily generated via aerial photo (1:15,000) interpretation (Leckie and Gillis 1995) and were last updated in the Morice TSA in the mid 1990s. Stand age is the age, in

years, of the forest stand on the date of polygon classification, and is based on ring counts from a bored core or is estimated from aerial photographs. Tree species and percentages are given for the six most dominant forest species. The dominant tree species, occupies the largest percentage of the stand. Both the species and the percentage of the stand composed of this species are recorded. In the forest inventory, dbh is reported as the quadratic mean stand dbh and is calculated based on forest age and site index. The quadratic mean is the square root of the arithmetic mean of squared values and gives greater weight to large trees. Typically, the quadratic mean is used for calculating stand dbh as it relates to other forest characteristics more meaningfully than does the arithmetic mean (Curtis and Marshall 2000).

Ancillary Datasets

Additional ancillary datasets were available for the Morice TSA and used in this study. A digital elevation model (DEM) with grid cells of 100m by 100m was used to derive topographic variables (Province of British Columbia 1996). The DEM was created from 1:20,000 scale Terrain Research Information Management data using linear interpolation and is accurate within 10 m (Province of British Columbia 1996). A dataset of BEC zones was obtained for the Morice TSA. This dataset represents the BEC zones and subzones present in the study area. In previous work, subzone classifications, SBSmc and SBSdk, have shown marked trends in movement patterns indicative of mountain pine beetle dispersal processes (Robertson et al. in press).

Methods

Our methodological approach to mapping forest composition is based on the use of regression tree classification, a method capable of modeling complex, nonlinear relationships. We used ArcGIS® software (ESRI 2005) for all GIS operations and mapping, and used R (Ihaka and Gentleman 1996) for all statistical analysis. Research methods are categorized into four groups. First, the data pre-processing methods are described. Secondly, the regression tree classification method is outlined. Thirdly, the approach used for accuracy assessment is described. Finally, the methods applied to assess the impact of the regression tree model spatial resolution and input data on model accuracy are presented.

1) Data and Data Pre-Processing

Variables used with regression tree classification reflect the infestation conditions, terrain and climate factors, and forest conditions (Table 1). The heli-GPS monitoring data was used to incorporate infestation magnitude and duration into the model of remaining pine forest. The intensity of infestation at each location was estimated by counting the number of trees killed by mountain pine beetles recorded in the heli-GPS data within a 2000 m radius of each plot centroid (DEAD). This distance threshold is a conservative estimate of the maximum short range dispersal distance of a mountain pine beetle through unsuitable habitat (Barclay et al. 2005). This variable should therefore be an indicator of the size of the local beetle population. Infestation duration (ID) was also derived for each plot centroid in the study area. For each point, heli-GPS points within

200 m were located. The first year that a point was found within 200 m of the grid cell centroid was identified as the beginning of infestation at that location. The 200 m threshold was used here to ensure that beetles were within a distance immediately within attack range (Safranyik et al. 1992).

In order to incorporate both infestation intensity (the number of dead trees) and infestation duration, an index called impact factor (IF) was calculated for each plot. The IF was calculated as the number of dead trees divided by the ID at a particular location.

$$IF = \frac{\sum_{i=1}^{j} DEAD}{(j-i)} \quad [1]$$

Where i is the first year of the infestation data, and j is the latest year of infestation data. The IF was used to be representative of different stages of infestation. For example, with a finite number of trees at a location, infestation at a particular intensity can be only be sustained until all trees are infested. The IF is designed to differentiate new, intense infestations (high impact) where beetles have perhaps migrated from adjacent areas, from older infestations where local populations have built up from incipient levels over several years (medium / low impact). From an ecological perspective, high impact infestations may represent beetle population increase in an area due to dispersal, while low impact infestations are increasing due to environmental factors.(Moran effect). In reality, these influences on beetle populations are often difficult to decompose.

To include the impacts of terrain in the post-beetle forest composition model, a DEM was used to quantify elevation (EL) and to calculate the slope or inclination (SLP) and aspect

or orientation (ASP) of the surface. Slope has previously been identified as an important variable for predicting locations of mountain pine beetle infested trees at a landscape scale (Wulder et al. 2006b). Additionally, southern aspect sites are generally more productive for lodgepole pine than other aspects, and have been found to be associated with newly established infestations (Safranyik et al. 1974). A cosine transformation of aspect provided a linear scale for the 'northness' of the slope direction from -1 (due south), to +1 (due north). A sine transformation of aspect provided a linear scale for the 'eastness' of the slope orientation, from -1 (due west) to +1 (due east). Previous research has successfully used these transformations for regression modelling of aspect (Schwarz et al. 2003).

Climatic and terrain impacts were included in the model by calculating direct solar radiation (SWR) using the methods of Kumar et al. (1997). Consideration of terrain through solar radiation was useful given the complex terrain of the study area, where elevation ranges from approximately 545 m to 2730 m over an area of 1.5 million ha. Direct solar radiation has been demonstrated to be a significant variable for predicting locations of previously infested trees (red foliage) in pine stands (Coops et al. 2006). Calculations for potential direct solar radiation require parameters for terrain (DEM), latitude, day of the year, and time interval at which to make calculations. Kumar et al. (1997) suggest that short time intervals have greater accuracy, so calculations were made every 30 minutes. Consequently, for computational efficiency, two summer days (Aug. 1-2) were selected to be representative of spatial variation in direct solar radiation. These

dates approximate the period of mountain pine beetle emergence in the study area and the season when incoming solar radiation is greatest.

Variables on forest conditions known to influence forest susceptibility to the mountain pine beetle were also included. Data on forest age (AGE) was determined from the forest inventory data. Plot percentage pine (PPERP) and stem density (DENSE) were acquired from field plots. When assessing the impact of different input data on regression tree output using large area data sets with more general information, stand percentage pine (PERP) was used instead of plot percentage pine. Mean stand quadratic dbh (QDBH) was also utilized from the forest inventory data, and no density data were available. The true value for the variable being modeled, the percentage of pine infested (PPI), was always determined from the plot data.

2) Statistical Analysis: Regression Trees

Regression tree (RT) modelling was used to construct models of the proportion of infested trees (Breiman et al. 1984). The regression tree approach has traditionally been employed in fields such as medical diagnosis (Ridley et al. 1998) and machine learning (Quinlan 1986), but is increasingly being used in ecological analyses (De'ath and Fabricius 2000). There are attractive advantages of regression tree analysis over traditional regression: distributional assumptions of variables are not required, complex interactions between variables can be accurately handled, and the methods are robust with respect to outliers and missing values, all of which are common in ecological data (Prasad et al. 2006). Some examples of regression tree modeling applied in forestry

include predicting tree mortality (Dobbertin and Biging 1998; Fan et al. 2006) and forest health applications (Rosso and Hansen 2003; Negron and Popp 2004; Coops et al. 2006).

