DISSERTATION

QUANTIFYING AND MAPPING TREE MORTALITY DUE TO MOUNTAIN PINE BARK BEETLES VIA ANALYSES OF REMOTE SENSING DATA IN NORTHERN COLORADO

Submitted by

Hamza A.S Taleb

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2024

Doctoral Committee:

Advisor: Melinda Laituri

Steven Fassnacht

Stephen Leisz

Neil Grigg

ABSTRACT

QUANTIFYING AND MAPPING TREE MORTALITY DUE TO MOUNTAIN PINE BARK BEETLES VIA ANALYSES OF REMOTE SENSING DATA IN NORTHERN COLORADO

In the past two decades, Mountain Pine Bark Beetle (MPBB) infestations have become more pervasive due to increasing temperatures and drought conditions related to climate change causing regional-scale mortality. Insect effects on tree die-off, fuels, and fire behavior can vary widely. A key problem in understanding insect-fire relationships is the lack of empirical maps that show interrelated changes in the distribution of insect infestations and fire zones over space and time. This study demonstrates an approach to tracking and mapping the spread of MPBB by analyses of vegetation indices calculated from Landsat TM data in the study site in northwestern Colorado. These indices were used for calculations in the Random Forest (RF) classifier algorithm and the Support Vector Machine (SVM) classifier algorithm to determine the presence or absence of MPBB and to illustrate the changes in the distribution of infestations with time. A comparison was made between the accuracy of the two classification algorithms (RF and SVM) in tracking and mapping the spread of MPBB. R^2 has proved to be a reliable measure of accuracy of regression models. If the statistical accuracy of all the models, (RF vs. SVM and binary vs. regression) are compared, both the regression and binary models based on RF are more accurate. The results of this study can provide a useful tool for forest managers to make decisions about how changing conditions affect potential problems in forest management.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor Dr. Melinda Laituri for her unwavering support, guidance, and mentorship throughout my academic journey. Her expertise and dedication have been instrumental in shaping the direction of my research and helping me achieve my academic goals.

I am also immensely thankful to my committee members, Dr. Neil Grigg, Dr. Steven Fassnacht, and Dr. Stephen Leisz, for their valuable insights, constructive feedback, and the time they devoted to reviewing and improving my work. Their collective expertise enriched the quality of my research.

To my beloved wife Emtinan Bussairi, your boundless love, patience, and understanding have been my rock and motivation during this challenging endeavor. Your unwavering support and sacrifices have made it possible for me to pursue my academic dreams.

I am also grateful to my family for their continuous encouragement and belief in my abilities. Your support has been a source of strength throughout this academic journey.

Lastly, a special mention goes to my dear friend Dr. Sharon Stonecipher. Your friendship and encouragement provided a much-needed source of motivation and emotional support during the tough times. Thank you for always being there.

I am indebted to all of you for your roles in my academic and personal growth. Your support has been invaluable, and I am fortunate to have you in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 – Introduction To the Dissertation

1.1 Background

Water is one of the most important resources in western USA. It is scarce, and its quality and quantity are highly variable from year to year. The flux of water through mountain basins in western U.S depends on the interaction of climate, vegetation, and subsurface processes. Characterizing these factors, and their influence on snowpack and water supply, is essential for effective management of land and water resources. Hydrological processes in mountains are highly sensitive to changes in climate as both precipitation and temperature can exhibit abrupt changes over short distances due to changes in altitude and differing exposures to sun radiation and winds. These can have a large impact on the amount of accumulated snow and in the timing of accumulation versus melting, resulting in changes in hydrologic regimes (Moran-Tejeda, et al., 2015).

The hydrology of the western U.S. region, where the Rocky Mountains are located, is for the most part dominated by snowmelt and precipitation. This resource is affected by several processes: water is first stored as snow; it can be lost due to interception and evaporation, sublimation or transpiration; it can be released as snow melt runoff; or it can be infiltrated to form groundwater. In forest precipitation must pass through vegetation to reach the ground. Precipitation caught by forest canopy is termed interception. Some intercepted precipitation never reaches the ground because it is evaporated back to the atmosphere. Forest vegetation typically results in greater interception losses than most other types of surfaces because forests are composed of multiple layers that provide substantial opportunity to intercept and retain precipitation. As fire, pests, and drought modify forests in response to climate change important water fluxes become altered (Adams et al., 2012). Reduced canopy cover results in reduced interception, sublimation and transpiration (Bearup et al., 2014) from this above-ground storage reservoir, as well as higher solar radiation and wind speeds through the open area (Harpold et al., 2014). All of these processes will be exacerbated by increased temperatures associated with climate change.

More modern controls on water throughput include factors such as tree mortality as a result of pine bark beetle infestation and climate change. In the last decade, infestations of bark beetle species such as the Mountain Pine Beetle (MPB) have grown from endemic to epidemic levels, killing large areas of montane and sub-alpine forests. Increased tree mortality rates due to MPB infestation cause changes in canopy cover which have the potential to result in more rapid snow melt and drier soil in the summer (Mikkelson, et al., 2011). Annual water yield and peak flows have been shown to increase following insect infestation. Decreases in forest canopy also result in decreased transpiration, and thus reduce soil water/groundwater uptake and potentially increase groundwater recharge (Mikkelson, et al., ibid.). Climate change also affects MPBB, increasing infestations as temperature rises (Anderegg et al., 2015). To understand the effect of tree mortality on hydrologic modeling, it is first necessary to have a better understanding of controls on pine bark beetle die-offs.

The extent of MPBB mortality will likely increase under both of two well established climate-change scenarios (RCP 4.5 and 8.5) in Grand County. This suggests that the impacts of MPBB outbreaks on vegetation composition and structure, and ecosystem functioning are also likely to increase in the future (Liang et al. 2014). Outbreaks also have a direct effect on hydrology through their effects on snow accumulation and melting, increased or decreased precipitation, and evaporation and sublimation. In turn, climate and weather affect outbreaks of MPBB in three ways: 1- life cycle changes as a function of temperature; 2- cold-induced mortality; 3- drought stress on host trees. Temperature directly controls the life stage development rates of mountain pine bark beetles, i.e., the progression from egg to larvae to pupae to adult. In colder situations it can sometimes take two to three years to complete the life cycle of an MPBB (Creeden et al. 2014).

A key problem to understanding the distribution of insect-caused tree mortality (e.g., Mountain Pine Bark Beetle) is the lack of empirical maps that track infestations over time and space. Information on the locations, extents, and severities of Mountain Pine Bark Beetle (MPBB) disturbances need to be updated for decision making by ecologists and natural resource managers to understand the impacts and prioritize restoration activities for minimizing future beetle outbreaks. This study describes a technique for tracking tree mortality due to MPBB infestations by analyzing remote sensing (RS) data. The goal of this study is developing a tool for tracking and mapping tree mortality over time by using RS techniques to investigate controls on tree mortality, to aid in developing new tools in forest management to mitigate the future spread of MPBB, and to enable improved planning and control of future fires.

1.2 Objectives of This Study:

The purpose of this study is to investigate dynamic forest destruction patterns related to MPBB infestation and climate change (e.g., fire severity, drought) in the Upper Yampa Basin in Colorado using RS techniques. The goal of this study is developing a tool for tracking and mapping tree mortality over time by using RS techniques to investigate controls on tree mortality and determine whether changes in forest management techniques can be developed to limit the extent and severity of future widespread tree mortality. This study addresses the effect of previous MPBB outbreaks on fire severity in the future, especially as related to climate change. The main objective of this study is to develop methods that can quantify and track outbreaks of MPBB to facilitate improved understanding of their distribution, both geographically and over time, and relate these distributions to other factors such as climate change. The main approach is to use classifier algorithms such as Random Forest (RF) and Support Vector Machine (SVM) to analyze vegetation indices calculated from Landsat data, and to statistically analyze the resulting models to determine which are most accurate.

1.3 Organization of dissertation

This dissertation is organized into 7 chapters. Chapter 1 includes a general introduction on the purpose of this study, as well as a description of the study site and its characteristics, methods of field data collection, and a table showing various vegetation indices related to MPBB infestations. Chapter 2 includes a literature review of the MPBB impacts on pine forests in North America and Colorado in particular, and reviews of remote sensing data and algorithms used to track and monitor MPBB infestations. Chapter 3 illustrates how a time series analysis was developed and used to detect changes in forest health with time and to relate these changes to factors such as fires, clearcutting, and MPBB infestations. Chapter 4 discusses how data from a set of Landsat TM wavelength bands were used to calculate standard vegetation indices, which were then analyzed using the RF classifier algorithm to detect, map, and quantify the percentage of infestation of MPBB in the selected study sites in the Upper Yampa Basin. Chapter 5 discusses how the SVM classifier algorithm was used to analyze another set of vegetation indices to provide similar results as those seen in Chapter 4. Chapter 6 discusses a statistical analysis of the two classifier algorithms to compare accuracy. Chapter 7 is the general summary of results.

1.4 Study Area

The Upper Yampa Basin was selected for this study based on its characteristics and location. Important characteristics include: water supply, watershed characteristics, presence of well-developed pine forests, incursion of insect infestation, data availability, variability of elevation and orientation, and location in Northern Colorado (Figure 1.1). The Upper Yampa River is located in the Upper Colorado River Basin and is defined by USDA hydrologic unit code 14050001 and outlet gauge USGS 09246550, which are located at Lat 40 29 50 and Long 107 30 34 (USGS National Water Information System). This area is characterized by steep, high mountain ranges and associated mountain valleys. Elevations range from 1,980- 4,390m. Temperature regimes are mostly frigid (average annual maximum and minimum temperatures are 9.8 and – 10.2 ˚C) respectively. The predominant moisture regime is ustic (semi-arid). Precipitation varies between less than 9 inches annually along the semiarid lower region to greater than 50 inches along the upper western regions where the majority of 17” annual precipitation occurs as snow (Repass, 2005).

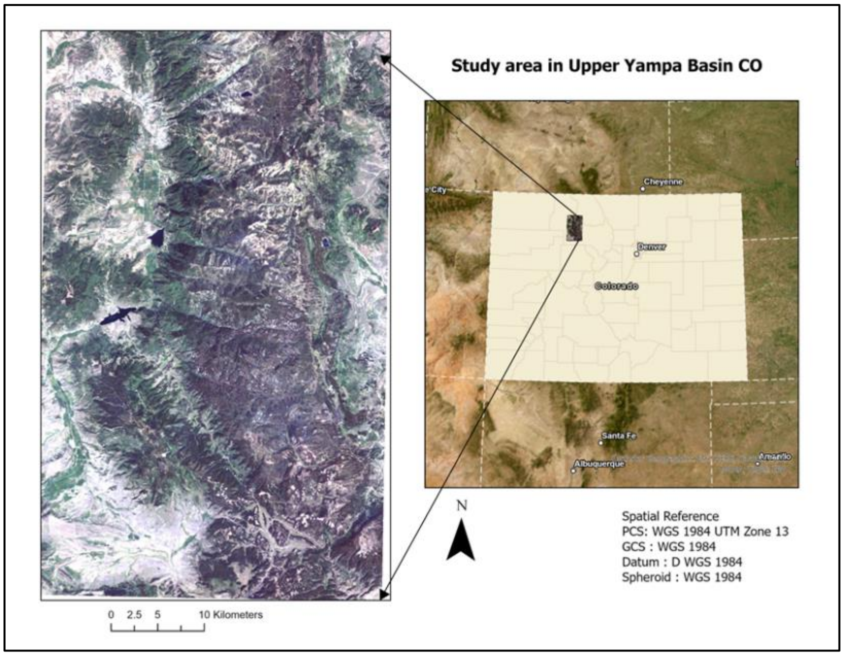


Figure 1.1 Study area

Vegetation characteristics are dominated by sagebrush-grass at lower elevations, and with increasing elevation, include coniferous forest to alpine tundra. Typical forest composition includes Aspen, Douglas fir, Engelmann Spruce, Lodgepole Pine, and various mesic mixes of mountain shrubs.

Weather characteristics across the basin vary as a function of elevation, topography, and orientation. In the lower elevations of the basin, summers are mild with one to three months having average temperatures greater than 10 ˚C. The coldest months average below -3 ˚C. Precipitation in lower elevation areas occurs uniformly throughout the year. As the elevation increases in the basin, summers become cooler and winters more severe. In the highest elevations of the basin, climate patterns are more complex and depend on elevation and exposure. The highest snowfall amounts in Colorado are found in the highest sections of the study area (USDA, 2023). These regions in the intermountain west are important hydrologically, even though they comprise only a small fraction of the area. The average snow accumulation in the Upper Yampa Basin in Colorado is approximately 14 inches during the winter. However, snowpack levels vary from year to year. In some cases, the snowpack on the mountain in this region can reach as high as 37 inches during seasonal peaks (https://snoflo.org/report/snow/colorado/yampa-5.1). Mountain regions with snow-dominated hydrology are a major source of water for runoff, groundwater recharge, and agricultural support in the US West.

). These subalpine forests lie at the highest forested elevations, about 2700-3600 m (Peet, 1981; Romme et al., 2009). Common tree species in the subalpine forest include Engelmann spruce (*Picea Engelmannii Parry ex Engelm*.), lodgepole pine (*Pinus contorta Dougl. ex Loud.*), quaking aspen (*Populus tremuloides Michx*.), subalpine fir (*Apies lasiocarpa (Hook.) Nutt.*), five-needle pines (i.e., white pine – *Pinus strobiformis Engelm*.), though forest composition may vary throughout the region. In subalpine forests, the most important tree-killing bark beetles are the mountain pine beetle (MPB; *Dendroctonus ponderosae Hopkins*), spruce beetle (SB*; Dendroctonus rufipennis (Kirby)*), and western balsam bark beetle (WBBB; *Dryocoetes confucius Swaine* (Table 1.1) (Andrus et al., 2021; Wilson et al., 2013). The primary hosts for MPB include lodgepole pine and five-needle pines. Engelmann spruce is typically attacked by SB, and subalpine fir is attacked by WBBB (Bentz et al., 2009). Other insects and pathogens may have contributed to a limited amount of tree mortality in the study area and were not included in this study.

Table 1.1 Bark beetles and their host species.

Insect agent name	**Tree host species**
Mountain Pine Beetle MPB (*Dendroctonus ponderosae*)	Lodgepole pine (pinus contorta) and five-needle pines (Pinus aristata, Pinus flexilis, Pinus strobiformis)
Spruce Beetle SB *(Dendroctonus rufipennis*)	Engelmann spruce (Picea engelmannii), blue spruce (*Picea pungens*), spruce species (Picea spp).
Western Balsam Bark Beetle WBBB (*Dryocoetes confusus, Armillaria spp., and others)*	Subalpine fir/ corkbark fir (Abies lasiocarpa)

1.5 Study Site and its relation to Mountain Pine Bark Beetle

Mountain pine bark beetles (MPBB Dendroctonus Ponderosa) are native to pine forest ecosystems in western North America. However, in the past 2.5 decades, MPBB has been growing to epidemic levels where the area infested far exceeds the extent previously documented in the past 125 years (Spruce et al., 2019). The host trees that MPBBs prefer to infest, attack, and kill are either stressed or older trees with larger diameters. The impacts of MPBB are especially obvious in Lodgepole Pine Forests in the Rocky Mountain region with a severe epidemic occurring in North-central Colorado (Meddens et al., 2013), which started around 2002 and increased to reach the peak period of tree mortality in 2005 (Liang et al., 2014; Meddens et al., 2012). The data covering this infestation event is plentiful and readily accessible from public sources making this a good area for the study. This study will investigate the time period from 2000 to 2010.

1.6 Field data collection

Field data collection: several trips (7/28/2018, 8/20/2020,8/25/2021, 8/02/2022) were made to collect visual data on locations of: Fire zones, clear-cut zones, healthy conifers, infested conifers, and non-conifer vegetation. GPS Garmin was used to mark locations.

Dominant tree species at higher elevation are Engelmann spruce, subalpine fir, and limber pine. At middle elevations (approximately 2500-3100 m), lodgepole pine is the dominant species along with other occasional species such as aspen. Ponderosa pine and Douglas fir occur mostly on drier sites at lower elevations (Meddens et al., 2013).

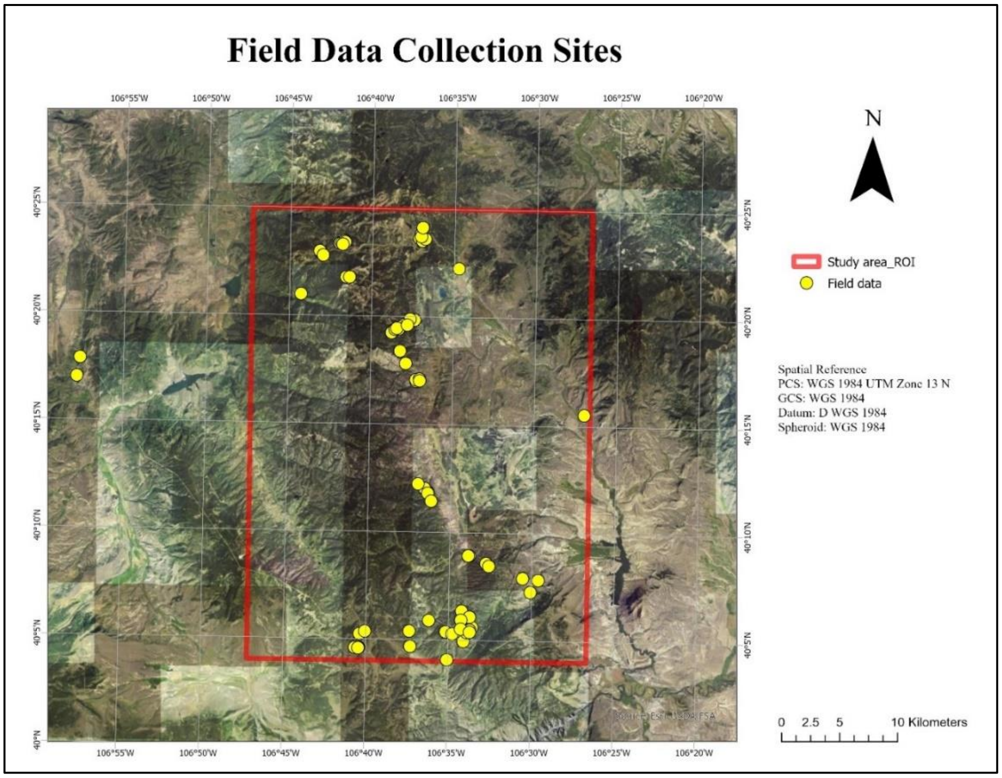


Figure 1.2 Field data collection sites

1.7 Vegetation indices related to MPBB infestations

Table 1.2 vegetation indices

Note: Tasseled-Cap transformations have coefficients which are only applicable to specific sensors. It should be noted that these coefficients are specific to Landsat 5 (TM) and ETM+.

Index	**Formula**	**Notes**	**reference**
Normalized Burn Ratio (NBR)	(NIR-SWIR2) / (NIR+SWIR2	Moisture stress & needle structure	(Meigs et al.,2011)
Normalized Difference Vegetation Index (NDVI)	(NIR-RED) / (NIR+RED)	Sensitive to conifer tree health	Tucker (1979)
Band5/ Band4 (B5/B4)	B5/B4	Sensitive to conifer tree health	Vogelmann (1990)
Tasseled-Cap Brightness (TCB)	0.2043 * B1 + 0.4158 * B2 + 0.5524 * B3 + 0.5741 * B4 + 0.3124 * B5 + 0.2303 * B7	Sensitive to surface brightness	Crist (1985)
Tasseled-Cap Greenness (TCG)	-0.1603 * B1 - 0.2819 - B2 * 0.4934 * B3 + 0.7940 * B4 - 0.0002 * B5 - 0.1446 * B7	Moisture stress	Crist (1985)
Tasseled-Cap Wetness (TCW)	0.0315 * B1 + 0.2021 * B2 + 0.3102 * B3 + 0.1594 * B4 - 0.6806 * B5 - 0.6109 * B7	Sensitive to vegetation greenness	Crist (1985)
Normalized Difference Moisture Index (NDMI)	(NIR-SWIR1) / (NIR+SWIR1)	Sensitive to canopy water content	Wilson and Sader (2002)
Band5/Band4 anomaly from undisturbed mean (Diff B5/B4)	Anomaly B5/B4 – B54 undisturbed	Differentiate predominantly live vs. dead	Bright et al., 2020

Normalized Burn Ratio (NBR) was used for time series analyses. NBR is particularly useful because sensitivity analyses have shown its reliability for capturing disturbance processes (Cohen et al., 2010; Kennedy et al., 2010), and it is well known by the fire science community (Eidenshink et al., 2007). Other subsets were used for Random Forest and Support Vector Machine analyses including Band 4 and 5 (Landsat TM), Normalized burn ratio (NBR), band5/band4 ratio (B54), Tasseled Cap Greenness (TCG), Mean band5/band4 from 2000-2009, anomaly band5/band4 (AnomB54), subtract Anomaly B54 (year) from Mean B54 (DiffB54), and DEM.

Chapter 2 - Background and literature review

2.1 Literature Review

2.1.1 Hydrology and MPBB

The hydrology of the western U.S. region, where the Rocky Mountains are located, is for the most part dominated by snowmelt and precipitation. This resource is affected by several processes: water is first stored as snow; it can be lost due to interception and evaporation, sublimation, or transpiration; it can be released as snow melt runoff; or it can be infiltrated to form groundwater. More modern controls on water throughput include factors such as tree mortality because of pine bark beetle infestation and climate change. In the last decade, infestations of bark beetle species such as the Mountain Pine Bark Beetle (MPBB) have grown from endemic to epidemic levels, killing large areas of montane and sub-alpine forests. Increased tree mortality rates due to MPBB infestation causes changes in canopy cover which have the potential to result in more rapid snow melt and dryer soil in the summer (Mikkelson, et al., 2011). Annual water yield and peak flows have been shown to increase following insect infestation. Decreases in forest canopy also result in decreased transpiration, and thus reduce soil water/groundwater uptake and potentially increase groundwater recharge (Mikkelson, et al., ibid.). Climate change also influences MPBB, increasing infestations as temperature rises.

2.1.2 Mountain Pine Bark Beetle in North America

In North America, bark beetle populations have increased in their geographic extent and range resulting in the largest-scale tree die-off ever recorded from an insect infestation in multiple wooded regions (Spruce et al., 2019). A variety of beetle species have caused widespread tree mortality, with each species focusing on their favored host tree (Hicke et al., 2016). One of the most common species is the Mountain Pine Bark Beetles (MPBB, Dendroctonus ponderosae), which is also one of the most destructive native bark beetles in North America (Hicke et al., 2020), and which attacks several Pinus species. As of 2009, MPBB has affected 3.6 million ha in the United States (USDA Forest Service, 2010). In western North

America, bark beetles have killed large areas of trees, the size of which varies from year to year and is likely increasing (Hicke et al., 2016; Meddens et al., 2012; Meigs et al., 2011). The tree-level damage caused by these beetles can be characterized by several infestation stages. Trees experience continuously increasing water stress (Foster et al., 2017) because larvae feed on the phloem tissue and introduce a fungus that disrupts movement of water and nutrients within the trunk (Safranyik & Carroll, 2006). Most bark beetles reproduce in the phloem or fluid-nutrient transport systems of the host trees. The need to kill the host to produce a successful brood (Christiansen et al., 1987) causes the beetles to use a "mass attack" strategy in which many beetles attack the same tree at the same time to overwhelm its defenses. After a successful attack, the bark beetle lays eggs and introduces fungi for the emerging larvae to live on trunk (Safranyik & Carroll, 2006). These fungi block the flow of the water and nutrients hastening the death of the trees. Bark beetle life cycles range from two weeks to three years depending on beetle species and temperature (Wood, 1982). For the MPBB, the life span is approximately 3 years. Tree die-off leads to changes in forest structure and land cover and influences other forest disturbances including wildfires, biogeochemical cycles (nitrogen, carbon, and water), and biogeophysical characteristics (leaf area, albedo) (Hart & Preston, 2020).

The tree-level damage caused by these beetles can be characterized by several visible infestation stages. The period immediately following MPBB attack, but before symptoms are readily visible in the tree crown, is called the "green stage" (Wulder et al., 2009). Green-stage trees experience increased water stress (Foster et al., 2017) because the larvae have begun to feed on the phloem tissue and to introduce the fungus. The "red stage" follows the green-stage as green chlorophyll pigment is significantly lost and foliage turns red. Host trees ultimately drop all needles after 3-5 years, transitioning to the "gray stage". As infested trees begin to die, the percentage of each visible color changes. The initial stage, called the "green stage" is characterized by yellowing of the needles. As the dying needles dry out, they turn red over the course of a year or more, reaching what is referred to as the "red stage" of tree mortality. The "grey stage" then occurs 3-5 years after the initial infestation when the needles drop off the trees and bare dead branches dominate. The red and grey stages provide pronounced color changes which are visible with remote sensing technology (Wulder et al., 2006).

2.1.3 Drought

Another factor that affects tree demographics is drought, which affects moisture availability. Where trees are less adaptive to drought, such as in warmer and drier sites, tree mortality increases rapidly during a drought period, whereas normally in wetter and cooler sites trees can withstand short drought periods more easily (Anderegg et al., 2019). Warming and drought also affect insect outbreak disturbances and have been linked to tree die-off rates in many places around the world, (e.g., Europe, Canada, Australia) and, under changing climate conditions predicted for the future, are expected to drive severe, extensive tree mortality events (Anderegg et al., 2015). Warm temperatures directly affect insect outbreaks through increasing developmental and survival rates of pupae that drive population growth success (Weed et al., 2013). Typically, MPBB attack weakened trees based on characteristics of host trees, such as density, abundance, age, size, physiology, and their distribution patterns across the landscape - factors which are known to impact insect population growth, and which in turn eventually increase tree mortality under drought conditions (Raffa et al., 2008).

Andrus et al., (2020) set up permanent study plots ranging across southwestern Colorado, specifically the San Juan Mountains, southern Rocky Mountains, in a subalpine conifer setting located across a range of site conditions, stand ages, and species compositions and tallied changes in the health of individually marked trees 9 times over 37 years. They found that tree mortality was primarily attributed to stress from unfavorable climate changes and nutrient/ light competition with lesser effects caused by bark beetle activity, wind damage and wildlife impacts. Tree mortality due to climate stress and bark beetle attacks tripled across old stands between the initial census in 1982 and the most recent censuses in 2019. The highest rate of tree mortality correlated with warmer maximum summer temperatures, and greater summer moisture deficits. Drought affected older stands of trees more than smaller younger trees. The prediction is that climate warming will continue to increase background mortality rate in subalpine forests. Combined with increases in bark beetle activity and the declining frequency of cool-moist years suitable for seedling establishment, an increasing rate of tree mortality is likely (Andrus et al., 2020).

Koll et al., (2019) showed that pine forests under water-stressed conditions are less able to use their normal defense systems against bark beetle attacks. Bark beetle attacks and biota carried by beetles typically induce a general increase in concentration of phloem-based mono- and sesqui-terpenes, whereas water stress did not. Bark beetle attacks also induced an increase in resin flow for unstressed trees but not for water stressed. Mortality was highest for beetle-attacked and water-stressed trees (Kolb et al., 2019).

Stephenson et al., (2019) illustrated in a set of figures that for ponderosas, larger pine trees are more seriously affected by drought and showed consistently higher mortality rates under drought conditions. They also showed these patterns are not the same in other trees species. These results were based on statistical measurement of growth rate and tree diameter vs. tree mortality. The authors also discuss how current forest management programs are aimed at reducing tree density and leaving clumps of trees, which are more like historical conditions. However, these residual clumps may make the forest more susceptible to mountain pine beetle attacks. Land managers need to be aware of how MPBB will respond to their restoration treatments, to prevent and mitigate tree mortality due to MPBB attacks which could negate restoration effects. Widely spaced clumps of trees may be beneficial for limiting the spread of fires, but they make infestations by MPBB easier (Stephenson et al., 2019).

2.1.4 Climate Change

Currently one of the main factors affecting mountain pine bark beetle die-off is climate change. Recent climate conditions have facilitated insect outbreaks by increasing tree vulnerability and altering insect dynamics (Bright et al., 2020). Climate affects outbreaks of MPBB in three ways: 1- life cycle changes of beetles as a function of temperature; 2- cold-induced mortality of beetles; 3- drought stress on host trees. Temperature directly controls the life stage development rates of mountain pine bark beetles, i.e., the progression from egg to larvae to pupae to adult. In colder situations it can sometimes take two to three years to complete the life cycle of an MPBB (Creeden et al., 2014). The direct effects of climate conditions (e.g., extreme temperatures, drought) often lead to tree stress which might drive trees beyond critical physiological thresholds for survival therefore increasing tree susceptibility to MPBB attacks (Adams et al., 2017; Andrus et al., 2021). Insect infestation plays a key role in ecosystem dynamics by controlling succession in renewing old and unhealthy forests (Senf et al., 2015). Although there is some debate about this, topographic factors including aspect, slopes, and elevation may also have major effects on local variability in temperature and moisture (Oke, 2002), which in turn affect a tree's ability to muster its defenses. Nelson et al., 2007, reported that warmer and drier slopes on southerly aspects at low elevation might decrease tree defenses against beetle attack while colder and wetter slopes on northerly aspect at high elevation increase tree defenses.

Expected increases in drought frequency and intensity associated with climate change suggest bark beetles will have increasing opportunities and success in attacking water-stressed conifers in future (Fettig et al., 2013; Andrus et al., 2021; Huang et al., 2020). Bark beetle populations can transition from endemic to epidemic during drought times resulting in landscape-scale tree mortality (Raffa et al., 2008).

Numerous climatological variables associated with drought conditions have been linked to bark beetle outbreaks, including vapor pressure deficit (VPD), climate water deficit (CWD), high previous-year Summer and Fall temperatures, and various multi-decadal ocean current pattern oscillations, such as the pacific decadal oscillation (Cooper et al., 2018), which causes the La Niña/ El Niño weather cycles in the western US. All these climate conditions decrease the availability of water to the host tree, limiting both its ability to grow and its ability to defend itself from attack beetle (Kane & Kolb, 2010). Differences in growth between beetle-killed and surviving trees may suggest a differences in allocation of resources, i.e., a difference in amount of carbon allocated for growth versus defensive compounds (Cooper et al., 2018; Herms & Mattson, 1992), when drought limits photosynthetic activity (Cooper et al., 2018).

