Risk rating for mountain pine beetle infestation of lodgepole pine forests over large areas with ordinal regression modelling

Robertson, C.*[1], Wulder, M.A.[2], Nelson, T.A.[1] and White, J.C.[2]

*corresponding author:

Colin Robertson

Spatial Pattern Analysis & Research (SPAR) Laboratory, Dept of Geography, University of Victoria, PO Box 3050, Victoria, BC V8W 3P5, Canada

Email: colinr23@gmail.com, Ph. 1-250-885-8065

1) Spatial Pattern Analysis & Research (SPAR) Laboratory, Dept of Geography, University of Victoria, PO Box 3050, Victoria, BC V8W 3P5, Canada

2) Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, 506 West Burnside, Victoria, BC V8Z 1M5, Canada

Submitted to: Forest Ecology and Management

Submission date: Feb 2008

Abstract

The mountain pine beetle *Dendroctonus ponderosae* Hopkins is endemic to lodgepole pine *Pinus contorta* var. *latifolia* Engelmann forests in western Canada. However, the current beetle epidemic in this area highlights the challenges faced by forest managers tasked with prioritizing stands for mitigation activities such as salvage harvesting, thinning, and direct control methods. In western Canada, the operational risk rating system for mountain pine beetle is based on biological knowledge gained from a rich legacy of stand-scale field studies. Due to the large spatial and temporal extents of the current epidemic, new research into large-area mountain pine beetle processes has revealed further insights into the landscape-scale characteristics of beetle infested forests. In this research, we evaluate the potential for this new knowledge to augment an established system for rating the short-term risk of tree mortality in a stand due to mountain pine beetle. New variables explored for utility in risk rating include direct shortwave radiation, site index, diameter at breast height, the temporal trends in local beetle populations, Biogeoclimatic Ecosystem Classification and beetle-host interaction variables. Proportional odds ordinal regression was used to develop a model for the Vanderhoof Forest District in west-central British Columbia. Prediction on independent data was assessed with the area under the receiver operator curve (AUC), indicating good discriminatory power (AUC = 0.84) for predicting damage due to mountain pine beetle.

Key words: mountain pine beetle, risk rating, landscape scale, monitoring, infestation

Introduction

Management of lodgepole pine *Pinus contorta* var. *latifolia* Engelmann forests in western Canada in recent years has been dominated by a large scale mountain pine beetle *Dedroctonus ponderosae* Hopkins epidemic that has impacted more than 9 million ha (Westfall, 2007). Management of any forest disturbance requires decision support tools that enable managers to predict future forest scenarios, set priorities, and evaluate management strategies. In the context of mountain pine beetle, forest managers must know the ability of a forest stand to support an epidemic mountain pine beetle population (i.e., susceptibility), and the possibility of host tree mortality as a result of an existing beetle infestation (i.e., risk). Susceptibility is determined using stand and site characteristics, independent of surrounding beetle population levels. Conversely, risk is determined by considering the susceptibility of the stand in the context of the local beetle population within the stand and in the vicinity of the stand (Bentz et al., 1993). A risk rating system is a specific decision support tool that is used to identify those forest stands on the landscape that are at greatest risk of timber losses as a result of a mountain pine beetle infestation (Shore et al. 2006).

Mountain pine beetle risk rating

Many risk rating systems for mountain pine beetle have been developed over the past three decades. Safranyik et al. (1975) initially used weather station data to model and map the beetle outbreak hazard in western Canada. Amman et al. (1977) used stand characteristics such as elevation, age, and diameter at breast height (dbh) to develop a three-class risk classification system (i.e., low, moderate, high). Other risk rating systems were developed that also relied on stand characteristics and adopted similar approaches to rating stand risk as a categorical variable (e.g., Mahoney 1978; Berryman 1978a; Stuart 1984; Anhold and Jenkins 1987). Bentz et al. (1993) evaluated the accuracy of three categorical risk rating systems (Amman et al. 1977; Mahoney 1978; Berryman 1978a) and one continuous variable risk rating system (Schenk et al. 1980). All of the systems evaluated by Bentz et al. (1993) were found to provide poor estimates of pine mortality, primarily because they failed to consider spatial relationships between host stands and beetle populations. Furthermore, Bentz et al. (1993) concluded that the empirical development of these risk rating systems limits their portability to other geographic areas.

Shore and Safranyik (1992) introduced a continuous variable risk rating system which incorporated those elements of previous systems that had a strong theoretical basis for characterizing susceptibility and risk. Shore and Safranyik (1992) define risk as the short-term expectation of volume loss due to mountain pine beetle attack. There are two components to the Shore and Safranyik risk rating system: stand susceptibility, defined as the inherent characteristics of a stand of trees that affect its likelihood of attack by mountain pine beetle; and, beetle pressure, which is a measure of the size and proximity of the mountain pine beetle population to the stand. Susceptibility in Shore and Safranyik (1992) is determined using stand density, age, composition, and geographic location. Each of these variables has a direct link to biological processes associated with mountain pine beetle. Stand density effects tree competition for light and nutrients, so less dense stands tend to have larger, more vigorous trees; however, low density stands have a negative impact on the microclimate that is required to facilitate pheromone mediated attacks, landing, and emergence rates (Bartos and Amman 1989). Intermediate stand densities of 750 to 1500 stems/ha are thought to be more conducive to beetle-induced tree mortality (Anhold and Jenkins 1987), giving rise to a nonlinear relationship between risk and stand density. Stand age relates to the beetle's preference for large diameter trees (Safrankyik et al. 1974), and stand age has an inverse relationship with tree vigor after maturity, which determines a tree's ability to resist infection by beetle-introduced fungi (Shrimpton 1973). Once physiological maturity has been reached, trees become susceptible to attack, although the rate and likelihood of attack is impacted by other variables such as climate (Shrimpton and Thomson 1983). Stand composition is included in the Shore and Safranyik model because stand risk relates to the amount of near term volume loss, so greater amounts of pine contribute to higher risk ratings. The composition variable is measured as the percentage of a stand's basal area that is composed of large diameter pine. The location factor incorporates the impact of geographic location on beetle survival. At higher elevations and northern latitudes, beetles are exposed to colder temperatures, thereby increasing winter mortality and disrupting the beetle's development cycle (Amman 1973).

Beetle pressure in Shore and Safranyik (1992) is based on the number of beetles within and proximal to the stand. Beetle population is estimated by the number of infested trees and the likelihood of attack is incorporated by the distance between the stand being rated and the infestation. The maximum distance at which beetles from surrounding areas can enter the stand being assessed is 3 km. Susceptibility variables and beetle pressure are combined multiplicatively to determine the overall risk rating for a stand of trees. Recent model refinements have replaced discrete look-up tables with continuous equations for each of the susceptibility variables, the beetle population variable, and for the risk calculation in an effort to reduce the impact of class boundaries on final risk assessments (Shore et al. 2006). However, the fundamental elements of the Shore and Safranyik model remain unchanged: four equally weighted susceptibility variables combined with a spatial measure of the beetle population to determine a relative ranking of stand risk.

With the explosive growth of mountain pine beetle populations in western Canada, the risk rating system has become a recommended forest planning tool (British Columbia Ministry of Forests 1995). However, there are considerable limitations to implementing the Shore and Safranyik system over large areas. Firstly, the relationships in the model are derived from field research over relatively small geographic areas (Shore and Safranyik 1992; Shore et al. 2000). Extrapolating these relationships to new geographic regions may neglect regional variations in mountain pine beetle processes. Secondly, the data inputs required for operational modelling across large areas often do not exist and substitute variables available in forest inventory data are generally poor replacements (Nelson et al. 2006). Hence, there is an information need for a decision support tool capable of assessing risk over large areas.

Recent research

Research conducted over the last decade may be able to provide enhancements to existing risk rating systems. The growth of geographic information systems (GIS) as a tool for managing complex spatial and attribute information combined with increasing efficiencies in automated and semi-automated data collection technologies has enabled forest managers and researchers to link theoretical and empirical knowledge of ecological processes (e.g., Blackburn and Milton 1996). Additionally, more advanced analytical methods for spatial data are being developed that facilitate spatially explicit analysis of forest disturbances across large spatial and temporal scales (e.g., Nelson and Boots 2005). These new results in mountain pine beetle research may enhance risk rating systems.

Landscape scale red attack modelling

A number of studies have investigated the potential for locating and estimating the severity of mountain pine beetle red attack damage of forested landscapes (red attack is a term used to describe the characteristic fading of an attacked tree's foliage, which typically occurs within 6 to 8 months following attack). Variables useful for predicting red attack damage over large areas may also be helpful for predicting areas at risk of beetle attack. While remotely sensed data have been used in many studies for detecting and mapping red attack damage at a range of spatial scales (Sirois and Ahern 1988; Franklin et al. 2003; Skakun et al. 2003; White et al. 2005; Wulder et al. 2006b), here we highlight studies that have used imagery in conjunction with ancillary terrain and radiation information (White et al. 2006; Wulder et al. 2006a; Coops et al. 2006). Elevation, slope, and direct radiation have all been significant predictors in logistic regression models of red attack damage (see Table 1). Additionally, Coops et al. (2006) investigated the relationship between probability of attack, forest structure, and forest susceptibility variables using regression tree models and found site index and slope to be most important for explaining variation in the probability of red attack. Negron and Popp

2004) used a similar approach for ponderosa pine (*Pinus ponderosa* Lawson) and found stand density index (SDI) and quadratic mean diameter to be most important for estimating the plot scale probability of infestation.

Data uncertainty, accuracy, and large-area spatial analysis

Large-area application of forest risk models rely on operational data which often have varying levels of accuracy and completeness. For instance, detailed variables such as stand density and basal area by species are not available in most provincial or national forest inventory products. Nelson et al. (2006) explored the impact of operationally available data representative of large areas on the Shore and Safranyik susceptibility model. The authors found that the surrogate variables, due to lack of detail, were only moderately correlated with the true variables, thereby highlighting the importance of including operationally available variables in landscape scale models. In an investigation of the accuracy of the Shore and Safranyik risk model computed with similar large-area data, Dymond et al. (2006) found the risk index to be within 30% to 43% true positive for high risk (>5), and 93% accurate for low risk (>0<5) for predicting presence of infestation. However, the severity of infestation was not evaluated against the risk index. Wulder et al. (2006c) compared two methods for estimating the beetle pressure component of risk, the distance-based method used in Shore and Safranyik (1992) and a density-based method employing Voronoi polygons; the density-based estimate was found to have greater correspondence with infestation occurrence.