The RT approach works by recursively partitioning response variables into mutually exclusive groups. This process is referred to as "growing" the classifier (tree). At each stage, the predictor variables are assessed to determine which variable can be used to split the dependent variable into the two most homogeneous groups. The variable that achieves the most homogeneity is used to split the data, and a node is generated with two sub-branches. Each node represents the partition, among a number of candidate partitions in the set of predictor variables, that maximizes the homogeneity of the two resultant groups. The RT analysis tries to accurately classify the response variable into groups that minimize within-group variance and maximize between-group variance. Additionally, a RT should be sufficiently small (size measured by the number of final classes, called terminal nodes), so that the partitioning of the data is meaningful, and useful for prediction on new data. The general methodology proposed by Breiman et al. (1984) for growing an optimal RT is to continue splitting the response to form a very large RT, a model that has a large number of binary splits, and then reduce or prune nodes, using cross validation to select an optimal RT size.

Deviance, the sum of squares about each candidate group mean, was used as the partitioning criterion to grow trees into sequences of binary decisions on the predictor variables. The mean deviance across all terminal nodes, residual mean deviance (RMD), was used to assess the fit of a RT to the data (Ripley 1996). The RMD quantifies the

average heterogeneity in the terminal node classes. Minimizing the RMD is a way of reducing group impurity, but does not select optimal trees because it does not consider tree size. The RMD is biased downward: it will always be minimized with the largest trees because the terminal node classes will be composed of the fewest observations (Breiman et al. 1984). This is analogous to stepwise linear regression, where adding variables always increases the R^2 for the model. Pruning addresses this problem for RTs by estimating the 'unbiased deviance'.

We used ten-fold cross validation as a pruning method to obtain estimates of the true deviance for each tree size ('honest estimates' in Breiman et al. 1984). In the cross validation procedure, the dataset is split into ten groups, RTs are grown based on nine groups and tested on the 10^{th} group. This is performed on each of the groups separately and deviances are averaged over all groups. Following the procedure recommended for prediction by Ripley (1996), we repeated the cross validation ten times. The largest and smallest RTs that minimized the cross validated deviance were selected for testing. We used two sizes as a way of balancing our objectives of description and prediction.

When the RT model was implemented in this study, it was carried out using a random sample of 66% of the field plot locations for 2003 and 2004 (n = 254). Once rules were generated for field plots, which had known PPI, the model could be applied to classify the entire study area using the data available. Each time a RT model is generated all variables are used as input; however, only some are utilized by the model for classification (Table 2).

3) Accuracy Assessment

The 34% of data (n = 121) not used for building the RT model were reserved for assessing the accuracy of the model output. Accuracy was categorized as +/- 25% and +/- 15%. Observed and predicted values of pine mortality were also converted into the intensity classes used in province wide assessments of forest damage in B.C. (see Westfall 2007). Additionally, prediction maps were assessed visually.

4 Assessing the Impact of Model Resolution and Input Data

To assess the impact of spatial scale on our RT model output, analysis was repeated with model output resolution equal to the original field plots (0.02 ha), as well as, 100 ha grid cells, 2500 ha grid cells and 10000 ha grid cells. These resolutions represent potential cell sizes for mapping at the landscape, regional and national levels. The mean forest inventory polygon area in the Morice TSA is 20 ha, so grid cell sizes were selected larger than this threshold to avoid repeated observations from the same polygon.

To assess our methodology for application in large-area mapping, we investigated the impact of data on the analysis by using both field plot and large-area FIP data sets to attribute the RT model. A key difference in these data sets is that for the plot data, percentage of pine and stem density were calculated based on field observations within 0.02 ha plots. In the FIP data, percentage pine is estimated for forest inventory polygons where pine is included as one of the six dominant species. Percentage pine measurements will differ between the FIP and plot data for two reasons. Firstly, the

percentage pine in the FIP data is estimated, based on photo interpretation and ground-truthing, and is therefore subject to greater error than the plot data. Secondly, the FIP data are collected at a stand scale while field data are collected for plots that are sub-stand in size. As such, one would expect forest attribute values from these two data sources to differ. Another important difference between FIP and field data is that density information is only available for the field plots. For analysis conducted with the FIP data a surrogate variable used for susceptibility modelling, quadratic mean dbh, is used (Nelson et al., 2006a).

For field plots, 78% of the random sample (198 of 254) used for training RTs were located in unique forest inventory polygons. For RTs generated from plot data, the values for each variable represent the mean of the values recorded in the field plots, while for RTs based on FIP data the values represented the mean of the all of the values within the grid cell. For example, in one 100 ha grid cell which contained three field plots, the plot RT value for elevation would be the average of the elevation values at all three plots, while for FIP RT, the elevation would be the average of all DEM values within the grid cell (i.e., 100 100m pixels). This is illustrated in Figure 2. In all regression trees, the dependent variable was the plot level PPI.

Results

Phase 1: Modelling infestation intensity for field plots.

For plot RTs at the 0.02 ha scale, a RT was grown to an initial size of 17 terminal nodes. Residual mean deviance for the initial RT was 0.03. Cross validation selected optimal tree sizes in the range of three to six terminal nodes. Over all of the test sample plots (n = 121), the six node RT predicted infestation intensity within 25% of the actual value for 60% of the data. The three node RT predicted within 25% accuracy for 58% of the data (Table 3a). Classes with a very high mortality due to infestation tended to be underestimated due to the averaging of class values that occurs when RTs are pruned. When plots with greater than 85% infestation were excluded from the test samples, accuracies increased to within 25%, 70% of the time, for both the three and six node RT (Table 4a).

In plot RTs classified at the 0.02 ha scale the most important variables for partitioning PPI were plot density (DENSE) and the percentage of pine observed in the field plot (PPERP). The six node RT is presented in Figure 3. The left side reveals that plots where stem density is below 375 stems / ha are classified as either 100% infested, or 69% infested, depending on the percentage of the trees that are pine being below or above 45%. The right side decisions reveal that the lowest class, 21% dead pine, is found where greater than 40% of the plot's trees are pine, density is above 375 but below 675 stems / ha, and fewer than 132 dead trees are within a 2 km radius of the plot.

Phase 2: Modelling infestation intensity using regression trees and operational data.

Using FIP and topographic data averaged to a 1 km grid cell resolution, RTs were grown to classify average plot infestation level for plots within each grid cell. The results of this analysis are presented as dataset FIP-2 in Table 3b. The initial RT had 21 terminal nodes, RMD = 0.032. Cross validation selected optimal sizes within the range of 4 (RMD = 0.056) to 9 nodes (RMD = 0.046). Important variables for classification of infestation level were IF, AGE, PERP and SWR (Table 4). The nine node RT is presented in Figure 4. Percentage of pine was less important for variable splits for FIP data compared to the plot data. It is worth noting that the Pearson's correlation coefficient between the observed plot percentage of pine (PPERP) and the percentage of pine (PERP) reported in the forest inventory polygon the plot was located within indicated only a moderate association (r = 0.32).