Climate change indirectly influences MPBB growth and spread because of its impact on tree physiology and growth. Severe drought condition led to breakdown of hydraulic conductance in the tree xylem due to air blockage and loss of nutritional carbohydrates resulting in tree death (Adams et al., 2017). Lack of precipitation also severely limits development of chemical defenses. Decreases in tree growth occur during hot and dry conditions as carbohydrate reserves are depleted due to closure of stoma in the pine needles (McDowell et al., 2007). Water stress reduces net photosynthetic rate, lowers resin flow, especially when occurring at the same time bark beetle attack, and alters various terpene concentrations which represent tree defenses (Kolb et al., 2019). As a result, arid conditions typically serve as a trigger for MPBB outbreaks and encourage spreading of beetles to new locations until all host trees are exhausted or cool weather events kill-off MPBB (Safranyik & Linton, 1991). Various fungi commonly related to MPBB infestations are also affected by changes in temperature and soil moisture condition (Ojeda Alayon et al., 2017). These fungi are temperature sensitive and typically show peak growth between 25 °C – 30 °C (Ojeda et al., 2017) two of the most common fungi can survive extreme cold temperatures whereas one does not (Rice et al., 2008). Water stress decreases lesion linkage in lodgepole pines (Arango-Velez et al., 2016). Water deficit can also alter the concentration of individual monoterpenes but not the total monoterpene concentration. Some of these changes in composition of volatile emissions may facilitate increased colonization by MPBB (Lusebrink et al., 2013, 2016).

Expansion by MPBB into regions that had not been previously infested because of the effects of climate change present a clear danger to pine ecosystems in North America. Indigenous previously untouched lodgepole forests are more susceptible to attack and are more suitable for MPBB reproduction than lodgepole pines from regions which have experienced outbreaks on consistent bases because the former possesses insufficiently developed defenses to resist MPBB attacks owing to the lack of evolutionary history (Clark et al., 2010). Range shifts of host pines in response to climate change make risk assessments even more complex. Geographic displacement of forest ecosystems and over all range contractions of key pine species are projected to occur by the end of this century (Fettig et al., 2013). Multiple factors will control the base and direction of such range changes, including factors such as a tree’s adaptive and migratory potential, sensitivity of trees of different ages (e.g., saplings vs. older trees) to climate changes especially at range edges, and locational ecosystem characteristics (Monahan & Fisichelli, 2014). In addition, insects will show faster adaptive responses to climate change relative to host trees due to the insect's shorter life cycles. Impacts due to other causes of tree mortality such as disease and wildfire in pine ecosystems create further uncertainty in predicting MPBB distribution under climate change (Sambaraju et al., 2012).

2.1.5 Wildfires

Several factors such as increasing drought, innate tree conditions (e.g., stand density, age, vigor, etc.), and land cover connectedness control how susceptible forests are to both insects and wildfire disturbances (Hart & Preston, 2020; Meigs et al., 2011). MPBB infestation that causes tree mortality is expected to change fire activity by altering flammability and structure of fuels (Jenkins et al., 2014). As discussed above, initially following MPBB attack, tree leaves enter the "red stage", which refers to a change in needle color from green to reddish brown. This change reflects declines in leaf moisture content and alterations in chemical components that raise flammability (Jolly et al., 2012). Several studies have developed predictive models of fire behavior that suggest that these changes might increase and enhance crown fire (Hoffman et al., 2015). Approximately three years after MPBB attack, infected trees drop their needles, and trees have entered the gray "stage", where the needles are now completely dead. Additional fuels build up on forest ground due to falling needles and branches and affect understory vegetation (Hart and Preston, 2020). These changes in forest characteristics following MPBB infestation lead to increase fire susceptibility, however fire behavior modeling has shown that in addition to these changes in fuel characteristics, weather conditions, temporal and spatial patterns of tree die-off, and surface fire-line intensity speed are all important factors in fire behavior in MPBB affected trees (Hoffman et al., 2015). Wayman & Safford (2021) collected data on 180 plots within two recent fire areas in CA, US. Their analyses show that pre-fire tree mortality was influential on all measures of wildfire severity, including basal area killed by fire, Rd NBR, and canopy torch, although it was less influential than fire weather, i.e., relative humidity. This was especially true in forests dominated by pines and less true in fir/cedar dominated forest. This study is particularly helpful for forest managers where they are prioritizing fuels reduction treatments in areas of high tree mortality (Wayman & Safford, 2021).

Jenkins et al. (2008) discuss the common assumption that bark beetle-caused tree mortality in conifer forest increases fire hazard and alters potential fire behavior by altering the quantity and quality of forest fuels. Early in an epidemic there is a net increase in the amount of highly inflammatory surface fuels as compared to those found with older endemic stands. In fact, in post-epidemic stands large, dead, woody fuels and light surface fuels dominate. The authors concluded that rates of spread, and fire-line intensities were higher in red-stage epidemic stands than in late gray-stage endemic stands (Jenkins et al., 2008).

2.2 Mapping Mountain Pine Beetle Infestation

In the past, US-Forest Service aerial detection surveys (ADS) have been the standard method for tracking and mapping infestations since 1947 (Meigs et al., 2015; Williams & Birdsey, 2003). Trained interpreters fly over forest in specialized camera-equipped airplanes on an annual basis to locate infested trees. Polygons are digitized around corresponding areas on a map. While these surveys have provided much needed information, they have had several shortcomings. Interpretation is arbitrary and based on skill of the interpreter and the scale of absolute lines drawn to separate disturbed and undisturbed forest, when in reality outbreak patterns are very heterogenous, in both the presence/absence and degree of disturbance (Meddens et al.,2011). Aerial surveying is also costly in terms of time and resources and aviation work is hazardous. A 2012 study found that US ADS underestimate infestation extent by a factor ranging from 3-20 (Coops et al., 2006). ADS is valued because of its widespread data availability, both in terms of time and aerial extent, but the datasets have many inconsistencies where exclusions make them difficult to analyze, and risks and costs make them difficult to maintain (Bode et al., 2018).

Satellite imaging offers many advantages over ADS for monitoring forest disturbances. For example, Landsat imagery has emerged as the most widely used dataset for these purposes because of its extended temporal availability and relatively fine spatial grain (30 m spatial resolutions, Senf et al., 2017). Because of this fine spatial resolution, Landsat information can be used at a scale that is informative for ecological research and management (Wulder et al., 2006). Studies (e.g., Vogelmann et al., 2012) using Landsat data for insect disturbance mapping in coniferous forests typically use spectral information from one or two images.

Meddens et al. (2013) evaluated the relative ability of single-date versus multidate Landsat imagery to estimate the percentage of the pixel that contained trees in red stage; trees in gray stage were excluded. They used 4-band National Agriculture Imagery Program (NAIP) imagery, which has 0.3 m resolution, to develop aggregated 30 m "superpixels" as reference data which can be correlated with Landsat data. Success was related to homogeneity in the reference data, i.e., pixels with high percentages of red-stage trees had high classification accuracies, however pixels with small percentages were not classified with equally high accuracy. This study represented an important step in analysis using Landsat data, but Meddens, et al. (2013), were unable to detect pixels with fewer than 25% trees in the red stage with acceptable accuracy.

One of the most widely used vegetation indices for examination of progressive forest defoliation is the Normalized Difference Vegetation Index (NDVI) (Silva et al., 2013). However, some studies showed that other vegetation indices surpassed the accuracy of the NDVI. Other studies suggested that SWIR-based vegetation indices were more suitable for defoliation monitoring than NDVI (Townsend et al., 2012). The red-edge-based vegetation indices were also found to be very useful in defoliation assessments (Coops et al., 2006). Ye et al. (2021) completed a rigorous statistical analysis for parameter optimization to determine the most critical indices for examining MPBB infestations and determined that the suite of vegetation indices include: Disturbance Index (DI), Red-Green Index (RGI), Short-Wave Infrared1 (SWIR1), and Normalized Difference Moisture Index (NDMI) (Ye et al., 2021). This suite of vegetation indices showed the highest performance evaluation numbers of all analyzed suites. But it was stated by Rullan-Silva, et al. (2013) that performance of vegetation indices for monitoring forest defoliation may vary from site to site. Thus, it is recommended to explore different vegetation indices in each particular case. These vegetation indices will be discussed in more detail in Chapter 3.

2.2.1 Time series

Remote sensing has had an important role in characterizing forest disturbances, and systematically collected, global remote sensing datasets are key to emerging forest monitoring programs. Over the past 15 years there have been significant improvements in disturbance mapping algorithms that use Landsat data, taking advantage of high-quality data made available beginning in 2008 (Wulder et al., 2012). These algorithms perform high temporal density analyses that have the potential to yield temporal and spatially detailed disturbance maps. These algorithms differ in their specificity with respect to factors such as detecting abrupt or gradual disturbance trends, having different sensitivities to disturbance magnitudes based on thresholding, and limiting analyses to specific land cover types (Cohen et al., 2017).

Landsat satellite data is most useful when analyzed using time series-based approaches which can quantify diverse patterns and dynamics of changes in forests over time. Spectral trajectory-based methods are designed to detect linear or complex changes in spectral values over time and assess the degree of variation from that trajectory to define disturbance events. Early studies compared data from several years in sequence to examine multi-temporal trends over time (Cohen et al., 2002). Unfortunately change analyses based on two-data change detection methods do not fully show the interrelationships among many multitemporal images and may not be able to separate from background noise the subtle or long-duration changes that are likely to occur as result of climate change (Hicke et al., 2006). Recognizing this limitation, newer methods have been developed that simultaneously examine the signal of multiple images to improve the signal-to-noise ratio. One of the most successful approaches has been to use time-series fitting algorithms to separate longer-duration signals from-year-to-year noise caused by phenology, sun angle, sensor drift, atmospheric effects, and geometric misregistration (Vogelmann et al., 2010; Vogelmann et al., 2009).

The opening of the USGS Landsat archive greatly facilitated the capacity to generate time series analyses of Landsat data (Wulder et al., 2012). In particular, Landsat image time series (LTS), which combine data from successive time steps, are increasingly used to track the effects of MPBB infestation (Bright et al., 2020), by illustrating gradual changes in the spectral signal (Meigs et al., 2011) that reflect decreasing tree vigor, top-kill, and increasing mortality. Thus, many recent studies suggest that Landsat time series can be used to characterize the complex spatial and temporal dynamics of inset outbreak (Assal et al., 2014; Meddens & Hicke, 2014; Meigs et al., 2011).

Google Earth Engine (GEE) is a cloud-based geospatial processing platform which offers a large set of user-friendly application programming interfaces (API) that can be used to analyze freely available satellite images to produce statistics, maps, and graphical representation of investigated phenomena through parallel time analyses (Cohen et al., 2017; Sidhu et al., 2018). One of the most useful subroutine algorithms in GEE is LandTrendr (Kennedy et al., 2010; Kennedy et al., 2018). Early time series algorithms used idealized temporal models of land cover change, but LandTrendr uses a new strategy of arbitrary temporal segmentation, using straight line segments to model the important features of trajectory changes to eliminate noise and define disturbance events. The data determine the shape of the change trajectory. The result of temporal segmentation is a simplified model of the spectral trajectory, where the starting and ending points of segments are vertices whose time position and spectral value provide the essential information needed to produce maps of changes (Kennedy et al., 2014; Kennedy et al., 2010).

In subsequent studies Meddens and Hicke (2014) used Landsat time series to develop a continuous measure of red-stage tree mortality. They used information including Tasseled Cap Components (TCC), spectral bands (Franklin et al., 2003), spectral mixture analyses (Radeloff et al., 1999), and vegetation indices based on Near-Infrared (NIR) and shortwave-infrared (SWIR) reflectance (Franklin et al., 2003; Meigs et al., 2011). Since approaches based on single year or binary (comparing two years) maps are deficient in characterizing the complex ecological dynamics of insect outbreaks, a more comprehensive mapping approach was developed utilizing as many points in time as possible to characterize both the disturbance magnitude and duration (Goodwin et al., 2008). This process will be discussed in more detail in Chapter 3.

2.3 Data classification systems

Most of the scientific work at regional scales concerning MPBB infestations relies on mapping efforts. The need to locate and define very large areas of infestation has led to the development of techniques based on multispectral imaging. Spectral characteristics of foliage from healthy trees and from trees in various stages of MPBB attack have been shown to be sufficiently different in several different wavelength bands to be useful for discriminating of classes with similar degrees of infestation (Ahern, 1988). The most important differences involve spectral wetness indices exhibited between the green-, red-, or gray-stage trees (Cheng et al., 2010; Goodwin et al., 2008; Jewett et al., 2011). These studies have been most successful in identifying the presence or absence of each stage of attack, although spatial resolution of the sensor is typically a limiting factor. Homogeneous regions are easily classified and mapped, however, mapping heterogeneous regions with a mix of healthy and infested trees has been scale dependent (Bentz & Endreson, 2004; Franklin et al., 2003). The pixel size definition of moderate-resolution satellite imagery, such Landsat data (30m x 30m), is not precise enough, because a single pixel can include several whole and partial trees, some healthy and some not. The ability to quantify the percentage of tree mortality within a Landsat-sized pixel is important because it allows better understanding of the early stages of outbreak, when the percentage of tree mortality is low (Long & Lawrence, 2016).

The development of an MPBB infestation is composed of a series of thresholds that advance cyclically through four crucial stages of outbreak (Raffa et al., 2008). The most common phase is the “endemic phase” in which MPBB infestations occur across the landscape at such low densities that they are difficult to detect. The next stage, the stand-level incipient outbreak phase, is characterized by clusters of attacked stands found at both short and long distances from the original infested trees. This phase typically occurs when stands are compromised by other tree ailments or environmental conditions which affect tree resistance (Raffa & Berryman, 1987; Safranyik & Carroll, 2006). The infested stands eventually overlap and develop into a traveling wave in a landscape-level outbreak phase (Lundquist & Reich, 2014). This phase typically develops when weather and climate, landscape, and structure, and degree of MPBB dispersal further compromise the trees' resistance, thus increasing MPBB proliferation. The infestation returns to endemic levels in the outbreak collapse phase, when available hosts become scarce or environmental factors change limiting further population growth (Bode et al., 2018).

A variety of Landsat-defined Forest disturbance tracking packages have been developed over the years including vegetation change tracker, exponentially weighted moving average change detection, multi-index integrated change analysis, and continuous change detection and classification (Cohen et al., 2017). In addition, regression analysis can be used to identify gradual trends (Vogelman et al., 2012). However, only Landsat-based detection of trends can be used to monitor both abrupt change and gradual trends (Kennedy et al., 2010). In order to analyze the amount of data made available by Landsat programs, a variety of classification methods have been developed (e.g., Vegetation Change Tracker (VCT) (Huang et al., 2010), Landsat-based detection of Trends in Disturbance and Recovery (LandTrendr) (Kennedy et al., 2010), Breaks for Additive Seasonal and Trend (BFAST) (Verbesselt et al., 2010), and Continuous Change Detection and Classification (CCDC) (Zhu & Woodcock, 2014) using a variety of mapping techniques (e.g., ArcGIS pro and GEE) based on a variety of measured and calculated spectral indices (see Chapter 3, Table 3.2). A question to be considered is how to improve detection of early beetle-infection signals from Landsat time series compared with other algorithms and data sets.

Classification methods range from unsupervised algorithms such as maximum likelihood, to machine learning algorithms such as Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector Machine (SVM), and ensemble classification techniques (Breiman, 2001). Machine learning algorithms have proved to be more accurate and efficient than conventional parametric algorithms when faced with large complex data sets and have been used for large area mapping (Hansen et al., 1996; Rogan et al., 2003). These algorithms are efficient and effective because they do not rely on any assumptions about data distribution. These ensemble learning algorithms use the same base classifier to produce multiple classifications of the same data (Breiman, 2001), or use a combination of different base classifiers to generate multiple classifications of the same data or to target different subsets of the data (Mountrakis et al., 2009). The collection of multiple classifiers of the same data are combined using a variety of rule-based approaches, such as maximum voting, product, sum, or Bayesian rule.

2.3.1 Random Forest (RF)

Another algorithm called the Random Forest (RF) technique has been developed to use multi-spectral and hyperspectral satellite imagery (Ghimire et al., 2010; Lawrence et al., 2006), and lidar and radar data (Guo et al., 2011; Latifi et al., 2010). Ensemble learning techniques such as RF have been developed that provide higher accuracy than other machine learning algorithms because it produces a group of classifiers that performs more accurately than single classifiers (Ghimire et al., 2010; Pal, 2005).

RF is based on decision tree classifiers, and RF develops many different classification trees. To classify a new feature, the input vector is classified with each of the trees in the forest. Each tree gives a classification defining that input vector, and this process is called the tree “voting”. Then, the classifier algorithm chooses the decision tree having the most “votes” over all the trees in the forest. The Random Forest RF technique was chosen because it has been demonstrated to be effective in accurately mapping landcover across complex and heterogeneous landscapes, and because it is relatively insensitive to noise, making it suitable for use in complex and dynamically changing forest environments (Rodriguez-Galiano et al., 2012). This method will be discussed in Chapter 4.

2.3.2 Support Vector Machine (SVM)

The Support Vector Machine (SVM) classifier was initially devised within the framework of classification models by Cortes and Vapnik (1995). Subsequently, it was expanded to encompass regression by Drucker, Burges, Kaufman, Smola, and Vapnik (1997), as well as by Smola and Schölkopf (2004). The primary objective of the Support Vector Machine (SVM) classification algorithm is to identify a hyperplane that effectively partitions the dataset into a predetermined and distinct number of classes, utilizing a set of training samples. Note that this hyperplane is calculated in feature space, not image space. The utilization of the optimal separation hyperplane aims to decrease misclassifications, which are acquired during the training phase (Mountrakis et al., 2011). The fundamental concept of Support Vector Machines (SVM) as regards regression modeling is that observations whose residuals fall within a user-defined threshold are not considered in the regression fit. Conversely, data with residuals exceeding the threshold contribute to the regression fit in a linearly proportional manner. Therefore, due to the exclusion of squared residuals, the impact of significant outliers on the regression equation is restricted, while samples with minimal residuals do not influence the regression equation (Kuhn et al., 2013). This study utilized the Support Vector Machine (SVM) with a radial kernel, and various values of tuning parameters were examined: Epsilon, Gamma, and Cost.

Migas-Mzur et al. (2021) used the SVM algorithm (Cortes & Vapnik, 1995) to map bark beetle outbreaks in Tatra Mountains of Poland. (Karan & Samadder, 2016), used SVM-based image classification technique to assess the change in land use pattern due to mining in India. Zhao et al. (2015) tested the SVM algorithm in separating forest disturbance types including wildfires, harvest, and other disturbances such as insect outbreaks, tornados, snow damage and drought induced mortality in the Greater Yellowstone Ecosystem. This method will be discussed in more detail in Chapter 5.

2.4 Comparison of RF vs SVM

RF is comprised of an ensemble of a predetermined number (n) of decision trees, which are used to find the final outcome. Each tree is independently determined using a bootstrap sample of the data. Each model in the ensemble is used to generate a prediction for a new sample and these predictions are averaged to find the forest's overall prediction. A voting mechanism permits the algorithm to choose the class that is most prevalent or popular among the available options. In the context of regression analysis, the responses of trees are combined by averaging to derive an estimation of the dependent variable. The RF classifier randomly selects a defined number of predictors (mtry) at each split in the decision trees to reduce correlation effects among trees (Breiman, 2001; Hultquist et al., 2014). The number of trees created for the forest in this study was set to 600 for both regression and classification models. The mtry parameter was tuned from the following values: 2 to 32 (see Chapter 4).

The Support Vector Machine (SVM) classifier was initially devised within the framework of classification models by Cortes and Vapnik (1995). Subsequently, it was expanded to encompass regression by Drucker, Burges, Kaufman, Smola, and Vapnik (1997), as well as by Smola and Schölkopf (2004). The primary objective of the Support Vector Machine (SVM) classification algorithm is to identify a hyperplane that effectively partitions the dataset into a predetermined and distinct number of classes, utilizing a set of training samples. The utilization of the optimal separation hyperplane aims to decrease misclassifications, which are acquired during the training phase (Mountrakis et al., 2011). The fundamental concept of Support Vector Machines (SVM) as regards regression modeling is that observations whose residuals fall within a user-defined threshold are not considered in the regression fit. Conversely, data with residuals exceeding the threshold contribute to the regression fit in a linearly proportional manner. Therefore, due to the exclusion of squared residuals, the impact of significant outliers on the regression equation is restricted, while samples with minimal residuals do not influence the regression equation (Kuhn et al., 2013). This study utilized the Support Vector Machine (SVM) with a radial kernel, and various values of tuning parameters were examined: Epsilon, Gamma, and Cost (see Chapter 5).

Chapter 3 - Time series analysis based on LandTrendr using Normalized Burn Ratio (NBR) to track forest disturbances

3.1 Introduction

The purpose of this chapter is to discuss the possibility of improving generalized mapping of insect-related effects in forests by integrating Landsat time series change detection data, annual forest health detection surveys, and ground-based measurements to investigate the cumulative effects of insect impacts on tree die-off in conifer forests. Specific objectives are: 1. Develop and characterize spectral trajectories in Landsat time series and differentiate those associated with MPBB disturbances of varying duration and severity from other types of disturbances such as fires or clear cutting; 2. Correlate these spectral trajectories to ground-based measurements of insect-caused tree mortality to assess biophysical drivers of the observed spectral changes.

3.2 Study Area

This study examines the effects of tree-killing bark beetles in subalpine forests of Colorado in the Upper Yampa Basin (Figure 3.1). This is a mountainous region with a complex topography (3299m - 2291m) covering 696.92 km2. This area has a temperate continental climate with warm summers (ranging from 12.5-32.9 C˚, mean July maximum 24.1 C˚), cold winters (ranging from -21.3- -5.4 C˚, January mean minimum -12.3C˚), and bimodal distribution of precipitation with peaks in March-May and July-August. Precipitation ranges from 214 to 1875 mm with an average annual precipitation of 625 mm (PRISM Climate Group, Oregon State University, 2018).

3.3 Methodology

Three broad strategies define the LandTrendr approach. First, changes in cover condition must be tracked in a manner that is durable across years, not just within a single year. Cloud cover can cause data collection gaps in the16-day satellite repeat cycle, and, as a result, the annual time step has proved to be the best suited temporal scale of most of the Landsat archive data. Year-to-year variability is considered noise. Second, LandTrendr uses a purely pixel-based structure for analysis. Multiple images per year are fed to the algorithm, but each pixel's temporal trajectory is constructed from the series of single best values for each pixel for each year. This avoids errors caused by clouds, cloud shadows, and Landsat 7 scan-line corrector failure. Lastly, this algorithm allows both for the temporal smoothing of spectral noise in long duration signals and for the unsmoothed capture of spectral changes for abrupt events like other deviation-seeking strategies (Huang et al., 2010; Senf et al., 2015; Zhu, 2017). By allowing arbitrary segmentation, there is an increased risk of over- and under-fitting, therefore, the algorithm requires a series of control parameters (training data) and filtering steps designed to reduce overfitting while still capturing the desired features of the trajectory (Huang et al., 2010; Senf et al., 2015).

In the western US, the most readily available maps to assess landscape and regional insect dynamics are forest health Aerial Detection Surveys (ADS), which have been collected annually since 1947 (Meigs et al., 2015; Williams & Birdsey, 2003). Although valuable for coarse-scale assessments, ADS data show critical uncertainties related to variability of methodology, personnel, and observation conditions (Meddens & Hicke, 2014; Meigs et al., 2011). More importantly, ADS data delineates insect effects within large polygons that also contain live trees creating a blurred view of insect activity. As a result, insect activity may appear to be widespread and homogeneous across entire regions, whereas actual insect influence on tree mortality is highly variable at finer scales. ADS data tends to overestimate the extent of insect impacts at a regional scale while strongly underestimating actual impacts at the forest stand scale (Meigs et al., 2015). In addition, the ADS emphasis on detecting insect activity each year limits the potential to quantify lasting vegetation changes that occur over multiple years. Other geospatial and field datasets are necessary to permit accurate regional analysis and address forest management concerns (Meddens & Hicke, 2014; Meigs et al., 2015).

3.3.1 Vegetation indices derived from Landsat

Landsat satellite archive represents an ideal complementary dataset. Due to its moderate spatial resolution (30-m grain), broad temporal scope (1984-present), and consistent, seamless coverage, the Landsat TM sensor captures a variety of disturbance processes at forest stand, landscape, and regional scales (Cohen & Goward, 2004; Kennedy et al., 2014). Like the ADS data, Landsat-based estimates have limitations. Specifically, satellite data lacks any inherent link to specific disturbance agents, and the raw reflectance data have limited biological meaning without being linked to ground-based observations. To improve generalized mapping of insect-related effects in forest, Landsat time series are integrated with annual forest health Aerial Detection Surveys (ADS), and ground-based measurements to calibrate and verify disturbance changes with time identified by LandTrendr (Cohen et al., 2017; Meigs et al., 2011). Use of the LandTrendr temporal segmentation algorithm to develop a time series of normalized burn ratio (NBR) values at each pixel simplifies an often-noisy yearly time series into a simplified series of segments which capture the most important features of the trends while avoiding most false changes (Kennedy et al., 2010; Meigs et al., 2011). For example, disturbance segments can be defined as those segments experiencing a decline in a spectral index such as NBR over time. To minimize errors due to background variation and statistical noise, only those segments would be accepted whose change in NBR exceeds a given threshold value. Disturbance trajectories were classified according to whether they are associated with a long-duration decline (≥6 years) or short duration (<6 years) (Meigs et al., 2011).

When attempting to detect low magnitude disturbances with Landsat data, the signal associated with spectral change due to disturbance may be masked by noise associated with normal variations in atmospheric and geometric conditions. Kennedy et al (2010) found that different spectral indices had varying abilities for accurate detection of subtle disturbance signals. A variety of indices have been developed based on a combination of values from various Landsat spectral bands, each of which provides a different kind of information about changes in vegetation (Table 3.1) (Ye et al., 2021). Kennedy (2010) found that the Normalized Difference Vegetation Index (NDVI) performed less effectively than the NBR and Tassel Cap Wetness (TCW) indices. The spectral index referred to as the NBR (Coops et al., 2020; Key & Benson, 2006) contrasts Landsat bands 4 and 7 (Near Infrared (NIR) and shortwave infrared (SWIR):

$$\text{NBR} = (\text{NIR}-\text{SWIR})/(\text{NIR}+\text{SWIR}). \qquad \text{equation 1}$$

NBR is particularly useful because sensitivity analyses have shown its reliability for capturing disturbance processes (Cohen et al., 2010; Kennedy et al., 2010), and it is well known by the fire science community (Eidenshink et al., 2007) (Figure 3.1).

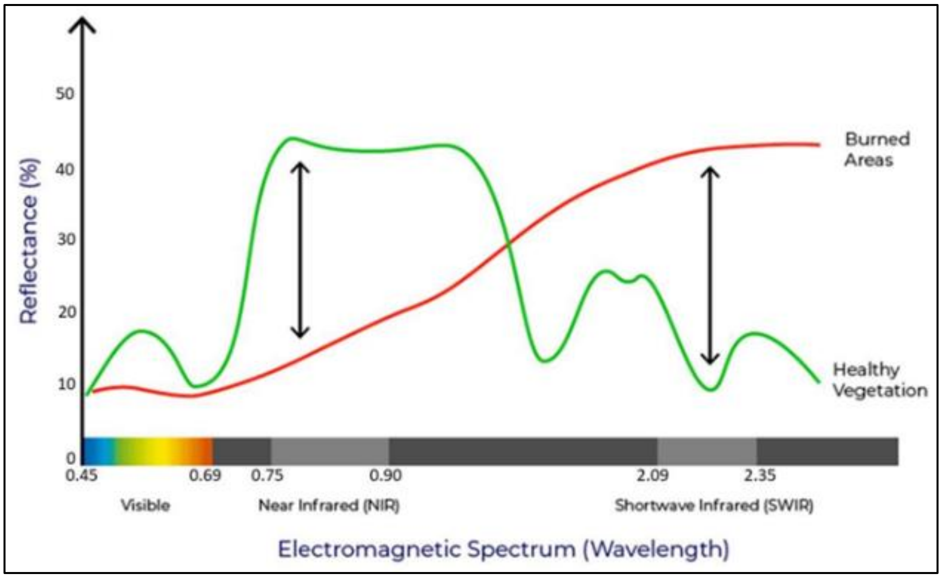


Figure 3.1 Comparison of the spectral response of healthy vegetation and burned areas (Source: USDA Forest Service).

Healthy vegetation shows a very high reflectance in the NIR, and low reflectance in the SWIR portion of the spectrum, so NBR is high in a healthy forest. As infestation progresses NBR will decrease. Data from NIR and SWIR bands were used to calculate NBR.

Table 3.1 Vegetation indices (VIs)

Vegetation indices (VIs) calculated from Landsat bands including red, green, and blue bands plus near infrared (NIR) and short-wave infrared (SWIR).