Epidemiology and population dynamics

The current state of knowledge on biology and epidemiology of the mountain pine beetle is succinctly reviewed in Safranyik and Carroll (2006). New research indicates that the four stage population cycle of the mountain pine beetle (endemic, incipient-epidemic, epidemic, and post-epidemic) is based on complex interactions with the host tree and an assemblage of secondary bark beetles. Due to the nature of large area sampling of mountain pine beetle populations (identification of red attack damage by aerial survey), it is difficult to detect beetle populations at endemic levels, where spatially disparate assemblages of only a few trees are infested. The most important aspect of the population cycle for modelling risk is the epidemic threshold; the point at which the growth in the beetle population exceeds the stand's ability to resist mass attacks of its large, healthy trees (Berryman 1982b). It is the large diameter trees that promote exponential growth in brood production, thereby enabling an epidemic when suitable host and climatic conditions exist (Safranyik et al. 1974). In population dynamics, this is the shift from the one stable equilibrium (endemic) to another stable equilibrium (epidemic) (Berryman 1978b). The critical threshold between these equilibria is governed by events that increase beetle populations (i.e., warm winter temperatures) or decrease stand resistance (i.e., drought). For modelling the risk of mountain pine beetle damage, it is important to represent the dynamic nature of this critical population threshold in order to accurately forecast future infestation advance or collapse (Raffa and Berryman 1986; Bentz et al.1993; Logan et al.1998; Nelson et al. 2007).

Range expansion, geographic variation and novel habitats

A key contributor to sustaining the beetle epidemic currently on-going in western Canada is an increase in climatically suitable habitat, enabling beetle range expansion (Carroll et al. 2006). Historically, the range of the primary host species exceeded the limits to mountain pine beetle range imposed by climate conditions. This has important implications for modelling risk as it is common for processes (i.e., dispersal) at range margins, to differ from those in traditionally colonized areas (Thomas et al. 2001). Indeed, latitudinal variation in mountain pine beetle developmental rates has been shown by Bentz et al. (2001). This perhaps supports other recent research pointing to large scale spatial synchrony of epidemic beetle populations (Aukema et al. 2006). Mountain pine beetle spatial processes, such as dispersal, pheromone dynamics, and host selection are influenced, if not determined, by environmental factors. It might therefore be expected that spatial patterns representing these processes will vary in different environments. Robertson et al. (2007) found spatial patterns of dispersal processes to have different frequencies in different Biogeoclimatic Ecosystem Classification (BEC) subzones. It may be useful for large scale models of risk to identify regional variations in the spatial patterns used to represent dynamic beetle processes (i.e., Wu et al. 2005).

Goals and objectives of this research

Knowledge gained from recent mountain pine beetle research may be incorporated into the next generation of risk models. In this study, we aim to encapsulate some of this knowledge with the goal of developing a risk model with the following characteristics:

- An operationally viable model that may be used with existing large-area data.

- A robust model that improves upon existing models for characterizing risk of immediate damage by mountain pine beetles over large geographic areas.

Study Area and Data

The study area for our model development is the Vanderhoof Forest District, located in central British Columbia, Canada (Figure 1). The approximately 1.38 million ha Vanderhoof Forest District has experienced substantial losses of lodgepole pine due to mountain pine beetle attack. Forest inventory data is available for this area and mountain pine beetle monitoring data is spatially and temporally exhaustive, providing a comprehensive view of beetle population levels. This area is dominated by forests of lodgepole pine and spruce (*Picea engelmannii x glauca* Moench. Voss), with a median age of 105 years. There are three main BEC zones represented in Vanderhoof, ranging in elevation from 680 m to 1800 m: Sub Boreal Spruce (SBS), Sub Boreal Pine Spruce (SBPS), and Engelmann Spruce-Subalpine Fir (ESSF). The Vanderhoof Forest District is proximal to the epicenter of the current outbreak, which is thought to be just to the west near Entiako Park and Protected Area and Tweedsmuir Provincial Park (Aukema et al. 2006).

Forest inventory information for the Vanderhoof Forest District conforms to the current provincial vegetation resource inventory (VRI) standards. VRI is a seamless spatial coverage of forest stands where attribute information is estimated by a combination of aerial photo interpretation and field plot verification (British Columbia Ministry of Sustainable Resource Management 2002). This dataset was last updated with harvest and natural disturbances in 2002. Forest stands in the VRI, defined as homogenous units by photointerpreters, made up the unit of analysis for all modelling. Attributes from the inventory that were used in modelling included stand composition, age, density, dbh, crown closure, and site index.

Mountain pine beetle populations were estimated from aerial overview survey (AOS) data collected in the study area as part of a province-wide forest health survey conducted annually throughout British Columbia (British Columbia Ministry of Forests 2000). Broad areas of red attack damage are delineated on 1:100,000 or 1:250,000 basemaps by trained observers (Wulder et al. 2006c). Severity codes are assigned to these areas to indicate the proportion of the area that is infested. These broad AOS data are collected primarily for strategic purposes and are used to direct the subsequent acquisition of more detailed survey information. In Vanderhoof, detailed surveys were conducted using helicopters equipped with GPS receivers, where locations of mountain pine beetle attack were recorded and assigned a severity. Information contained in the AOS and helicopter-GPS surveys were combined to produce a raster layer indicating the cumulative area of infestation in the Vanderhoof Forest District for each year from 1999 to 2005. The area of infestation values were averaged at the forest stand level. All model predictions were made for the year 2005 based on data up to and including 2004.

The BEC system stratifies landscapes based on vegetation, soils, and climatic and site characteristics. BEC zones are characterized by a common regional climate. Subzones, the finest spatial unit in the BEC system, represent geographically related ecosystems

(Eng and Meidinger 1999). The BEC data used in this analysis were mapped at a scale of 1:20,000 in 2003. The main BEC zone in the Vanderhoof Forest District is the Sub Boreal Spruce (SBS), making up 84% of the total area. Figure 1 illustrates where different subzones are located within the study area.

Elevation base data in the form of a digital elevation model were used for all topographic variables. This data set was obtained from the Government of Canada geographic data portal GeoBase, and conforms to Canadian Digital Elevation Data (CDED) 1:50,000 standards. Elevation values are referenced to the Canadian Vertical Geodetic Datum 1928 (DVGD28). The data was resampled to 100 m grid cell resolution. The DEM was used to generate direct solar radiation using the methods of Kumar et al. (1997).

Methods

Modelling risk for large areas

Our approach to modelling stands at risk of mountain pine beetle attack is based on Shore and Safranyik (1992). Our intent is to investigate the potential for model improvements based upon knowledge gained from recent research, while also overcoming some of the previously noted limitations when determining risk over large areas. Risk in Shore and Safranyik (1992) model is a continuous value indicating the relative risk of volume loss due to beetles and is based on basal area. One of the more difficult issues when dealing with the data constraints of landscape wide forest inventory is the selection of a dependent variable. Since basal area is often not available in forest inventories, our dependent variable was calculated as the average value of percent mortality pixels for each polygon, scaled by the percentage of pine associated with the polygon. These values were then linked back to damage classes (Table 2), so they become the AOS severity classes scaled by the amount of pine in each VRI polygon, hereafter referred to as damage level. VRI polygons that did not contain any pine were excluded from the analysis.

The modelling framework for the inclusion and structure of covariate variables was based on the breakdown of the mountain pine beetle-lodgepole pine system described by Raffa and Berryman (1986) and others, as either host/stand variables or beetle population variables. Simulations in Raffa and Berryman (1986) demonstrate how tree and beetle interactions influence the overall beetle population. Extending this idea to the stand scale, we capture this interaction by combining stand variables with beetle population variables (interaction variables). A list of new variables used in this analysis is presented in Table 3.

Stand Resistance Variables

Stand resistance typically indicates stand vigour, which is inversely related to the ability of mountain pine beetles to overcome a tree's defences. New variables related to the stand's ability to defend against beetle attacks included BEC subzone the stand is located within and the amount of annual direct shortwave radiation (SWR) (2006a). Other stand resistance variables that are known to be intimately linked with the state of the beetle population, such as site index, crown closure, and dbh, are included instead as interaction variables.

Beetle Population Variables

The integrated beetle population data contains cumulative area infested levels for each year from 1999 to 2005. From this data we estimated the spatial and temporal trends in beetle population levels. Beetle population variables provide information about the local characteristics of the infestation. Beetle population variables used in modelling included percent pine infested in 2004 (PPI), infestation in 2004 (INC0304), infestation in 2003 (INC0203), duration of infestation (INFDUR) and the number of red attacked trees in 2004 (NUMRED).

Interaction Variables

Stand variables whose impacts on beetle populations are tied to the population state were modelled as interaction variables. We used a sigmoid function to scale interaction variables, w, on a scale $[0,x]$, where x is the percentage of pine in the stand.

$$w = \frac{x}{1 + c^{((t + zr/2 - z)/zr))}} \qquad (1)$$

The threshold value of the variable z is denoted as t and the range of z that defines how quickly w reaches x is defined by zr. A constant c defines initial values and was determined experimentally from the data available. This sigmoid function represents the threshold nature of mountain pine beetle population interaction with variable z by defining two parameters: the initial increase of the weight as variable z increases, and the value of z at which exponential increase in weight occurs. The beetle population variable defines the slope of the curve (*zr*) and the threshold value (*t*) of *z* is defined based on previous mountain pine beetle research. The highest weights are associated with stand conditions that promote beetle brood production when the population state is at epidemic levels. Variables modelled as interactions were crown closure (R_C), site index (R_SI), and quadratic mean diameter at breast height (R_QDBH). Figure 2 presents the details of the interaction variable curves.