When RT estimates were converted to the forest health rating classes, accuracy of the six node RT was 19%, and 31% for the three node RT (Table 2b). The 9 node RT was accurate to within 25% at 48% of the test samples. The 4 node RT predicted moderately better, at 55%. Prediction accuracy increased to 56% and 66% when the highest levels of infestation were excluded (Table 4b).

Phase 3: Assessing the impact of spatial resolution on regression tree variables and model accuracy.

Generally, as the grid cell resolution became coarser, and large geographic areas were aggregated, accuracies of the regression trees increased. This was evident in both the plot

data and the FIP data (Table 3a and 3b). Highest overall accuracies were achieved using a spatial extent of 10 km (or 10 000 ha). However, at high levels of aggregation, regression trees also tended to become unstable. For example, using FIP data represented with a cell size of 10 000 ha, infestation levels were predicted within 15%, 48% of the time using a 4 node tree, yet using the 6 node tree on the same data reduced the accuracy to 9% (Figure 2b). Also, at the 2500 ha resolution, the FIP data could not generate a tree better than random, so the unpruned regression tree is presented as FIP-3. For both the PLOT and FIP data, the 0.02 ha and 100 ha resolutions appear to have achieved the best, most robust results.

Across all spatial scales, for the plot data, DENSE appears to be the most important variable classifying infestation intensity (Table 2). It is intuitive that at a tree level, greater density will reduce the proportion of trees that are killed at any location. However, density information is rarely recorded in operational forest inventory data, such as the FIP data used in this analysis. Interestingly, our surrogate variable for density, stand quadratic mean dbh, was not included in any of the FIP RTs. The most common variable used in the FIP RTs was PERP. Other variables common to RTs computed with plot and FIP data include: DEAD (include in 9 out of 15 trees), IF (6/15), EL (6/15), SWR (5/15), and ID (3/15).

Mapping the percentage of pine remaining in the Morice TSA

Due to data availability, mapping the percentage of pine remaining in the Morice TSA, required the use of the FIP RTs. We selected 100 ha grid cells as an appropriate spatial

grain for mapping. The RTs at this scale were also composed of appropriate rules for mapping. For accuracy levels other than the FHI classes, the 4 node RT performed better. We used this RT for mapping pine mortality across all of the Morice TSA. We applied the modelled PPI values to the FIP data to summarize the percentage of pine remaining in each forest inventory polygon. This map is presented in Figure 5. Additionally, Figure 6 presents the percentage change caused by mountain pine beetle between 1995 and 2004.

Summary and Conclusions

Identifying the changes occurring in forests impacted by epidemic mountain pine beetle populations is crucial for forest management. The approach to mapping forest composition taken in this research was useful for identifying the important variables in predicting spatial variation in mountain pine beetle caused pine mortality. These were selected from a range of variables (Table 2) known to impact a forest's susceptibility to mountain pine beetle infestation. Across all RTs, the important predictor variables for partitioning PPI were percentage pine and the number of dead trees within a 2 km radius. Lower numbers of pine trees surrounded by high numbers of dead pine trees naturally corresponds to a high level of pine mortality. Interestingly, at smaller spatial grains IF was also an important variable for classifying the percentage of pine mortality. For example, in the FIP RTs, IF was included in all RTs computed at spatial resolutions of 0.02 ha and 100 ha (Table 2). This suggests that locally, IF may be a useful measure of fast moving, highly intense infestations. However, at broader spatial scales, perhaps larger scale climatic and environmental variables are more relevant for overall pine

mortality, such as SWR and EL (i.e., FIP RT at 10 000 ha). Additionally, as local heterogeneity was smoothed out at larger spatial grains, the predictive power of the RTs improved. Considering the large spatial and temporal extents of the infestation in the Morice TSA, it is perhaps not surprising that the accuracy of RTs for prediction improved as the spatial grain was increased (i.e., see Wiens 1989). As larger areas of forest are grouped into individual grid cells, the variation of forest conditions within each grid cell increases, and variation between grid cells declines. It is important to note that patterns observed, and predicted at larger spatial grains, may not necessarily hold at finer spatial grains because mountain pine beetle spatial processes operating *within* and *across* forest landscapes may differ. An explicit acknowledgement of pattern and scale is important when interpreting model results.

RT analysis provided an approach to modelling the impact of environmental and population variables on mountain pine beetle caused pine mortality that is flexible enough to be applied for operational, large-area mapping. The decision rules output from RT analysis are easily implemented for forest planning, especially RT results that leverage existing data sources, such as FIP, DEMs and heli-GPS data. All of these variables can be derived within standard GIS software and commonly available data sources.

Figure 4 highlights one of the shortcomings of the RT modelling, namely that the output rules cannot classify outside the range of the input values. In the Figure 4 RT, all locations where IF is less than 47.22 will be classified as 38% infested (PPI). In

situations where beetles are not present, the IF will be 0 and the PPI will also be 0. Therefore, care should be taken when applying the rules to new data for prediction that the input data are fully representative of all locations being classified. In this analysis, RTs were grown on data where beetles were known to be present for at least one year, so the results of RT rules are limited to similar areas.

The results of the FIP RTs indicate that for operational data two key issues are important when modelling forest composition using RT. Firstly, an explicit consideration of spatial scale is critical to understand the impact of different scale dependent processes on the dependent variable. In our analysis, environmental and climatic variables, such as EL and SWR were more often important at larger spatial scales, while a local population variable (IF) was more important at smaller spatial grains. A major limitation to our RT analysis at large spatial grains was that infestation severity (PPI) information was not known for larger grid cells, so was calculated as the average of the PPI values recorded in the plots within each grid cell. This is a common problem in large-area research because of the difficulty associated with making accurate measurements across large, heterogeneous landscapes (Nelson et al. 2006a). Secondly, the impact of data quality and availability is also important. For example, while DENSE was always the most important variable for plot RTs (which is expected as it is highly correlated with PPI), the surrogate variable for plot RTs, QDBH, was never included in cross-validated RTs. This suggests that QDBH may be a poor surrogate for stem density and should be used as such cautiously.

Accuracy of classifications were within +/- 25% percentage of pine infested between 54-91% of the time, and within +/- 15% percentage of pine infested between 9-64% of the time, depending on spatial scale and data of the regression tree. RT accuracy for predicting forest health intensity classes ranged from 22-61%. All RTs performed better when the top 15% of the most intensely infested areas were excluded from the analysis (Table 4a-b). For operational mapping, the 100 ha scale, or 1 km by 1 km grid cells, are appropriate for landscape scale mapping of infestation severity, given the accuracy and scale of the forest inventory data.

The objective of this research was to develop and assess a methodology for large-area mapping of the impact of intense beetle infestation on forest composition. In this respect, our analysis has shown that using regression tree modelling is a flexible approach to mapping the changing characteristics of forests, given operational data constraints. This methodology allowed us to determine important variables modelling the spatial variation of infestation impact on forest composition, and provided a means for mapping the response variable. The map of the amount of change in percentage pine due to mountain pine beetle caused pine mortality (Figure 6) demonstrates the utility of visualizing the model results across the landscape, and allows for intuitive interpretation of spatial variability over a large area. Thus, we feel that the methodology, when applied at an appropriate spatial scale, is a useful approach for large area, operational forest mapping.