Index	Formula	Indicated change	Expected beetle	Stage of interest	Reference
Red-green Index (RGI)	RGI= RED/GREEN	correlation	MPB	Red stage	Coops et al. (2006); Hicke and Logan (2009); Meddens et al. (2013);
Normalized Difference Moisture Index (NDMI)	NDMI = NIR − SWIR1/ NIR + SWIR2	Moisture stress	MPB	Red stage	Goodwin et al. (2008); Meddens et al. (2013)
Normalized Burn Ratio (NBR)	NBR = NIR − SWIR2 /NIR + SWIR2	Moisture stress & needle structure	MPB	Red stage	Meigs et al. (2011); Liang et al. (2014); Senf et al. (2015);
Tasseled Cap Transformation Greenness (TCG)	TCG = − 0.2848 * BLUE − 0.2435 * GREEN − 0.5436 * RED + 0.7243 * NIR + 0.0840 * SWIR1 − 0.1800 * SWIR2 (Landsat 5/7) OR	Needle structure	MPB	Red stage	Meddens et al. (2013); Hais et al. (2016)
Tasseled Cap Transformation Wetness (TCW)	TCW = 0.1509 * BLUE + 0.1793 * GREEN + 0.3299 * RED + 0.3406 * NIR – 0.7112 * SWIR1 – 0.4572 * SWIR2 (Landsat 5/7)	Moisture stress	MPB	Red stage	Wulder et al. (2006); Coops et al. (2009); Skakun et al. (2003); Meddens et al. (2013);
Disturbance Index (DI)	DI = TCB – TCW – TWG	Moisture stress & needle structure	SB	Gray Stage	Hart and Veblen (2015); DeRose et al. (2011)
(TCB: Tasseled Cap Transformation Brightness)					
Normalized Difference Vegetation Index (NDVI)	NDVI = NIR − RED /NIR + RED	Needle structure	SB	Red and Gray Stage	Meddens et al. (2011); Meddens et al. (2013); Assal et al. (2014); Coops et al. (2006); Hart and Veblen (2015);

3.3.2 Data acquisition and processing

Landsat surface reflectance cloud-masked composite collections are commonly used for generating baseline land cover maps. For this study 10 georectified multispectral annual Landsat TM/ETM+ images with < 10% cloud cover from June 1-August 30 2000-2010 were acquired (Figure 3.2, step 1) from USGS archive and processed through GEE. Multiple images representing Level-2 (USGS Landsat Science Products) Surface Reflectance data were used in years when clouds obscured part of the study area. All these images were preprocessed by GEE using the latest generation orthorectification, radiometric calibration, and atmospheric correction protocols, including cloud and cloud shadow masking using the CFmask pixel_qa band protocol (LaSRC) (Claverie et al., 2015; Roy et al., 2016; Zhu et al., 2015) (Figure 3.2, steps 2 and 3). Annual image collections from L4, L5, and L7 satellites were used for LandTrendr spectral-temporal segmentation and for generating GEE-LandTrendr-based indices (Figure

3.2, step 4). Median value composites of each Landsat multispectral band were calculated for each summer season, June 1st to August 30 (2000 to 2010). For this temporal stack of normalized Landsat imagery, a spectral index referred to as the NBR (Key & Benson, 2006) was calculated, which contrasts Landsat bands 4 and 7 (Near Infrared (NIR) and shortwave infrared (SWIR)) (Table 3.2). GEE was used to access high-resolution images from the National Agriculture Image Program (NAIP) (Figure 3.2, step 8) for validation purposes (refer to Appendix 1). NAIP images were used to validate the final classification of NBR values. NBR-based maps were overlaid with high resolution NAIP images to confirm the interpretation of environmental conditions on the ground (Figure 3.2, steps 5,6, and 7).

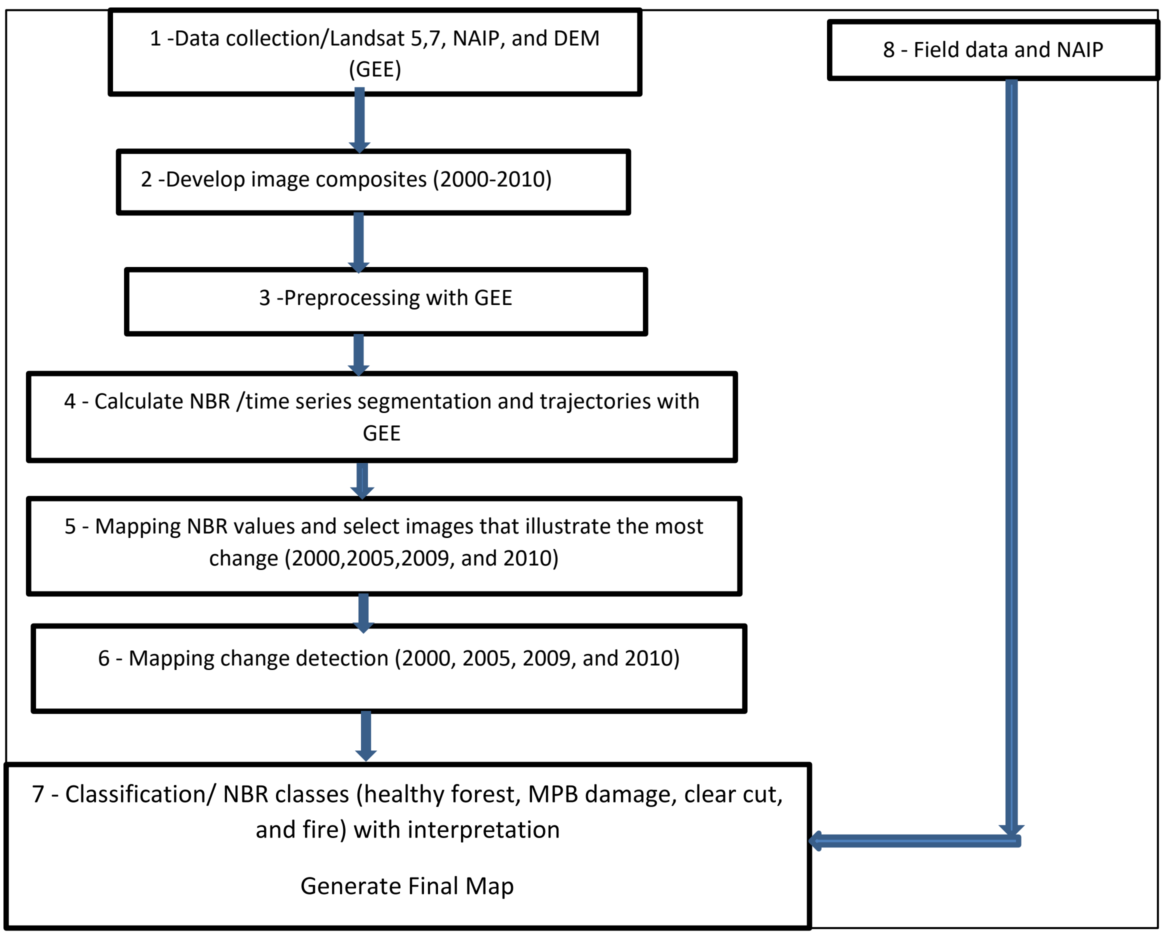


Figure 3.2 Flowchart for data analysis

Flowchart showing multiple steps in data collection, processing, analysis, and interpretation to produce final maps showing distribution of MPB damage across the study area.

Table 3.2 Landsat sensors bands TM5 and ETM+7 and their spectral and temporal characteristics

Bands	**Sensor TM5 characteristics**	**Sensor ETM+7 characteristics**
Band 1 (blue)	0.45 - 0.52 µm	0.45 - 0.515 µm
Band 2 (green)	0.52 - 0.60 µm	0.525 - 0.605 µm
Band 3 (red)	0.63 - 0.69 µm	0.63 - 0.69 µm
Band 4 (NIR)	0.76 - 0.90 µm	0.75 - 0.90 µm
Band 5 (NIR) or (SWIR1)	1.55 - 1.75 µm	1.55 - 1.75 µm
Band 6 (thermal)	10.40 - 12.50 µm	10.40 - 12.5 µm
Band 7 (SWIR2)	2.08 - 2.35 µm	2.09 2.35 µm
Spatial resolution (meter)	30 m	30 m
Temporal resolution (days)	16 days	16 days

3.3.3 Data selection, collection, and processing

10 points or regions of interest (ROI) (Table 3.3) were selected based on NBR values and high-resolution image such as NAIP to represent forested areas that had remained healthy or that had undergone different histories, such as fire, clear cutting, or insect infestation, and that could be tracked via time series analyses. Knowledge of health status and occurrence of destructive events in each ROI was based on NBR values (+1 to -1) calculated for 2000, 2005, 2009, and 2010. High NBR values between +1 and 0.4 represent healthy trees and low values below 0.4 to -1 represent disturbed area. In addition, the trajectories of time series analyses provide information on the type of disturbance with steep trajectories suggesting instantaneous events such as clearcutting or fires, while gradual sloped trajectories indicate infestation development. Two bands of satellite data (Band 4 and Band 7) were sampled for each site for each year in the time series (2000-2010) and processed to calculate NBR values using Equation (1) as discussed above.

Data were chosen from the years 2005 and 2009 because these data could be compared with fine resolution National Agricultural Image Program (NAIP) data for validation. Four sites were selected to represent different historical scenarios that produced different segmentation patterns.

Table 3.3 Locations of selected regions of interest (ROI) with associated NBR values

Sites	Longitude	Latitude	2000_NBR	2005_NBR	2009_NBR	2010_NBR
roi1	-106.618	40.20595	0.391009	0.4568	0.445248	0.437522
roi2	-106.667	40.26364	0.290929	0.270615	0.08264	0.033195
roi3	-106.636	40.08625	0.410266	0.305918	0.15532	0.096116
roi4	-106.62	40.07812	0.409326	0.405859	0.394936	0.370411
roi5	-106.589	40.09076	0.341366	0.309205	0.162187	0.109406
roi6	-106.566	40.08945	0.418903	0.430997	0.232012	0.128792
roi7	-106.573	40.11041	0.405055	0.374149	0.227177	0.090189
roi8	-106.675	40.26439	0.319671	0.303155	0.083898	-0.01429
roi9	-106.701	40.29755	0.500028	-0.34381	-0.1388	-0.14245
roi 10	-106.641	40.09607	0.35697	0.314189	-0.09	-0.19169

3.4 Data analysis via LandTrendr and disturbance mapping

The LandTrendr segmentation approach (Kennedy et al., 2010; Meigs et al., 2015) was chosen to map and characterize forest disturbance between 2000-2010. The analysis program follows three main steps: 1. Create an annual time series of cloud-free observations; 2. Fit time series trajectories of NBR values for each pixel; and 3. Derive a set of metrics from each trajectory to describe the disturbance and recovery characteristics of each pixel. All Landsat images were plotted in their original UTM projection (WGS84) and mosaicked to provide the final disturbance metrics for the study site in the NAD83, UTM zone 13N projection.

Once a consistent annual time series is created, the LandTrendr algorithm fits spectral trajectories to the time series by dividing it into a series of connected linear segments. The initial spectral trajectory represents the average NBR values prior to the commencement of the disturbance and is indicative of the baseline status of the forest in that region of interest for the years prior to the disturbance. The severity or magnitude is the change in the NBR values associated with the disturbance event from the year of commencement to the year in which the disturbance ends (defined when the fitted segment trajectory ends, and new segment begins). The persistence or duration of the disturbance is the time between the initial and the end year of the disturbance. Fig. 3.3 shows a simplified hypothetical example of a disturbance spectral trajectory change in response to some disturbance (Coops et al., 2020).

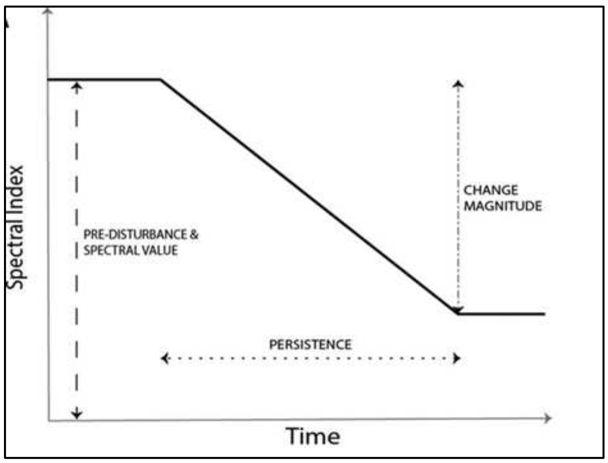


Figure 3.3 Spectral trajectory

Schematic diagram of the three attributes extracted from a spectral trajectory by the LandTrendr segmentation process. The diagram shows the initial spectral value, the magnitude of change in the spectral value, and the length of time over which this change took place as a function of the persistence of the disturbance (Coops et al., 2020).

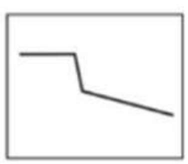
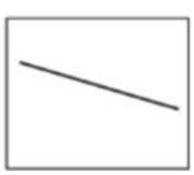
In this study, calculated NBR data were used to derive the segmentation and fitting models for the forest disturbance history. Segmentation distills a noisy annual time series into a simplified set of vertices and segments to capture the most important features of a trajectory while omitting most false changes (Kennedy et al., 2010; Senf et al., 2015). Segments were identified where NBR declined or increased between vertices. For each disturbance segment, the absolute change in NBR was determined as an estimate of disturbance magnitude. In addition, the onset year, duration in years, and slope of each segment (where slope is total NBR difference magnitude (NBRMax - NBRMin) of a segment divided by duration) were mapped. This last variable, NBR slope, can be used to distinguish abrupt, severe disturbances like stand-replacing wild-fires and logging events from more gradual, subtle changes due to insect activity (Meigs et al., 2015; Senf et al., 2015). These distinct trajectory differences permit classification of spatial and temporal patterns of disturbances. LandTrendr disturbance and recovery metrics were classified as harvest/fire disturbances, insect disturbances, and undisturbed areas.

Stands affected by MPBB do not always show the simple decrease in NBR shown in (Figure 3.3). Some NBR trajectories associated with MPB exhibit sharp short-term decline in NBR (e.g., Table 3.4a) followed by a slower decrease which suggests a sharp mortality event followed by subsequent slower mortality and dead tree fall. Most NBR plots associated with MPB mortality show the same sort of slow gradual decline NBR shown in (Table 3.4b).

Other NBR plots show a gradual decline in NBR (e.g., Table 3.4c), sometimes followed by a gradual increase in NBR; these areas likely experienced an MPB mortality event followed by recovery. In general, these areas likely represent healthy stands of lodgepole pine which resulted in lower total mortality and more robust regrowth (Kennedy et al., 2010; Senf et al., 2015).

Table 3.4 Shows conceptual frameworks for time series trajectories. (Meigs et al., 2011)

Note - Spectral trajectories of insect-affected stands generally exhibited one of the three patterns shown in this table. The patterns are generally related to environmental conditions within those stands.

Spectral trajectory	**Interpretation**	**Insect mortality agent**	**Environmental conditions**
 a	Stable, rapid mortality, slow mortality	mountain pine beetle	several disturbance events, such as beetle infestation followed by tree fall
 b	long term slow mortality	Mountain pine beetle and Western spruce budworm	long-term presence of insects.
 c	slow mortality and regrowth	mountain pine beetle and Western spruce budworm	relatively healthy area exhibiting rapid regrowth following insect infestation

Following the data acquisition, pre-processing, processing, and classification steps described above, it was possible to leverage the aerial surveys, Landsat spectral trajectories, and field data to create new maps of MPB impacts on tree mortality. To minimize commission errors due to background variation and statistical noise, only segments whose change in NBR exceeded a threshold value of 50 units (NBR*1000) were used, with adjustments to higher values for shorter duration segments. Disturbance segments were classified according to whether they were associated with a long-duration decline (defined as lasting five or more years) or short-duration decline (less than three years). The length of a given disturbance was defined based on marked change points in trajectory slopes (beginning year to end year). All geospatial processing, statistical analyses, and data visualizations were performed with the ArcGIS Pro 3.1.0 and R programming language.

3.4.1 Statistical analysis plan

Data from all 10 Regions of Interest (ROI) (Table 3.3) were examined for 4 selected years (2000, 2005, 2009, 2010). The first date marked the beginning of the study period (2000). One data point was selected in the middle of the study period (2005) which also had NAIP data available. The last data point marked the year at the end of the study period (2010), while one additional point was selected from 2009 because it also had NAIP data available. To facilitate the interpretation of 4 data sets, they were assessed by means of a one-way ANOVA followed by a Tukey's pairwise comparison. Statistical significance was determined at $P< 0.05$. $P > 0.05$ is the probability that the null hypothesis is true. 1 minus the P value is the probability that the alternative hypothesis is true. A statistically significant test result ($P \leq 0.05$) means that the test hypothesis is false or should be rejected.[1] A P value greater than 0.05 means that no effect was observed. All statistical analyses were performed in R. A mixed model was fitted using R version 4.2.1 (2022-06-23 ucrt) with the NLME package (Pinheiro J. & package version 3.1-162). The response

[1] Goodman, S. (2008). A dirty dozen: twelve p-value misconceptions. Seminars in hematology,

value for the linear regression was NBR. Years (2000, 2005, 2009, 2010) were included as a fixed effect. Sites (1-10) were included as a random effect to account for repeated measures at these sites. Residual diagnostics plots were used to verify model assumptions. Table 3.4 shows conceptual frameworks for time series trajectories.

3.5 Results

Statistical results from one-way repeated measures show that there is significant difference in NBR values over time based on P values. The Tukey's pairwise comparison showed that there is a significant difference on NBR average between NBR 2000 and NBR 2009, between NBR 2000 and NBR 2010, and between NBR 2005 and NBR 2010 with P. value 0.0039, 0.0002, and 0.0183, respectively Table 3.5.

Table 3.5 Pairwise comparison

Contrast	**Estimate**	**SE**	**df**	**t.ratio**	**p.value**
time2000 - time2005	0.1016	0.0601	27	1.691	0.3477
time2000 - time2009	0.2289	0.0601	27	3.808	0.0039
time2000 - time2010	0.2926	0.0601	27	4.869	0.0002
time2005 - time2009	0.1272	0.0601	27	2.117	0.1732
time2005 - time2010	0.1910	0.0601	27	3.178	0.0183
time2009 - time2010	0.0637	0.0601	27	1.061	0.7158

3.5.1 Time series trajectory segmentation and interpretation

The most important points to note are the year of change detection, the change detection duration, and the change detection magnitude. Three thresholds were defined based on NBR values that separated 4 main differing trajectories: stable healthy forest, insect infestation, destruction by fire, and clear cutting (Fig. 3.4). For example, healthy forests show NBR's of > 0.3. An area that was affected by insect infestation shows a decrease in NBR from above 0.3 down to 0.05 from 2005-2010. An area affected by forest fire shows a drop in NBR from above 0.3 down to -0.45 from 2001 to 2002. An area that was clear cut shows a decrease from 0.3 down to – 0.2 from 2008 to 2010. All these values were calculated from Band 4 and Band 7 L5 and L7 data using Equation (1) discussed above.

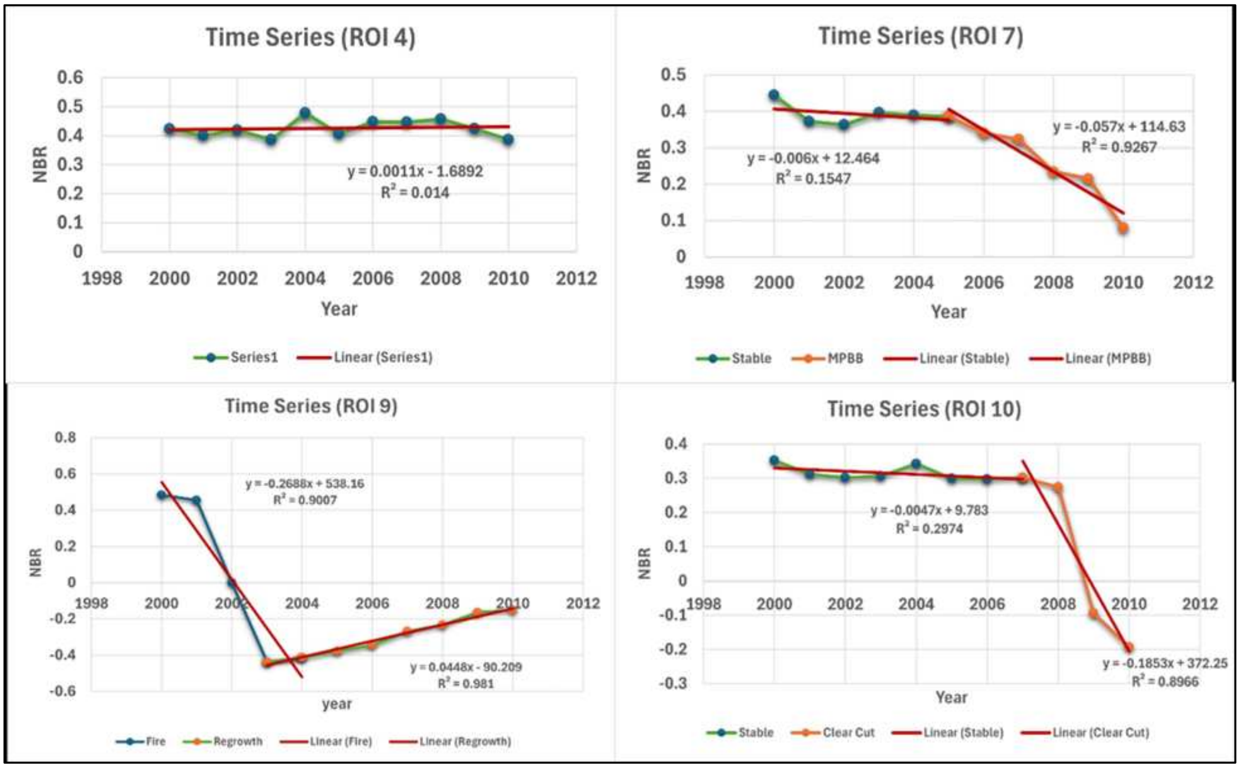


Figure 3.4 NBR time series for 4 sampling locations.

This figure shows changes in NBR values with time to show the different patterns of change in sites where the trees remained healthy (ROI4, a) or showed insect infestation (ROI 7, b), versus sites which experienced abrupt changes such as fire (ROI 9, c) or clear cutting (ROI 10, d). NBR value ranges and their general interpretation are listed in (Table 3.6).

Table 3.6 NBR ranges, classes, and general interpretation

Estimated percent damage	NBR ranges	Class number	general interpretation
100	-0.4 – 0	0	Unvegetated or destruction by fire or clear cut
80	0.01 – 0.1	1	Severe damage MPB
60	0.11 – 0.2	2	Moderately severe damage MPB
40	0.21 – 0.3	3	Minor damage MPB
20	0.31 – 0.4	4	Stressed forest
0	0.41 – 0.9	5	Stable healthy forest

3.5.2 Disturbance mapping over time

LandTrendr disturbance metrics can be used to sort forest changes into three main classifications: 1) undisturbed forest, 2) insect disturbances, 3) fire disturbances and clear-cut disturbances. Figure 3.5 shows examples of Landsat spectral trajectories for each classification.

Note that wherever there has been a disturbance the NBR value drops. Figure 3.5A shows that most of the NBR values for relatively undisturbed areas plot above the 0.4 level (Table 3.6). Slightly disturbed zones (the middle part of the trajectory shown in Fig 3.5B) exhibit NBR values between 0.3-0.4. These zones likely represent areas where trees are stressed due to poor environmental conditions. Moderately disturbed areas are divided into two zones which show NBR values between 0.1 -0.3 and 0.2- 0.3, respectively, which represents areas showing the onset and development of MPB infestations. In contrast, severely disturbed areas show NBR values ranging from - 0.45 to 0.01, reflecting the almost complete to complete denudation of forested areas by fire or clear-cut harvesting. A map showing a depiction of disturbance severity was generated to highlight areas that show disturbance based on NBR values (Fig 3.6). Red areas in the NBR image (2010) denote areas where the maximum amount of disturbance occurred. To improve computational efficiency, a small area was selected which showed the clearest development of infestation over time. This data subset permitted classification the of the type of disturbance that occurred without running the entire data set.

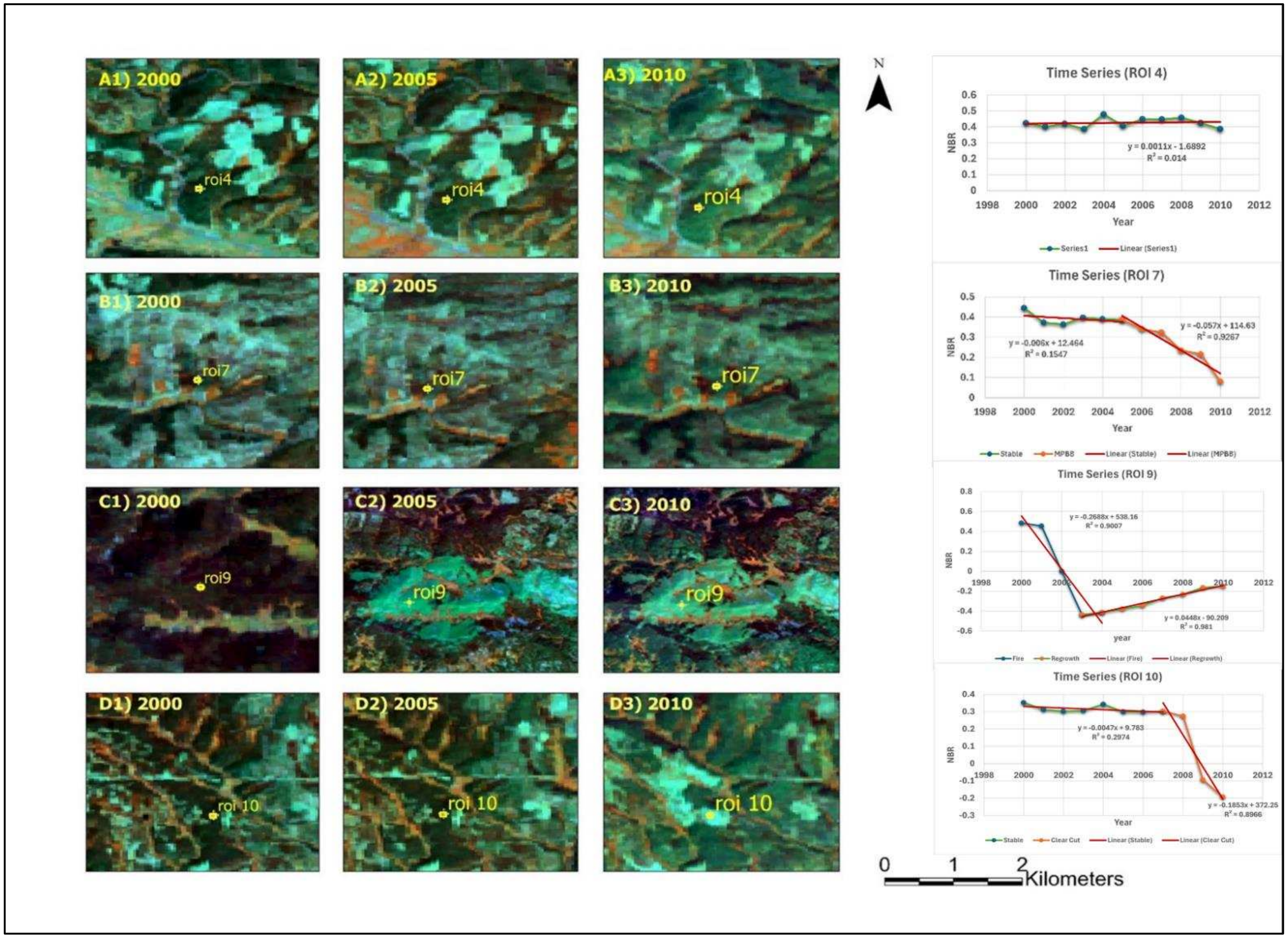


Figure 3.5 Comparison of LandTrendr fitted trajectories with corresponding Landsat Images.

In this figure, row A shows three map views of region of interest (ROI) 4 which is an area in which the forest remained undisturbed from 2000 to 2010. Row A contains 3 satellite photos, one each from the years 2000, 2005, and 2010, and the NBR time series trajectory for this site. Row B contains a similar set of images for ROI 7, row C contains images from ROI 9, and row D contains images from ROI 10. Row B shows a gradual decline in the NBR time series, whereas row C shows a sharp drop in the NBR time series, and row D also shows a sharp but lower magnitude drop in the NBR time series.

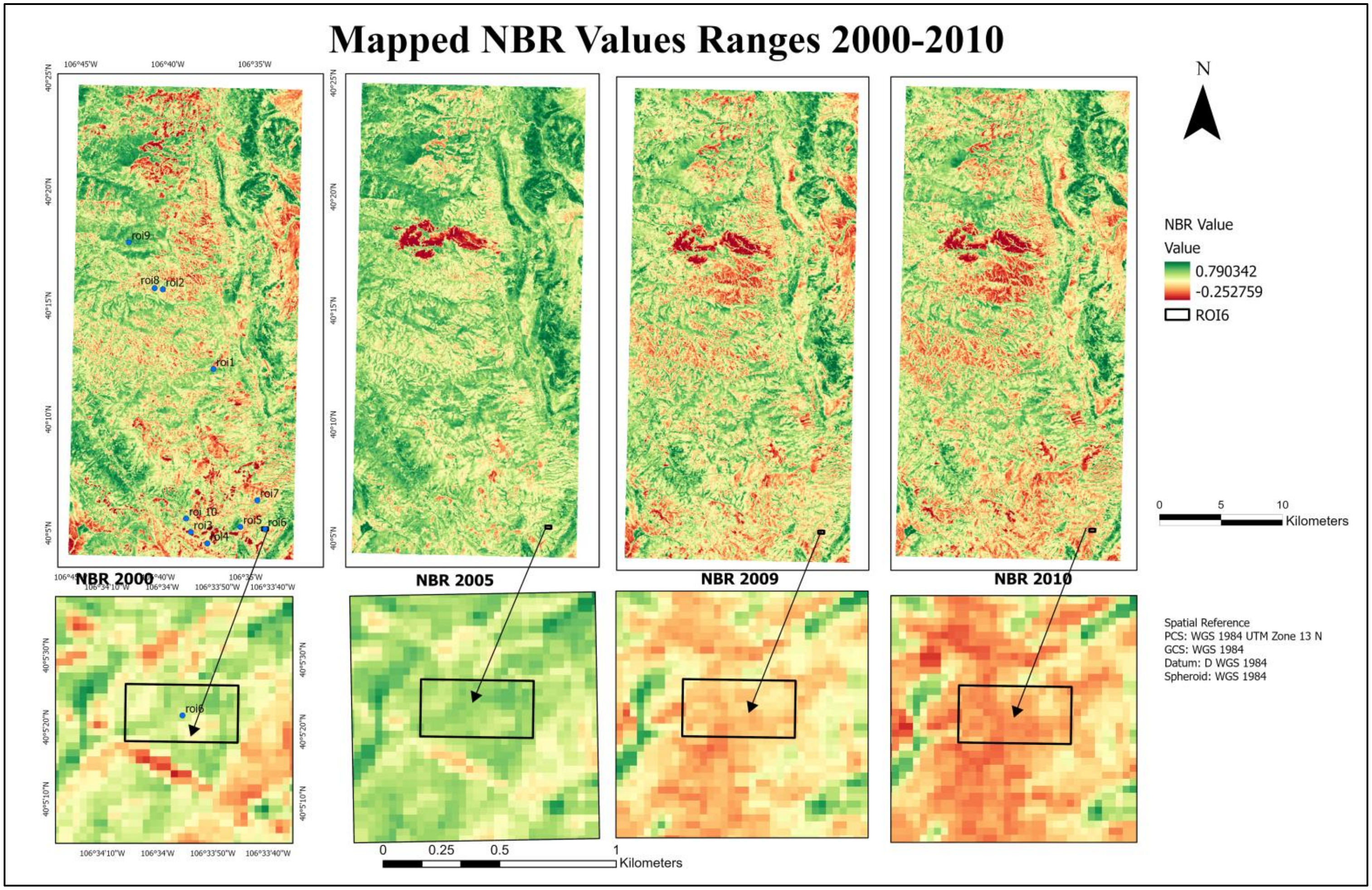


Figure 3.6 Mapped NBR value ranges 2000-2010

In Fig 3.6, NBR values from ROI 6 ranging from -0.2 to 0.8 were plotted for the four years analyzed in this study to show changes in landscape characteristics with time. Note for example the large area showing a very low NBR value in 2005. This area represents an area that experienced a large fire in 2001. The images from 2009 and 2010 are very similar because they both show the effects of long-term stress due to drought and MPB infestation. Note, however, how different they are from 2005, which represents a relatively wet year.