Ordinal Logistic Regression

The dependent variable for our analysis, damage level, represents the amount of pine mortality due to mountain pine beetle in six distinct classes (none, trace, low, moderate, severe, very severe). The ordered nature of this variable warrants the use of an ordinal regression model. An ordinal model is an extension of binary logistic regression, an approach which has been successfully employed for predicting locations of mountain pine beetle red attack damage (Wulder et al. 2006a, Coops et al. 2006). In the ordinal model, as in binary logistic, the combination of linear predictor variables relates to the expectation of the dependent variable through a link function, usually the logit function. Since there are multiple ordered responses, multiple equations need to be resolved in ordinal regression (Guisan and Harrell 2000).

There are two commonly used ordinal regression models that deal with dependent variables derived from a continuous phenomenon that have been categorized. The proportional odds (PO) model (Walker and Duncan 1967) is based on cumulative probabilities, and the continuation ratio (CR) model (Armstrong and Sloan 1989) is based on conditional probabilities. CR models are suited for situations where the dependent variable Y must pass through one category to reach the next (Guisan and Harrell 2000). The standard PO model assumes that the slopes of each independent variable X are equal for all levels of Y. Since we are modelling the dynamics between Y and X explicitly through the interaction variables and the progression of damage level need not pass through one level to reach the next (given our temporal resolution), we selected the PO model, defined as follows.

$$P(Y \geq j \mid X) = \frac{1}{(1 + \exp[-(a_j + Xb)])} \quad (3)$$

Or rather, that the probability of the observed Y falling in a class greater than or equal to class j given the explanatory variables in X is similar to a logistic model where $1 = Y \geq j$ and $0 = Y < j$ for all levels of j in the ordinal dependent variable Y. Thus in the case where j = 1, the PO model is equivalent to the logistic model. For additional levels of j, the coefficients, b, stay the same while the intercept term, a, varies. All models were developed using functions in the Design package (Harrell 2001) for the statistical software R (Ihaka and Gentelman 1996). The PO model assumes that the independent variables vary linearly with the ordinal response variable. To check this assumption, we plotted the mean of each of the final predictor variables for each damage level against the expected value under ordinality. Confirmation of the PO assumption can be determined if the observed means are similar to the expected.

Despite the more relaxed assumptions for ordinal regression (similar to logistic) compared to ordinary least squares (OLS), one remaining requirement is that observations are independent. Since we know mountain pine beetle processes are structured spatially, we might also expect damage level to be spatially autocorrelated, and therefore there is the potential for inflated parameter estimates due to reduced degrees of freedom (Cliff and Ord 1981; Legendre 1993). Following the approach laid out in Bigler et al. (2005), we applied the Huber-White covariance estimator for cluster-correlated data, where each stand polygon was treated as a cluster (Huber 1967; White 1982). Correlated responses were corrected using the Huber-White method as implemented in the Design package for R (Harrell 2001). All reported model parameter estimates are corrected versions.

Model Development, Selection and Validation

Models were developed using a random sample of approximately 50% of the polygons in the Vanderhoof VRI data (n = 23,683). Edge polygons were excluded so as to avoid any edge effects in the neighborhood effect variable. Significance of predictor variables was assessed with the Wald x^2, which tests the null hypothesis that the coefficient is zero. The relative importance of variables was determined with x^2-df on significant variables in the model. The overall fit of the model was assessed with the Nagelkerke R-Square (R^2) (Nagelkerke 1991). Furthermore, model fit was also assessed with a measure of association for ordinal data, Goodman-Kruskal Gamma (Goodman and Kruskal 1954), which accounts for the ordered nature of the data (i.e., L predicted as M is less wrong than L predicted as VS); calculated as:

$$\gamma = (P - Q)/(P + Q) \quad (4)$$

Where P is the number of concordant pairs and Q is the number of discordant pairs, so γ ranges from 0 to 1 where 1 indicates perfect prediction. Predictions no better than random will have γ near 0. In our analysis, P corresponds to the number of pairs classified in a class adjacent to the true damage level, while Q is the number of pairs where the predicted class is not adjacent to the true damage level. Ties, or correctly predicted classes, are ignored. Gamma can be interpreted as contribution of the independent variables in reducing errors incurred when predicting the response randomly.

To assess our model as a means of developing a large-area risk index, we compared a baseline model where damage level is related to the Shore and Safranyik Pine Risk Index (PRI), a modification of the 1992 risk rating system, to the best model obtained by adding the additional variables outlined in Table 3. Additionally, the component variables of PRI were also included separately during model development to assess their individual impact on risk. This allowed us to determine the relative enhancements additional variables may play in risk rating for large areas. Potential collinearity in variables was assessed with Pearson's r for each pairwise combination of variables. For values of $r > 0.7$, we tested variables separately during model development. As a final test to determine the impact of the interaction variables on model results, they were replaced by untransformed versions and models were re-estimated and assessed.

Model testing on independent data (i.e., the remaining 50% of the VRI polygons; n = 23,729) was also undertaken to assess the predictive accuracy of the adjusted model. The final adjusted model used for prediction used significant variables with coefficients >= 0.01. Prediction accuracy was assessed for each level of the dependent variable to assess the variation in model sensitivity to different damage levels. Prediction accuracy was measured by calculating sensitivity, specificity and classification accuracy for each damage level based on a 50% probability threshold.

We also assessed prediction accuracy with the area (AUC) under the receiver operator characteristic (ROC) curve for each set of predictions (Fielding and Bell 1997; Manel et al. 2001; Allouche et al. 2006; Coudun and Gegout 2007; Wunder et al. 2007). The ROC is a plot of the true positive rate on the *y* axis against the false positive rate on the *x* axis. The AUC is a threshold independent measure of overall accuracy with values from 0.5, indicating a random model, to 1, indicating perfect prediction. The AUC avoids the need for an arbitrary selection of a single probability threshold and provides information about the nature of both sensitivity and specificity (Manel et al 2001; Allouch et al. 2006). ROC plots for each predicted damage level are presented to assess the trade off between sensitivity and specificity. Hosmer and Lemeshow (2000) suggest the following general guide to interpreting the AUC: a value >0.9 is outstanding prediction, 0.8 – 0.9 is an excellent model, and 0.7 – 0.8 is an acceptable model.

Results

Baseline model

The fit of the baseline model is summarized in Table 4a. Overall, PRI in 2004 related weakly to forest damage level in 2005. PRI was a significant predictor variable (x^2 =

2603.66, $p < 0.0001$) of forest damage, yet the model explained a small proportion of the overall variability ($R^2 = 0.11$). The ordinal association measure was low, with $\gamma = 0.26$. The adjusted parameter estimate for PRI indicates a very slight increase (1.54%) in the odds of an increase in damage level will occur with one unit increase in PRI. The baseline model was not assessed for predictive accuracy on the independent data.

Adjusted model

Pearson's correlation matrix of all new variables revealed collinearity in NUMRED and INC0304. NUMRED was subsequently dropped from model development. The fit of the adjusted model incorporating additional variables is summarized in Table 4b. With all variables, the model was moderately successful (Nagelkerke $R^2 = 0.47$, $\gamma = 0.57$) at explaining the variation in damage level. The relative importance of model variables is presented in Figure 3. Increased probability of an increase in damage level of forests in 2005 was associated mostly with R_C, INC0203, BEC subzones in the ESSF zone, MSxv, SBPSdc, SBSdw, SBSmc, R_SI and BP. The plots of the independent variable means against the expected for each damage level indicated that the PO assumption was justified for most variables (Figure 4). Parameter estimates are provided in Table 4b. Re-estimating the full model with interaction variables replaced by the untransformed variables from the VRI (site index, crown closure, quadratic mean dbh) yielded a weaker model with much less explanatory power (Nagelkerke $R^2 = 0.21$, $\gamma = 0.37$).

Prediction Accuracy

The adjusted model performed reasonably well when using a 50% probability of damage threshold for the cutoff (Table 5). The percentage of correctly classified cases ranged between 75%-93%. Examining the prediction performance for positive against negative cases revealed widely varying prediction accuracies. The sensitivity, or true positive rate, ranged from 1%-98%, while the specificity, or true negative rate, ranged from 48%-100%. Sensitivity declined with increasing damage level, while specificity showed the opposite trend. The variability in these measures however demonstrates their sensitivity to prevalence, common to most measures derived from contingency tables (Allouche et al 2006).

Predictions made on cumulative damage levels yielded an average AUC of 0.84. This falls in the range of an ‘excellent’ predictor. The ROC curves for each damage level are presented in Figure 5. Predictions of polygons with at least a trace amount of damage, analogous to predicting red attack presence, yielded an AUC of 0.93. Predictions made at other damage levels ranged from 0.80 to 0.83. All of these scores indicate good prediction accuracy model for the adjusted model.

The spatial distributions of predicted probabilities for each cumulative damage level are presented in Figure 6. Polygons in each map indicate the probability value that it is equal to or above the specified damage level. There is large spatial variation in damage level. A large patch of non susceptible pine is discriminated fairly well at the low damage level. Prediction probabilities for very severe are quite low, indicating underprediction for areas where very severe damage existed. However, overall predictive accuracy for the very severe class was high because so much of the area was classified as true negative (Table 5). The 50% probability threshold used to derive contingency table scores in Table 5 was too low to capture the predictions of very severe damage output by the model. This highlights the difficulty of using a fixed probability threshold for predicting multiple ordinal classes. Depending on infestation level or user information need, the threshold may be made more or less conservative. The probability values may also be mapped to spatially identify trends in infestation levels and as an aid to planning or mitigation activities.

