AIMS Geosciences, 12(2): 360–387.
DOI: 10.3934/geosci.2026014
Received: 18 August 2025
Revised: 04 January 2026
Accepted: 10 February 2026
Published: 01 April 2026

https://www.aimspress.com/journal/geosciences

Research article

A quantitative framework for modeling the spatiotemporal dynamics of mountain pine bark beetle infestations utilizing Landsat time-series data and Random Forest

Hamza Taleb[1,*] and Melinda Laituri[2]

[1] Graduate Degree Program in Ecology (GDPE) at Colorado State University (CSU), Fort Collins, Colorado, USA
[2] Department of Ecosystem Science and Sustainability (ESS) at Colorado State University, USA

* **Correspondence:** Email: aaj@rams.colostate.edu; Tel: +17205875159.

Abstract: Mountain pine bark beetles (MPBB, *Dendroctonus ponderosae*) are a primary driver of tree mortality in pine-dominant North American forests. To characterize the spatiotemporal dynamics of MPBB infestations in lodgepole pine forests of north central Colorado, we analyzed Landsat spectral trends for 2005 and 2009. Using a stratified random sampling design (N = 1,021), we classified land cover trajectories to identify stable forests, infestation, fire, clear-cutting, and regrowth sites. A Random Forest (RF) classifier was developed to detect infestation presence and predict mortality severity. The model achieved high classification accuracies of 96% (2005) and 97% (2009), while the regression for mortality severity yielded a strong fit (R^2 = 0.878) with a low Root Mean Square Error (RMSE = 0.1425). A rigorous topographic analysis revealed that infestation risk is strongly non-random: South-facing slopes exhibited 9.0 times higher odds of infestation compared to north-facing slopes, likely driven by solar insolation and water stress. However, limitations remain in detecting low-severity mortality (<25% canopy loss) and distinguishing species-specific responses in mixed stands. A simple binary classification of "disturbed" versus "undisturbed," as is the primary output of many earlier studies, is insufficient for prioritizing management actions. To address these challenges, we propose integrating multi-source data fusion (e.g., Landsat and Sentinel-2), leveraging UAV-based sub-pixel validation, and utilizing phenological metrics from Harmonized Landsat-Sentinel (HLS) data. These advanced approaches, combined with the RF modeling demonstrated, offer a pathway for more precise, early-warning monitoring of forest health in complex topographic landscapes. In addition, this

approach can be used for future studies designed to track the location of trees that have developed self-immunity to the beetles with the hope of reforesting with seedlings of these resistant trees.

Keywords: tree mortality; MPBB; NBR; time series; Landsat TM; Random Forest; topographic analysis; data fusion; sentinel-2

1. Introduction

Forest ecosystems worldwide are experiencing an increase in the frequency and severity of disturbances, driven largely by climate change [1,2]. Among the most significant of these disturbances are eruptions of native bark beetles, which have resulted in extensive tree mortality across millions of hectares and are considered a primary driver of landscape-scale ecological change [3,4]. The mountain pine bark beetle (MPBB, *Dendroctonus ponderosae*) in western North America is a classic example of such an eruptive species. Recent outbreaks, particularly within lodgepole pine (*Pinus contorta*) ecosystems, have precipitated significant ecological and economic consequences, including substantial losses in timber value, increased costs for wildfire mitigation, and fundamental alterations to ecosystem structure and function [5–7]. The proliferation of MPBB populations is attributed to a confluence of factors, the increased prevalence of susceptible, mature hosts, a legacy of historical fire suppression, and regional climate change. Long-term warming trends have enabled beetle populations to expand into higher-elevation and higher-latitude habitats that were climatically unsuitable [8–10]. Warmer winters reduce overwintering mortality, a key population bottleneck, while accelerated developmental cycles can lead to two generations per year in some areas, dramatically increasing reproductive output [11,12]. Furthermore, prolonged drought conditions elevate the physiological stress on host trees, compromising their natural defensive capabilities, such as the production of resinous compounds, and rendering them vulnerable to successful colonization by beetles [13–15]. The complexity of these coupled biotic and abiotic drivers necessitates the development and application of advanced methodologies for monitoring and modeling infestation dynamics to inform effective forest management.