The data from the first map (2000) was added to data from the last map (2010) to produce a binary change map which shows the areas with the maximum amount of change (Figure 3.7). This allowed selection of the areas with the most change for further analysis to categorize the cause of the change.

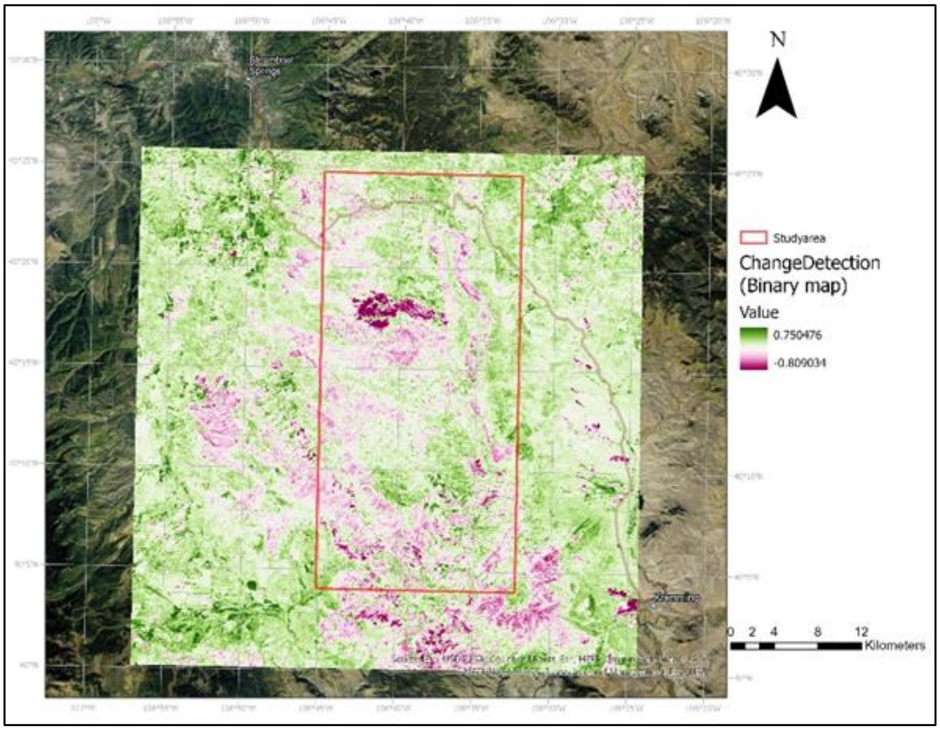


Figure 3.7 Binary map showing maximum change in NBR 2000-2010

In figure 3.7, the areas colored purple are areas which showed the maximum change in NBR between 2000 and 2010. In contrast, the areas colored green showed the least change in NBR, while the white-colored areas showed a moderate amount of change.

The next set of maps (Fig. 3.8) show the change in disturbed zones between 2000 and 2010, with a finer discrimination of the degree and likely cause of the disturbances. These four maps show changes and disturbances that were classified into six divisions based on NBR values (Table 3.6).

The estimated percentage damage is related to a given NBR range associated with a general interpretation of the degree of damage, and likely cause of that damage. Each range has been given a class number from (0 - 5).

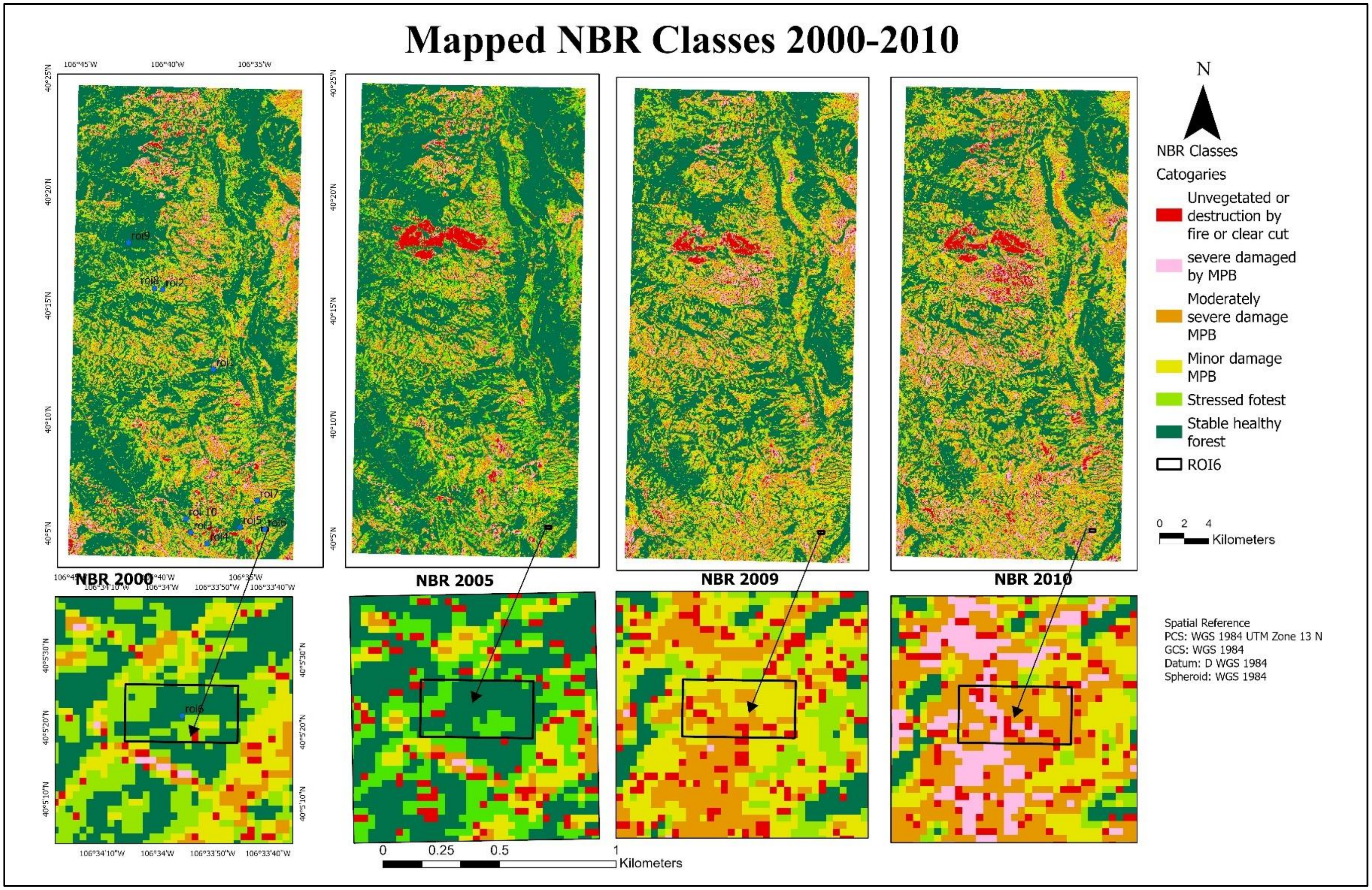


Figure 3.8 Mapped NBR classes 2000-2010

The regional distribution of 6 NBR classes are differentiated on the basis of the amount of healthy forest existing in that area.

The maps in Fig 3.8 show the change in distribution across the study area of 6 NBR classes from 2000-2010. For numeric NBR values that define the classes see Table 3.6. The dark green areas (class 5) represent stable healthy forests, whereas the lighter green areas (class 4) represent forested areas experiencing stress from drought. In contrast, red areas (class 0) represent areas which were unvegetated or severely damaged by fire or clear cutting within this same time period. The other 3 classes show the range of minor MPB damage (class 3) to severe MPB damage (class 1). Class 2 represents areas where MPB damage was moderately severe. These maps could provide invaluable information for forest management programs by defining the location and movement with time of MPB infestations.

3.6 Discussion

The results of this study suggest that insect disturbances can be distinguished from undisturbed areas and more intense disturbances such as clear-cut and fire (Kennedy et al., 2010; Meigs et al., 2011; Senf et al., 2015). The goal was to develop an accurate measure of MPB-caused tree mortality using Landsat spectral indices and imagery data combined with finer-resolution reference data such as NAIP (Figure 3.2, step 8). Continuous measures of tree mortality provide more information than a simple determination of disturbed versus undisturbed, increasing the usefulness of these data for forest management purposes. This study sought to develop a remote sensing analysis procedure to distinguish a variety of temporal trajectories of spectral response (Figure 3.4) to insect activity as well as other common disturbances. LandTrendr algorithm results were used in conjunction with aerial survey data (NAIP) and ground check surveys (see Appendix 1) by the author to validate interpretation of Landsat time series (spectral trajectories) to characterize changes in the history of an area over time. These changes were plotted in a series of maps to show the change in land cover over time. For example, comparison of Fig. 3.4 time series plots B, C, and D shows the marked difference in trajectory for areas affected by MPB, fires, and clear cutting, respectively. The trajectory plots also clearly show the timing of events, with the fire occurring in 2002, the clear cut occurring in 2008, and MPB infestation beginning in 2005 and continuing through 2010 (Meddens et al., 2013). The results can be statistically analyzed to produce binary maps (Fig. 3.7) that show the areas with the maximum amount of change. This allows selection of the areas with the most change for further analysis.

The goal of these analyses is to produce a data set and maps that characterize the cause, degree of, and duration of disturbances of healthy forests. These summary results are documented in Tables 3.7 and 3.8. Table 3.7 shows the total change in area for each class defined in Table 3.5 for each year selected for this study. Table 3.8 shows the percentage of the total area (69,692 ha) of the study represented by each class of the time frame of the study. Note that the range of cumulative mortality never reached 100% even in areas with major outbreaks, suggesting that there are still living trees in these areas, as confirmed by field studies. These living remainders are what permit recovery in even the most severely damaged areas (e.g., Figure 3.5C). Note that although the long slow mortality representing MPB attack is shown in Fig. 3.5B is a common pattern, it is not the only pattern of change shown by MPB infestation. Meigs et al. (2011) discusses that attack by MPB can show some variety in pattern depending on factors such as the relative health of the trees in the given area, whether is one beetle attack or multiple attacks or if the present of the insects is long term or there may be multiple insect agents affecting different host within the area. These variations in history are reflected in the different spectral trajectories, as seen in Table 3.4. The subtle variations can be resolved by reference to other data such as NAIP, imagery on Google Earth Pro platform, and further field verification. The statistical results in Table 3.5 can be used to select areas that need closer examination on the ground by the Forest Service. This will allow the Forest Service to focus on aerial data gathering to areas of particular interest to lower the costs and time associated with more random or generalized flight patterns.

Table 3.7 Total change for each class within each area

This table shows the total change for each class within each area in 2000, 2005, 2009, and 2010. Class 0 shows an unvegetated area; Class 1 shows severe damage (MPBB); Class 2 shows moderately severe damage; Class 3 shows minor damage; Class 4 shows stressed forest; and Class 5 shows stable, healthy forest.

classes	**Area-ha/2000**	**Area-ha/2005**	**Area-ha/2009**	**Area-ha/2010**
0	3433	5074	5412	6583
1	2541	1618	2891	4573
2	5809	4321	6780	8807
3	9357	8312	10999	11140
4	12281	12176	13046	11683
5	36271	38191	30564	26907

Table 3.8 Percent change for each class within each area

Classes	**% of change/2000**	**% of change/2005**	**% of change/2009**	**% of change/2010**
0	4.9	7.3	7.8	9.4
1	3.6	2.3	4.2	6.6
2	8.3	6.2	9.7	12.6
3	13.4	11.9	15.8	13
4	17.6	17.5	18.9	16.8
5	52.04	54.8	43.9	38.6

3.7 Conclusions

In the past two decades, MPB's have become more pervasive due to increasing temperatures and drought conditions related to climate change causing regional-scale mortality. As discussed above, a variety of studies have suggested that insect effects on tree die-off, fuels, and fire behavior can vary widely. A key problem with developing a better understanding of insect-fire relationships is the lack of empirical maps that can show interrelated changes in the distribution of insect infestations and fire zones reliably over space and time. Although remote sensing data has frequently been used to map forest fires with well-established techniques familiar to forest managers, the science of mapping insect infestations is relatively new and lacking in well-developed and documented spatiotemporal data sets that can be used to track these infestations and quantify the negative impacts of the infestations as well as other various agents related to forest health in a consistent fashion. This paper documents the development of such a technique based on time series statistical analyses of NBR spectral index trajectories. This approach appears to produce reliable repeatable results which can be used by forestry agents to improve their ability to manage, maintain, or improve the health of regional forests.

Chapter 4 - Quantifying and mapping of tree mortality as a function of time using Random Forest (RF)

4.1 Introduction

This research examines an integrative approach to model land cover changes at a regional level, using as a case study the simulation of spatiotemporal dynamics of Mountain Pine Bark Beetle (MPBB) infestations in lodgepole pine forests of north central Colorado. The main goals of this work are:

1 – Model MPBB spread in the study area between 2005-2009 using the Random Forest RF classifier and testing its performance.

2 – Evaluate the usefulness of selected predictor variables which describe the influence of local topography and to determine the extent and speed of spread of MPB infestation.

3 - Predict land cover changes and compare these predictions with NAIP and field data from 2009.

4.2 Methodology

4.2.1 Random Forest (RF)

RF is a decision tree algorithm which basically randomly selects a sample of observations and a sample of variables many times (default value of 500) to produce a large number of a small classification trees which are then aggregated through the process of "bootstrap aggregating" or bagging. The idea is to use multiple versions of a predictor or classifier to make a final decision on classification based on a majority vote rule to determine the final category (Breiman, 2001). In bagging, it has been proven that as the number of predictors increases, accuracy also increases until it reaches a certain point at which it will begin to drop off. Finding the optimal number of predictors to generate the result will yield the highest accuracy. Each tree in the forest is created in the following manner. Given a training set, a random subset is sampled and used to create a decision tree. Then about 1/3 of the bootstrap sample is left out and considered to be out-of-bag (OOB) data. Future trees are created with new subsets in which part of the data has been replaced with new data. In addition, not every feature is used to construct each tree; a random selection of features is evaluated for each node. The OOB data are used to get a classification error as trees are added to the forest and to measure the input feature importance. After the forest is completed, a new case can be classified by taking a majority vote among all the trees in the forest resembling the bootstrap aggregating idea (Kulkarni & Lowe, 2016).

The error rate of the forest can be measured by using two different values. The error rate depends on two calculations: correlation between any two trees in the forest and the strength, or error rate of each tree. For example, if we have M input variables, we can select m of them at random to grow a tree. As m increases, correlation and individual tree accuracy also increases, and some optimal m will give the lowest error rate. Each tree will be grown by choosing a new subset of the m variables. RF also measures important variables by using OOB data. Each variable is randomly permuted and the permuted OOB cases are sent through the tree forming process again. Subtracting the number of correctly classified cases using permuted date from the number of correctly classified cases using non-permuted data will give the importance value for each m variable. These values are different for each tree but the average of each value over all trees gives a raw importance score for each variable (Cutler et al., 2007).

4.2.2 Study Plan

This study follows the steps listed in flow chart (Figure 4.1). The first step requires the collection of three different types of data (Table 4.1): Digital Elevation Model (DEM) data, 6 bands of data from Landsat 4-5 Thematic Mapper (TM), 7 bands of data from Landsat 7 Enhanced Thematic Mapper Plus (ETM+) (Table 4.1).

The next steps are developing the annual image composites, preprocessing, converting numerical pixel values to surface reflectance numbers, and digitizing images by using GEE (Step 2). DEM data is extracted and used to calculate slope, and aspect. These data are used to examine relationships between the orientation of a certain tract and the degree of infestation to see if infestations are preferentially

Table 4.1 Spectral region of Landsat TM5 and ETM+

NIR: Near Infrared; SWIR: Short Wave Infrared; Pan: Panchromatic; TIRS: Thermal Infrared Sensor.

Bands	**Landsat 5 (TM) spectral regions**	**Landsat 7 ETM+ spectral regions**
Band 1 (blue)	045 – 0.52	0.45 - 515
Band 2 (green)	0.52 - .60	0.525 – 0.605
Band 3 (red)	0.63 0.69	0.63 – 0.69
Band 4(NIR)	0.76 – 0.90	0.75 – 0.90
Band 5 (SWIR1)	1.55 – 1.75	1.55 – 1.75
Band 6 (thermal)	10.40 – 12.50	10.40 – 12.50
Band 7 (SWIR2)	2.08 -2.35	2.09 – 2.35

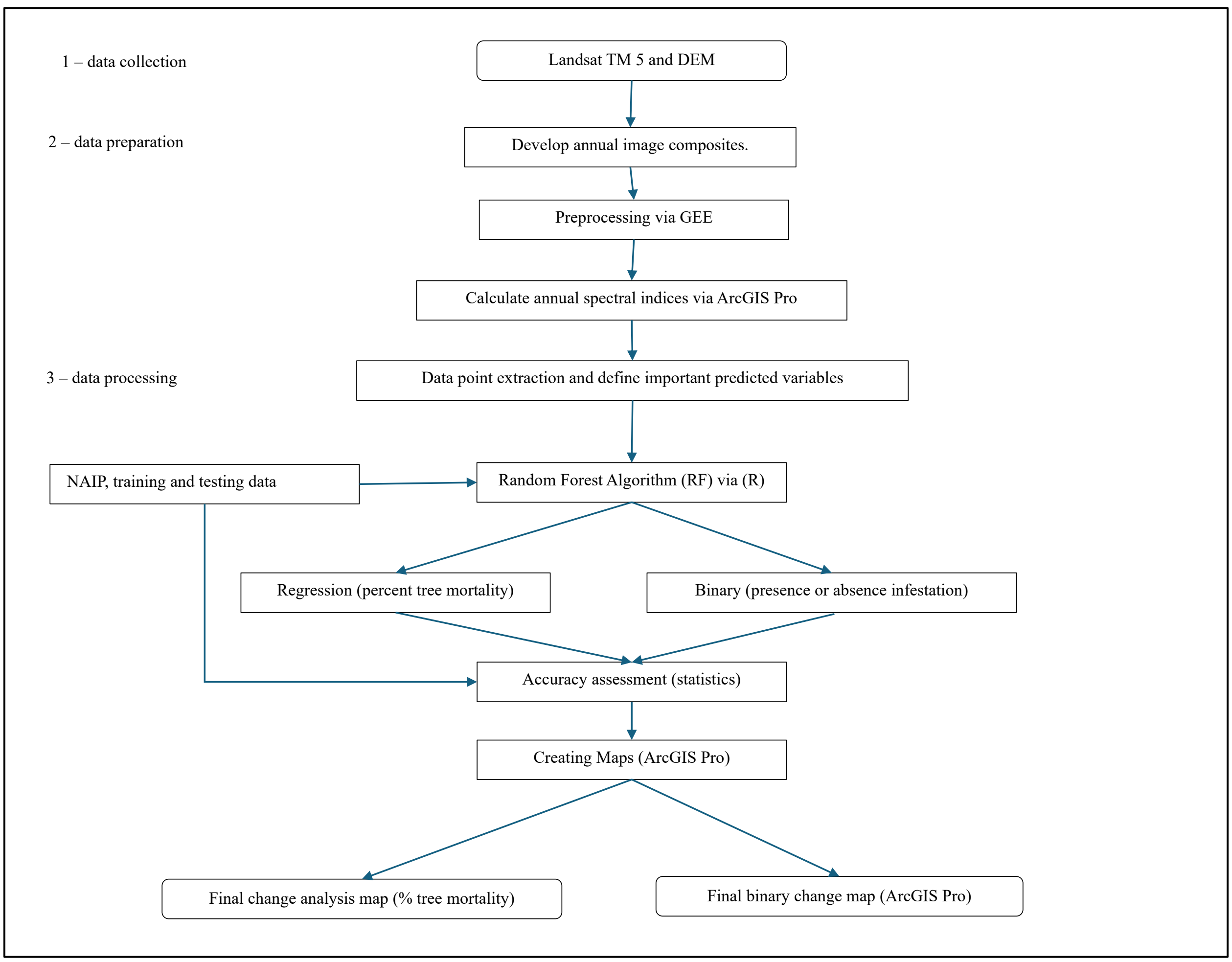


Figure 4.1: Flowchart of methodology

developed on slopes facing a certain direction. A subset of this data will be collected to act as training data for the analysis algorithm using ArcGIS Pro 3.1.0.

Step 3 involves the processing of the data to calculate annual spectral indices (Table 4.2), segmentation of yearly time series by Google Earth Engine (GEE) and ArcGIS Pro (Chapter 3, Figure 3.4). At this point, training data selected by ArcGIS is used as input data to RF to provide definitions for classifying various features such as trees, grass, soil (feature extraction). Data points defining trees will be separated into infested vs. non-infested categories. The accuracy of this separation will be assessed by comparison with field data and NAIP images collected for validation. Collection of field data involves checking for presence or absence of infested trees and is documented with photographs, and field notes. These observations are also compared with NAIP images for the study area. Final maps of the change analysis are then prepared.

Table 4.2 Vegetation Indices

These vegetation indices were calculated from data derived from the various Landsat bands indicated in the formula column.

Index	**Formula**	**Notes**	**reference**
Normalized Burn Ratio (NBR)	(NIR-SWIR2) / (NIR+SWIR2	Moisture stress & needle structure	(Meigs et al.,2011)
Normalized Difference Vegetation Index (NDVI)	(NIR-RED) / (NIR+RED)	Sensitive to conifer tree healthy	Tucker (1979)
Band5/ Band4 (B5/B4)	B5/B4	Sensitive to conifer tree healthy	Vogelmann (1990)
Tasseled-Cap Brightness (TCB)	0.2043 * B1 + 0.4158 * B2 + 0.5524 * B3 + 0.5741 * B4 + 0.3124 * B5 + 0.2303 * B7	Sensitive to surface brightness	Crist (1985)
Tasseled-Cap Greenness (TCG)	-0.1603 * B1 - 0.2819 - B2 * 0.4934 * B3 + 0.7940 * B4 - 0.0002 * B5 - 0.1446 * B7	Moisture stress	Crist (1985)
Tasseled-Cap Wetness (TCW)	0.0315 * B1 + 0.2021 * B2 + 0.3102 * B3 + 0.1594 * B4 - 0.6806 * B5 - 0.6109 * B7	Sensitive to vegetation greenness	Crist (1985)
Normalized Difference Moisture Index (NDMI)	(NIR-SWIR1) / (NIR+SWIR1)	Sensitive to canopy water content	Wilson and Sader (2002)

4.2.3 Landsat data collection and preparation

High-resolution multispectral imagery was acquired through GEE of the Upper Yampa Basin study area from 2000-2010. Surface reflectance data from Landsat imagery of the study area was preprocessed, subset selected, and downloaded from GEE. Study data were selected from images available from May-

September, 2000-2010, which were preprocessed through GEE to eliminate cloud cover problems. All images were orthorectified into the coordinate system, terrain-corrected, and atmospherically corrected for surface reflectance, and transformed from Landsat visible-NIR-SWIR bands. These images have been found to be most effective in remote sensing of Bark Beetle-induced tree mortality (Meddens et al., 2013). The annual Landsat image time series for the years 2000-2010 was created using functions from the LandTrendr algorithm in GEE (Refer Chapter 3). Annual time series of individual bands and indices were created using the "transformSRcollection" and "collectionToBandstack" functions in order to calculate a number of indices based on 7 bands from Landsat TM. Data from these bands are used to calculate a variety of indices that provide information on vegetation health. (Table 4.2). These indices provide information about factors such as moisture stress and needle structure that help identify both the type of tree (conifer vs. deciduous) and the moisture level in the canopy which is an indicator of tree health (Hart & Veblen, 2015; Kennedy et al., 2018; Meddens & Hicke, 2014; Migas-Mazur et al., 2021).

The iterative method of Meddens et al. (2013) was used to create time series of B5/B4 and NBR anomalies from undisturbed means. To create undisturbed mean B5/B4 and NBR values, images of B5/B4 and NBR for all years in the study were combined and mean values for each pixel were calculated. Pixels that were more than one standard deviation from the mean in the direction of disturbances (positive B5/B4 and negative for NBR) were identified as disturbed pixels, and mean images without disturbed pixels were calculated. This process was repeated multiple times until the mean stabilized to create undisturbed mean B5/B4 and NBR images. The basic assumption of this approach is that forested pixels were undisturbed at some points during the study period.

Time series of B5/B4 and NBR means for the study period were created by subtracting undisturbed mean B5/B4 and NBR images from raw annual B5/B4 and NBR images, respectively. B5/B4 and NBR anomalies were used for differentiation of predominantly live (undisturbed) or predominantly dead (disturbed) tree cover. This preprocessing was performed in GEE, and output from GEE was then further analyzed, classified, and a predictive model developed by use of the R Statistical Computing program and ArcGIS Pro 3.1.0.

4.2.4 Reference and training data

Reference data were selected at the 30 m plot level from the available very high-resolution (VHR) satellite imagery in Google Earth Pro for training and validation purposes. In addition, multispectral aerial imagery (spatial resolution 1 x 1 m) was collected from National Agricultural Imagery Program (NAIP) (Aerial Photography Field Office (AFPO), 2012) for manual interpretation. NAIP images are repeated every 2 to 5 years providing a continuous long-term record. There are 7 years of NAIP images with a resolution of 2m or finer (Colorado NAIP from years 2004, 2005, 2006, 2009, 2011, 2013, and 2015) available from GEE. In addition, field data of confirmed MPBB infested trees were collected and additional data over the study area was derived from Insect and Disease Detection Surveys (IDDS) (Colorado Forest Service, from years 2000, 2005, and 2009). These field data were used as a general reference, however, there was not always a direct one-to-one correlation between reference data points and study data points because of accessibility issues (private property vs. national forest) or topographic or ground cover effects. Mortality fraction estimates were calculated by generating 10-by-10 regular grids in each 30 m plot for reference and counting the number of grids intersecting with a tree with a red or grey stage canopy. For validation in this study, a random sample of training data comprised of field measured ground truth, comparison with NAIP and GEE images, and available IDDS data was used. In addition, the data were divided into two sets, one being the main dataset used in model development (training data), and the second, a subset of data points to be used for validation (testing data). 70 % of total data points were selected to use as training data to determine the distribution of healthy versus unhealthy forest and degree of disturbance. The remaining 30 % of the data were used for validation.

4.2.5 Multi-spectral classification

A classification scheme separating non-forest, disturbed forest, and undisturbed forest was developed following the classification method of Meddens, et al. (2013). Forest was separated from non-forest using Band 2 (B2), forest exhibited a range of values from 0.033 to 0.063. Values below 0.033 represent non-forested areas such as water, whereas values above 0.063 represent soil, rocks, and other vegetation. To separate out grassy areas, Tasseled-Cap Greenness (TCG) values were used, where values ranged from 0.11 to 0.38. A second index that can be used to separate out barren soil, burned zones, or clear-cut areas is the Tasseled-Cap Brightness in a range from 0.34 to 0.65. Forested areas were subdivided into undisturbed or disturbed using NBR values, where undisturbed areas show values ranging from 0.41 to 0.63, whereas disturbed areas exhibit that use ranging from 0.013 to 0.4. Classification accuracy was assessed using NAIP and Testing data.

4.2.6 Developing a model for estimating presence or absence and the percent tree mortality by using regression techniques

The approach followed in this study was to use aggregated high-resolution data from NAIP, Google Earth Pro, and Landsat-derived variables including spectral bands and indices considered as potential explanatory variables to ultimately permit development of models predicting percent tree mortality. 1021 data samples (pixels) were hand selected and separated into disturbed (162) and undisturbed (859) categories. 70 % of data points from each category were separated to use as training data for developing a prediction model. The remaining 30% of data points from each category would be used to test the model. Non-parametric Random Forest (RF) modeling (Breiman, 2001; Liaw & Wiener, 2002) was chosen to avoid possible violations of normal linear regression (Schwantes et al., 2016). The first model developed was designed to separate areas which showed evidence of infestation from areas which had not been attacked. In RF, the ‘rf.modelSel’ function of the ‘rfUtilities’ was used to select the best indices for the model. The ‘rf.modelSel’ function creates scaled variable importance scores for each variable, and tests models that include variables which show correlation scores above various important thresholds. In RF, all explanatory indices were used to determine which were the most important indices for accurately determining percentage of disturbance by correlation methods. The indices that showed less than 70% correlation within response variable were chosen to build the model because higher correlations increased the forest error rate in RF analyses. The model that explains the most variations in the response variables using the fewest predictor variables is the most accurate (note – the response variable here denotes the presence or absence or percentage of infestation, while the predictor variables are NBR, B5/B4, TCG, band4, band 5, MeanB5/B4, DiffB5/B4, DEM, and AnomB5/B4) (Table 4.3).