Discussion

The improvement of the adjusted model over the baseline model highlights the importance of additional variables for assessing the risk of mountain pine beetle infestation over large areas. The components of PRI such as tree age, location, stand density, and stand composition are derived from many field experiments and have been validated empirically (Safranyik et al 1974 and others), yet when used for modelling stand risk in this analysis, the PRI did not perform well, either as an index, or when broken out into its individual factors. Only one (BP) out of the five most important predictors of damage level was from the baseline model (Figure 3). Reasons for this disparity are likely based on data quality rather than biological factors associated with the beetle. Specifically, performance is likely impacted by how well different variables are represented in the forest inventory and the nature of mountain pine beetle aerial detection surveys. For example, crown closure may be more accurately estimated than stems per hectare via air photo interpretation. Additionally, the nonlinear relationship between risk and stem density in PRI, representing intraspecific competition yielding weaker, less productive beetle broods (Shore and Safranyik 1992), may not be important for modelling the overall trend of the infestation over large areas. Stands with many moderately productive trees or fewer very productive trees, as long as they are above the diameter threshold, will contribute to an increased beetle population the following year.

The objective of this research was to incorporate knowledge gained from recent mountain pine beetle research, and new methodological approaches, to identify variables important for predicting risk of mountain pine beetle infestation over large areas. The adjusted model (Nagelkerke R^2 = 0.47, γ = 0.57) demonstrated an improved fit over the baseline model (Nagelkerke R^2 = 0.11, γ = 0.26), indicating the importance of some of the new variables in predicting risk over large areas. The adjusted model performed well when used for predicting mountain pine beetle damage level on independent data (AUC = 0.84). The ROC curves for model predictions at different cumulative damage levels in Figure 5 show that the model predicts best for at least trace amounts, which is similar to a binary logistic model predicting locations of red attack. Interestingly, the next best predictions were for polygons of with at least severe and at least very severe damage levels.

The use of ordinal regression modelling provides a framework for exploring the probability of occurrence of different levels of mountain pine beetle infestation. The PO ordinal model specifies constant regression coefficients and regression intercepts that vary with different levels of the dependent variable. We applied the adjusted model to independent data and assessed prediction accuracy for different damage levels. Using the AUC as our measure of predictive accuracy avoided the need to select an arbitrary probability threshold. This is important for mountain pine beetle risk assessment because the main source of beetle population data, aerial overview surveys, is collected in ordinal mortality classes. We were able to explore the variation in these classes, and build a new model of risk for large areas that predicts new occurrences of mortality classes. In our analysis, mortality classes were scaled by the percentage of pine in each polygon to yield damage level.

Probabilities derived from the adjusted model for cumulative levels of damage were mapped for the entire study area (Figure 6). Each map shows the probability that each polygon is at greater than or equal to a given damage level. Examining the difference between these maps sheds light onto the spatial distribution of mountain pine beetle risk over a very large area. While beyond the scope of this current research, spatial investigation of the differences between these probability maps may yield insights into the importance of various model input variables (Wulder et al. 2007).

The interaction variables R_C and R_SI were both important predictors of damage level. The s-curve functions defining the interaction variables appear to have captured the dynamics of damage level increases. When interaction variables were replaced with untransformed versions, the model fit declined markedly. R_C was the most important variable in the model (Figure 3). Increases in R_C were associated with an increased probability of an increase in damage level. This supports results from Powell et al. (2000), which found crown closure to be positively related to stand risk to mountain pine beetle attack. The interaction between beetle population level and crown closure indicates that high crown closure, high percentage of pine, and a large increase in infestation in the year previous is positively associated with increases in damage level. Perhaps modelling crown closure in this way captures the period during which stand resistance declines, yet phloem thickness remains sufficient to sustain the epidemic (Berryman 1982b), or indicates that beetle populations have grown to the point where even an effective host resistance is no defense against beetle mass-attack (Safranyik and Carroll 2006) . For predicting risk over large areas, R_C may be a more useful variable for risk rating than variables based on stand density. This is important because crown closure is more commonly available than stand density in forest inventory data

The site index interaction variable, R_SI, was also an important predictor of damage level. Site index has been identified in previous models of red attack damage as an important variable at the stand level (Coops et al. 2006). In our model, R_SI was inversely related to the increases in beetle damage. This suggests that high site index combined with high percentage of pine is associated with lower odds of a stand sustaining an increased damage level. This is expected if site index accurately reflects the site conditions of the stand and contributes to increased tree vigor and concomitant stand resistance to beetle attack. Furthermore, site index is a difficult variable to include in models of large area risk because it generally refers to site productivity of the leading species. Thus, where lodgepole pine is not the leading species, site index does not represent site productivity for lodgepole pine. In the Vanderhoof Forest District, lodgepole pine is a leading species in over 50% of stands. Species composition is less of an issue in our model because the upper bound of R_SI is defined by the percentage of pine, so stands where site index describes a non-pine leading species will have a maximum R_SI value less than 0.50.

BEC subzones were also important factors for predicting damage level. Negative associations were found in the SBPSmc and the SBSdk. The SBPSmc occurs along the south eastern edge and south western edges of the study area (Figure 1). This zone is dominated by dry lodgepole pine forests, and the site productivity in the SBPS is generally low. Stressed and stunted trees in this zone may be limiting the broods of beetles, and the cold climate of the SBPS may be causing greater mortality of beetles in this area relative to other zones (Steen and Demarchi 1991). The SBSdk occurs throughout study area at lower elevations, following the Nechako river valley. Here, conditions for lodgepole pine are more favourable with higher site productivity, yet mixed forests are also more common (Meidinger et al. 1991). Since a high damage level is related to the amount of pine in each stand, the risk level may not increase past a certain level in these areas. All other subzones had positive associations with damage level. The importance of climate and location in the mountain pine beetle system is well known. The BEC subzones in our model of stand risk offer a suitably scaled stratification of the landscape for exploring regional variation in mountain pine beetle processes, and subsequently, for predicting the future course of epidemics.

Interestingly the most important beetle population variable in the model was the increase in infestation in 2003, two years previous to the prediction year. Over large areas, locations previously infested are more likely to be infested than locations that have not already subjected to attack (White et al. 2006). It is possible that the winter conditions (brood over-wintering success) in 2002 were similar to those in 2004, resulting in similar outcomes.

Mitigating the spread of mountain pine beetles into new areas may be facilitated by focusing efforts on known areas of infestation to effect local reductions in beetle populations (as well as the proportion of beetles eligible for long range transport). The effort and expense in seeking endemic infestation levels at the leading edge, which may be less “dangerous” as fewer beetles exist locally for mass attack, and the proportion of beetles that do engage in long range transport (estimated at 2.5% by Safranyik et al. 1992) will not be sufficient in number to overwhelm the defenses of a mature pine (especially in climatically difficult, new range, environments). Much research has indicated that short range dispersal is the primary means of infestation spread over large areas. Perhaps the largest, most intense landscape infestations should be the focus of management activities aimed at reducing the spread of beetle populations into new areas.

Modelling beetle-host interactions at the stand level with interaction variables also improved our model. Incorporating ecological complexity into models of risk requires a trade off with accuracy. Our formulation captures some of the dynamics of this ecological complexity, yet remains relatively simple to implement with operational data. Future risk assessment models may benefit from exploring the concept of interaction variables further.

The inclusion of BEC subzones in the model provides evidence that incorporating information about local ecological characteristics is vital to any model of mountain pine beetle infestation. For British Columbia, the BEC system offers a suitably scaled classification which could serve this purpose. However in order to incorporate the full BEC system into a mountain pine beetle risk model, detailed analysis of the suitability of each subzone for mountain pine beetle is required. One option would be to map mountain pine beetle climate suitability classes (Carroll et al. 2006) to BEC subzones and use these as a component in a large area risk index. Alternatively, regional variation in mountain pine beetle populations may be too complex to model accurately with one model. A better approach may be to develop region specific models of risk.

Conclusion

In this research we developed a model using operationally available data for predicting mountain pine beetle damage over a large area. Our objective was to identify new variables related to mountain pine beetle risk, with the aim of ultimately developing an index, similar to the Shore and Safranyik PRI, which could easily be applied to large areas throughout the mountain pine beetle range. The results of our model have several implications for predicting the spatial pattern of mountain pine beetle infestations.

The addition of variables describing the temporal trend in beetle population improved our ability to predict future patterns of infestation. Detection and mapping of beetle red attack damage levels have been topics of extensive research and many sources of data are available to accurately characterize past infestations (AOS, satellite remotely sensed data, helicopter-GPS, and air photo interpreted products). Most beetle monitoring programs have been collecting data for many years and for many locations, multiple years of survey data have been compiled. Additionally, the use of various remotely sensed data sources facilitates retrospective analyses for spatial and/or temporal gaps in the historical survey record. The integration of various sources of survey data allows temporal and spatial trends in the beetle infestation to be extracted and used for risk assessment. In our analysis, the beetle population from two years previous to the prediction year was most important for discriminating damage level.

Through this research we have demonstrated the utility of including readily available data sets for representing beetle population trend, interaction with hosts, and an stratification of the landscape to aid in operational risk rating of mountain pine beetle infestation. We follow Shore et al. (2006), in acknowledging the stochastic nature of the prediction mountain pine beetle infestation likelihood. The complex interactions between forest conditions, climate, and insects have many random or poorly understood elements, and as a result, any improvement in model predictive ability, especially those facilitated through readily available additional support data, are welcomed and of use to the science and management communities interested in mountain pine beetle movement and impacts.

Acknowledgements

This research was funded by the Government of Canada as part of the Mountain Pine Beetle Program administered by Natural Resources Canada, Canadian Forest Service, with additional information available at: http://mpb.cfs.nrcan.gc.ca/.

References

O. Allouche, T. Asaf, K. Ronen, 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43, 1223-1232.

G.D. Amman, 1973. Population changes of the mountain pine beetle in relation to elevation. *Environmental Entomology*, 2, 541-547.

Amman, G.D., McGregor, M.D., Cahill, D.B., Klein, W.H., 1977. Guidelines for reducing losses of lodgepole pine to the mountain pine beetle in unmanaged stands in the Rocky Mountains. USDA For. Serv. Gen. Tech. Rep. INT-136.

J. Anhold, M. Jenkins, 1987. Potential mountain pine beetle (Coleoptera: scolytidae) attack of lodgepole pine as described by stand density index. *Environmental Entomology*, 16, 738-742.