In addressing our objective, variables useful for modelling post-beetle forest composition were identified. Although all variables input to the model were known to impact spatial

variation in infestation severity, only some were used by the RT classifier for modelling. The variables identified as important for classification of infestation severity were mainly stem density, percentage pine, number of dead trees within 2 km, and IF. These variables all relate to either the number of susceptible trees, or the number of already infested trees nearby. Other variables known to influence the impact of mountain pine beetle to a forest, such as aspect, slope, BEC classification, and dbh were rarely or never used by the RT models. What appears to be important to the development of severe beetle infestations is the availability of adjacent suitable hosts, which facilitate short range dispersal. In addition to other "beetle proofing" silviculture practices such as varying age class distribution and uniform spacing, planning future forest composition to be less susceptible to severe infestation should perhaps focus on interspersing patches of susceptible trees by at least distances greater than 2 km. While this would not stop the spread of the infestation, as the actual short range dispersal distance is likely to be variable, and wind-driven dispersal would still occur, it would at a minimum mitigate the effect of high impact infestations (intense, fast moving).

Acknowledgements

This project was funded by the Government of Canada through the Mountain Pine Beetle Initiative, a 6 year, $40 million Program administered by Natural Resources Canada, Canadian Forest Service. Publication does not necessarily signify that the contents of this report reflect the views or policies of Natural Resources Canada—Canadian Forest Service.

References

Amman, G.D., 1973. Population changes of the mountain pine beetle in relation to elevation. Environ. Entomol. 2(4): 541-547.

Barclay, H.J., Li, C., Benson, L., Taylor, S. and Shore, T., 2005. Effects of fire return rates on traversability of lodgepole pine forests for mountain pine beetle (Coleoptera: Scolytidae) and the use of patch metrics to estimate traversability. Can. Entomol. 137: 566-583.

Bonesi, L., Rushton, S. and Macdonald, D., 2002. The combined effect of environmental factors and neighbouring populations on the distribution and abundance of Arvicola terrestris. An approach using rule-based models. Oikos 99(2): 220-230.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., 1984. Classification and Regresssion Trees. Wadsworth, Belmont, CA, 358 pp.

Cole, W.E. and Amman, G.D., 1980. Mountain pine beetle dynamics in lodgepole pine forests. Part I: Course of an infestation. USDA For. Serv. Gen. Tech. Rep. INT-89 56 p.

Coops, N.C., Wulder, M.A. and White, J.C., 2006. Integrating remotely sensed and ancillary data sources to characterize a mountain pine beetle infestation. Rem. Sens. Environ. 30(2): 83-97.

Curtis, R.P. and Marshall, D.D., 2000. Why quadratic mean diameter? West. J. Appl. For. 15(3): 137-139.

De'ath, G. and Fabricius, K.E., 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology 81(11): 3178-3192.

Dobbertin, M. and Biging, G.S., 1998. Using the non-parametric classifier CART to model forest tree mortality. For. Sci. 44(4): 507-516.

Eng, M. and Meidinger, D., 1999. A method for large-scale biogeoclimatic mapping in British Columbia. B.C. Ministry of Forests.

ESRI (2005) ArcGIS 9.0.1 Environmental Systems Research Institute, Inc. Redlands, California.

Fan, Z., Kabrick, J.M. and Shifley, S.R., 2006. Classification and regression tree based survival analysis in oak-dominated forests of Missouri's Ozark highlands. Can. J. For. Res. 36(1740-1748).

Furniss, M. and Furniss, R., 1972. Scolytids (Coleoptera) on snowfields above timberline in Oregon and Washington. Can. Entomol. 104: 1471-1477.

Geiszler, D., Gallucci, V. and Gara, R., 1980, Modeling the dynamics of mountain pine beetle aggregation in a lodgepole pine stand. Oecoligia, 46, pp. 244-253.

Ihaka, I. and Gentelman, R., 1996. R: a language for data analysis and graphics. J. Comput. Graph. Stat. 5: 299-314.

Kumar, L., Skidmore, A.K. and Knowles, E., 1997. Modelling topographic variation in solar radiation in a GIS environment. Int. J. Geogr. Inform. Sci. 11: 475-497.

Leckie, D. and Gillis, M., 1995. Forest inventory in Canada with emphasis on map production. For. Chron. 71(1): 74-88.

Meidinger, D., Pojar, J. and Harper, W.L., 1991. Sub-Boreal Spruce Zone. In: D. Meidinger and J. Pojar (Editors), Ecosystems of British Columbia. B.C. Ministry of Forests, Victoria, pp. 209-221.

Negron, J.F. and Popp, J.B., 2004. Probability of ponderosa pine infestation by mountain pine beetle in the Colorado Front Range. For. Ecol. Manag. 191: 17-27.

Nelson, T., Boots, B., Wulder, M.A., Shore, T., Safranyik, L., and Ebata, T., 2006a. The impact of data on the rating of forest susceptible to the mountain pine beetle infestations. Can. J. For. Res. 36(11): 2815-2825.

Nelson, T., Boots, B. and Wulder, M.A., 2006b. Large-area mountain pine beetle infestations: Spatial data representation and accuracy. For. Chron. 82(2): 243-252.

Prasad, A.M., Iverson, L.R. and Liaw, A., 2006. Newer classification and regression tree techniques: Bagging and random forest for ecological prediction. Ecosystems 9: 181-199.

Province of British Columbia, 1996. Gridded DEM specifications, Ministry of Sustainable Resource Management, Victoria.

Quinlan, J.R., 1998. Decision trees and multi-valued attributes. In: J.E. Hayes, D. Michie and J. Richards (Editors), Machine Intelligence 11. Clarendon Press, Oxford, pp. 305-318.

Ridley, S., Jones, S., Shahani, A., Brampton, W., Nielsen, M. and Rowan, K., 1998. Classification trees: A possible method for iso-resource grouping in intensive care. Anaesthesia 53: 833-840.

Ripley, B.D., 1996. Pattern recognition and neural networks. Cambridge University Press, Cambridge, UK, 403 pp.

Robertson, C., Nelson, T. and Boots, B., in press. Mountain Pine Beetle Dispersal: The spatial-temporal interaction of infestations. For. Sci.

Rosso, P.H. and Hansen, E.M., 2003. Predicting Swiss needle cast disease distribution and severity in young douglas-fir plantations in coastal Oregon. Phytopathology 93(7): 790-798.

Safranyik, L. and Carroll, A., 2006, The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. In The mountain pine beetle, a synthesis of biology, management, and impacts on lodgepole pine, L. Safranyik and B. Wilson (Eds.), pp. 3-66 (Victoria, BC: Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre).