Temporal anomalies represent the difference between any spectral index value at a given time and the multi-temporal, pre-disturbance calculated mean of that spectral index. Time series Landsat data for forested pixels were selected from years preceding bark beetle disturbance and were then used to establish a baseline from which to subsequently distinguish spectral deviations to detect forest disturbance. The magnitude of change between the multi-temporal means and the value at a given time was used to discriminate between disturbed and undisturbed forest pixels. Therefore, this model is designed to separate areas showing any signs of infestation from undisturbed areas. The regression model predicts percent mortality. To give various mortality severities (0-37%) equal weight, pixels were selected at random, and percentage values for disturbed trees (red and gray stage) were manually counted in each pixel represented in NAIP and Landsat photos. These estimates were then corroborated with NBR and NDVI indices.

To successfully apply the Random Forest algorithm (RF), it is essential to configure two key parameters: the number of trees (ntree) and the number of features considered in each split (mtry). Multiple investigations have indicated that favorable results can be attained by utilizing the default values (Duro et al., 2012; Liaw & Wiener, 2002). Liaw and Wiener (2002) discussed that the inclusion of a large number of trees in a model will yield a consistent estimation of variable importance. Furthermore, Breiman, (2001) noted that employing too many trees may not improve results, although it does not adversely impact the model. In addition, (von Holt et al., 2022; Zhao et al., 2015) asserted that even using a minimum number of trees (ntree = 200), Random Forest (RF) was still able to attain precise outcomes. Several authors discussed employing the default value mtry = $\sqrt{p}$ for the mtry parameter, where p represents the number of predictor variables (Duro et al., 2012; Noi and Kappas, 20117). In this study, an investigation was conducted to determine the most accurate RF model for classification (Zhao et al., 2015). To do this, a series of analyses were performed, testing, and evaluating various values for two parameters: ntree, which included values of 100, 200, 500, and 600; and mtry, which encompassed values ranging from 1 to 10 with an increment of 1.

Table 4.3 Predictor variables for binary and regression models

Band or index	Binary Model	Regression Model
NBR	yes	Yes
B5/B4	yes	Yes
TCG	yes	Yes
Mean B5/B4	No	Yes
Band5	No	Yes
DEM	yes	Yes
Band4	yes	No
Anomaly B5/B4	yes	No
Diff B5/B4	yes	No

4.3 Results

Using 1021 data samples (pixels) hand selected and separated into disturbed (162) and undisturbed (859) categories, two different models were developed, the binary (presence-absence) model, and the regression (percentage of infestation) model. Each of the two models used a different set of vegetation indices for calculation selected in R. Each model was then analyzed by a variety of statistical methods to assess their accuracy.

4.3.1 Binary classification model

Data from two years in the time series (see Chapter 3) were selected representing pre-infestation versus maximum infestation periods. Classification was done using R. R is a programming toolbox and open-source software environment primarily designed for statistical computing and graphics. It provides a wide range of statistical and graphical techniques for data analysis, data visualization, and statistical modeling (Wickham & Grolemund, 2017). In this study, R was used to analyze the two classes of data with the purpose of separating the study area into disturbed and undisturbed areas, and to compare the predicted pixels with the actual data points (Refer Appendix D – R code).

Table 4.4 Relative size distribution of undisturbed versus disturbed areas into 2005 vs 2009

Year	Undisturbed (area km^2)	Disturbed (area km^2)	Accuracy
2005	488.45	208.12	96%
2009	288.72	272.49	97%

Accuracy for this calculation was based on confusion matrix analysis.[2] A confusion matrix is a tabular representation of the performance of a classification system showing the counts of true positive, true negative, false positive and false negative predictions for each class in the classification (Sammut and Webb, 2011). It represents a comparison between reference data and predictions. Reference data represent actual pixels that can be verified as representing undisturbed (0) or disturbed (1) data points. To maximize classification accuracy, four types of explanatory variables proved to be the most reliable, TCB, TCG, B2, and NBR. TCB values were used to separate out barren soil, fire-damaged zones, and clear-cut zones. TCG combined with B2 data were used to separate out grassy areas, shrubs, and deciduous trees such as aspens. Thus, the maximum-corrected binary model gives a simplified view of the presence or absence of infestations.

In these tables, the diagonal elements (top left to bottom right) represent the correctly classified instances (true positives and true negatives). The opposite-diagonal elements represent the misclassified instances (false positive and false negative) (Zerrouki and Bouchaffra 2014).

[2] Sammut, C., & Webb, G. I. (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.

Tables 4.5 Confusion matrices for 2005 and 2009

RF 2005		**Reference**		
		0	**1**	**total**
Prediction	**0**	251	7	258
	1	4	37	41
	total	255	44	299
Overall accuracy = **0.96** and Kappa = **0.85**				

RF 2009		**Reference**		
		0	**1**	**total**
Prediction	**0**	253	4	257
	1	2	40	42
	total	255	44	299
Overall accuracy = **0.97** and Kappa = **0.92**				

The accuracy of this classification system can be calculated where:

1) Accuracy = (sum of diagonal elements) / (total number of samples)

Another method of measuring statistical agreement between two different classifications is called the Kappa coefficient. The Kappa coefficient takes into account both the agreement that could be expected by chance and be actual agreement observed (Cohen,1960). It essentially tells you how much better the agreement is compared to what would be expected purely by chance. The formula for calculating the Kappa coefficient is:

2) K = (Pa -Pe) / (1 – Pe)

Where Pa is overall accuracy and Pe is the probability of agreement by chance (Stehman 1996).[3] The Kappa statistic serves as a more rigorous estimate of accuracy than the overall accuracy because it penalizes for agreement that might be expected to occur by chance (i.e., it decreases the overall accuracy

[3] Stehman, S. (1996). Estimating the kappa coefficient and its variance under stratified random sampling. *Photogrammetric engineering and remote sensing*, *62*(4), 401-407.

by an amount proportional to the probability of the chance). The Kappa coefficient can range from -1 to 1. A value of 1 indicates perfect agreement. A value of 0 indicates agreement that's no better than random chance. A value less than 0 indicates that the agreement is worse than would be expected by chance. The Kappa value for the year 2005 is 0.84; the Kappa value for 2009 is 0.92.

Output results of the binary model were represented in visual form in Figure 4.2 comparing the distribution of infested vs. non-infested sections of the study area for the years 2005 vs. 2009.

It is clear from this diagram that infestation has spread significantly from 2005 to 2009, which matches the trend seen in the time series analysis discussed in chapter 3. It appears that after looking at topographic information that most of the spread has been on south- and west-facing slopes. Although there is some debate about this, topographic factors including aspect, slopes, and elevation may also have major effects on local variability in temperature and moisture (Oke, 2002), which in turn affect a tree's ability to muster its defenses.

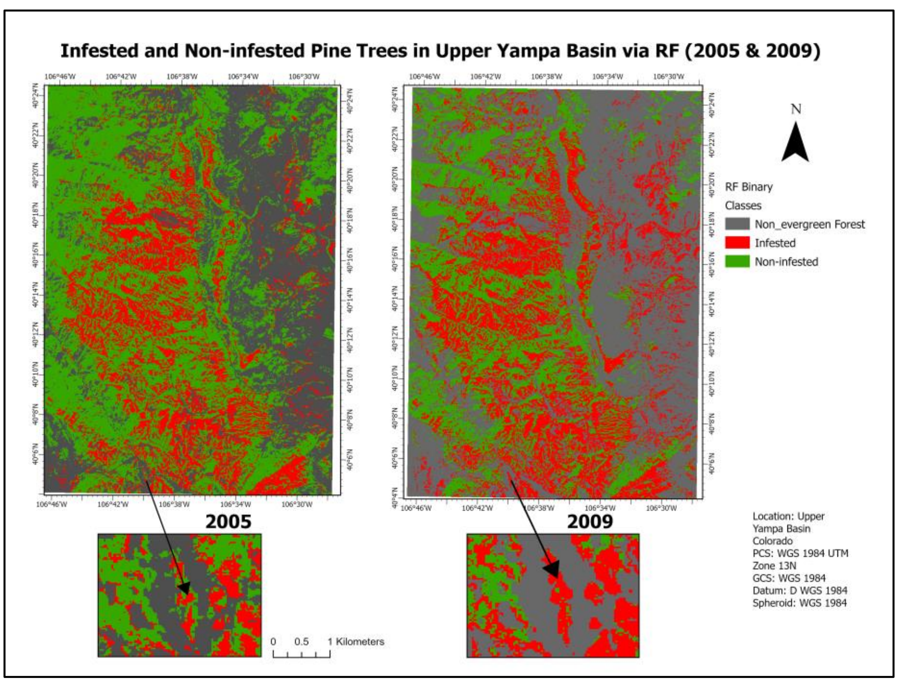


Figure 4.2 Binary classification

Binary classification model for 2005 and 2009 which shows spread of infestation.

(Nelson et al., 2007) reported that warmer and drier slopes on southerly aspects at low elevation might decrease tree defenses against beetle attack while colder and wetter slopes on northerly aspect at high elevation increase tree defenses.

3.3.2 Regression analysis

To predict not just the presence or absence of infestation, but rather the percentage of damage in any given area, a regression analysis is the most useful method.[4] Regression analysis is a statistical method used for modeling and analyzing the relationships between variables. It's commonly used for predicting a continuous target variable (also called the dependent variable) based on one or more predictor variables (also called independent variables). The goal of regression is to find the best-fitting line or curve that describes the relationship between the variables (Freedman, 2009). The Random Forest program uses a regression method that combines multiple decision trees to improve predictive accuracy and control overfitting. Decision trees can be used to partition the data into segments, and the average or median value of the target variable within each segment is used as the predicted value.

The “rf.modelSel” function in Random Forest creates scaled variable importance scores for each variable and selects the model that explains the most variation in response variable with the fewest predictor variables. In this study, originally ten predictor variables were chosen and correlated with one another, and, in particular, with NBR. In the next round of regression analysis, variables which had correlated too strongly with response variable were removed from the model because too many variables with high correlations will result in Random Forest reducing the number of decision trees and limiting the accuracy of the results (Thanh Noi & Kappas, 2017). Cross-validation in the “caret” package of R was used to assess RF model performance in predicting new predictor variables.

A final RF model predicting percent tree mortality was applied to the entire data sets for 2005 and 2009. Predictor variables that were used in this model were: NBR, B5/B4, TCG, MeanB5/B4, Band5, and

[4] Freedman, D. A. (2009). *Statistical models: theory and practice*. cambridge university press.

DEM. In the model, 500 decision trees used two data points at a time for analysis of the entire data set. The result shows a % var explained of 86.3, which means that the model was able to explain 86.3% of the total variability of the data set, where the total variability in the data set is the sum of the squared differences between each data point and the overall mean of the data set. Higher values of % var indicate that the model is capturing the underlying patterns in the data.

A second RF model was run on just the testing data for 2009 (Figure 4.3). Predictor variables were the same, however, the 500 decision trees used 4 mtry data points at a time for each run. A lower RMSE indicates that the model's predictions are closer to the actual values, which means the model is performing better. In this case, an R-squared value of 87.8% suggests that the model explains approximately 87.8% of the variability in the target variable (Figure 4.3), and the low RMSE of 0.1425 agrees that the model is performing very well. This indicates that the model is capturing a substantial portion of the underlying patterns in the data.

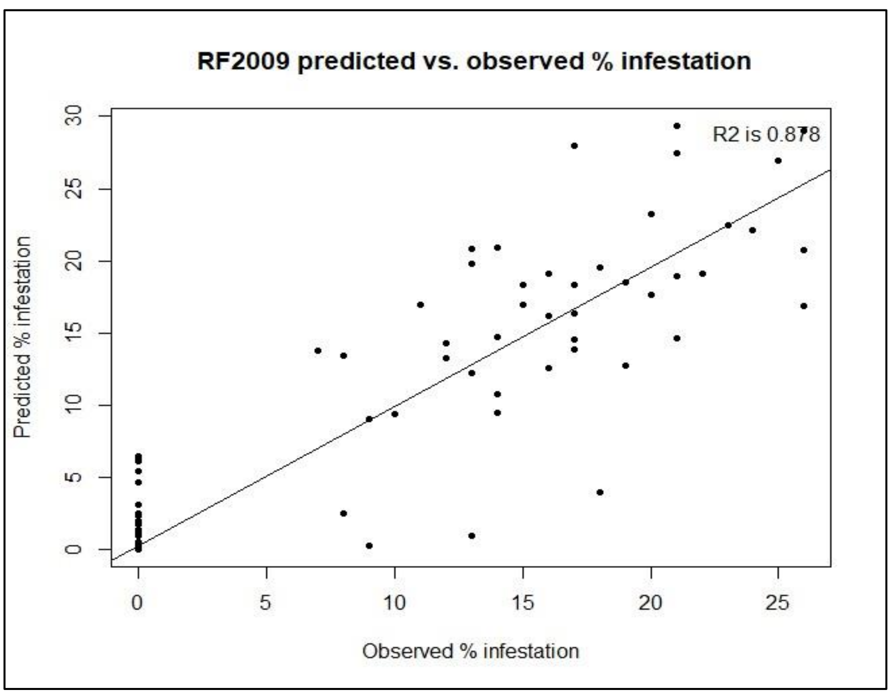


Figure 4.3 Scatter plot between predicted and observed percent of infestation

Note that the R^2 value for this scatter plot is 0.878 and RMSE is 0.1425.

R^2 is a statistical measure used in regression analysis to assess the goodness of fit of a regression model to the data. R^2 is a value between 0 and 1, where: 0 indicates that the model does not explain any of the variance in the dependent variable, meaning it is a poor fit; 1 indicates that the model perfectly explains all of the variance in the dependent variable, meaning it is an excellent fit.

The output from the RF training data regression model for 2009 was used to create a map view of the percentage of tree mortality (Figure 4.4). The map on the left side shows the distribution and percentage of damage across the entire study area. The small map on the right side is a close-up image showing the percentage damage map of a small area.

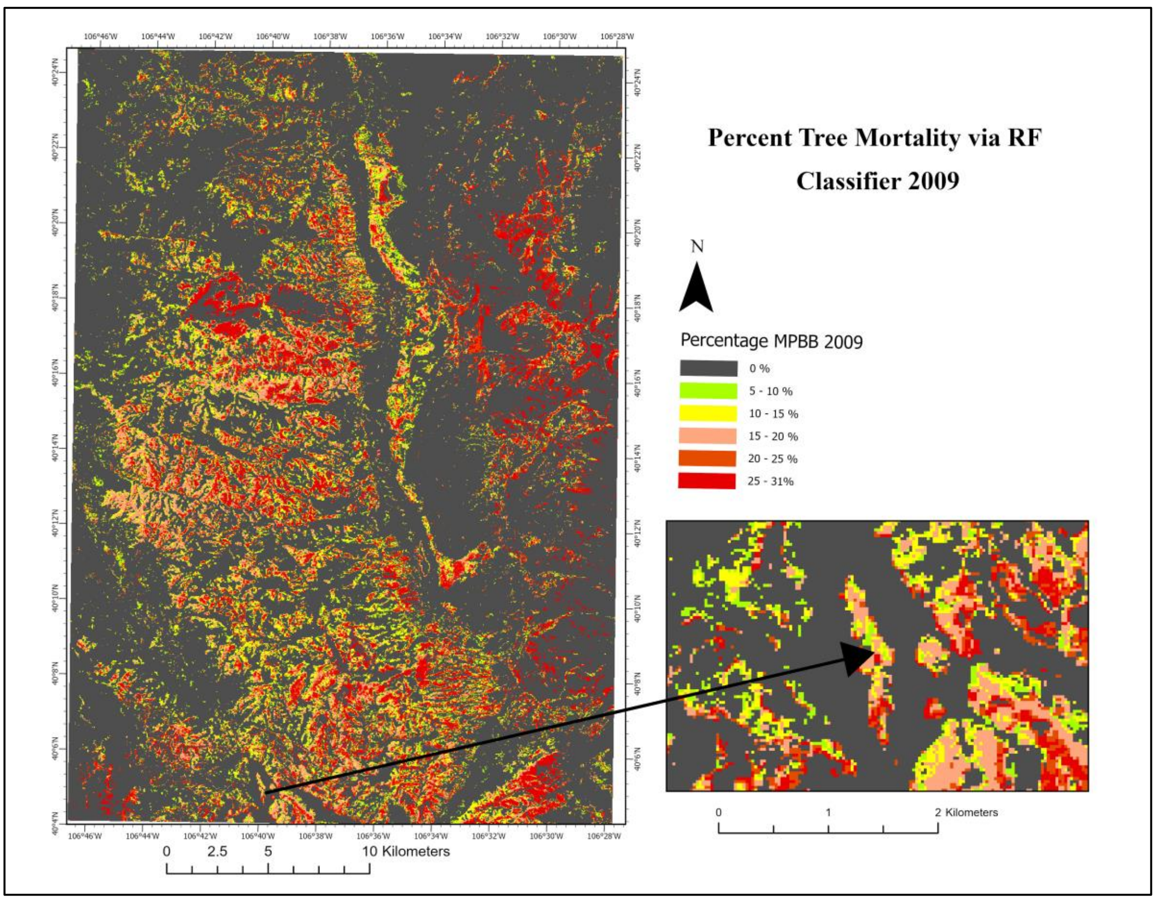


Figure 4.4 Percentage tree mortality

Percentage of tree mortality based on regression model in 2009 and comparison with NAIP image.

4.4 Discussion

The purpose of this study was to develop a method for determining the severity of insect-caused tree mortality in conifer forest in the Upper Yampa Basin and to track the spread of insect infestation with time. These goals were accomplished through the analysis of various Landsat spectral bands for the years 2005 and 2009. Non-forest, undisturbed and disturbed forest areas were separated with overall accuracies of 96% and 97%, for the two years respectively. NBR proved to be the most useful index for tracking and calculating the extent of insect damage. Other useful indices included B5/B4 and its anomalies, and TCG. B5/B4 appeared to be sensitive to changes to forest structure, rather than directly sensing changes to vegetation physiology. TCG is an indicator of moisture content in the forest canopy, and thus, a good indicator of the beginning of a die-off.

The main benefit of using RF methodology is that it has the potential to be applied across large spatial extents. Many studies (e.g., Bright et al., 2020) have shown that techniques similar to the one developed in this study can be used to separate undisturbed from disturbed forest in a variety of settings. This suggests that this approach can be applied across a large region, such as an entire wilderness area, national forest, state, or Forest Service Region. Caution should be used when applying these models outside areas where those models were calibrated, as accuracy might decrease in new areas (Cohen, et al., 2017, and 2018). The use of several calibration sites distributed across these larger areas, as well as applying these algorithms and models only within similar forest types, would help to minimize the danger of inappropriate extrapolation outside calibration areas.

Researchers such as Meddens, et al. (2013) have cautioned about the usefulness of using Landsat data for early stages of infestation or slowly progressing disturbances. Meddens indicated that Landsat data was useful for detecting severe insect infestations (i.e., killing >~ 25% of trees in the canopy within a given pixel), whereas finer spatial resolution data might be necessary to detect dispersed or single tree mortality across forested landscapes. He further indicated that the Landsat data cannot detect slowly progressing disturbances within a given pixel, (i.e., outbreaks that result in fewer than 25% infested trees at any one time, with high accuracy). Similarly, Bright et al. (2020) indicated that comparison of Landsat-drive mortality maps with Insect Detection Surveys (IDS) showed some discrepancies, especially in areas of low tree mortality. However, IDS data are a fundamentally different, vector-based representation of disturbance, so differences are expected. IDS polygons are large and highly generalized. To overcome these issues, in this study, comparison with high-resolution aerial images (NAIP) effectively served to bridge the scaling gap between individual tree mortality (ground data) and the 30-m resolution of Landsat imagery.

The combination of the RF models with NAIP images also proved useful in identifying other forest disturbances. Although the predominant disturbance in the study area was mountain bark pine beetle-associated tree mortality, other types of disturbances did exist within the study area. Tasseled-Cap Brightness (TCB) data was used to suggest areas where other disturbances such as clear-cut and fires, and the presence of these areas was confirmed through NAIP image examination so that they could be removed from areal beetle-kill calculation (Meigs et al., 2011). For all trees studied, mortality was much greater in climatically hotter and drier areas, i.e., lower elevations and latitude. For ponderosa pines, mortality was significantly lower in treated stands, whereas mortality was highest in stands with a high density of medium- to large-size trees, especially in areas with lower precipitation, suggesting that abundance of large closely spaced host trees for insect attack was an important determinant of pine mortality (Restaino et al., 2019).

Koll et al., (2019) showed that pine forests under water stressed conditions are less able to use their normal defense systems against bark beetle attacks. Bark beetle attacks and biota carried by beetles typically induce a general increase in concentration of phloem-based mono and sesqui-terpenes, whereas water stress did not. Bark beetle attacks also induced an increase in resin flow for unstressed trees but not for water stressed. Mortality was highest for beetle-attacked and water-stressed trees (Kolb et al., 2019).

4.5 Conclusion

The purpose of this study is to track and map the spread of MPBB in the Upper Yampa Basin. Several vegetation indices were calculated from Landsat TM data and put into the Random Forest classifier algorithm to determine first, the presence or absence of MPBB infestations, and second, to predict the percentage of attacked trees in each infested area. RF was successful in performing these analyses with high accuracy (96% - 2005 and 97% - 2009) that permitted the construction of regional maps showing the distribution of infested areas and changes in that distribution with time. The study showed that RF can be used to achieve the objectives of this study and provide the sort of information needed by forest managers to fight these infestations.

Chapter 5 - Analysis of remote sensing data using Support Vector Machine (SVM) statistical analysis to track tree mortality over time

The goal of this chapter is to investigate dynamic forest destruction patterns related to MPBB infestation in the Upper Yampa Basin in Colorado by using another remote sensing classifier, Support Vector Machine (SVM) (Noi and Kappas, 2017). This study will address the following research proposals:

1 – Determine whether analysis of remote sensing data via SVM techniques can be used to assist in tracking tree mortality over time.

2 – Define the most important predictor variables for binary and regression statistical analysis.

3 – Use the output from the SVM model to create reliable maps showing the distribution of infested areas geographically with time.

5.1 Study site

For more information see (Chapter 1 Figure 1.1).

5.2 Material and Methodology

5.2.1 Support Vector Machine (SVM)

The major asset of the Support Vector Machine (SVM) algorithm is that it is a powerful supervised statistical classification method that can handle multiple-band imagery having high resolution. In addition, it can manage large quantities of segmented satellite data with ease when compared to other classification techniques where attribute table management becomes difficult (Noi and Kappa, 2017). The SVM classifier program works through the principle of generation of a hyperplane that separates linearly separable classes (Szuster et al., 2011). The optimal separating hyperplane minimizes classification errors by maximizing the distance between itself and the planes representing the two classes. Note that the hyperplane is calculated in feature space, not image space (define). Support vectors, which are the data points nearest the hyperplane, are also the most important feature of the training samples.

SVMs were used originally for solving linear classification problems, however, they can also be used for non-linear decision functions via the kernel function (Pilario et al., 2019; Wang et al., 2017). The kernel function can be very important for SVM analysis because a kernel can be used to transform a non-linear problem into a linear problem and reduce the complexity of the mapping by projecting pixel vectors into a higher dimensional space where one can estimate the maximum hyperplane in the new space in order to improve linear separability of data (Boser et al., 1992; Wang et al., 2017). There are several types of kernel functions including: linear, polynomial, radial basis, and sigmoid.

The radial basis function (RBF) kernel is the most commonly used and shows the most reliable performance. There are two important parameters that need to be set when applying the SVM classifier with RBF: the regularization parameter (C) and kernel width parameter Υ (Ballanti et al., 2016). The C parameter decides the size of misclassification that allowed for non-separable training data. The kernel with parameter of the class dividing large value of C may lead to over fitting model, whereas increasing Υ value will affect the shape of the class-dividing hyperplane, which may affect the classification accuracy results (Li et al., 2016). The SVM has several advantages and disadvantages. One major advantage of SVM over other classification algorithms is that it is less affected by noise, correlated bands, and an unbalanced number or size of training data within each class (Noi and Kappas, 2017). SVM has been chosen because of its strengths such as: easily manages large feature space; insensitive to accuracy as a function of the number of spectral bands or dimensions investigated (Hughes effect); works with small training dataset; and does not overfit (can control misclassification) (Gómez et al., 2016). The main weaknesses of SVM include: needs parameters (regularization and kernel); poor performance with small feature space; computationally intense; and designed as binary approach, although variations exist (Gomez et al.,2016; Li et al., 2016).

5.2.2 Study plan

A well-structured study plan for this chapter consists of three essential steps: data collection, data preparation, and data preprocessing Figure 5.1. This study follows the steps listed in flowchart (Figure

5.1). The first step requires the collection of Digital Elevation Model (DEM) data and 6 bands of data from Landsat 4-5 Thematic Mapper (TM), 7 bands of data from Landsat 7 Enhanced Thematic Mapper Plus (ETM+) (Chapter 4 Table 4.1). The next steps are developing the annual image composites, preprocessing, converting numerical pixel values to surface reflectance numbers, and digitizing images by using GEE (Step 2). DEM data is extracted and used to calculate slope, and aspect. These data are used to examine relationships between the orientation of a certain tract and the degree of infestation to see if infestations are preferentially developed on slopes facing a certain direction. A subset of this data will be collected to act as training data for the analysis algorithm using ArcGIS Pro 3.1.0.

The next steps involve the processing of the data to calculate annual spectral indices (Chapter 4, Table 4.2), and segmentation of yearly time series by Google Earth Engine (GEE) and ArcGIS Pro (Chapter 3, Figure 3.4). At this point, training data selected by ArcGIS is used as input data to SVM to provide definitions for classifying various features such as trees, grass, soil (feature extraction). Data points defining trees will be separated into infested vs. non-infested categories. The accuracy of this separation will be assessed by comparison with field data and NAIP images collected for validation. Collection of field data involves checking for presence or absence of infested trees and is documented with photographs, and field notes. These observations are also compared with NAIP images for the study area. Final maps of the change analysis are then prepared.

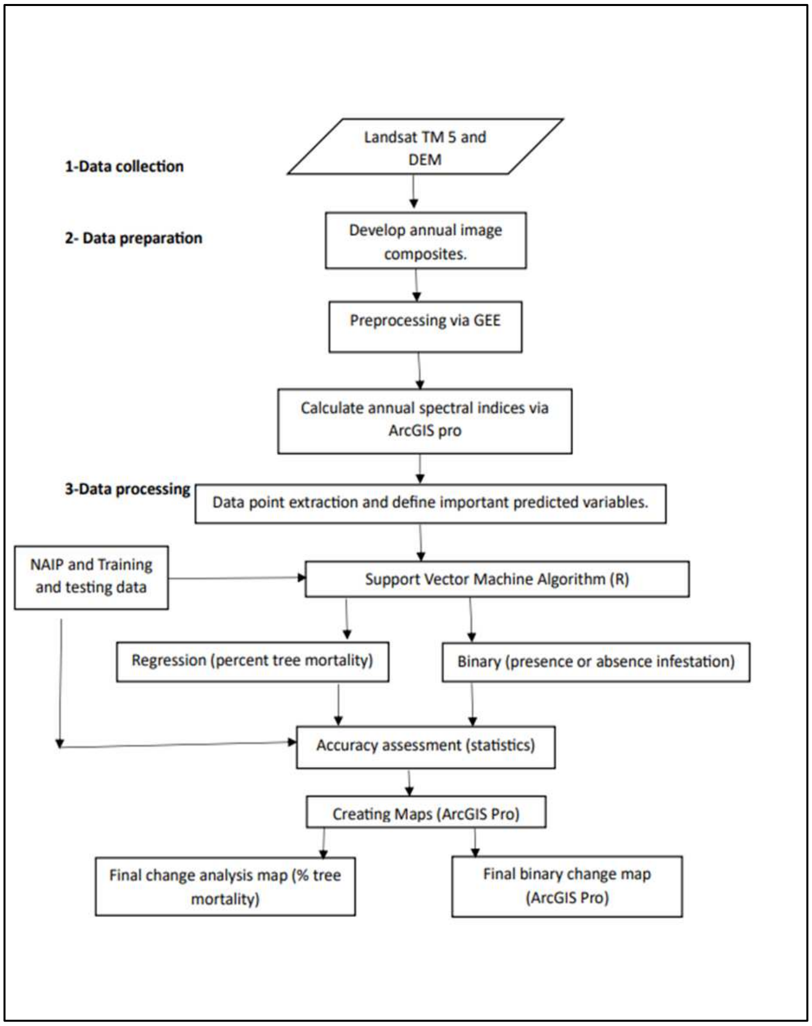


Figure 5.1 Flowchart

5.2.3 Steps in statistical analysis to create regression and binary models

5.2.3.1 Regression model (percentage)

To use SVM to create a regression model that estimates percentage of infestation, predictor variables (22 vegetation indices) were selected to evaluate which ones were most important in controlling tree mortality. The first step of analysis (Figure 5.1) was to manually examine Landsat data and develop a database with measured values for factors such as reflectance as well as information from 7 other wavelength bands that were used to calculate the 22 vegetation indices (predictor variables). The next step is correlating all the variables with one another, removing highly correlated predictor variables that provide the same type of information. In R, the ‘rf.modelSel’ function of the ‘rfUtilities’ was used to select the best indices for the model. For a pair of predictor variables with a correlation coefficient higher than 0.7, the variable with a lower correlation to the response variable (percentage of infestation) was dropped. In this study, the R package (e1071) used was the LIBSVM library by Chang and Lin (2011). 6 predictor variables were chosen to build the model: Normalized Burn Ratio (NBR), Band 5/ Band 4 (B5/B4), Tasseled-Cap Greenness (TCG), Mean B5/B4, Band 5, and Digital Elevation Model (DEM) (Bright et al., 2020, and Meng et al 2021) (refer to Table 4.2, Chapter 4 for more information on calculations). In the next step, the caret package in R (Kuhn, 2008) was employed to randomly select 70% of the data points (722) to use as training data. Then the SVM classification algorithm was trained on the 722 data points to predict a first pass model of percentage of infestation. Final accuracy assessment was evaluated using the remaining 30 % (299 data points) of the original data set. Following this, the model was tuned by using the three parameters in R that can give a measure of the accuracy of any given predicted value. These parameters include: Epsilon, which defines the acceptable margin of error for training data points that are on or within the margin boundary; gamma (γ) which controls the shape of the Radial Basis Function (RBF) kernel function that defines a similarity measure between data points in a higher-dimensional space; and Cost (C) which determines the trade-off between maximizing the margin and minimizing the classification error on the training data. A small gamma will create a broader and smoother decision boundary, making the model generalize more, whereas a larger gamma will create a more complex and wiggly decision boundary, leading to the model fitting the training data more closely (Hsu et al., 2010). A smaller C will result in a larger margin but may allow some misclassification. On the other hand, a larger C will lead to a narrower margin and fewer training errors (Li et al., 2016). The SVM algorithm with BRF radial kernel was applied and the following values of tuning parameters were tested: seven values of C (20,21, 22,23,24,25,26,27) and ten values of epsilon (0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1) and 10 values of gamma (0.1 to 10). The optimal regression model used the optimal tuning parameters: 2009- Epsilon = 0.1, Gamma = 0.1, and Cost = 4. Output of this regression model was converted into a map image of infestation percentage using the raster functions in R.