B.G. Armstrong, M. Sloan, 2001. Ordinal regression models for epidemiologic data. *American Journal of Epidemiology*, 129, 191-204.

B.H. Aukema, A.L. Carroll, J. Zhu, K.F. Raffa, T.A. Sickley, S.W. Taylor, 2006. Landscape level analysis of mountain pine beetle in British Columbia, Canada: spatiotemporal development and spatial synchrony within the present outbreak. *Ecography*, 29, 427-441.

Bartos, D.L., Amman, G.D., 1989. Microclimate: An alternative to tree vigor as a basis for mountain pine beetle infestations. Res. Pap. INT-400, USDA For. Serv. 10p.

B. Bentz, G. Amman, L. Logan, 1993. A critical assessment of risk classification systems for the mountain pine beetle. *Forest Ecology and Management*, 61, 349-366.

B.J. Bentz, J.A. Logan, J.C. Vandygriff, 2001. Latitudinal variation in Dendroctonus ponderosae (Colyoptera: Scolytidae) development time and adult size. *Canadian Entomologist*, 133, 375-387.

Berryman, A.A. 1978a. A synoptic model of the lodgepole pine/mountain pine beetle interactions and its potential application in forest management. In: Theory and practice of mountain pine beetle management in lodgepole pine forests, ed. A.A. Berryman, G.D. Amman, and R.W. Stark, 98-105. Moscow, ID: University of Idaho.

A.A. Berryman, 1978b. Towards a theory of insect epidemiology. *Researches on Population Ecology*, 19, 181-196.

Berryman, A.A. 1982a. Population dynamic of bark beetles. Pages 264-314 In: J. B. Mitton and K.B. Sturgeon, eds. Bark beetles in North American conifers: a system for the study of evolutionary biology. University of Texas Press, Austin, TX. 527 p.

A.A. Berryman, 1982b. Mountain pine beetle outbreaks in Rocky Mountain lodgepole pine forests. *Journal of Forestry*, 80, 410-413.

C. Bigler, D. Kulakowski, T.T. Veblen, 2005. Multiple disturbance interactions and drought influence fire severity in Rocky Mountain subalpine forests. *Ecology*, 86, 3018-3029.

G.A. Blackburn, E.J. Milton, 1996. Filling the gaps: remote sensing meets woodland ecology. *Global Ecology and Biogeography Letters*, 5, 175-191.

British Columbia Ministry of Forests. 1995. Bark Beetle Management Guidebook Table of Contents. Forest Practices Branch.

British Columbia Ministry of Forests. 2000. Forest health aerial overview survey standards for British Columbia, Version 2.0. Forest Practices Branch and Canadian Forest Service, Victoria, BC, 45 p.

British Columbia Ministry of Sustainable Resource Management. 2002. Vegetation Resources Inventory: Photo Interpretation Procedures, Version 2.4. http://www.for.gov.bc.ca/hts/vri/standards/.

Carroll, A.L., Regniere, J., Logan, J.A., Taylor, S.W., Bentz, B.J., Powell, J.A., 2006. Impacts of climate change on range expansion by the mountain pine beetle. Natural

Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC. Mountain Pine Beetle Initiative Working Paper 2006-14. 20 p.

Cliff, A., Ord, J.K., 1981. Spatial Processes Models and Applications. Pion Limited, London, 266p.

N.C. Coops, M.A. Wulder, J.C. White, 2006. Integrating remotely sensed and ancillary data sources to characterize a mountain pine beetle infestation. *Remote Sensing of Environment*, 30, 83-97.

C. Coudun, J.C. Gegout, 2007. Quantitative prediction of the distribution and abundance of Vaccinium myrtillus with climatic and edaphic factors. *Journal of Vegetation Science*, 18, 517-524.

C.C. Dymond, M.A. Wulder, T.L. Shore, T.A. Nelson, B. Boots, B. Riel, 2006. Evaluation of risk assessment of mountain pine beetle infestations. *Western Journal of Applied Forestry*, 21, 5-13.

M. Eng, D. Meidinger, 1999. A method for large-scale biogeoclimatic mapping in British Columbia. Province of British Columbia, Victoria, BC.

A.H. Fielding, J.F. Bell, 1997. A review of methods for the assessment of prediction error in conservation presence / absence models. *Environmental Conservation*, 24, 38-49.

S. Franklin, M.A. Wulder, R. Skakun, A. Carroll, 2003. Mountain pine beetle red-attack forest damage classification using stratified Landsat TM data in British Columbia, Canada. *Photogrammetric Engineering & Remote Sensing*, 69, 283-288.

L.A. Goodman, W.H. Kruskal, 1954. Measures of association for cross classifications. *Journal of the American Statistical Association*, 49, 732-764.

A. Guisan, F.E. Harrell, 2000. Ordinal Response Regression Models in Ecology. *Journal of Vegetation Science*, 11, 617-626.

Harrell, F.E. 2001. Regression Modeling Strategies: With applications to linear models, logistic regression, and survival analysis. Springer, New York, 600 p.

Hosmer, D.W., and Lemeshow, S. 2000. Applied Logistic Regression. 2nd edn. John Wiley, New York, 392 p.

Huber, PJ. 1967. The behaviour of maximum likelihood estimates under non-standard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1:221-233.

I. Ihaka, R. Gentelman, 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5, 299-314.

L. Kumar, A.K. Skidmore, E. Knowles, 1997. Modelling topographic variation in solar radiation in a GIS environment. *International Journal of Geographic Information Science*, 11, 475-497.

P. Legendre, 1993. Spatial autocorrelation: Trouble or new paradigm? *Ecology*, 74, 1659-1673.

J. Logan, P. White, B. Bentz, J.A. Powell, 1998. Model analysis of spatial patterns in mountain pine beetle outbreaks. *Theoretical Population Biology*, 53, 236-255.

J. Logan, B. Bentz, 1999. Model Analysis of Mountain Pine Beetle (Coleoptera: Scolytidae) Seasonality. *Environmental Entomology*, 28, 924-934.

Mahoney, R. 1978. Lodgepole pine-mountain pine beetle risk classification methods and their application. In Theory and practice of mountain pine beetle management in lodgepole pine forests, ed. D.M. Baumgartner, 106-113. Pullman, WA.

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

Meidinger, D., J. Pojar, and W.L. Harper. 1991. Sub-Boreal Spruce Zone. In Ecosystems of British Columbia, 209-221. Special Report Series 6, British Columbia Ministry of Forests, Victoria, BC.

N.J. Nagelkerke, 1991. A note on a general definition of the coefficient of determination. *Biometrika*, 78, 691-692.

J.F. Negron, J.B. Popp, 2004. Probability of ponderosa pine infestation by mountain pine beetle in the Colorado Front Range. *Forest Ecology and Management*, 191, 17-27.

T. Nelson, B. Boots, 2005. Identifying insect infestation hot spots: an approach using conditional spatial randomization. *Journal of Geographical Systems*, 7, 291-311.

T. Nelson, B. Boots, M.A. Wulder, T. Shore, L. Safranyik, T. Ebata, 2006. The impact of data on the rating of forest susceptibility to mountain pine beetle infestations. *Can. J. For. Res.*, 36, 2815-2825.

W.A. Nelson, A. Potapov, M.A. Lewis, A.E. Hundsdorfer, F. He, 2007. Balancing ecological complexity in predictive models: A reassessment of risk models in the mountain pine beetle system. *Journal of Applied Ecology*, 45, 248-257.

K.F. Raffa, A.A. Berryman, 1986. A mechanistic computer model of mountain pine beetle populations interacting with lodgepole pine stands and its implications for forest managers. *Forest Science*, 32, 789-805.

C. Robertson, T. Nelson, B. Boots, 2007. Mountain Pine Beetle Dispersal: The spatial-temporal interaction of infestations. *Forest Science*, 53, 395-405.

Safranyik, L, DM Shrimpton, and HS Whitney. 1974. Management of lodgepole pine to reduce losses from the mountain pine beetle. Forestry Technical Report 1. Pacific Forestry Centre, Canadian Forestry Service,Victoria, BC, 25 p.

Safranyik, L., Shrimpton, D. and Whitney, H. 1975. An interpretation of the interaction between lodegpole pine, the mountain pine beetle, and its associated blue stain fungi. Managment of Lodgepole Pine Ecosystems. In: Baumgartner, D.M. (ed.), Managment of Lodgepole Pine Ecosystems Symp. Proc., Washington State Univ. Cooeprative Extension Service, pp. 406-428.

Safranyik, L., 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, 3-66. Victoria, BC: Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre.

J.A. Schenk, R.L. Mahoney, J.A. Moore, D.L. Adams, 1980. A model for hazard rating lodgepole pine stands for mortality by mountain pine beetle. *Forest Ecology and Management*, 3, 57-58.

Shore, T., Safranyik, L. 1992. Susceptibility and risk rating systems for the mountain pine beetle in lodgepole pine stands. Information Report No. BC-X-336.Forestry Canada, Pacific and Yukon Region, Pacific Forestry Centre, Victoria, BC, p.12.

T.L. Shore, L. Safranyik, J.P. Lemieux, 2000. Suceptibility of lodgepole pine stands to the mountain pine beetle: testing of a rating systems. *Canadian Journal of Forest Research*, 20: 44-49.

Shore, T.L., B.G. Riel, L. Safranyik, and A. Fall. 2006. Decision support systems. In: The mountain pine beetle: A synthesis of biology, management, and impacts on lodgepole pine, 193-230. Victoria, B.C.: Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre.

D.M. Shrimpton, 1973. Age-and size-related response of lodgepole pine to inoculation with Europhium clavigerum. *Canadian Journal of Botany*, 51, 1155-1160.

D.M. Shrimpton, A.J. Thomson, 1983. Growth characteristics of lodgepole pine associated with the start of mountain pine beetle outbreaks. *Canadian Journal of Forest Research*, 13, 137-144.

J. Sirois, F.J. Ahern, 1989. An investigation of SPOT HRV for detecting recent mountain pine beetle mortality. *Canadian Journal of Remote Sensing*, 14, 104-110.