Safranyik, L., Linton, D., Silversides, R. and McMullen, L., 1992. Dispersal of released mountain pine beetles under the canopy of a mature lodgepole pine stand. J. Appl. Entomol. 113: 441-450.

Safranyik, L., Shrimpton, D. and Whitney, H., 1974. Management of lodgepole pine to reduce losses from the mountain pine beetle. Canadian Forest Service, Pacific Forestry Centre, Victoria, B.C.

Schwarz, P.A., Fahey, T.J. and McCulloch, C.E., 2003. Factors controlling spatial variation of tree species abundance in a forested landscape. Ecology, 84(7): 1862-1878.

Shore, T.L., Safranyik, L., Hawkes, B.C. and Taylor, S.W., 2006, Effects of the mountain pine beetle on lodgepole pine stand structure and dynamics. In The mountain pine beetle:

A synthesis of biology, management, and impacts on lodgepole pine, L. Safranyik and B. Wilson (Eds.), pp. 95-114 (Victoria, BC: Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre).

Shore, T. and Safranyik, L., 1992. Susceptibility and risk rating systems for the mountain pine beetle in lodgepole pine stands. BC-X-336, Canadian Forest Service, Pacific Forestry Centre, Victoria, B.C.

Taylor, S. and Carroll, A., 2003. Disturbance, forest age and mountain pine beetle outbreak dynamics in BC: A historical perspective. In: T. Shore, J.E. Brooks and J.E. Stone (Eds.), Mountain Pine Beetle Symposium: Challenges and Solutions, 30-31 October 2003, Kelowna, BC, Canada. Natural Resources Canada. Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, Information Report BC-X-399, pp41-51, 298.

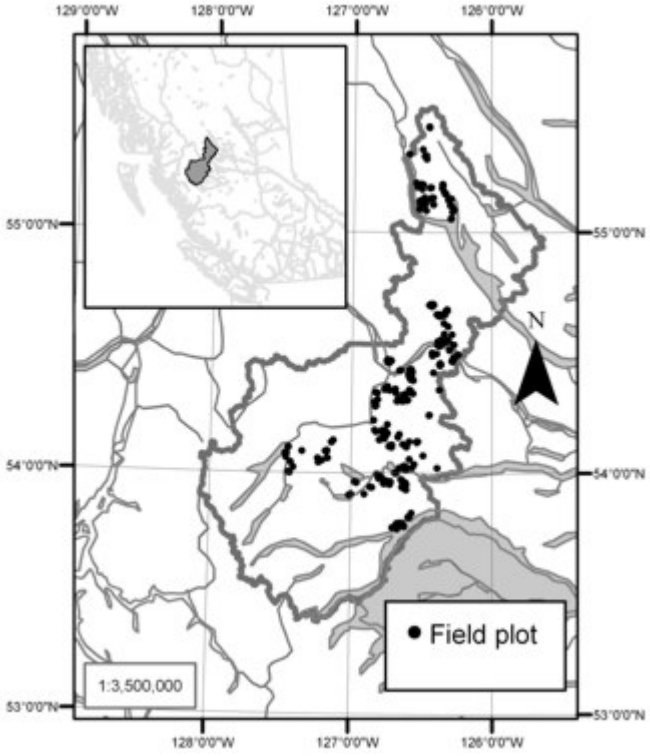
Westfall, J., 2007. 2006 Summary of Forest Health Conditions in British Columbia, Ministry of Forests, Forest Practices Branch.

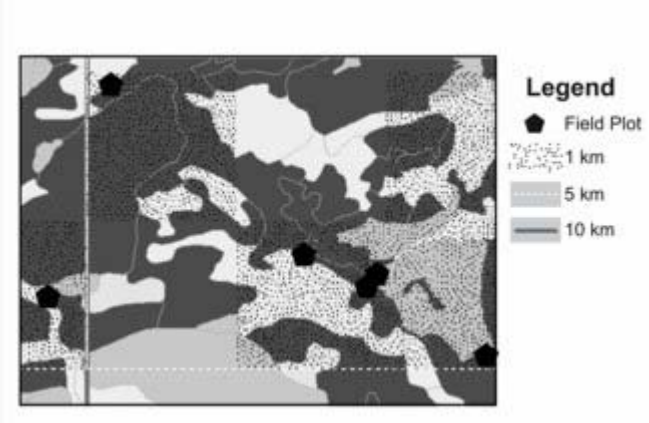
Wiens, J., 1989. Spatial scaling in ecology. Funct. Ecol. 3: 385-397.

Woodward, F.I., 1987, Climate and plant distribution (Cambridge: Cambridge University Press).

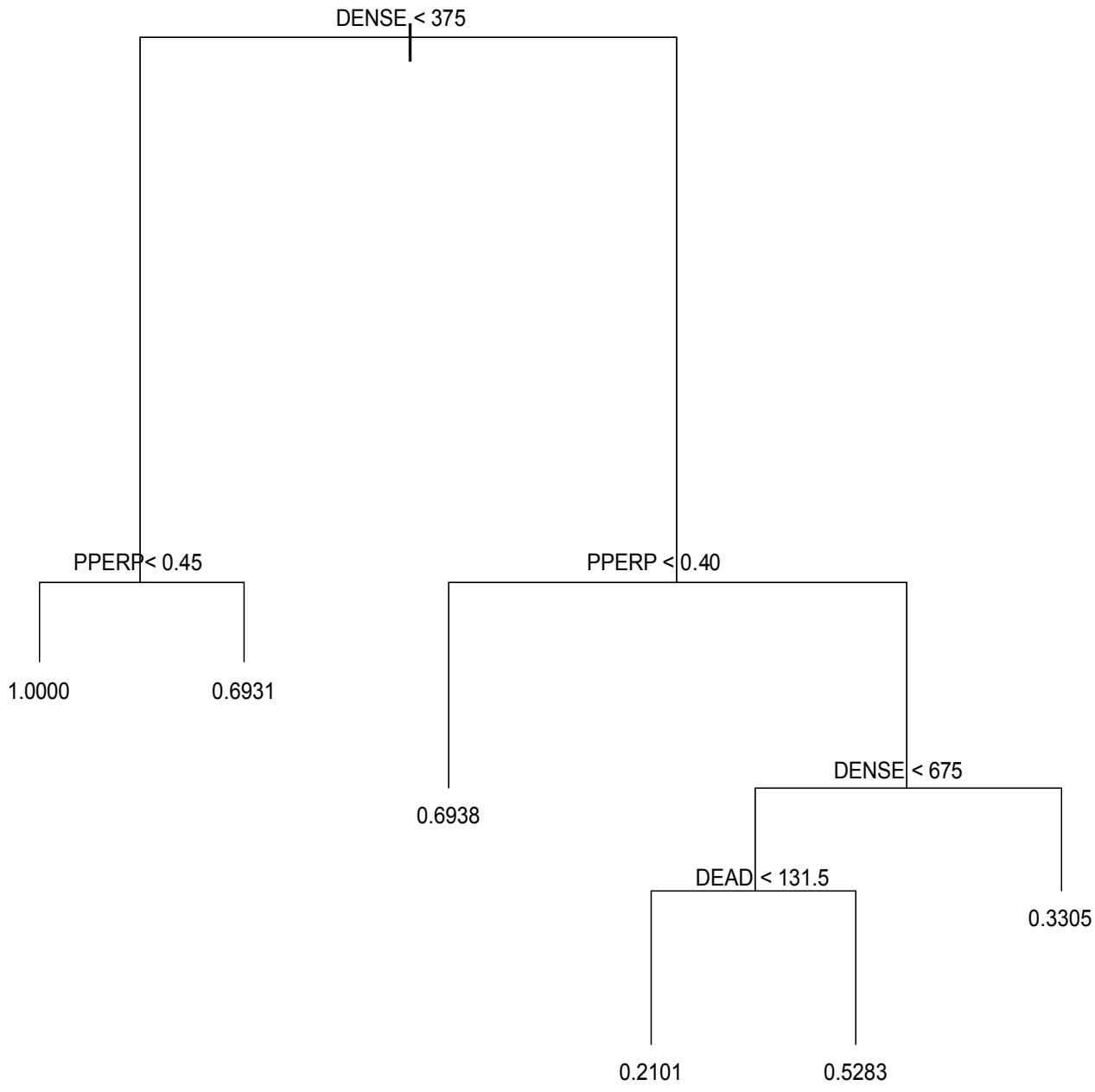
Wulder, M.A., Dymond, C.C., White, J.C. and Erickson, B., 2006. Detection, mapping and monitoring of the mountain pine beetle. In: L. Safranyik and B. Wilson (Editors), The mountain pine beetle: A synthesis of biology, management, and impacts on lodgepole pine. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, pp. 123-154.