5.2.3.2 Binary model (classification)

The simple purpose of the binary model is to classify the dataset into infested vs. non-infested areas. Thus, the response value of this model is therefore the presence or absence of infestation. For selection of the important predictor variables, the same procedure as was used in the regression analysis was employed to select the most important predictor variables, which for this model were: NBR, B5/B4, TCG, DEM, band4, AnomB5/B4, and DiffB5/B4. The optimal binary model used the optimal tuning parameters: 2005- Epsilon = 0, Gamma = 0.1, and Cost = 4; 2009- Epsilon = 0, Gamma = 0.1, and Cost = 8. The same software packages in R were used to generate the final output map which showed the distribution of infestations.

The main problem with the resulting map is that it includes all possible categories including other categories that do not represent infestations, and thus they may be incorrectly mapped. For example, soil, grass, shrubs, and deciduous trees may also be mapped as infested areas. To remedy this problem, the output data was transferred into the ArcGIS Pro (3.1.0) to extract these misclassified categories by using B2, TCG, and TCB. For example, B2 was used to separate forest from non-forest. TCG was used to separate grass, and TCB was used to separate out barren soil, burned zones, or clear-cut areas. The final output map simply showed the distribution of infested vs. non-infested trees in the Upper Yampa Basin. Two maps were created comparing this distribution for the years 2005 and 2009. See discussion below associated with Figure 5.3.

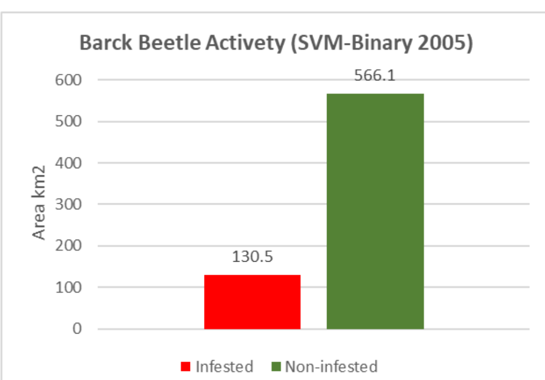
To check the accuracy of the models, performance was assessed by use of a 10-fold cross-validation method repeated 5 times. Thus, the testing data was divided into 50 different subsets. The normalized root mean squared error (RMSE) and R-squared (R^2) were used as performance metrics for the regression model. For the binary models, the overall accuracy and Kappa coefficients were calculated from the normal equations.

5.3 Results

Binary and regression analysis of training data allowed selection of the most important and most reliable predictor variables to estimate the distribution and severity of tree mortality due to mountain pine bark beetle attack. We compare the results of the two techniques separately and then compare the types of information and accuracy of the results.

5.3.1 Binary model (classification)

The primary focus of the binary model is to quantitatively separate infested vs. non-infested sections of the study area. As Figure 5.2 shows, the ratio of infested to non-infested areas has markedly increased in 2009 relative to 2005. This aligns with the trend seen in the time series analysis (Chapter 3). When the data used to create Figure 5.2 is plotted in map form (Figure 5.3), the change in distribution and extent of infested zones is readily visible. The smaller detailed views compare the binary view vs. the NAIP image of the same area to validate the results. Note that is impossible to identify the change in degree of infestation simply by looking at the NAIP view (see Chapter 4).



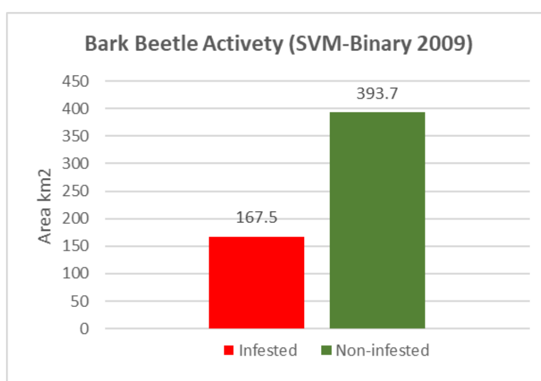


Figure 5.2 MPBB activity comparing the years 2005 and 2009.

Areal extent of infestation comparing 2005 vs. 2009 as calculated by SVM binary model.

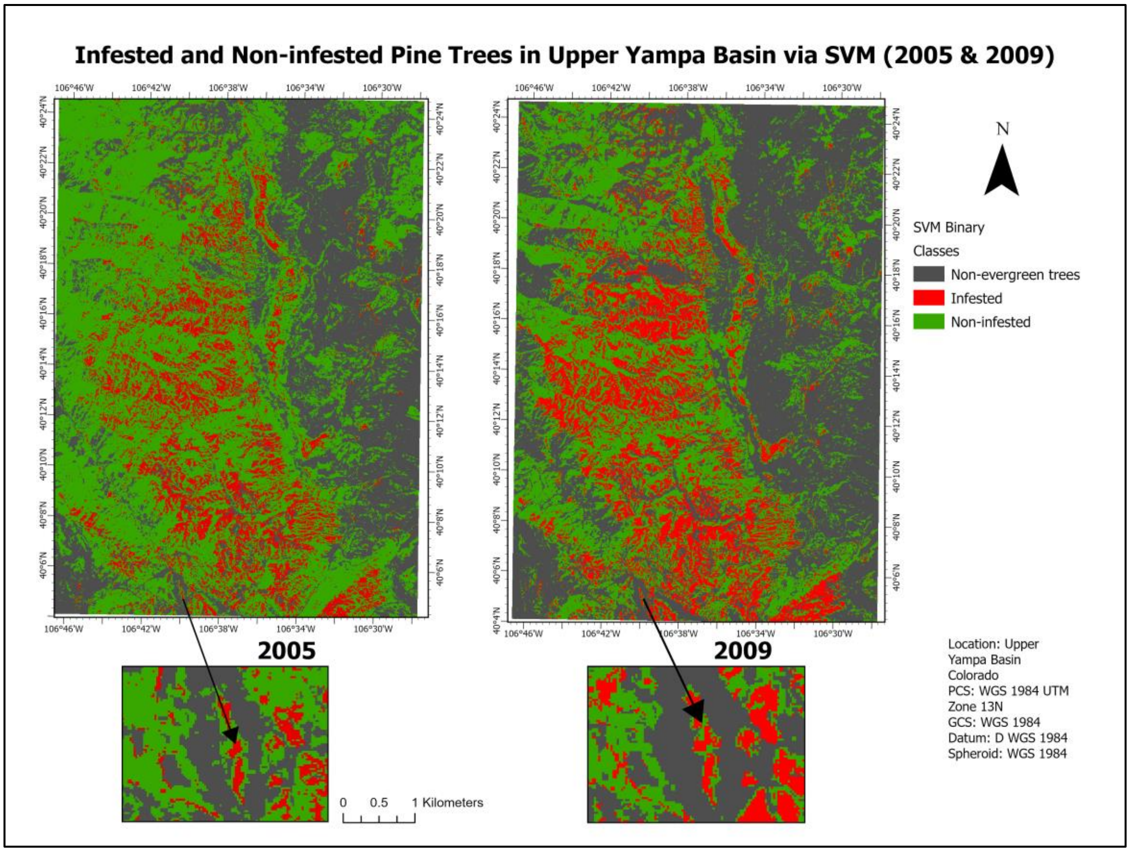


Figure 5.3 Binary classification model showing the spread of MPBB infestation from 2005 to 2009.

Table 5.1 Confusion Matrices for SVM 2005 and 2009

SVM 2005		**Reference**		
		0	**1**	**total**
Prediction	**0**	248	6	254
	1	7	38	45
	total	255	44	299
Overall accuracy = **0.95** and Kappa = **0.83**				

SVM 2009		**Reference**		
		0	**1**	**total**
Prediction	**0**	253	5	258
	1	2	39	41
	total	255	44	299
Overall accuracy = **0.97** and Kappa = **0.9**				

The confusion matrix (Table 5.1) is a way of measuring how many data points from the testing dataset were classified correctly vs. how many were incorrectly classified. The data points in the 0 column and row represent non-infested samples, whereas the data points in the 1 column and row represent infested samples. To determine whether the classification identified by model was correct or not, each data point was compared with the classification determined for each pixel in the NAIP images. Table (5.1-a) shows that for 2005, 248 non-infested samples were correctly identified, whereas only 6 samples were misclassified as non-infested when they were actually infested, and 7 infested samples were misclassified as non-infested, whereas 38 infested samples were correctly identified. Similarly, for 2009, 253 non-infested samples were correctly identified, whereas only 5 samples were misclassified as non-infested when they were actually infested, and 2 infested samples were misclassified as non-infested, while 39 infested samples were correctly identified. The accuracies for the binary 2005 and 2009 classifications are 0.95 and 0.97, respectively. As mentioned above, the second measure of accuracy of the binary model is the Kappa value. The Kappa value for 2005 is 0.83 and 2009 is 0.9.

5.3.2 Regression Model (percentage tree mortality)

The purpose of doing a regression analysis within the SVM model is to estimate the percentage of infestation at any given point. Output data from SVM was input into ArcGIS Pro 3.1.0 to visualize the percentage data represented by each pixel (Table 5.2).

Table 5.2 Numeric output from ArcGIS Pro 3.1.0

Cumulative area affected by infestation calculated from the percentage of infestation times the number of pixels represented by that percentage times the area of each pixel

ID	Percentage of infested	Number of pixels	Area (km^2)	Area (%)
1	0%	1208055	861.2	84.2
2	5 - 10 %	38749	27.6	2.7
3	10 - 15%	67838	48.4	4.7
4	15 - 20%	61739	44.0	4.3
5	20 - 25 %	40903	29.2	2.9
6	25 - 31 %	17864	12.7	1.2

.

Table 5.2 shows the numeric output from ArcGIS Pro 3.1.0 which has counted the number of pixels within each class of infestation percentage. The pixel count is converted into an area measure by multiplying the count * the area represented by each pixel (712.89 m2).

Figure 5.4 shows the relative proportion of each percentage range of tree mortality as function of area.

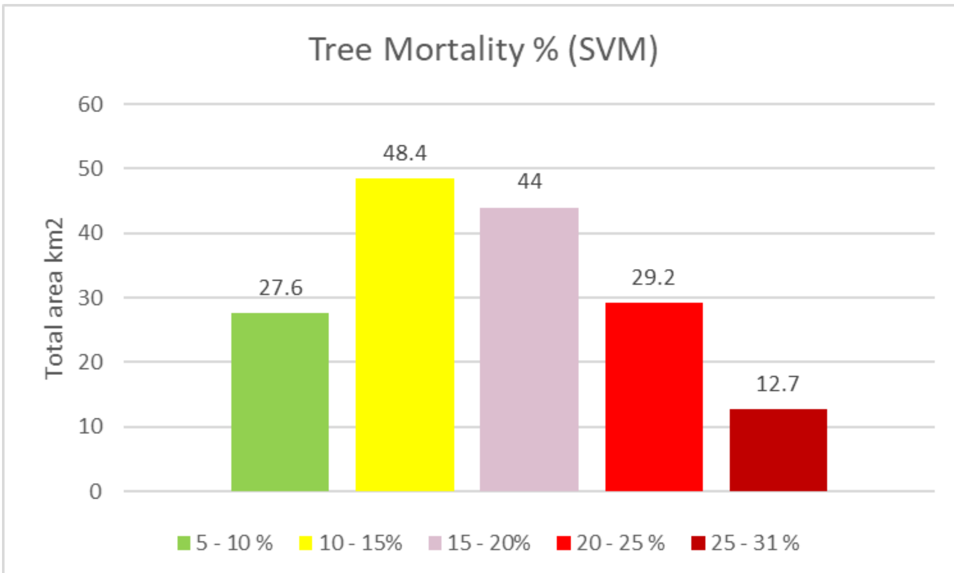


Figure 5.4 Tree mortality areal percentage

Chart of the relative area represented by each tree mortality percentage class.

Since each pixel data point is matched with its position coordinates, ArcGIS Pro can convert the areal percentage data into a map that shows the geographic distribution of each percentage class.

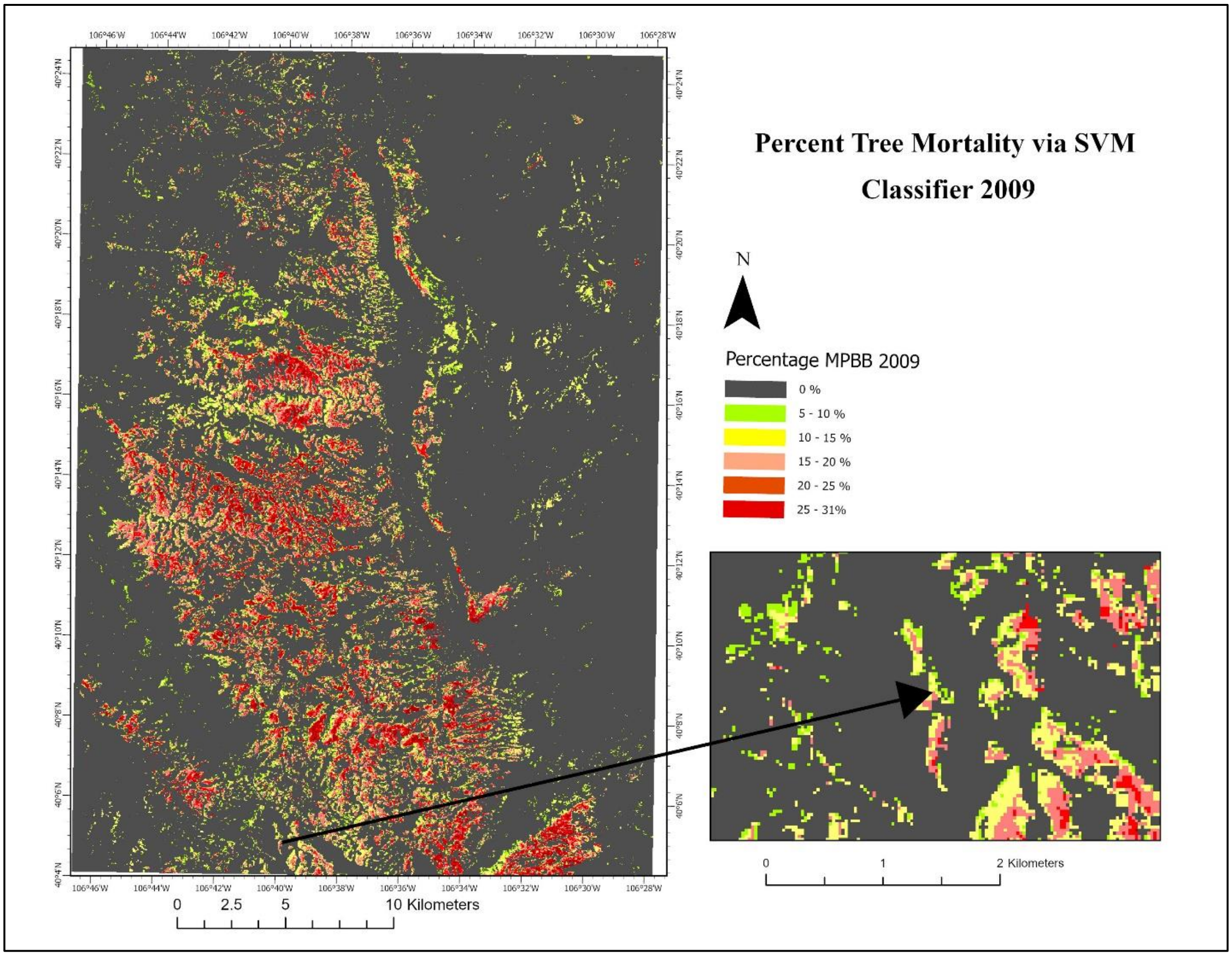


Figure 5.5 Percent tree mortality via SVM classifier, Regression model

Percentage of tree mortality as calculated vis SVM classifier for 2009.

Figure 5.5-a illustrates the distribution of the degree of infestation across the study area in 2009. It appears that after looking at topographic information that most of the spread has been on south- and west-facing slopes. Although there is some debate about this, topographic factors including aspect, slopes, and elevation may also have major effects on local variability in temperature and moisture (Oke, 2002), which in turn affect a tree's ability to muster its defenses. (Nelson et al., 2007) reported that warmer and drier slopes on southerly aspects at low elevation might decrease tree defenses against beetle attack while colder and wetter slopes on northerly aspect at high elevation increase tree defenses.

The inset image shows a higher resolution image of the distribution over a small area from the model percentage map. Accuracy for image 5.5 was measured by calculating R2 = 0.86 and RMSE is 0.113. In this case, an R-squared value of 86% suggests that the model explains approximately 86% of the variability in the target variable (Figure 5.6), and the low RMSE of 0.113 agrees that the model is performing very well and indicates that the model is capturing a substantial portion of the underlying patterns in the data.

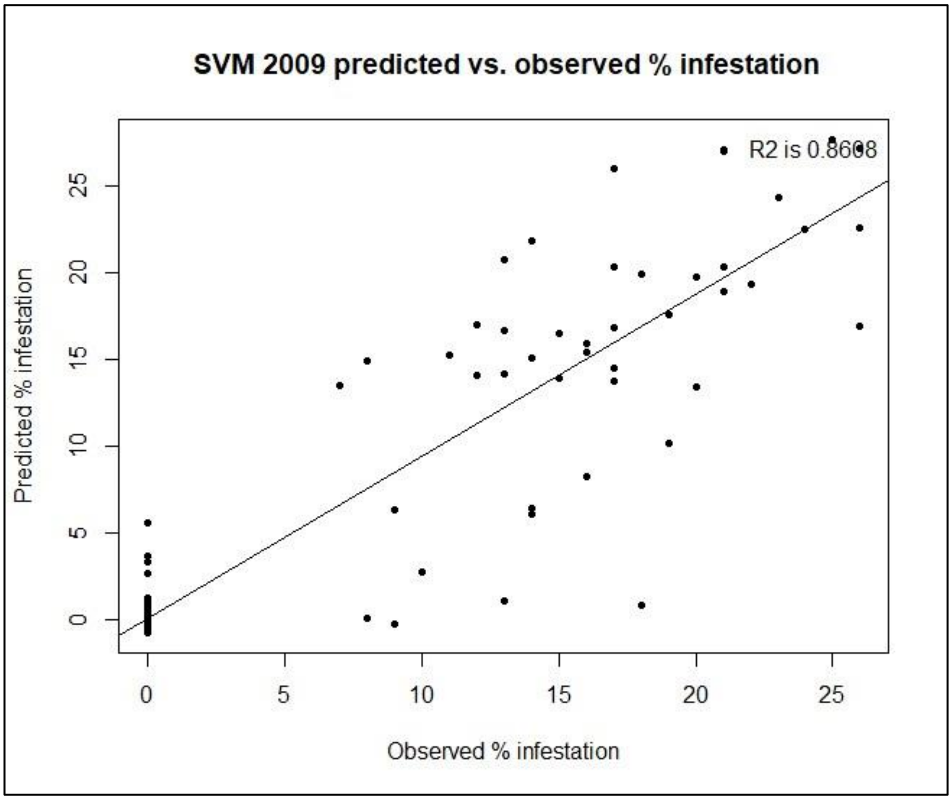


Figure 5.6 SVM regression model performance

5.4 Discussion

The purpose of the SVM technique is to classify Landsat data based on key attributes and calculate the percentage for each class. Two different statistical approaches were used to analyze the classes defined by the SVM technique. The binary approach was used to separate infested from non-infested data points, whereas the regression approach provided information of the relative percentage of infestation per pixel. The binary approach can be used to map the presence or absence of pine bark beetles across a given area, whereas the regression approach can illustrate the relative severity of infestation at any given point. If these maps are generated over a sequence of years, the distribution and direction of movement can be tracked. This is the kind of information that is essential to managing the war on mountain pine beetle infestations. Forest managers can use this information to plan their approach to controlling and minimizing further damage to the forests.

5.5 Conclusions

The purpose of this study is to track and map the spread of MPBB in the Upper Yampa Basin, Colorado. Several vegetation indices calculated from Landsat TM data were put into the Support Vector Machine (VSM) classifier algorithm to determine the presence or absence of MPBB infestations comparing the years 2005 and 2009. For the year 2009 it was also possible to calculate the percentage of attacked trees in each infested area. For the year 2005, NAIP images were insufficiently detailed to permit differentiation between healthy and slightly infested trees. SVM was successful in performing these classifications with high accuracy (0.95 - 2005 and 0.97 - 2009) that permitted the construction of regional maps showing the distribution of infested areas and changes in that distribution with time. The study showed that SVM can be used to achieve the objectives of this study and provide the sort of information needed by forest managers to fight these infestations.

Chapter 6 - Comparing the accuracy of RF and SVM algorithms in tracking and mapping MPBB

6.1 Introduction

The selection of an appropriate model for classification is essential to achieve accurate forest disturbance mapping outcomes. Numerous classification algorithms have been utilized for the purpose of mapping beetle infestation in forests. Among these algorithms, Random Forest (RF) and Support Vector Machine (SVM) are widely employed in the remote sensing community due to their ability to effectively classify high-dimensional data. For this reason, these same two Machine Learning (ML) methods were selected for creating regression and binary models for the study site in (Chapter 1) The purpose of this study is to determine whether tree mortality due to changes in environment as a function of time be tracked by comparative analysis of various statistical algorithms applied to RS.

The results from the two classifier algorithms (RF and SVM) were presented in chapters 4 and 5. In this chapter these results will be compared by regression and binary statistical analyses methods. Each method produced a different result data set with different attendant measures of accuracy. For example, the accuracy of the regression method is measured in terms of R^2, which is a statistical measure used in regression analysis to assess how well the independent variable(s) explain the variability in the dependent variable. In other words, it quantifies the proportion of the variance in the dependent variable that can be predicted or explained by the independent variable(s).

The selection of an appropriate model for classification is essential to achieve accurate forest disturbances mapping outcomes. Thus, an important part of this process is determining and comparing accuracy measurement for each model. Accuracy assessment for both models, RF and SVM, was estimated by using confusion matrices such as Cohen's Kappa (k), overall accuracy (OA), producer's accuracy (PA), and user's accuracy (UA), matrix for the binary model. For the regression model, R^2 was used to estimate the performance.

6.2 Statistical analysis methods

Kappa has become one of the most popular and widespread metrics developed to test and assess the reliability and accuracy of various rating algorithms. Recent studies have shown that Kappa shows some limitations since it gives information that is misleading for practical decision making. For example, Delgado and Tibau, (2019) showed that if marginal probabilities are very small, the distribution of a misclassification can also affect Kappa with the result that worse classification results can achieve higher Kappa values, which would therefore provide misleading results. Other studies, such as Stein et al., 2005, also mention the drawbacks of Kappa in remote sensing by noting that a single Kappa value is incapable of expressing an overall assessment of a given classification approach.

Various authors, such as Chicco and Jurman (2021) and Chicco et al., (2021) have made a strong case for using the Matthews Correlation Coefficient (MCC) to evaluate the performance of classification models, particularly in situations where the classes are imbalanced because it takes into account true positives, true negatives, false positives, and false negatives, providing a balanced assessment of a model's classification accuracy (Chicco, Tötsch, et al., 2021; Chicco, Warrens, et al., 2021). The MCC produces high scores only if the prediction obtained reliable results for all confusion matrix categories proportionally, both to the size of positive elements and the size of negative elements in the dataset.

The MCC ranges from -1 to +1, where +1 represents perfect prediction, 0 represents random prediction, and -1 represents total disagreement between prediction and actual labels.

The formula for calculating the Matthews Correlation Coefficient is as follows:

$$MCC = (TP * TN - FP * FN) / \sqrt{((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))}$$

Where:

- TP: True Positives (correctly predicted positive instances)
- TN: True Negatives (correctly predicted negative instances)

- FP: False Positives (incorrectly predicted positive instances)
- FN: False Negatives (incorrectly predicted negative instances)

The MCC provides a more comprehensive evaluation of a classification model's performance compared to metrics like accuracy, especially when dealing with imbalanced datasets where one class may dominate the other. It considers both the true and false predictions of both classes, giving a balanced view of how well the model is performing across all aspects.

6.3 Methodology

Regression and classification models of defoliation by MPBB infestation in the Upper Yampa Basin were built using Vegetation Indices (VIs) of two Landsat TM data from 2005 and 2009 (refer to chapters 4 and 5). Confusion matrices were analyzed by several different statistical methods including: accuracy, Kappa, R^2, and MCC.

The confusion matrix results from RF in 2005 and 2009, in Table 6.1-a and 6.1-b, show differentiation between true positives and true negatives indicating that RF did a reliable job in classification.

Table 6.1-a and 6.1-b Confusion Matrices for RF 2005 and 2009

a) RF 2005		**Reference**		
		0	**1**	**total**
Prediction	**0**	251	7	258
	1	4	37	41
	total	255	44	299
Overall accuracy = **0.96** and Kappa = **0.85**				

b) RF 2009		**Reference**		
		0	**1**	**total**
Prediction	**0**	253	4	257
	1	2	40	42
	total	255	44	299
Overall accuracy = **0.98** and Kappa = **0.92**				

Table 6.2 illustrates that all the different methods of calculating accuracy based on RF models (binary and regression) agree that the accuracy of these models ranged between 0.85 and 0.98, suggesting that RF correctly classified all the data.

Table 6.2 Comparison analysis of accuracy by various methods (RF)

RF year	**Accuracy**	**Kappa**	**R^2**	**MCC**
Binary 2005	0.96	0.85	No	0.85
Binary 2009	0.98	0.92	No	0.92
Regression 2009	No	No	0.88	No

The confusion matrix results from SVM in 2005 and 2009, in Table 6.3-a and 6.3-b, show well defined differentiation between true positives and true negatives indicating that SVM did a reliable job in classification overall.

Table 6.3-a and 6.3-b Confusion Matrices for SVM 2005 and 2009

a) SVM 2005		**Reference**		
		0	**1**	**total**
Prediction	**0**	248	6	254
	1	7	38	45
	total	255	44	299
Overall accuracy = **0.96** and Kappa = **0.83**				

b) SVM 2009		**Reference**		
		0	**1**	**total**
Prediction	**0**	253	5	258
	1	2	39	41
	total	255	44	299
Overall accuracy = **0.98** and Kappa = **0.97**				

Table 6.4 Comparison analysis of accuracy by various methods (SVM)

SVM year	Accuracy	Kappa	R^2	MCC
Binary 2005	0.96	0.83	No	0.83
Binary 2009	0.98	0.97	No	0.9
Regression2009	No	No	0.86	No

Table 6.4 illustrates that all the different methods of calculating accuracy based on SVM models (binary and regression) agree that the accuracy of these models ranged between 0.83 and 0.98, suggesting that SVM correctly classified all the data.

Table 6.5 SVM binary model

SVM2005				
Classes	**ID**	**Count**	**Area**	**Percent**
non-evergreen	0	458073	326.5557	31.91751
Infested	1	183014	130.4689	12.75201
Non-infested	2	794104	566.1088	55.3314

Table 6.6 RF binary model

RF 2005				
classes	**ID**	**Count**	**Area**	**Percent**
non-evergreen	0	458068	326.5521	0.319172
infested	1	291940	208.1211	0.203417
non-infested	2	685170	488.4508	0.477411

6.4 Discussion

To illustrate the usefulness of these approaches, we compare data from RF 2005 binary model with SVM 2005 binary model, paying particular attention to the various accuracy measures. For example, examining the accuracy and Kappa values for the binary 2005 model, notice there is no difference in the accuracy values, and a slight decrease in the Kappa value for the SVM model (0.85 for RF binary vs. 0.83 for SVM binary model). These differences may not seem significant, but the actual area data reveals differences that are significant. For example, in 2005 the SVM model projects 130 km^2 as infected, whereas RF model for 2005 projects 208 km^2 as infected (Tables 6.5 and 6.6). Similarly for 2009, the accuracy and Kappa values for the binary model have little difference between the RF and SVM accuracy values (0.98 for both), and a slight decrease in the Kappa value for the RF model (0.92 for RF binary vs. 0.97 for SVM binary model). However, the difference in the real extent of infestation is more significant, with the SVM model predicting 167 km^2 and the RF 272 km^2. These differences in areal extent are far more significant than the accuracy models would suggest. A visual comparison of Figure 5.1 and Figure 5.2 shows a far greater extent of infestation both in terms of areal extent and location of infestation with new areas appearing in the eastern part of the study area.