R. Skakun, M.A. Wulder, S. Franklin, 2003. Sensitivity of the Thematic Mapper enhanced wetness difference index to detect mountain pine beetle red-attack damage. *Remote Sensing of Environment*, 86, 433-43.

O. Steen, D.A. Demarchi, 1991. Sub-Boreal Pine-Spruce Zone. In Ecosystems of British Columbia, 195-208. Spec. Rep. Ser. 6: B.C. Ministry of Forests.

J.D. Stuart, 1984. Hazard rating of lodgepole pine stands to mountain pine beetle outbreaks in southcentral Oregon. *Can. J. For. Res.*, 14, 666-671.

C.D. Thomas, E.J. Bodsworth, R.J. Wilson, A.D. Simmons, Z.G. Davies, M. Musche, 2001. Ecological and evolutionary processes at expanding range margins. *Nature*, 411, 577-581.

S.H. Walker, D.B. Duncan, 1967. Estimation of the probability of an event as a function of several independent variables. *Biometrika*, 54, 167-179.

H. White, 1982. Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1-25.

J.C. White, M.A. Wulder, D. Grills, 2006. Detection and mapping mountain pine beetle red-attack damage with SPOT-5 10-m multispectral imagery. *BC Journal of Ecosystems and Management*, 7, 105-118.

H.X. Wu, C.C. Ying, H.B. Ju, 2005. Predicting site productivity and pest hazard in lodgepole pine using biogeoclimatic system and geographic variables in British Columbia. *Ann. For. Sci.*, 62, 31-42.

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

M.A. Wulder, J.C. White, C.C. Dymond, T. Nelson, B. Boots, T.L. Shore, 2006b. Calculating the risk of mountain pine beetle attack: A comparison of distance- and density-based estimates of beetle pressure. *Journal of Environmental Informatics*, 8, 58-69.

M.A. Wulder, C.C. Dymond, J.C. White, D.G. Leckie, A.L. Carroll, 2006c. Surveying mountain pine beetle damage of forests: A review of remote sensing opportunities. *Forest Ecology and Management*, 221, 27-41.

M.A. Wulder, J.C. White, N.C. Coops, T. Nelson, B. Boots, 2007. Using local spatial autocorrelation to compare outputs from a forest growth model. *Ecological Modelling*, 209, 264-276.

J. Wunder, B. Reineking, J-F. Matter, C. Bigler, H. Bugmann, 2007. Predicting tree death for Fagus sylvatica and Abies alba using permanent plot data. *Journal of Vegetation Science*, 18, 525-534.

Table 1 Overview of recent studies and variables for predicting mountain pine beetle red attack damage.

Publication	Scale	Primary Data Source	Dependent Variable	Significant Predictors	Method	Accuracy (95% CI)
Wulder et al. 2006	landscape	Landsat7, DEM 30m	Red attack presence	+elevation, -slope	Logistic	86% (5%)
White et al. 2006	landscape	SPOT5 10m, DEM 25m	Red attack presence	+direct radiation, -elevation	Logistic	71% (9%)
Coops et al. 2006	landscape	Landsat7, Landsat5, DEM 30m	Red attack presence	+direct radiation, +elevation	Logistic	69% (N/A)
Coops et al. 2006	landscape	VRI	Probability of red attack	Site index, slope, basal area, crown closure	CART	N/A
Coops et al. 2006	landscape	VRI	Probability of red attack	Location factor, age factor, basal area factor, density factor	CART	N/A
Negron and Popp 2004	Plot	Field plot	Probability of red attack	Stand density index, quadratic mean diameter	CART	N/A

Table 2 – Damage level classes are based on the amount of infestation and the percentage of pine. Percent area mortality in the AOS data is multiplied by the percentage of pine in each VRI polygon.

Damage Level	Criteria
None	0
Trace	<= 1
Low	>1 <=10
Medium	>10 <=29
Severity	>29 <=49
Very Severe	>49

Table 3 – New Variables included in model development.

Variable name	Category	Rationale
BEC Subzone (BEC)	Stand Resistance	The beetle life cycle is principally determined by temperature. BEC zones may represent suitably scaled delineations for identifying variations in the productivity of beetle populations.
Annual direct shortwave radiation (SWR)	Stand Resistance	Changing climate may impact the effect of elevation on beetle populations. Direct radiation may indicate variability in the productivity of beetle populations more accurately than elevation alone.
Percent pine infested in 2004 (PPI04)	Beetle Population	The proportion of pine infested represents both beetles and the amount of pine remaining in the stand.
Damage level in 2004 (INC0304)	Beetle Population	Aerial increase in infestation in 2004 as recorded in aerial surveys.
Damage level in 2003 (INC0203)	Beetle Population	Aerial increase in infestation in 2003 as recorded in aerial surveys.
Number of years infestation present in 2004 (INFDUR)	Beetle Population	Indicator of the infestation trend.
Number of red attacked trees in stand in 2004 (NUMRED)	Beetle Population	Estimate of the local beetle population.
Site index (R_SI)	Interaction	High site indices represent productive stands and under epidemic conditions, productive beetle populations.
Crown closure (R_C)	Interaction	Stand canopies determine the amount of radiation received on the bole and impact pheromone mediated dispersal.
Quadratic mean diameter at breast height (R_QDBH)	Interaction	Mountain pine beetles attack large diameter pine disproportionately. Tree diameters above 25.4 cm are thought to be infestation sources, while those below this threshold are infestation sinks (Safranyik et al. 1974)

Table 4 – Model coefficients and estimates.

a) Baseline Model. Nagelkerke $R^2 = 0.11$, K-W $\gamma = 0.26$.

Parameter	coefficient	Wald	Sig.	S.E.
Y >= T	0.8941	37.73	0.000	0.024
Y >= L	-0.3184	-14.55	0.000	0.022
Y >= M	-1.4177	-61.07	0.000	0.023
Y >= S	-2.4484	-91.21	0.000	0.027
Y >= VS	-3.7628	-107.51	0.000	0.035
PRI	0.0153	51.03	0.000	0.000

b) Adjusted Model. Nagelkerke $R^2 = 0.47$, K-W $\gamma = 0.57$.

Parameter	coefficient	Wald	Sig.	S.E.
Y >= T	-0.8820	-12.83	0.000	0.069
Y >= L	-2.5582	-35.88	0.000	0.071
Y >= M	-3.9838	-54.66	0.000	0.073
Y >= S	-5.3274	-70.54	0.000	0.076
Y >= VS	-6.9153	-85.44	0.000	0.080
PRI	-0.0047	-9.45	0.000	0.000
BP	2.3129	30.87	0.000	0.075
Inc0203	-6.1958	-40.34	0.000	0.154
Inc0304	-1.1121	-5.15	0.000	0.216
R_QDBH	1.1073	9.67	0.000	0.114
R_C	4.8708	67.80	0.000	0.072
R_SI	-2.6736	-30.27	0.000	0.088
ESSFmvp	1.1219	2.20	0.028	0.510
ESSFxv	1.1601	3.11	0.002	0.373
MSxv	0.4191	1.89	0.059	0.222
SBPSdc	0.2374	2.00	0.046	0.119
SBPSmc	-0.6227	-10.57	0.000	0.059
SBSdk	-0.6039	-12.89	0.000	0.047
SBSdw	0.3819	8.47	0.000	0.045
SBSmc	0.2810	7.10	0.000	0.040

Table 5 – Prediction accuracy for different damage levels using 50% probability threshold. Sensitivity is the percentage of correctly predicted true positives, and specificity is the percentage of correctly predicted true negatives. Correct classification rate is the total percentage of correctly predicted cases.

Damage level	**Sensitivity**	**Specificity**	**Correct Classification Rate**
T	98%	48%	91%
L	84%	60%	76%
M	67%	80%	75%
S	38%	92%	81%
VS	1%	100%	93%

Figure Captions.

Figure 1 – Biogeoclimatic ecosystem classification (BEC) subzones of the Vanderhoof Forest District.

Figure 2 – Interaction variable curves. For each, the maximum weight is defined by the percentage of pine in the stand. Additional parameters are based on the beetle population in 2004 (bp2004), the associated VRI variables, and threshold values. The threshold value indicates when rapid increase begins and x - bp2004 defines the number of units in the VRI variable required to reach the max weight. If bp2004 is high, the weight increases faster than if bp2004 is low. Site index threshold was based on the most important value for predicting red attack damage in regression tree analysis in Coops et al. 2006. Quadratic mean diameter threshold was based on a similar minimum dbh threshold of attack in Safranyik et al. 1974 (10cm), and so that the weight increase is proportional to the percentage mortality increase for each unit increase in diameter (increase of 5 at peak growth period). Crown closure threshold of 20 was set experimentally as no theoretical values were found.

Figure 3 – Variables in new model. The measure of variable importance in the model, $x^2 - df$, takes into consideration the extra degrees of freedom in BEC subzone variable (categorical predictors), and is equivalent to the x^2 test for continuous predictor variables.

Figure 4 – Plots of independent variable means for each ordinal class for adjusted model. The PO assumption is confirmed if the observed (solid) matches the predicted (dotted) under the assumption of proportional odds. This can be roughly confirmed for 4 out of the 6 variables, suggesting the PO model was appropriate.

Figure 5 – ROC curves for 2005 mountain pine beetle forest damage levels predicted from separate model alpha coefficients.

Figure 6 – Maps of predicted probabilities for each damage level.

Covering letter

Colin Robertson
Department of Geography
University of Victoria
Mailing Address:
University of Victoria
Dept of Geography, PO Box 3050
Victoria, BC
V8W 3P5

February 4, 2008

Prof. P.M. Attiwill
Editor
Forest Ecology and Management

Dear Editor,

Please find attached our manuscript, titled: Risk rating for mountain pine beetle infestation of lodgepole pine forests over large areas with ordinal regression modelling.

This manuscript reports on an ordinal regression based approach to characterizing stands at risk of mountain pine beetle infestation using large-area, operational data. We review recent trends in mountain pine beetle research and incorporate relevant variables into a risk model.