Wulder, M.A., White, J.C., Bentz, B., Alvarez, M.F. and Coops, N.C., 2006, Estimating the probability of mountain pine beetle red-attack damage. Remote Sensing of Environment, 101, pp. 150-166.

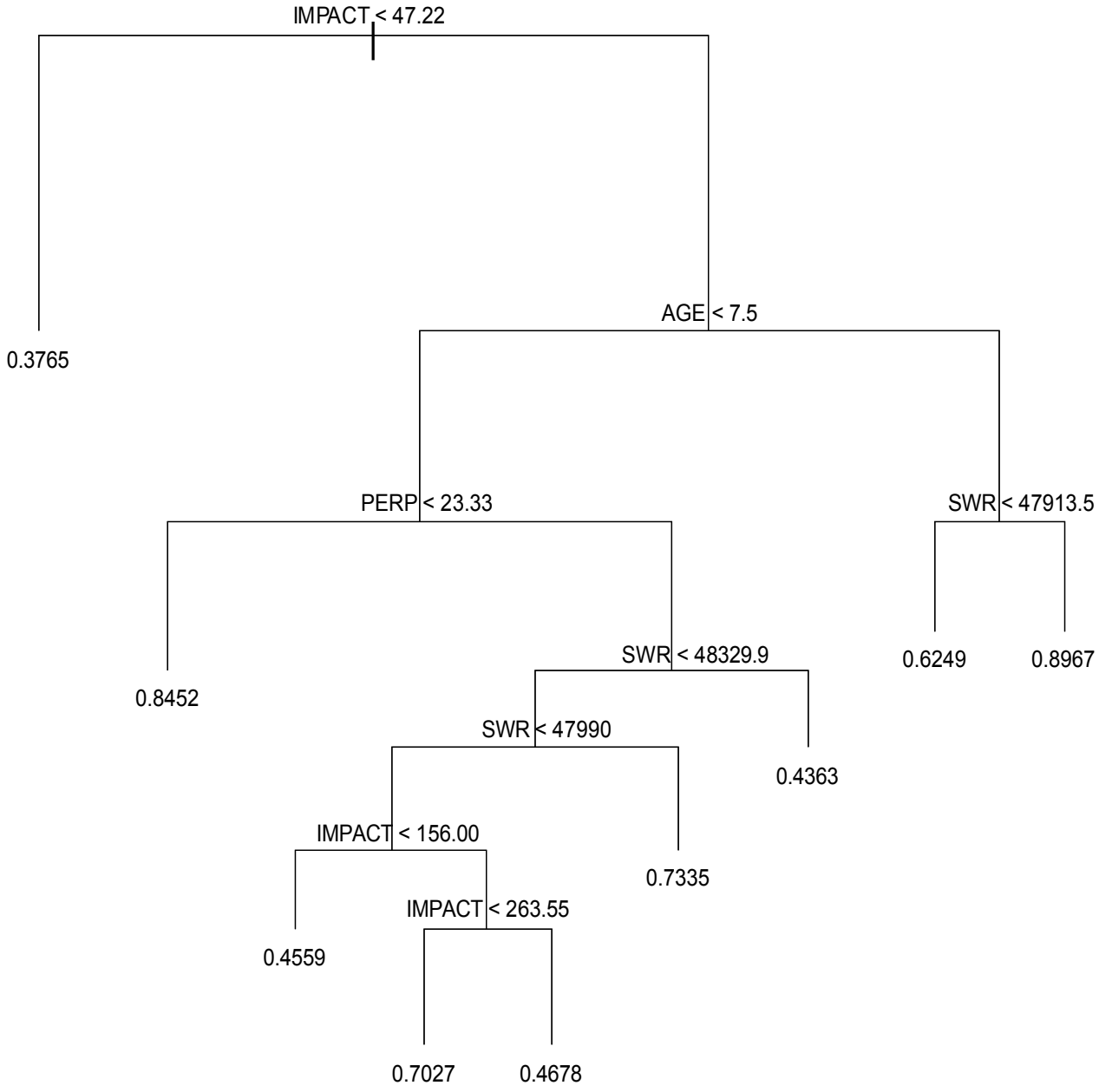

129°0'0"W
128°0'0"W
127°0'0"W
126°0'0"W
55°0'0"N
54°0'0"N
53°0'0"N
N
Field plot
1:3,500,000