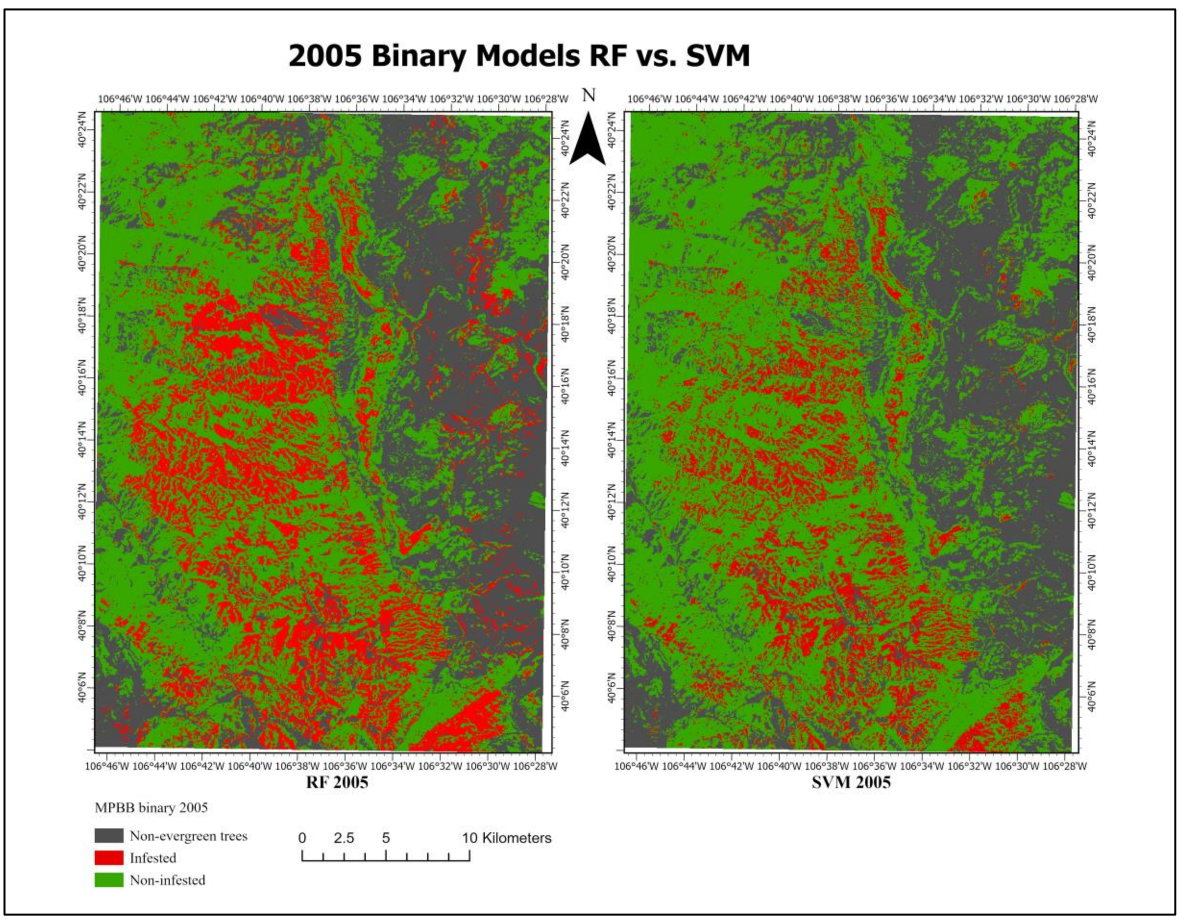


Figure 6.1 2005 binary model SVM vs. RF

A comparison of accuracy and areal extent for the regression models also shows significant differences. For example, the RF regression model (percent of tree mortality) for 2009 shows an R2 of 0.88 while the SVM regression model for 2009 shows an R2 of 0.86, but areal comparisons of the most severely infested areas (25 – 31%) show a marked difference with RF model predicting an areal extent of 72.3 km2 vs. 12.7 km2 predicted by the SVM model Figure 6.3.

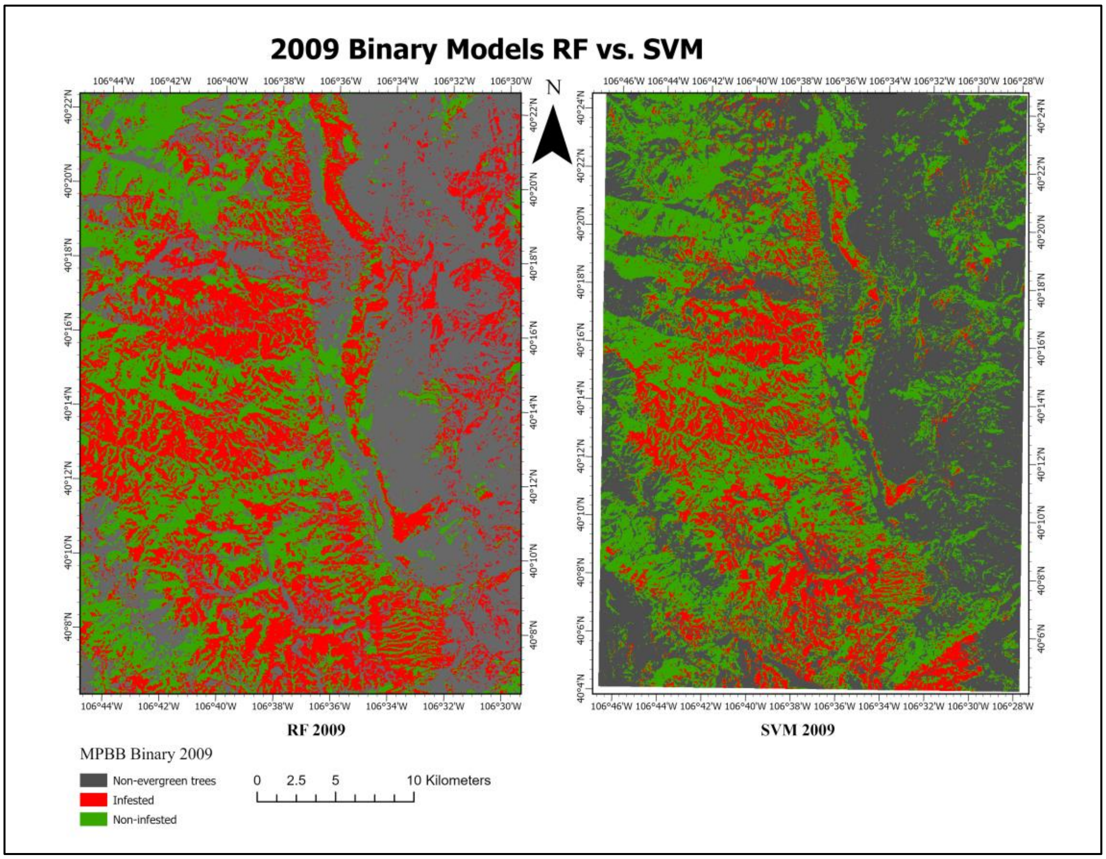


Figure 6.2 2009 Binary models SVM vs. RF

As the deciding factor in determining accuracy, MCC values for the binary RF 2005, and 2009 vs. SVM 2005 and 2009 supports the suggestion that RF is more accurate than SVM (Table 6.4). Li et al. (2013) also found that, while both RF and SVM could identify changes, in dominant forest type in their 20-year study, RF was slightly superior to SVM. In addition, Figure 6.4 shows that RF was far more sensitive at capturing burned over and sparsely treed areas. Mellor et al. (2018) found a trend for the area proportional allocation of training samples, in which the greater a given class area is, the more training samples are needed to produce the best classification accuracy. They showed that if the training sample size is small (less than 500 pixel per class) the difference in accuracy of imbalanced and balanced data would be large. However, when the training sample was large enough, the performance of RF of

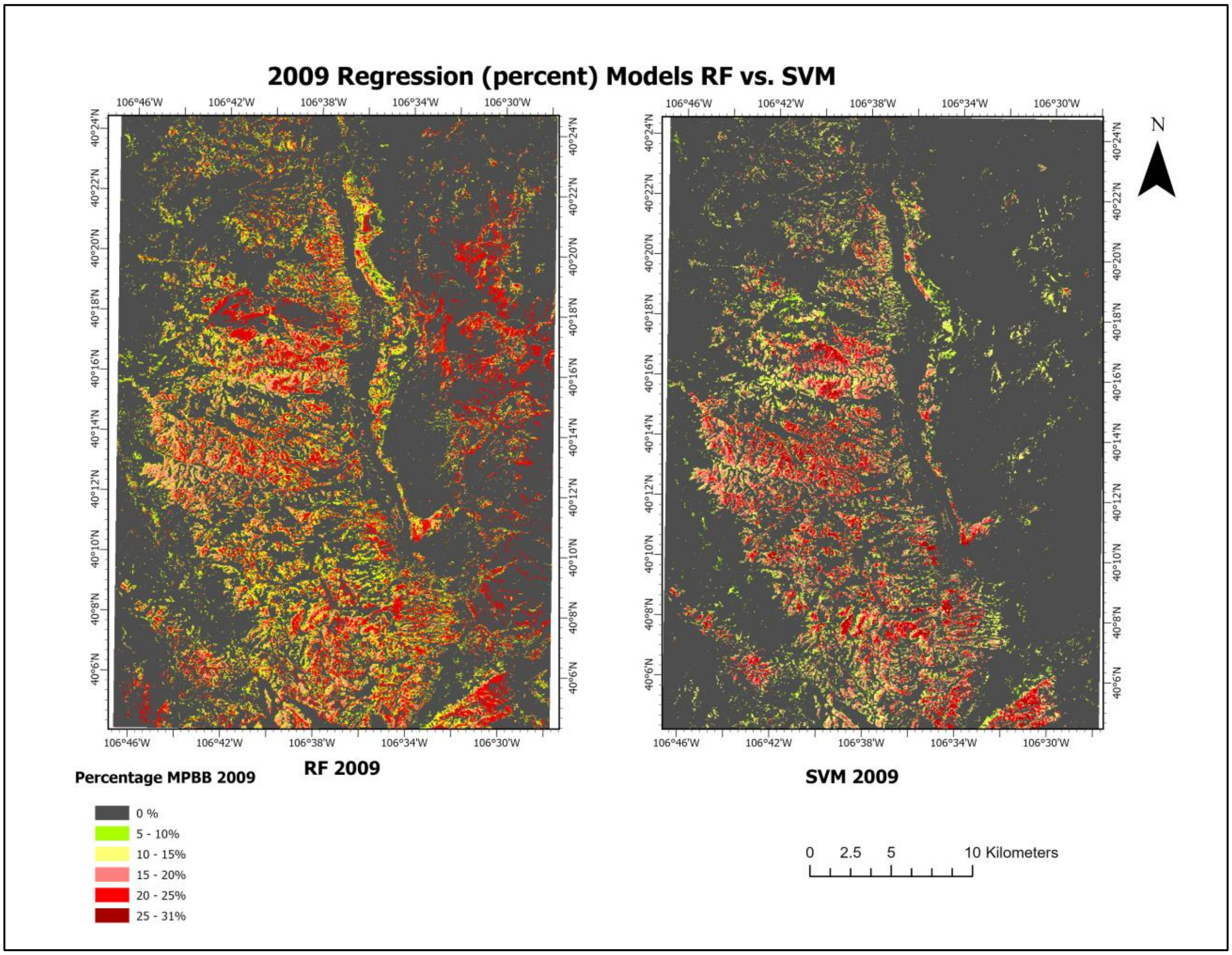


Figure 6.3 Regression models for RF vs. SVM

imbalanced and balanced training sample was similar. In this study, 1021 pixels were used to predict disturbed areas. Land managers need to be aware of how MPB will respond to their restoration treatments, to prevent and mitigate tree mortality due to MPB attacks which could negate restoration effects. Widely spaced clumps of trees may be beneficial for limiting the spread of fires, but they make infestations by MPB easier (Stephenson et al., 2019).

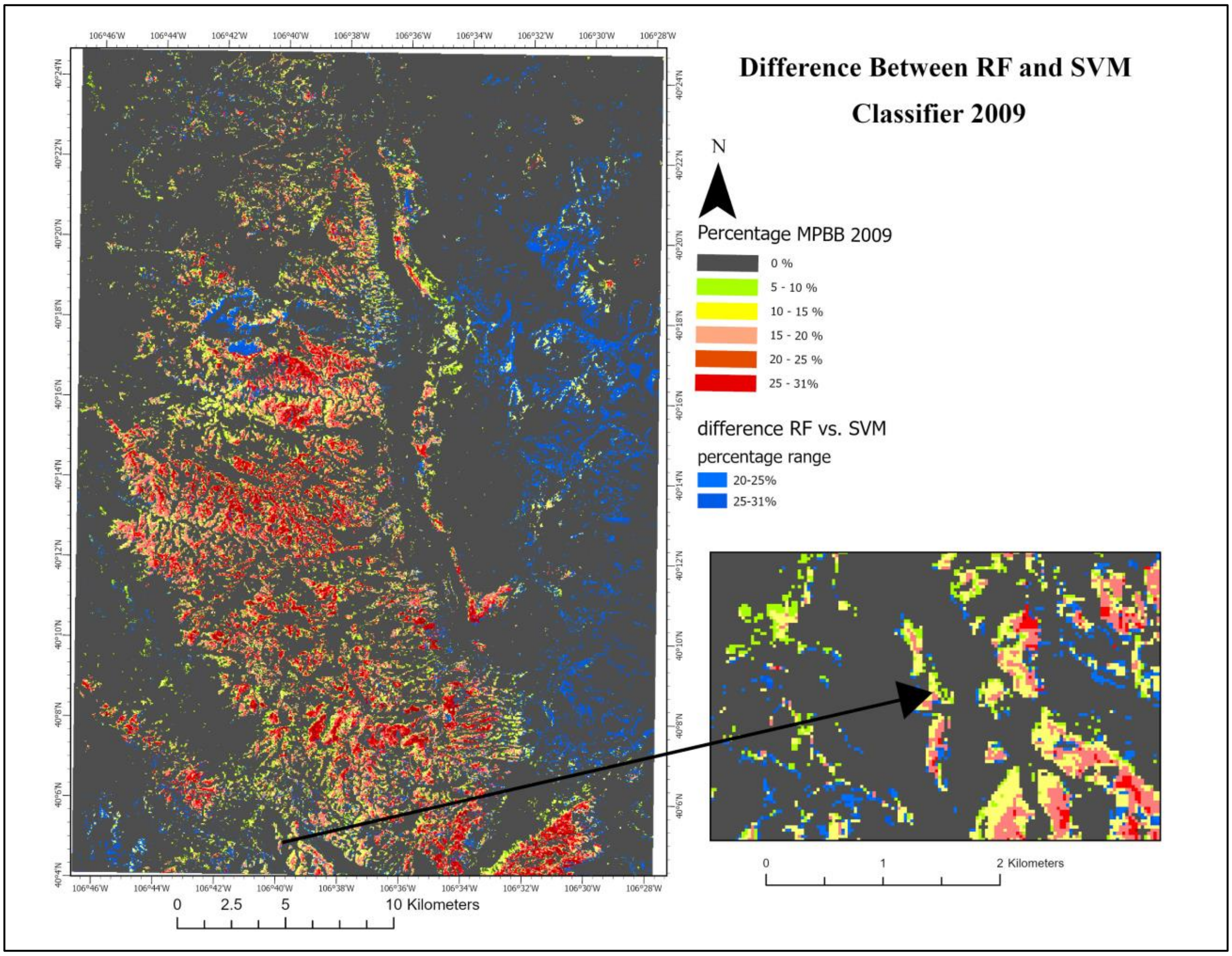


Figure 6.4 Sensitivity difference between RF vs. SVM classifier

In this figure, the blue colored area represents either a small burned-over area to the west, or a large area on the east side of the study area which still contains some trees, but in which the trees are very space. The presence of those trees means that these areas could be repopulated with trees in future more easily. RF identified this area that was missed by SVM.

6.5 Conclusions

The purpose of this study is to compare the accuracy of two classification algorithms (RF and SVM) to track and map the spread of MPBB in the Upper Yampa Basin, Colorado. Two statistical measures of accuracy, Kappa and R^2, were used to compare the performance of the two classification algorithms.

Kappa was used to analyze the accuracy for binary models. But it was determined if marginal probabilities are very small, the distribution of a misclassification can affect Kappa with the result that worse classification results can achieve higher Kappa values, which would therefore provide misleading results. Because of this limitation, the Matthews Correlation Coefficient (MCC) has proven more reliable in determining the accuracy of binary models. For the regression models, R^2 was used to assess the performance of each model. The higher the R^2 value, the more accurate the regression model. R2 has proved a reliable measure of accuracy in regression models.

If the statistical accuracy of all the models run, RF vs. SVM and binary vs. regression, are compared, both the regression and binary models based on RF are more accurate.

Chapter 7 – Discussion and Conclusions

7.1 Summary discussion

In the past two decades, MPBB's have become more pervasive due to increasing temperatures and drought conditions related to climate change causing regional-scale mortality. As discussed in Chapter 3, a variety of studies have suggested that insect effects on tree die-off, fuels, and fire behavior can vary widely. A key problem with developing a better understanding of insect-fire relationships is the lack of empirical maps that can show interrelated changes in the distribution of insect infestations and fire zones reliably over space and time. Although remote sensing data has frequently been used to map forest fires with well-established techniques familiar to forest managers, the science of mapping insect infestations is relatively new and lacking in well-developed and documented spatiotemporal data sets that can be used to track these infestations and quantify the negative impacts of the infestations as well as other various agents related to forest health in a consistent fashion. This study documents the development of such a technique based on time series statistical analyses of NBR spectral index trajectories. This approach appears to produce reliable repeatable results which can be used by forestry agents to improve their ability to manage, maintain, or improve the health of regional forests.

Chapter 4 shows the development of an approach to tracking and mapping the spread of MPBB by analyses of several vegetation indices which are calculated from Landsat TM data. These indices were put into the Random Forest (RF) classifier algorithm to determine first, the presence or absence of MPBB infestations, and second, to measure the percentage of attacked trees in each infested area. RF was successful in performing these analyses with high accuracy which permitted the construction of regional maps showing the distribution of infested areas and changes in that distribution with time. The study showed that RF can be used to achieve the stated objectives of the study, and to provide the sort of information needed by forest managers to fight these infestations.

In Chapter 5, a different set of vegetation indices were input into the Support Vector Machine (SVM) classifier algorithm to determine the presence or absence of MPBB and to illustrate the changes in the distribution of infestations with time, comparing the years 2005 and 2009. For the year 2009 it was also possible to calculate the percentage of trees attacked in each infested area. SVM was successful in performing this classification with a high enough accuracy that permitted the construction of regional maps showing the distribution of infested area and changes in that distribution with time similar to those constructed through the use of RF in Chapter 4.

In chapter 6, a comparison was made between the accuracy of two classification algorithms (RF and SVM) in tracking and mapping the spread of MPBB infestations. Two statistical measures of accuracy, Kappa and R^2, were used to compare the performance of the two algorithms. Kappa was used to analyze the accuracy for binary models. It was determined that if marginal probabilities are very small, the distribution of misclassified data points can affect Kappa with the result that worse classification results can achieve higher Kappa values, which would therefore provide misleading results. Because of this limitation, the Matthews Correlation Coefficient (MCC) has proven more reliable in determining the accuracy of binary models. For the regression models, R^2 was used to assess the performance of each model. The higher the R^2 value, the more accurate the regression model. R^2 has proved to be a reliable measure of accuracy of regression models. If the statistical accuracy of all the models run, RF vs. SVM and binary vs. regression, are compared, both the regression and binary models based on RF are more accurate.

7.2 Limitations

These results suggest that any single model of temporal change will not be able to capture the full range of insect-induced changes in heterogeneous landscapes. Any remote sensing approach requires the capacity to detect the full variety of short- and long- duration changes, including both losses and gains in the spectral indices, as well as the potential for verification by other types of data sets such as high-resolution aerial data sets. In addition, it must be noted that the signal of spectral change likely depends on overall site productivity. More productive sites have more vegetation to lose, potentially improving detection of changes, but they typically also have greater understory vegetation that may lessen the spectral impact of change. These sites also have potential for robust regrowth and recovery which may affect detectability thresholds along productivity gradients.

7.3 Conclusions

The results of this study can provide a useful new tool for forest managers who must make decisions about how best to deal with changing conditions. (Udall & Overpeck, 2017) discuss the relationship between climate change and increasing drought in the Colorado River Basin, predicting temperature-induced decreases in river flow up to -30% by the middle of the next century and – 55% by the end of the century. Given the likely increase in fire and insect activity in western North American Forest due to climate change (Meigs et al.,2011), the accurate characterization of insect's effects on forest, fuels, and subsequent wildfires will become increasingly important. The advantage of being able to map and track infestations can allow forest managers to improve mitigation effort by allowing them to: address low-severity infestations before they intensify, monitor intensifying infestations with previously identified outbreaks, and develop a plan for trying to get ahead of the spread of infestations (Bode et al.,2018).

If the statistical accuracy of all the models run, RF vs. SVM and binary vs. regression, are compared, both the regression and binary models based on RF are more accurate. The results of this study can provide a useful new tool for forest managers who must make decisions about how best to deal with changing conditions. The advantage of being able to map and track infestations can allow forest managers to improve mitigation efforts by allowing them to: address low-severity infestations before they intensify, monitor intensifying infestations with previously identified outbreaks, and develop a plan for trying to get ahead of the spread of infestations. Genetic analyses of trees that have survived MPBB infestation show that some trees are now developing immunity to infestation. If these resistant trees can be tracked and mapped similarly to the infested trees, this may also allow development of programs for reforestation with genetically immune trees to insure the health of future forests.

An intriguing possibility for a new application of this mapping ability is related to a study by a group of genetic botanists lead by Diana Six at the University of Montana (Six et al., 2018). Dr. Six and her group discussed the idea that, with climate change encouraging aggressive invasion of bark beetles, it is increasingly important to understand the ability of trees to adapt and develop ways to resist infestation (Millar et al., 2007; Ramsfield et al., 2016). Their study showed that some trees in infested areas were genetically different than those selected and killed by the beetles, and that these trees may aid in adaptation and persistence of genetically resistant traits. If these resistant trees can be tracked and mapped similarly to the infested trees, this may also allow development of programs for encouraging forest adaptation. For example, retaining and mapping surviving trees as primary seed sources could allow for the spread of beetle resistant trees. In contrast, common practices such as clearcutting and replanting burned over areas with general seed stock could destroy the resistant trees and set the stage for another cycle of infestation and fire associated with non-resistant trees.

References

Adams, H. D., Zeppel, M. J., Anderegg, W. R., Hartmann, H., Landhäusser, S. M., Tissue, D. T., Huxman, T. E., Hudson, P. J., Franz, T. E., & Allen, C. D. (2017). A multi-species synthesis of physiological mechanisms in drought-induced tree mortality. *Nature ecology & evolution*, *1*(9), 1285-1291.

Ahern, F. (1988). The effects of bark beetle stress on the foliar spectral reflectance of lodgepole pine. *International journal of remote sensing*, *9*(9), 1451-1468.

Anderegg, W. R., Anderegg, L. D., Kerr, K. L., & Trugman, A. T. (2019). Widespread drought-induced tree mortality at dry range edges indicates that climate stress exceeds species' compensating mechanisms. *Global change biology*, *25*(11), 3793-3802.

Anderegg, W. R., Hicke, J. A., Fisher, R. A., Allen, C. D., Aukema, J., Bentz, B., Hood, S., Lichstein, J. W., Macalady, A. K., & McDowell, N. (2015). Tree mortality from drought, insects, and their interactions in a changing climate. *New Phytologist*, *208*(3), 674-683.

Andrus, R. A., Chai, R. K., Harvey, B. J., Rodman, K. C., & Veblen, T. T. (2021). Increasing rates of subalpine tree mortality linked to warmer and drier summers. *Journal of Ecology*, *109*(5), 2203-2218.

Andrus, R. A., Hart, S. J., & Veblen, T. T. (2020). Forest recovery following synchronous outbreaks of spruce and western balsam bark beetle is slowed by ungulate browsing. *Ecology*, *101*(5), e02998.

Arango-Velez, A., El Kayal, W., Copeland, C. C., Zaharia, L. I., Lusebrink, I., & Cooke, J. E. (2016). Differences in defence responses of Pinus contorta and Pinus banksiana to the mountain pine beetle fungal associate Grosmannia clavigera are affected by water deficit. *Plant, Cell & Environment*, *39*(4), 726-744.

Assal, T. J., Sibold, J., & Reich, R. (2014). Modeling a historical mountain pine beetle outbreak using Landsat MSS and multiple lines of evidence. *Remote Sensing of Environment*, *155*, 275-288.

Ballanti, L., Blesius, L., Hines, E., & Kruse, B. (2016). Tree species classification using hyperspectral imagery: A comparison of two classifiers. *Remote Sensing*, *8*(6), 445.

Banskota, A., Kayastha, N., Falkowski, M. J., Wulder, M. A., Froese, R. E., & White, J. C. (2014). Forest monitoring using Landsat time series data: A review. *Canadian Journal of Remote Sensing*, *40*(5), 362-384.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, *114*, 24-31.

Bentz, B., & Endreson, D. (2004). Evaluating satellite imagery for estimating mountain pine beetle-caused lodgepole pine mortality: current status. In: Shore, TL; Brooks, JE; Stone, JE, eds. Mountain Pine Beetle Symposium: Challenges and Solutions; October 30-31, 2003; Kelowna, British Columbia. Information Report BC-X-399. Victoria, British Columbia: Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre. p. 154-163.,

Bentz, B., Logan, J., MacMahon, J., Allen, C. D., Ayres, M., Berg, E., Carroll, A., Hansen, M., Hicke, J., & Joyce, L. (2009). Bark beetle outbreaks in western North America: Causes and consequences. Bark Beetle Symposium; Snowbird, Utah; November, 2005. Salt Lake City, UT: University of Utah Press. 42 p.,

Bentz, B. J., Régnière, J., Fettig, C. J., Hansen, E. M., Hayes, J. L., Hicke, J. A., Kelsey, R. G., Negrón, J. F., & Seybold, S. J. (2010). Climate change and bark beetles of the western United States and Canada: direct and indirect effects. *Bioscience*, *60*(8), 602-613.

Bentz, B. J. (2016). Forest health in a changing world: Effects of globalization and climate change on forest insect and pathogen impacts.

Bode, E. T., Lawrence, R. L., Powell, S. L., Savage, S. L., & Trowbridge, A. M. (2018). Time-series approach for mapping mountain pine beetle infestation extent and severity in the US Central Rocky Mountains. *Journal of Applied Remote Sensing*, *12*(4), 046030.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory,

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Bright, B. C., Hudak, A. T., Egan, J. M., Jorgensen, C. L., Rex, F. E., Hicke, J. A., & Meddens, A. J. (2020). Using satellite imagery to evaluate bark beetle-caused tree mortality reported in aerial surveys in a mixed conifer forest in northern Idaho, USA. *Forests*, *11*(5), 529.

Chapman, T. B., Veblen, T. T., & Schoennagel, T. (2012). Spatiotemporal patterns of mountain pine beetle activity in the southern Rocky Mountains. *Ecology*, *93*(10), 2175-2185.

Cheng, T., Rivard, B., Sánchez-Azofeifa, G., Feng, J., & Calvo-Polanco, M. (2010). Continuous wavelet analysis for the detection of green attack damage due to mountain pine beetle infestation. *Remote Sensing of Environment*, *114*(4), 899-910.

Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData mining*, *14*(1), 1-22.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier score in binary classification assessment. *IEEE Access*, *9*, 78368-78381.

Christiansen, E., Waring, R. H., & Berryman, A. A. (1987). Resistance of conifers to bark beetle attack: searching for general relationships. *Forest Ecology and Management*, *22*(1-2), 89-106.

Clark, E. L., Carroll, A. L., & Huber, D. P. (2010). Differences in the constitutive terpene profile of lodgepole pine across a geographical range in British Columbia, and correlation with historical attack by mountain pine beetle. *The Canadian Entomologist*, *142*(6), 557-573.

Claverie, M., Vermote, E. F., Franch, B., & Masek, J. G. (2015). Evaluation of the Landsat-5 TM and Landsat-7 ETM+ surface reflectance products. *Remote Sensing of Environment*, *169*, 390-403.

Cohen, W. B., & Goward, S. N. (2004). Landsat's role in ecological applications of remote sensing. *Bioscience*, *54*(6), 535-545.

Cohen, W. B., Healey, S. P., Yang, Z., Stehman, S. V., Brewer, C. K., Brooks, E. B., Gorelick, N., Huang, C., Hughes, M. J., & Kennedy, R. E. (2017). How similar are forest disturbance maps derived from different Landsat time series algorithms? *Forests*, *8*(4), 98.

Cohen, W. B., Spies, T. A., Alig, R. J., Oetter, D. R., Maiersperger, T. K., & Fiorella, M. (2002). Characterizing 23 years (1972–95) of stand replacement disturbance in western Oregon forests with Landsat imagery. *Ecosystems*, *5*, 122-137.

Cohen, W. B., Yang, Z., & Kennedy, R. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync—Tools for calibration and validation. *Remote Sensing of Environment*, *114*(12), 2911-2924.

Cooper, L. A., Reed, C. C., & Ballantyne, A. P. (2018). Mountain pine beetle attack faster growing lodgepole pine at low elevations in western Montana, USA. *Forest Ecology and Management*, *427*, 200-207.

Coops, N. C., Johnson, M., Wulder, M. A., & White, J. C. (2006). Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation. *Remote Sensing of Environment*, *103*(1), 67-80.

Coops, N. C., Shang, C., Wulder, M. A., White, J. C., & Hermosilla, T. (2020). Change in forest condition: Characterizing non-stand replacing disturbances using time series satellite imagery. *Forest Ecology and Management*, *474*, 118370.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273-297.

Creeden, E. P., Hicke, J. A., & Buotte, P. C. (2014). Climate, weather, and recent mountain pine beetle outbreaks in the western United States. *Forest Ecology and Management*, *312*, 239-251.

Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783-2792.
Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., & Vapnik, V. (1997). Advances in neural information processing systems. *NIPS 1996*, *155*.
Duro, D. C., Franklin, S. E., & Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, *118*, 259-272.
Edburg, S. L., Hicke, J. A., Brooks, P. D., Pendall, E. G., Ewers, B. E., Norton, U., Gochis, D., Gutmann, E. D., & Meddens, A. J. (2012). Cascading impacts of bark beetle-caused tree mortality on coupled biogeophysical and biogeochemical processes. *Frontiers in Ecology and the Environment*, *10*(8), 416-424.
Egan, J. M., Kaiden, J., Lestina, J., Stasey, A., & Jenne, J. L. (2019). *Techniques to enhance assessment and reporting of pest damage estimated with aerial detection surveys*.
Eidenshink, J., Schwind, B., Brewer, K., Zhu, Z., Quayle, B., & Howard, S. (2007). A project for monitoring trends in burn severity. Fire Ecology 3 (1): 3-21. *Fire Ecology Special Issue Vol*, *3*(4).
Fettig, C., Reid, M., Bentz, B., Servanto, S., Spittlehouse, D., & Wang, T. (2013). Changing Climates, Changing Forests: A Western North American Perspective. *Journal of Forestry -Washington-*, *111*, 214-218. https://doi.org/10.5849/jof.12-085
Foster, A. C., Walter, J. A., Shugart, H. H., Sibold, J., & Negron, J. (2017). Spectral evidence of early-stage spruce beetle infestation in Engelmann spruce. *Forest Ecology and Management*, *384*, 347-357.
Franklin, S., Wulder, M., Skakun, R., & Carroll, A. L. (2003). Mountain pine beetle red-attack forest damage classification using stratified Landsat TM data in British Columbia, Canada. *Photogrammetric Engineering & Remote Sensing*, *69*(3), 283-288.
Freedman, D. A. (2009). *Statistical models: theory and practice*. cambridge university press.
Ghimire, B., Rogan, J., & Miller, J. (2010). Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters*, *1*(1), 45-54.
Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS journal of photogrammetry and remote sensing*, *116*, 55-72.
Goodman, S. (2008). A dirty dozen: twelve p-value misconceptions. Seminars in hematology,
Goodwin, N. R., Coops, N. C., Wulder, M. A., Gillanders, S., Schroeder, T. A., & Nelson, T. (2008). Estimation of insect infestation dynamics using a temporal sequence of Landsat data. *Remote Sensing of Environment*, *112*(9), 3680-3689.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18-27.
Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS journal of photogrammetry and remote sensing*, *66*(1), 56-66.
Hansen, M., Dubayah, R., & DeFries, R. (1996). Classification trees: an alternative to traditional land cover classifiers. *International journal of remote sensing*, *17*(5), 1075-1081.
Hansen, M. C., & Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, *122*, 66-74.
Hart, S., & Preston, D. (2020). Fire weather drives daily area burned and observations of fire behavior in mountain pine beetle affected landscapes. *Environmental Research Letters*, *15*(5), 054007.
Hart, S. J., & Veblen, T. T. (2015). Detection of spruce beetle-induced tree mortality using high-and medium-resolution remotely sensed imagery. *Remote Sensing of Environment*, *168*, 134-145.
Hart, S. J., Veblen, T. T., Mietkiewicz, N., & Kulakowski, D. (2015). Negative feedbacks on bark beetle outbreaks: widespread and severe spruce beetle infestation restricts subsequent infestation. *PLoS One*, *10*(5), e0127975.

Herms, D. A., & Mattson, W. J. (1992). The dilemma of plants: to grow or defend. *The quarterly review of biology*, *67*(3), 283-335.
Hicke, J. A., Logan, J. A., Powell, J., & Ojima, D. S. (2006). Changing temperatures influence suitability for modeled mountain pine beetle (Dendroctonus ponderosae) outbreaks in the western United States. *Journal of Geophysical Research: Biogeosciences*, *111*(G2). https://doi.org/https://doi.org/10.1029/2005JG000101
Hicke, J. A., Meddens, A. J., & Kolden, C. A. (2016). Recent tree mortality in the western United States from bark beetles and forest fires. *Forest Science*, *62*(2), 141-153.
Hicke, J. A., Xu, B., Meddens, A. J., & Egan, J. M. (2020). Characterizing recent bark beetle-caused tree mortality in the western United States from aerial surveys. *Forest Ecology and Management*, *475*, 118402.
Hoffman, C. M., Linn, R., Parsons, R., Sieg, C., & Winterkamp, J. (2015). Modeling spatial and temporal dynamics of wind flow and potential fire behavior following a mountain pine beetle outbreak in a lodgepole pine forest. *Agricultural and Forest Meteorology*, *204*, 79-93.
Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2010). A practical guide to support vector classification. Department of Computer Science, National Taiwan University. In: Taipei. Tech. Rep.
Huang, C., Goward, S. N., Masek, J. G., Thomas, N., Zhu, Z., & Vogelmann, J. E. (2010). An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sensing of Environment*, *114*(1), 183-198.
Huang, J., Kautz, M., Trowbridge, A. M., Hammerbacher, A., Raffa, K. F., Adams, H. D., Goodsman, D. W., Xu, C., Meddens, A. J., & Kandasamy, D. (2020). Tree defence and bark beetles in a drying world: carbon partitioning, functioning and modelling. *New Phytologist*, *225*(1), 26-36.
Hultquist, C., Chen, G., & Zhao, K. (2014). A comparison of Gaussian process regression, random forests and support vector regression for burn severity assessment in diseased forests. *Remote Sensing Letters*, *5*(8), 723-732.
Jenkins, M. J., Hebertson, E., Page, W., & Jorgensen, C. A. (2008). Bark beetles, fuels, fires and implications for forest management in the Intermountain West. *Forest Ecology and Management*, *254*(1), 16-34.
Jenkins, M. J., Runyon, J. B., Fettig, C. J., Page, W. G., & Bentz, B. J. (2014). Interactions among the mountain pine beetle, fires, and fuels. *Forest Science*, *60*(3), 489-501.
Jewett, J. T., Lawrence, R. L., Marshall, L. A., Gessler, P. E., Powell, S. L., & Savage, S. L. (2011). Spatiotemporal relationships between climate and whitebark pine mortality in the Greater Yellowstone Ecosystem. *Forest Science*, *57*(4), 320-335.
Jolly, W. M., Parsons, R. A., Hadlow, A. M., Cohn, G. M., McAllister, S. S., Popp, J. B., Hubbard, R. M., & Negron, J. F. (2012). Relationships between moisture, chemistry, and ignition of Pinus contorta needles during the early stages of mountain pine beetle attack. *Forest Ecology and Management*, *269*, 52-59.
Kane, J. M., & Kolb, T. E. (2010). Importance of resin ducts in reducing ponderosa pine mortality from bark beetle attack. *Oecologia*, *164*, 601-609.
Karan, S. K., & Samadder, S. R. (2016). Accuracy of land use change detection using support vector machine and maximum likelihood techniques for open-cast coal mining areas. *Environmental monitoring and assessment*, *188*, 1-13.
Kennedy, R. E., Andréfouët, S., Cohen, W. B., Gómez, C., Griffiths, P., Hais, M., Healey, S. P., Helmer, E. H., Hostert, P., & Lyons, M. B. (2014). Bringing an ecological view of change to Landsat-based remote sensing. *Frontiers in Ecology and the Environment*, *12*(6), 339-346.
Kennedy, R. E., Yang, Z., & Cohen, W. B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. *Remote Sensing of Environment*, *114*(12), 2897-2910.
Kennedy, R. E., Yang, Z., Gorelick, N., Braaten, J., Cavalcante, L., Cohen, W. B., & Healey, S. (2018). Implementation of the LandTrendr algorithm on google earth engine. *Remote Sensing*, *10*(5), 691.

Key, C., & Benson, N. (2006). *Landscape Assessment (LA) Sampling and Analysis Methods. USDA Forest Service Gen*.

Kolb, T., Keefover-Ring, K., Burr, S. J., Hofstetter, R., Gaylord, M., & Raffa, K. F. (2019). Drought-mediated changes in tree physiological processes weaken tree defenses to bark beetle attack. *Journal of chemical ecology*, *45*, 888-900.

Kuhn, M., Johnson, K., Kuhn, M., & Johnson, K. (2013). Over-fitting and model tuning. *Applied predictive modeling*, 61-92.

Kulkarni, A. D., & Lowe, B. (2016). Random forest algorithm for land cover classification.

Latifi, H., Nothdurft, A., & Koch, B. (2010). Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. *Forestry*, *83*(4), 395-407.

Lawrence, R. L., Wood, S. D., & Sheley, R. L. (2006). Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment*, *100*(3), 356-362.

Li, M., Ma, L., Blaschke, T., Cheng, L., & Tiede, D. (2016). A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. *International journal of applied earth observation and geoinformation*, *49*, 87-98.

Liang, L., Hawbaker, T. J., Chen, Y., Zhu, Z., & Gong, P. (2014). Characterizing recent and projecting future potential patterns of mountain pine beetle outbreaks in the Southern Rocky Mountains. *Applied Geography*, *55*, 165-175.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, *2*(3), 18-22.

Long, J. A., & Lawrence, R. L. (2016). Mapping percent tree mortality due to mountain pine beetle damage. *Forest Science*, *62*(4), 392-402.

Lundquist, J. E., & Reich, R. M. (2014). Landscape dynamics of mountain pine beetles. *Forest Science*, *60*(3), 464-475.

Lusebrink, I., Erbilgin, N., & Evenden, M. L. (2013). The lodgepole× jack pine hybrid zone in Alberta, Canada: a stepping stone for the mountain pine beetle on its journey east across the boreal forest? *Journal of chemical ecology*, *39*, 1209-1220.

Lusebrink, I., Erbilgin, N., & Evenden, M. L. (2016). The effect of water limitation on volatile emission, tree defense response, and brood success of Dendroctonus ponderosae in two pine hosts, lodgepole, and jack pine. *Frontiers in Ecology and Evolution*, *4*, 2.

Lynch, H. J., Renkin, R. A., Crabtree, R. L., & Moorcroft, P. R. (2006). The influence of previous mountain pine beetle (Dendroctonus ponderosae) activity on the 1988 Yellowstone fires. *Ecosystems*, *9*(8), 1318-1327.

McDowell, N. G., Adams, H. D., Bailey, J. D., & Kolb, T. E. (2007). The role of stand density on growth efficiency, leaf area index, and resin flow in southwestern ponderosa pine forests. *Canadian Journal of Forest Research*, *37*(2), 343-355.

Meddens, A. J., & Hicke, J. A. (2014). Spatial and temporal patterns of Landsat-based detection of tree mortality caused by a mountain pine beetle outbreak in Colorado, USA. *Forest Ecology and Management*, *322*, 78-88.

Meddens, A. J., Hicke, J. A., & Ferguson, C. A. (2012). Spatiotemporal patterns of observed bark beetle-caused tree mortality in British Columbia and the western United States. *Ecological Applications*, *22*(7), 1876-1891.

Meddens, A. J., Hicke, J. A., Vierling, L. A., & Hudak, A. T. (2013). Evaluating methods to detect bark beetle-caused tree mortality using single-date and multi-date Landsat imagery. *Remote Sensing of Environment*, *132*, 49-58.

Meigs, G. W., Kennedy, R. E., & Cohen, W. B. (2011). A Landsat time series approach to characterize bark beetle and defoliator impacts on tree mortality and surface fuels in conifer forests. *Remote Sensing of Environment*, *115*(12), 3707-3718.

Meigs, G. W., Kennedy, R. E., Gray, A. N., & Gregory, M. J. (2015). Spatiotemporal dynamics of recent mountain pine beetle and western spruce budworm outbreaks across the Pacific Northwest Region, USA. *Forest Ecology and Management*, *339*, 71-86.

Mellor, A., Haywood, A., Stone, C., & Jones, S. (2013). The performance of random forests in an operational setting for large area sclerophyll forest classification. *Remote Sensing*, *5*(6), 2838-2856.

Migas-Mazur, R., Kycko, M., Zwijacz-Kozica, T., & Zagajewski, B. (2021). Assessment of Sentinel-2 images, support vector machines and change detection algorithms for bark beetle outbreaks mapping in the Tatra mountains. *Remote Sensing*, *13*(16), 3314.

Millar, C. I., Stephenson, N. L., & Stephens, S. L. (2007). Climate change and forests of the future: managing in the face of uncertainty. Ecological applications, 17(8), 2145-2151.

Monahan, W. B., & Fisichelli, N. A. (2014). Climate exposure of US national parks in a new era of change. *PLoS One*, *9*(7), e101302.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS journal of photogrammetry and remote sensing*, *66*(3), 247-259.

Mountrakis, G., Watts, R., Luo, L., & Wang, J. (2009). Developing collaborative classifiers using an expert-based model. *Photogrammetric Engineering & Remote Sensing*, *75*(7), 831-843.

Negrón, J. F. (2020). Within-stand distribution of tree mortality caused by mountain pine beetle, Dendroctonus ponderosae Hopkins. *Insects*, *11*(2), 112.

Negrón, J. F., Bentz, B. J., Fettig, C. J., Gillette, N., Hansen, E. M., Hayes, J. L., Kelsey, R. G., Lundquist, J. E., Lynch, A. M., & Progar, R. A. (2008). US Forest Service bark beetle research in the western United States: Looking toward the future. *Journal of Forestry*, *106*(6), 325-331.

Nelson, T. A., Boots, B., Wulder, M. A., & Carroll, A. L. (2007). Environmental characteristics of mountain pine beetle infestation hot spots. *Journal of Ecosystems and Management*.

Ojeda Alayon, D. I., Tsui, C. K., Feau, N., Capron, A., Dhillon, B., Zhang, Y., Massoumi Alamouti, S., Boone, C. K., Carroll, A. L., & Cooke, J. E. (2017). Genetic and genomic evidence of niche partitioning and adaptive radiation in mountain pine beetle fungal symbionts. *Molecular Ecology*, *26*(7), 2077-2091.

Oke, T. R. (2002). *Boundary layer climates*. Routledge.

Otukei, J. R., & Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International journal of applied earth observation and geoinformation*, *12*, S27-S31.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International journal of remote sensing*, *26*(1), 217-222.

Peet, R. K. (1981). Forest vegetation of the Colorado Front Range: composition and dynamics. *Vegetatio*, *45*, 3-75.

Pilario, K. E., Shafiee, M., Cao, Y., Lao, L., & Yang, S.-H. (2019). A review of kernel methods for feature extraction in nonlinear process monitoring. *Processes*, *8*(1), 24.

Radeloff, V. C., Mladenoff, D. J., & Boyce, M. S. (1999). Detecting jack pine budworm defoliation using spectral mixture analysis: separating effects from determinants. *Remote Sensing of Environment*, *69*(2), 156-169.

Raffa, K. F., Aukema, B. H., Bentz, B. J., Carroll, A. L., Hicke, J. A., Turner, M. G., & Romme, W. H. (2008). Cross-scale drivers of natural disturbances prone to anthropogenic amplification: the dynamics of bark beetle eruptions. *Bioscience*, *58*(6), 501-517.

Raffa, K. F., & Berryman, A. A. (1987). Interacting selective pressures in conifer-bark beetle systems: a basis for reciprocal adaptations? *The American Naturalist*, *129*(2), 234-262.

Repass, J. M. (2005). *Assessment of Digital land cover maps for hydrological modeling of the Yampa River basin, Colorado, USA* Colorado State University].

Restaino, C., Young, D. J., Estes, B., Gross, S., Wuenschel, A., Meyer, M., & Safford, H. (2019). Forest structure and climate mediate drought-induced tree mortality in forests of the Sierra Nevada, USA. *Ecological Applications*, *29*(4), e01902.

Rice, A., Thormann, M., & Langor, D. (2008). Mountain pine beetle-associated blue-stain fungi are differentially adapted to boreal temperatures. *Forest Pathology*, *38*(2), 113-123.

Rodman, K. C., Andrus, R. A., Butkiewicz, C. L., Chapman, T. B., Gill, N. S., Harvey, B. J., Kulakowski, D., Tutland, N. J., Veblen, T. T., & Hart, S. J. (2021). Effects of Bark Beetle Outbreaks on Forest Landscape Pattern in the Southern Rocky Mountains, USA. *Remote Sensing*, *13*(6), 1089.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS journal of photogrammetry and remote sensing*, *67*, 93-104.

Rogan, J., Miller, J., Stow, D., Franklin, J., Levien, L., & Fischer, C. (2003). Land-cover change monitoring with classification trees using Landsat TM and ancillary data. *Photogrammetric Engineering & Remote Sensing*, *69*(7), 793-804.

Romme, W. H., Floyd, M. L., Hanna, D., Bartlett, E., Crist, M., Green, D., Grissino-Mayer, H., Lindsey, J., McGarigal, K., & Redders, J. (2009). Historical Range of Variability and Current Landscape Condition Analysis: South Central Highlands Section. *Southwestern Colorado & Northwestern New Mexico*, 1-186.

Roy, D. P., Kovalskyy, V., Zhang, H., Vermote, E. F., Yan, L., Kumar, S., & Egorov, A. (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment*, *185*, 57-70.

Safranyik, L., & Carroll, A. L. (2006). The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. *The mountain pine beetle: a synthesis of biology, management and impacts on lodgepole pine*, 3-66.

Safranyik, L., & Linton, D. (1991). Unseasonably low fall and winter temperatures affecting mountain pine beetle and pine engraver beetle populations and damage in the British Columbia Chilcotin Region. *Journal of the Entomological Society of British Columbia*, *88*, 17-21.

Sambaraju, K. R., Carroll, A. L., Zhu, J., Stahl, K., Moore, R. D., & Aukema, B. H. (2012). Climate change could alter the distribution of mountain pine beetle outbreaks in western Canada. *Ecography*, *35*(3), 211-223.

Sammut, C., & Webb, G. I. (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.

Schwantes, A. M., Swenson, J. J., & Jackson, R. B. (2016). Quantifying drought-induced tree mortality in the open canopy woodlands of central Texas. *Remote Sensing of Environment*, *181*, 54-64.

Senf, C., Pflugmacher, D., Wulder, M. A., & Hostert, P. (2015). Characterizing spectral–temporal patterns of defoliator and bark beetle disturbances using Landsat time series. *Remote Sensing of Environment*, *170*, 166-177.

Senf, C., Seidl, R., & Hostert, P. (2017). Remote sensing of forest insect disturbances: Current state and future directions. *International journal of applied earth observation and geoinformation*, *60*, 49-60.

Sidhu, N., Pebesma, E., & Câmara, G. (2018). Using Google Earth Engine to detect land cover change: Singapore as a use case. *European Journal of Remote Sensing*, *51*(1), 486-500.

Silva, C. R., Olthoff, A. E., de la Mata, J. A. D., & Alonso, A. P. (2013). Remote monitoring of forest insect defoliation. A review. *Forest Systems*, *22*(3), 377-391.

Simard, M., Romme, W. H., Griffin, J. M., & Turner, M. G. (2011). Do mountain pine beetle outbreaks change the probability of active crown fire in lodgepole pine forests? *Ecological Monographs*, *81*(1), 3-24.

Six, D. L., Vergobbi, C., & Cutter, M. (2018). Are survivors different? Genetic-based selection of trees by mountain pine beetle during a climate change-driven outbreak in a high-elevation pine forest. Frontiers in Plant Science, 9, 993.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*, 199-222.
Spruce, J. P., Hicke, J. A., Hargrove, W. W., Grulke, N. E., & Meddens, A. J. (2019). Use of MODIS NDVI products to map tree mortality levels in forests affected by mountain pine beetle outbreaks. *Forests*, *10*(9), 811.
Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Bray, M., & Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. *Advances in Space Research*, *50*(9), 1250-1265.
Stehman, S. (1996). Estimating the kappa coefficient and its variance under stratified random sampling. *Photogrammetric engineering and remote sensing*, *62*(4), 401-407.
Stephenson, N. L., Das, A. J., Ampersee, N. J., Bulaon, B. M., & Yee, J. L. (2019). Which trees die during drought? The key role of insect host-tree selection. *Journal of Ecology*, *107*(5), 2383-2401.
Szuster, B. W., Chen, Q., & Borger, M. (2011). A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Applied Geography*, *31*(2), 525-532.
Thanh Noi, P., & Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, *18*(1), 18.
Townsend, P. A., Singh, A., Foster, J. R., Rehberg, N. J., Kingdon, C. C., Eshleman, K. N., & Seagle, S. W. (2012). A general Landsat model to predict canopy defoliation in broadleaf deciduous forests. *Remote Sensing of Environment*, *119*, 255-265.
Udall, B., & Overpeck, J. (2017). The twenty-first century Colorado River hot drought and implications

for the future. Water Resources Research, 53(3), 2404-2418.

Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, *114*(1), 106-115.
Vogelmann, J. E., Kost, J. R., Tolk, B., Howard, S., Short, K., Chen, X., Huang, C., Pabst, K., & Rollins, M. G. (2010). Monitoring landscape change for LANDFIRE using multi-temporal satellite imagery and ancillary data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *4*(2), 252-264.
Vogelmann, J. E., Tolk, B., & Zhu, Z. (2009). Monitoring forest changes in the southwestern United States using multitemporal Landsat data. *Remote Sensing of Environment*, *113*(8), 1739-1748.
Vogelmann, J. E., Xian, G., Homer, C., & Tolk, B. (2012). Monitoring gradual ecosystem change using Landsat time series analyses: Case studies in selected forest and rangeland ecosystems. *Remote Sensing of Environment*, *122*, 92-105.
von Holt, B.-H. L., Westenberger, A., & König, I. R. (2022). Identification of representative trees in random forests based on a new tree-based distance measure. *bioRxiv*, 2022.2005. 2015.492004.
Wang, M., Wan, Y., Ye, Z., & Lai, X. (2017). Remote sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm. *Information Sciences*, *402*, 50-68.
Wayman, R. B., & Safford, H. D. (2021). Recent bark beetle outbreaks influence wildfire severity in mixed-conifer forests of the Sierra Nevada, California, USA. *Ecological Applications*, *31*(3), e02287.
Weed, A. S., Ayres, M. P., & Hicke, J. A. (2013). Consequences of climate change for biotic disturbances in North American forests. *Ecological Monographs*, *83*(4), 441-470.
Wickham, H., & Grolemund, G. (2017). R for data science: Import. *Tidy, transform, visualize, and model data*, *1*.

Williams, D. W., & Birdsey, R. A. (2003). Historical patterns of spruce budworm defoliation and bark beetle outbreaks in North American conifer forests [electronic resource].

Wilson, B. T., Lister, A. J., Riemann, R. I., & Griffith, D. M. (2013). Live tree species basal area of the contiguous United States (2000-2009).

Wood, S. L. (1982). The bark and ambrosia beetles of North and Central America (Coleoptera: Scolytidae), a taxonomic monograph. *The bark and ambrosia beetles of North and Central America (Coleoptera: Scolytidae), a taxonomic monograph.*

Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, *122*, 2-10.

Wulder, M. A., White, J. C., Carroll, A. L., & Coops, N. C. (2009). Challenges for the operational detection of mountain pine beetle green attack with remote sensing. *The Forestry Chronicle*, *85*(1), 32-38.

Ye, S., Rogan, J., Zhu, Z., Hawbaker, T. J., Hart, S. J., Andrus, R. A., Meddens, A. J., Hicke, J. A., Eastman, J. R., & Kulakowski, D. (2021). Detecting subtle change from dense Landsat time series: Case studies of mountain pine beetle and spruce beetle disturbance. *Remote Sensing of Environment*, *263*, 112560.

Zhao, F., Huang, C., & Zhu, Z. (2015). Use of vegetation change tracker and support vector machine to map disturbance types in greater yellowstone ecosystems in a 1984–2010 Landsat time series. *IEEE Geoscience and Remote Sensing Letters*, *12*(8), 1650-1654.

Zhao, Y., Huang, Y., Sun, X., Dong, G., Li, Y., & Ma, M. (2023). Forest Fire Mapping Using Multi-Source Remote Sensing Data: A Case Study in Chongqing. *Remote Sensing*, *15*(9), 2323.

Zhu, Z. (2017). Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS journal of photogrammetry and remote sensing*, *130*, 370-384.

Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment*, *144*, 152-171.

Zhu, Z., Woodcock, C. E., Holden, C., & Yang, Z. (2015). Generating synthetic Landsat images based on all available Landsat data: Predicting Landsat surface reflectance at any given time. *Remote Sensing of Environment*, *162*, 67-83.

https://wcc.sc.egov.usda.gov/reports/UpdateReport.html;jsessionid=HmvQ5ZapLP2AHZEUNPJtS0vpvpiLP2hwgbD4wZXc.nrcsprd0382?report=Upper+Colorado+River+Basin

Appendices

Appendix A - Background information

1 - Mountain Pine Beetle – Life cycle and Characteristics

The mountain pine beetle MPB (Dendroctonus Ponderosae) is a parasitic insect which is common in pine-forested areas of western North America. These species feed on several pine species including lodgepole, ponderosa, white bark, bristlecone, limber, and pinon pines, and in the process typically kill their hosts. The spread of beetle-killed forest areas commonly leads to subsequent wildfires, thus understanding their life cycles and feeding characteristics has become an important part of managing their destructive behavior.

Beetle development is controlled by temperature (Bentz et al., 1991). Newly developed adults utilize pheromones to attract other adults to stage mass attacks on healthy pines in sufficient numbers to overwhelm trees defenses. Therefore, weather factors, such as temperatures, especially increases in temperatures due to climate change, can cause considerable impacts on beetle development and mass attacks (Kunegel-Lion and Lewis 2020).

Temperature plays a critical role in the successful population growth of MPB by regulating the development and timing of each life stage. Typically, mountain pine beetles have a one-year life cycle, the most important part of which occurs during the summer when adults emerge and attack new host pines by drilling tunnels under the bark, where they mate, and females lay eggs. The eggs develop into larvae during the winter and then emerge as new adult following summer. Parent adults usually do not survive the winter. Each of the MPB life stages, including egg, larval, instars, pupa, and adult, has a set of minimum temperature requirements (>5.6 °C - > 15 °C) that are tied to the environmental conditions typically encountered at each stage (Bentz et al., 1991). For example, in colder regions, completing a life cycle can take 2 years, while, in contrast, some warmer areas show evidence of a second generation reaching maturation in the same summer. This relationship between life stages and temperature changes allows the simultaneous emergence of all adult beetles during a narrow temporal window in the summer (Bentze et al., ibid.). This simultaneous adult emergence allows the possibility of summer attacks and rapid colonization habits that are needed to overwhelm host defenses and kill the trees. Cooler conditions in inhospitable areas such as high-elevation pine forests delay MPB development and may result in beetle mortality and a decline local population because the different life stages are not suited to prevailing temperature trends. Beetle flights typically occur when summer temperatures are above 16 °C, but decline if there are prolonged periods of high temperatures of 38 C or above (Sambaraju and Goodman 2021). MPB flight is delayed and occurs over the shorter period of time in norther latitude, however, physiological changes to accommodate local environments may occur to enable successful colonization (Bleiker and Hezewijk, 2016). Egg laying and gallery construction are also influenced by temperature. The number of eggs laid per day, the number of eggs laid per unit length of gallery, and the daily gallery construction rate are all positively correlated with increases in temperature (Amman GD, 1972, and Sambaraju and Goodman, 2021). Winter temperatures are very important in determining the survival and spread of MPB. The species as a whole is intolerant to freezing but it has developed mechanisms to avoid cold injury in the larval stage. Temperatures below -23 °C kill eggs (Bleiker et al., 2017) but larva can survive to much lower temperatures than eggs, pupa, and adults (≤ 35 °C (Bentz and Mullins, 1999)). As temperature drops in the fall, cold-adaptive mechanisms are triggered in larva which include stopping food consumption, removing ice nucleators in bodily fluids, and accumulation of glycerol and other antifreeze compounds that allow them to survive otherwise freezing temperatures (Regniere and Bentz, 2007).

Mountain pine beetles attack trees in several ways. First, they burrow into the tree's bark and into the phloem. Additional beetles are then recruited to attack the trees via the release of pheromones by the initial attackers (Raffa, 1988). Beetles also introduce blue-stained fungus to trees which helps to kill the trees by blocking the xylem with fungal spores. The trees try to defend themselves by producing resin to physically expel the beetles and by producing defensive compounds designed to kill the beetles.

Unfortunately, these processes require a substantial investment in resources (Raffa and Berryman, 1983). Eventually, these mass attacks overwhelm the trees' defenses resulting in mortality of the host trees.

2 - Tree Defense Systems Against Beetle Attacks

Conifer defenses against bark beetles are complex and change with time. Defenses include: A) physical defenses such as the amount of resin flow from bark wounds (Strom et al., 2002); B) chemical defenses such as the composition and concentration of resin components, such as monoterpenes, sesquiterpenes, diterpenes, and phenolics (Raffa et al., 2017); C) inherent anatomical structures including the frequency and size of structures or tissues, such as resin ducts (Hood and Sala 2015), that produce and transport these defensive chemicals. Defense components can be (A) part of a tree's normal constitution, such as the normal defense system in tree tissues against bark beetle attack (Nebeker et al., 1992), or (B) induced, i.e., the changes in amount or character of these defenses in response to bark beetle attack including microorganisms such as the fungi brought in by bark beetles (Arango-Velez et al.,2018). Stress and physical damage to conifers often alter a tree's ability for induced defenses. One of the main stresses that affects a tree's defense system is drought or water stress (Arango-Velez et al.,2016). A variety of studies have shown that drought also alters the chemical composition of fluids in conifers resulting in a weakening tree defense by causing a reduction in the production of various compounds such as monoterpenes, etc., which act as a tree's antibodies (Keefover-Ring et al., 2016).

The size or maturity of a given tree appears to be significant in determining its ability to defend itself. According to Safranyik et al., (1974), MPB beetles did not attack trees with diameters less than 10 cm, and they showed that the rate of tree mortality increased 1.5 - 4 % with each centimeter increase in diameter.

Appendix B - Data Collection

https://code.earthengine.google.com/021b2d78514eace1c1458a3998167ed1

Appendix C - Time Series

https://code.earthengine.google.com/7b5fc7bb29fb835bbf30af549918c7de

Appendix D - Random Forest binary and regression 2005 and 2009

RF model binary (presence and absence of MPBB) for 2005 and 2009 and regression model for 2009

https://github.com/aajtaleb/Random-Forest-Model-05-and-09.git

Appendix E

Support Vector Machine binary and regression for 2005 and 2009

This is the SVM binary model (presence and absence of MPBB) 2005 and 2009 and regression for 2009

https://github.com/aajtaleb/support-vector-machine-25-and-09.git