Existing risk and susceptibility systems for mountain pine beetle were developed from stand scale field experiments and were designed with specific data requirements. Yet their use as models of risk across large landscapes, often with substitute variables, is cause for concern. In this research, we develop an operationally viable model for risk rating stands using standard inventory and forest health monitoring data. Further, our research encapsulates recent knowledge gained from the growing volume of large-scale forest disturbance research.

We appreciate consideration of our manuscript for publication in the *Forest Ecology and Management.*

Sincerely,

Colin Robertson

(Corresponding Author)

Cc: M.A. Wulder
T.A. Nelson
J.C. White

Figure1
Click here to download high resolution image

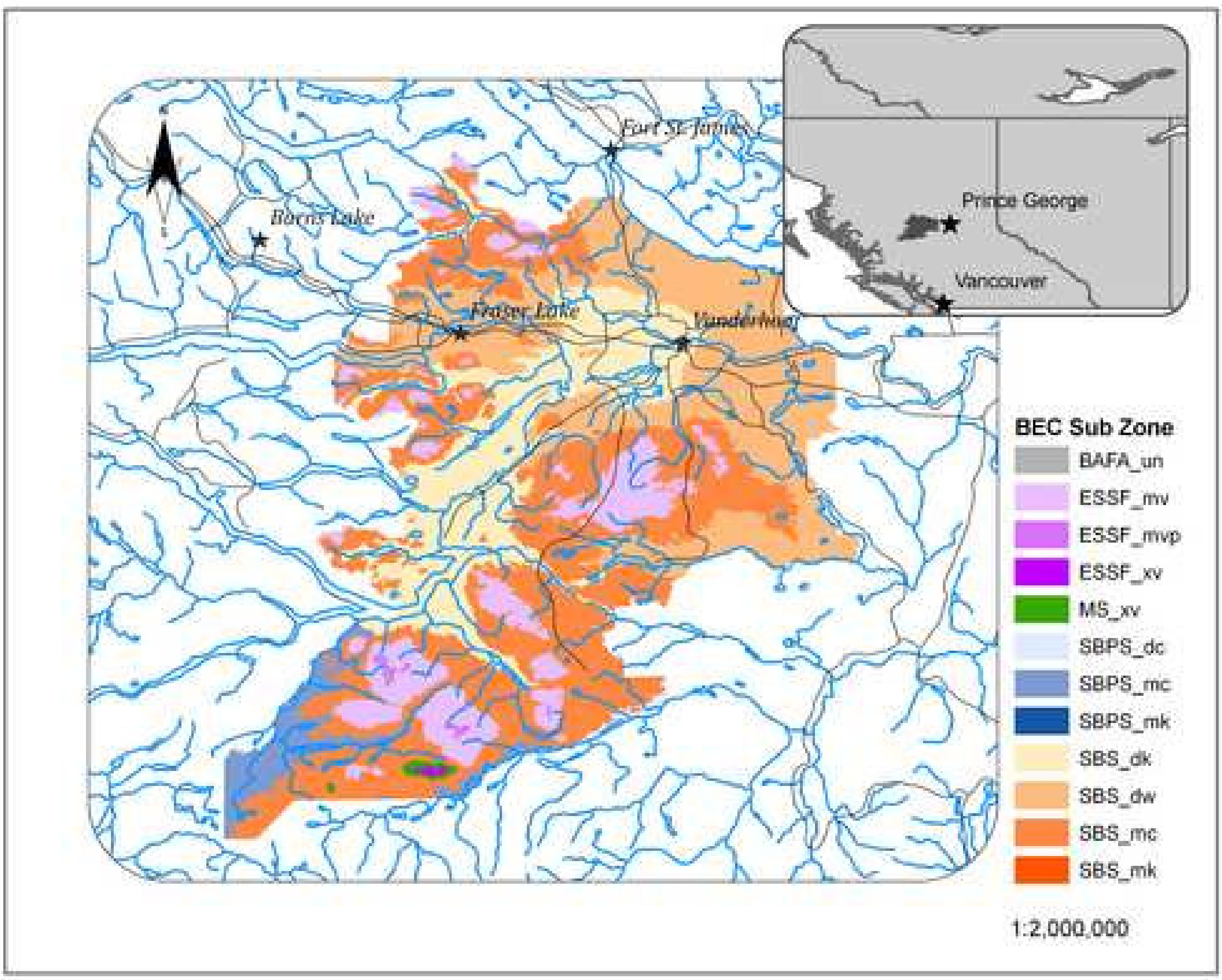

Figure2
Click here to download high resolution image

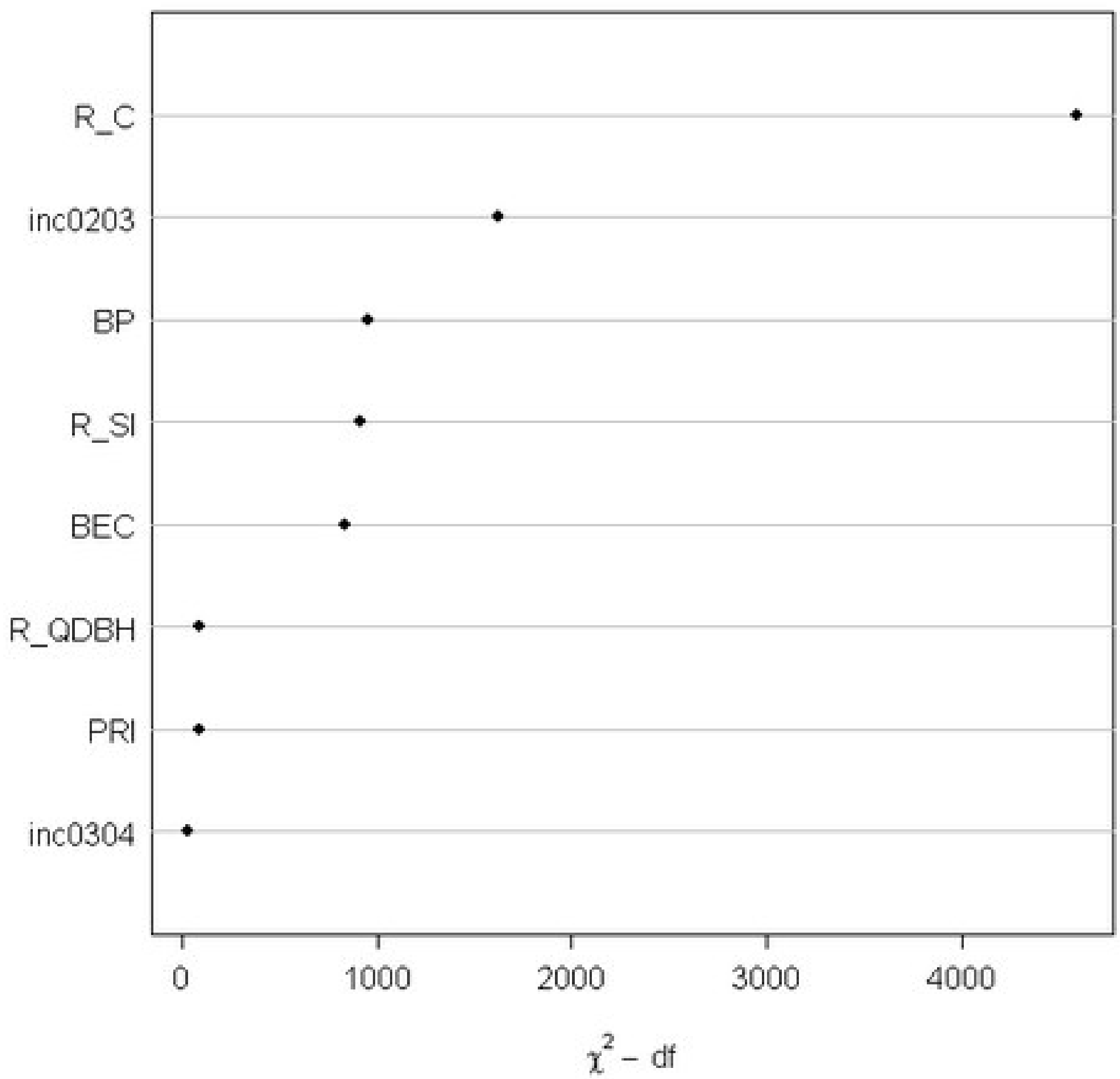

Figure3

Click here to download high resolution image

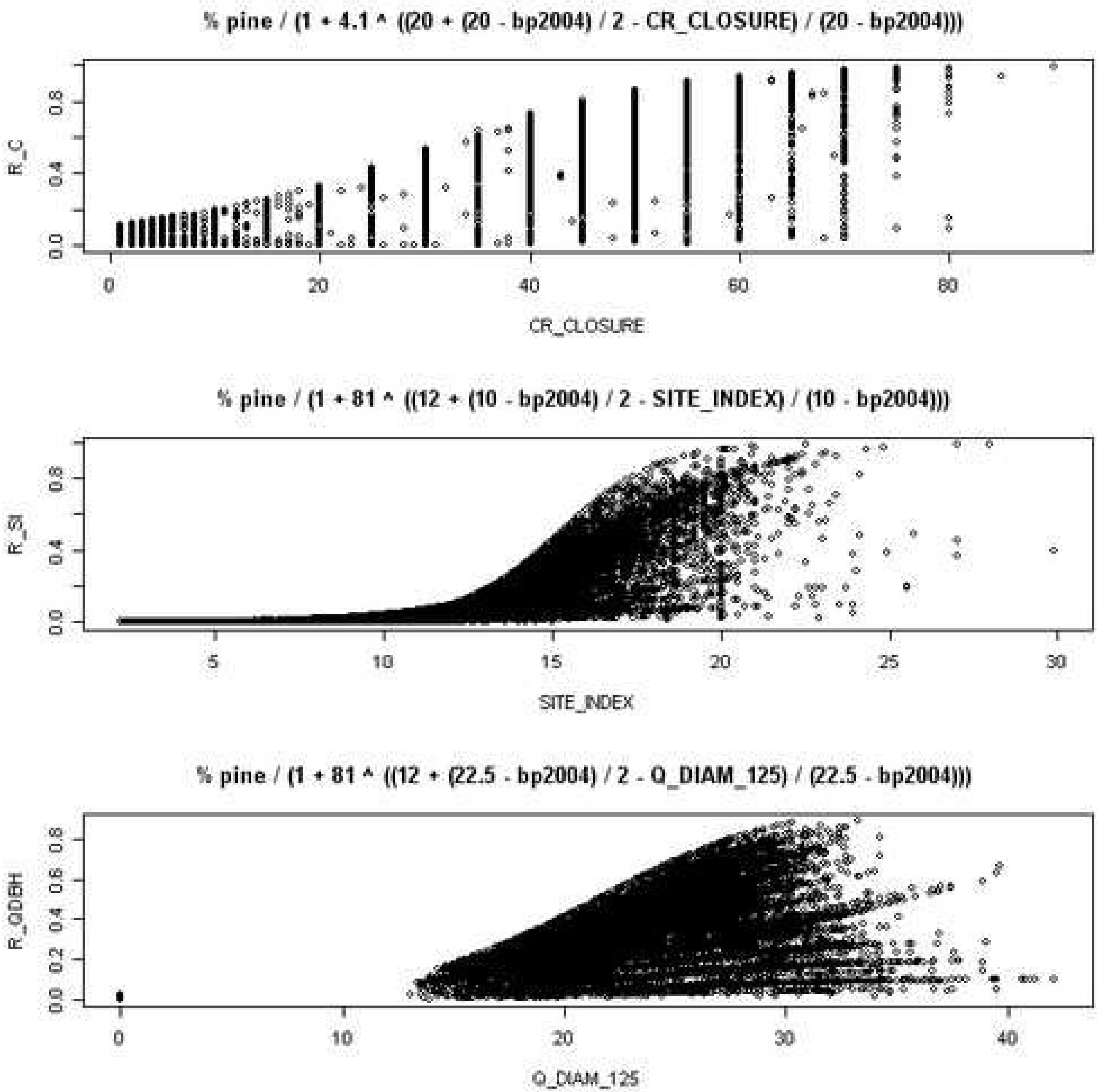

Figure4

Click here to download high resolution image

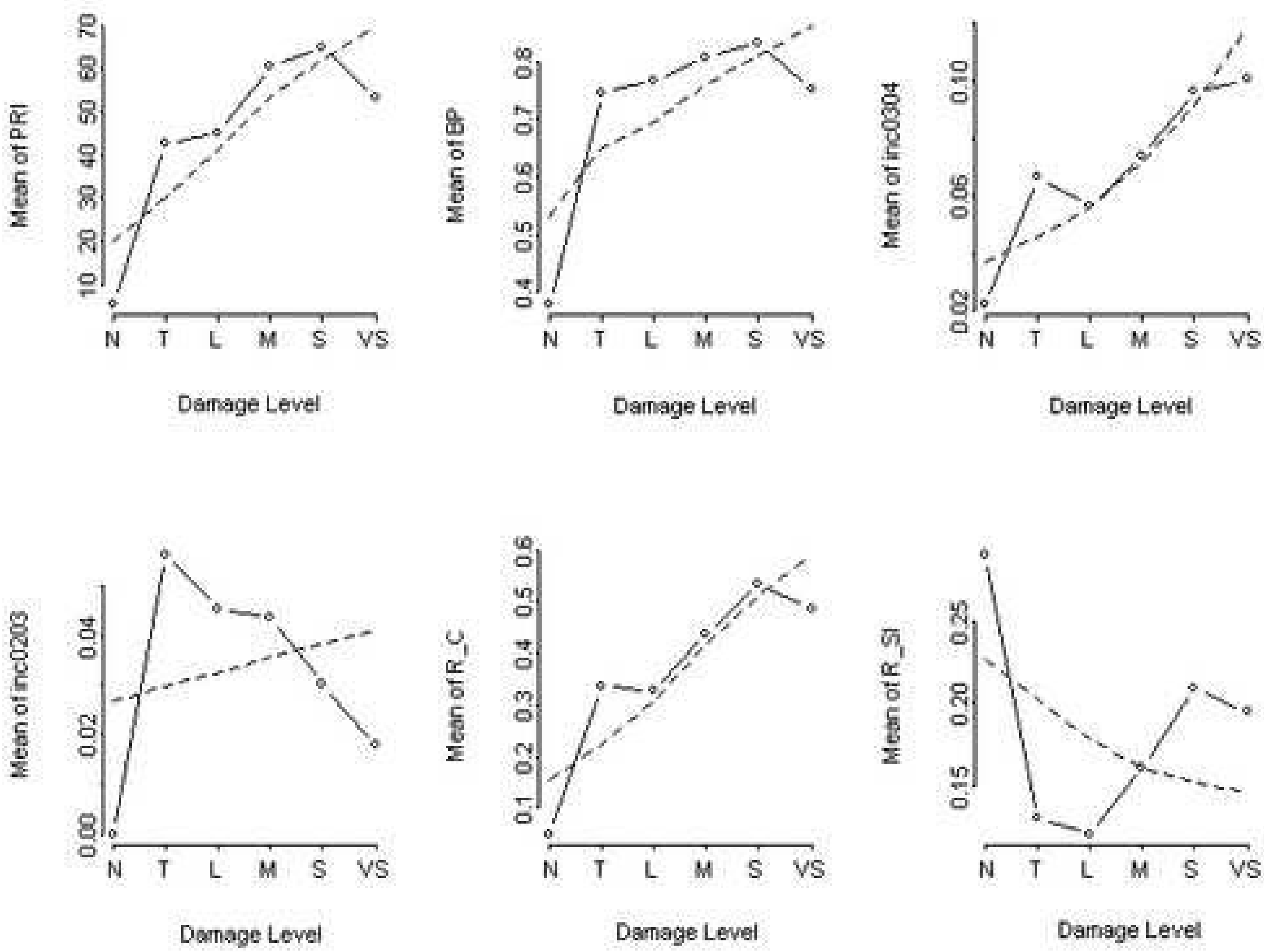

Figure5

Click here to download high resolution image

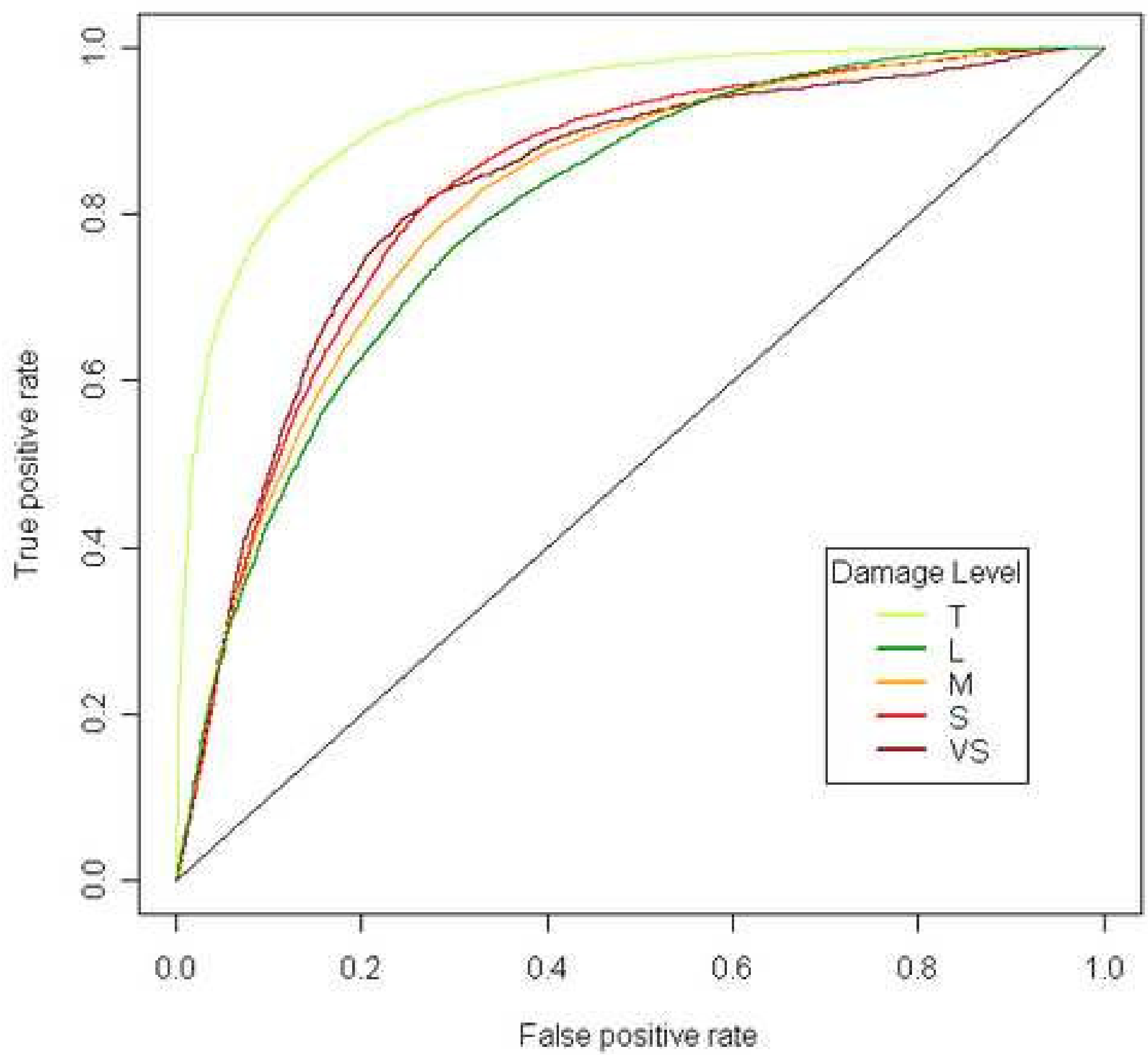

Figure6

Click here to download high resolution image