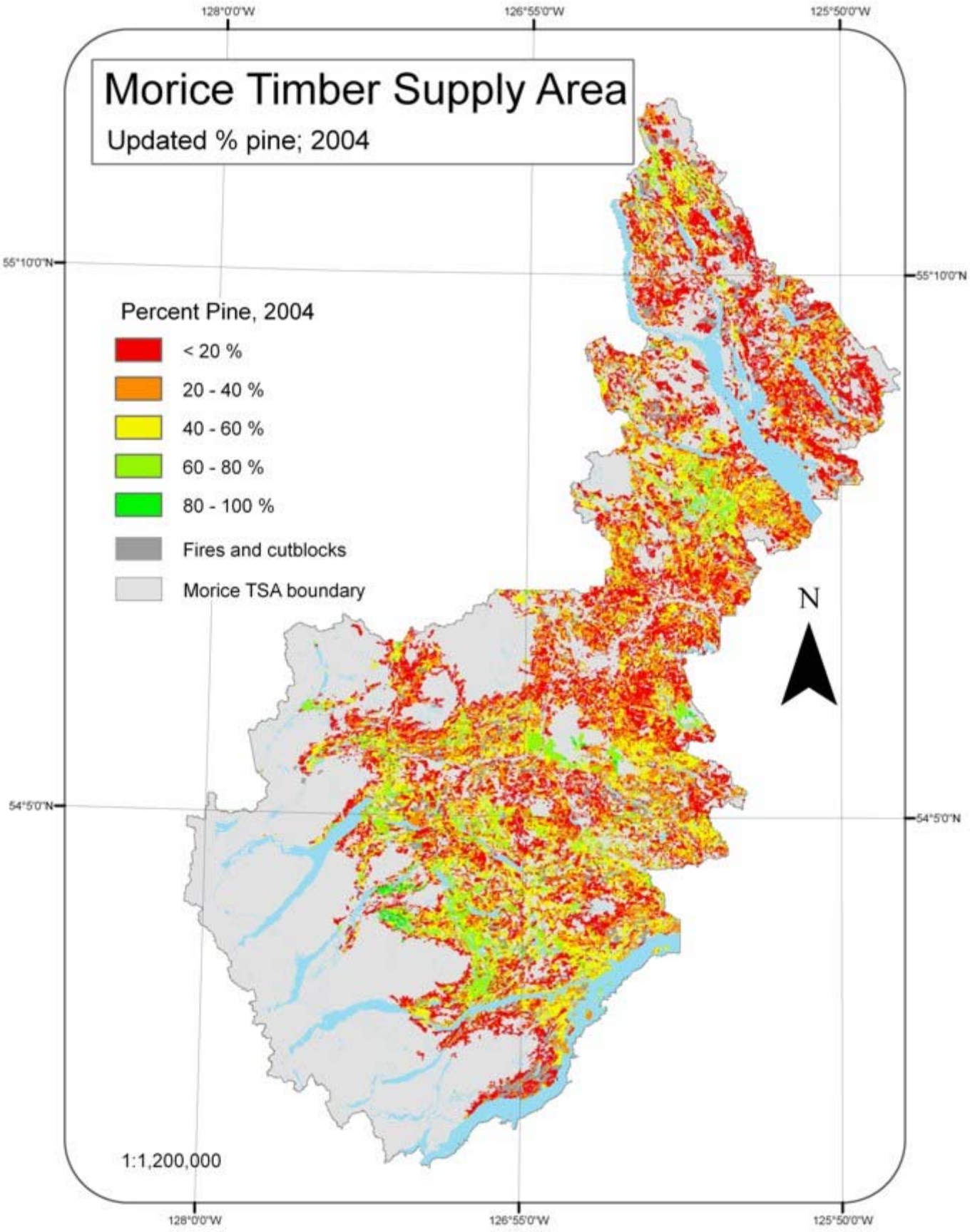

Legend
Field Plot
1 km
5 km
10 km

For Review Purposes Only/Aux fins d'examen seulement



For Review Purposes Only/Aux fins d'examen seulement




Morice Timber Supply Area
Updated % pine; 2004
128°0'0"W
126°55'0"W
125°50'0"W
55°10'0"N
54°5'0"N
Percent Pine, 2004
< 20 %
20 - 40 %
40 - 60 %
60 - 80 %
80 - 100 %
Fires and cutblocks
Morice TSA boundary
N
1:1,200,000

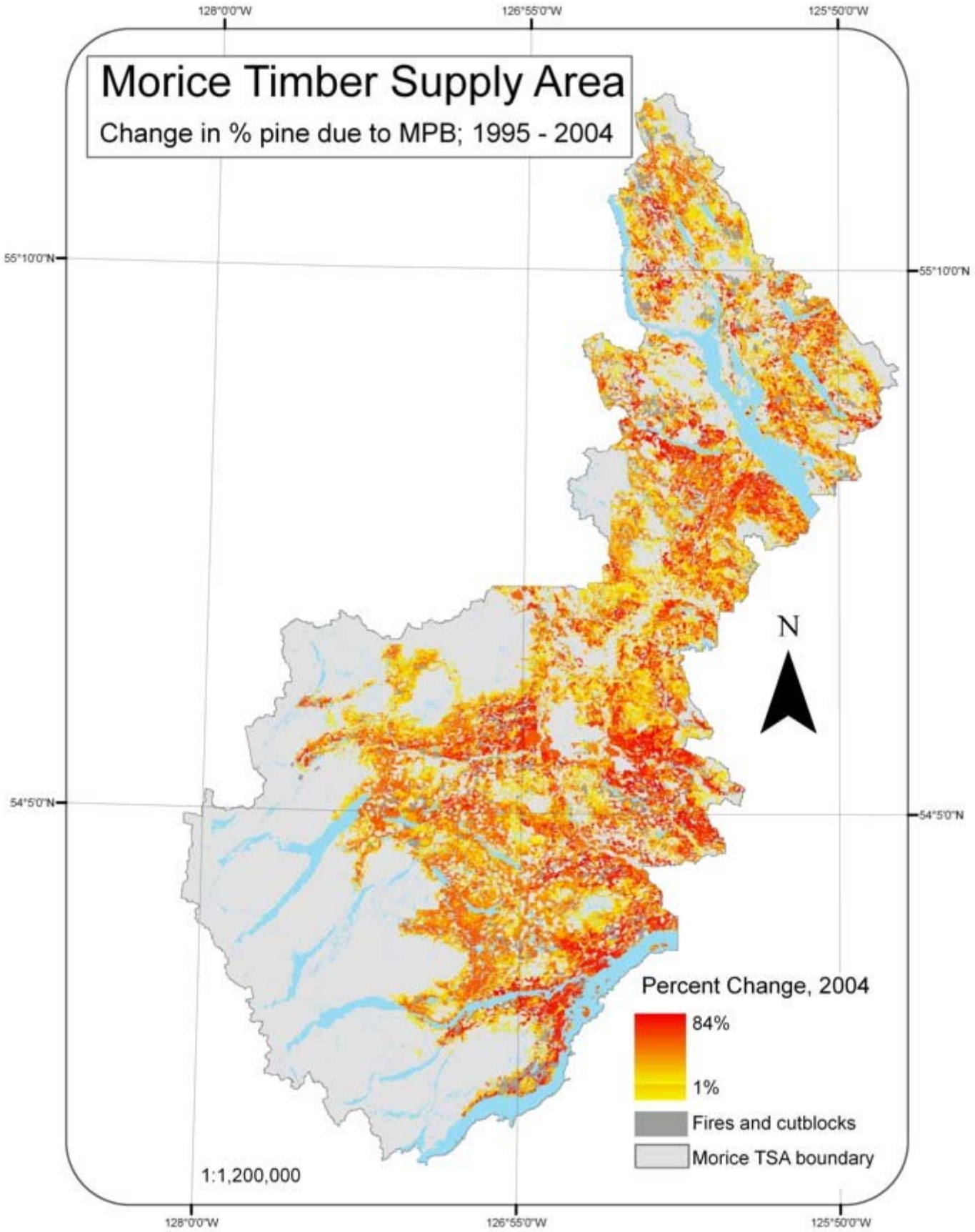

Morice Timber Supply Area
Change in % pine due to MPB; 1995 - 2004
128°0'0"W
126°55'0"W
125°50'0"W
55°10'0"N
54°5'0"N
N
Percent Change, 2004
84%
1%
Fires and cutblocks
Morice TSA boundary
1:1,200,000

Table 1. List and description of variables used in regression trees for predicting the percentage of pine infested at each plot (**PPI**). *Used only when field plot data were used as input to model. **Used when forest inventory data were used as input to model.

Variable	Description	Description
DEAD	# of dead trees within 2 km radius	The number of mountain pine beetle attacked trees within a 2 km radius.
ID	Infestation duration	The length of time since the first mountain pine beetle attacked tree was found within a 200 m radius of the sample plot.
IF	Impact factor	The number of mountain pine beetle attacked trees within a 2 km radius divided by the length of time beetles have been present.
EL	Elevation	Elevation as recorded in the digital elevation model.
SLP	Slope inclination	Surface inclination calculated from the digital elevation model.
ASP	Aspect orientation	Surface orientation calculated from the digital elevation model.
SWR	Short wave radiation	Direct shortwave radiation, in kJ/m^2 at 30 minute intervals.
AGE	Stand age class	Age class of the stand recorded in the forest inventory.
PPERP*	Percentage pine in the plot	The actual percentage of pine within the field plot.
DENSE*	Stem density in field plots (stems / ha)	The stem density within the field plot is directly related to the amount of pine infested when combined with percentage pine, and information about the local beetle population.
QDBH**	Stand quadratic mean dbh	Stand scale quadratic mean dbh is often used as a surrogate variable for stand density.
PERP**	Percentage pine in the stand	At a stand scale, the estimated percentage of pine extracted from forest inventory polygons.

Table 2. All regression trees used in the analysis.

Regression Tree	n train	Scale	Residual Mean Deviance	Variables used in tree
PLOT-1 6 node	254	200 m^2	.056	DENSE, PPERP, DEAD
PLOT-1 3 node	254	200 m^2	.046	DENSE, PPERP
FIP-1 9 node	254	200 m^2	.052	PERP, IF, EL, DEAD
FIP-1 7 node	254	200 m^2	.056	PERP, IF, EL, DEAD
PLOT-2 6 node	166	100 ha	.035	DENSE, IF, PPERP, DEAD
PLOT-2 5 node	166	100 ha	.038	DENSE, PPERP, DEAD
FIP-2 9 node	166	100 ha	.046	IF, AGE, PERP, SWR
FIP-2 4 node	166	100 ha	.056	IF, AGE, PEPR
PLOT-3 9 node	77	2500 ha	.019	DENSE, ID, PPERP, DEAD, EL, SLP, SWR
PLOT-3 6 node	77	2500 ha	.022	DENSE, ID, PPERP, DEAD
FIP-3 12 node	77	2500 ha	.035	EL, PERP, SWR, DEAD, IF
PLOT-4 4 node	43	10 000 ha	.027	DENSE, ID, PPERP
PLOT-4 2 node	43	10 000 ha	.047	DENSE
FIP-4 6 node	43	10 000 ha	.028	PERP, SWR, EL, DEAD
FIP-4 4 node	43	10 000 ha	.036	PERP, SWR, EL

Table 3. Accuracy of (a) plot level data and (b) stand level data on predicting infestation levels for regression trees with the highest (H) and lowest (L) number of terminal nodes after cross validation, at 25% accuracy, 15% accuracy and forest health intensity (FHI) classes.

(a)

Dataset	n train	n test	Scale	H +/- .25	L +/- .25	H +/- .15	L +/- .15	H FHI	L FHI
PLOT-1 (6/3)	254	121	200 m^2	60%	58%	26%	44%	23%	38%
PLOT -2 (6/5)	166	99	100 ha	68%	63%	24%	44%	32%	34%
PLOT -3 (9/6)	77	53	2500 ha	72%	75%	13%	55%	38%	42%
PLOT -4 (4/2)	43	33	10 000 ha	91%	70%	9%	64%	61%	36%

(b)

Dataset	n train	n test	Scale	H +/- .25	L +/- .25	H +/- .15	L +/- .15	H FHI	L FHI
FIP-1 (9/7)	254	121	200 m^2	55%	54%	26%	34%	27%	24%
FIP-2 (9/4)	166	99	100 ha	48%	55%	24%	34%	31%	22%
FIP-3 (12)	77	53	2500 ha	62%	na	13%	na	30%	na
FIP-4 (6/4)	43	33	10 000 ha	79%	70%	9%	48%	30%	48%

Table 4. Accuracy of (a) plot level data and (b) stand level data on predicting infestation levels when test cases with greater than 85% infestation are excluded for regression trees with the highest (H) and lowest (L) number of terminal nodes after cross validation, at 25% accuracy, 15% accuracy and forest health intensity (FHI) classes.

(a)

Dataset	n train	n test	Scale	H +/- .25	L +/- .25	H +/- .15	L +/- .15	H FHI	L FHI
PLOT-1 (6/3)	254	100	200 m^2	70%	70%	31%	50%	25%	33%
PLOT -2 (6/5)	166	79	100 ha	75%	76%	30%	53%	30%	28%
PLOT -3 (9/6)	77	43	2500 ha	84%	76%	16%	65%	35%	40%
PLOT -4 (4/2)	43	27	10 000 ha	93%	81%	11%	78%	52%	44%

(b)

Dataset	n train	n test	Scale	H +/- .25	L +/- .25	H +/- .15	L +/- .15	H FHI	L FHI
FIP-1 (9/7)	254	100	200 m^2	63%	63%	31%	41%	29%	27%
FIP-2 (9/4)	166	79	100 ha	56%	66%	30%	41%	29%	18%
FIP-3 (12)	77	43	2500 ha	70%	na	16%	na	30%	na
FIP-4 (6/4)	43	27	10 000 ha	85%	74%	11%	48%	26%	48%

Figure Captions.

Figure 1. Map of the study area and field plot locations.

Figure 2. Example of three grid sizes used in regression tree analysis for classifying percentage pine infested (PPI). Regression trees using data from field plots, at higher spatial grains, values were the average of the values at each plot. Regression trees using data from forest inventory polygons, values were the average of the values over the entire square.

Figure 3. 6 node regression tree using field plot data at plot scale (0.02 ha).

Figure 4. 9 node regression tree using forest inventory data at 100 ha scale

Figure 5. Updated percent pine map generated from percentage pine infested values modelled with regression tree analysis and forest inventory data.

Figure 6. Percent pine change map showing the magnitude of mountain pine beetle caused pine mortality between 1995 and 2004.