Historically, the systematic monitoring of forest disturbances has been conducted via the U.S. Forest Service's aerial detection surveys (ADS). While this method has generated an invaluable long-term observational record, it is constrained by several well-documented limitations, including the subjectivity inherent in manual interpretation, which can introduce observer bias and reduce data consistency [16,17], limited areal coverage due to the high costs of overflights, and the linearity of the arbitrary lines drawn to separate disturbed and undisturbed forest, when, in reality, outbreak patterns are very heterogenous in the presence/absence and degree of disturbance [18]. Most critically, validation studies have revealed that ADS can dramatically underestimate the true extent of infestation. For example, a landmark 2012 study by Meddens et al. revealed that when ADS-reported mortality areas were compared to those derived from high-resolution imagery, the ADS data underestimated the actual mortality area by a factor of 3.7 to 20.9 [19]. This means that for every 100 hectares of mortality mapped by ADS, the true affected area on the ground was between 370 and 2,090 hectares. This significant discrepancy is largely due to the inability of aerial observers to detect low-density mortality (i.e., scattered dead trees) and early-stage "green attack" infestations, which are spectrally indistinguishable from healthy trees to the human eye [18,20].

In response to these issues, automated algorithms that leverage the dense temporal information in the Landsat archive have become standard practice [21,22]. Algorithms like LandTrendr and BFAST apply temporal segmentation techniques to identify significant breakpoints in the spectral trajectory of each pixel, effectively flagging the timing, duration, and magnitude of change events [23,24]. These methods have proven powerful for mapping the spatial extent and timing of forest disturbances with high consistency and objectivity, providing valuable information on the binary presence or absence of change across the landscape [25,26]. The use of Short-Wave Infrared (SWIR)-based indices, such as the Normalized Burn Ratio (NBR), has been shown to be particularly effective for detecting changes in canopy water content associated with beetle-induced mortality [27].

Despite these advances, a critical knowledge gap persists in the operational monitoring of forest health. While methods excel at identifying *when* and *where* a disturbance occurs, they are often limited in their capacity to quantify the *magnitude* or *severity* of the event, particularly with respect to the percentage of tree mortality within a given pixel. For effective resource allocation, wildfire risk assessment, and ecological impact analysis, forest managers require information not only on the location of an infestation but also on its intensity. A simple binary classification of "disturbed" versus "undisturbed," as is the primary output of many change detection algorithms, is insufficient for prioritizing management actions, as it treats a pixel with 10% mortality the same as a pixel with 90% mortality. This lack of thematic detail is a direct consequence of the persistent challenge of sub-pixel spectral mixture, where the integrated spectral signal of a 30-meter Landsat pixel confounds the distinct signatures of healthy, infested, and non-host vegetation [28,29]. Resolving this mixture to derive quantitative estimates of canopy condition is a non-trivial problem, and its solution is a prerequisite for moving beyond simple disturbance mapping to true quantitative monitoring.

To address this challenge, researchers have increasingly turned to machine learning (ML) algorithms to model the complex relationships between spectral data and forest attributes [30,31]. Among these, the Random Forest (RF) algorithm has emerged as a particularly robust and effective tool for a wide range of classification and regression tasks in remote sensing [32,33]. In this study, we address the methodological gap by developing and implementing an integrative analytical framework that synergizes the temporal depth of the Landsat archive with the predictive power of the RF algorithm. The improved accuracy of this approach lies in its capability to overcome the limitations of prior methods. While studies have made important contributions to quantifying mortality, our approach represents a significant methodological advancement. For instance, Meddens et al. [34] utilized a parametric Maximum Likelihood Classifier to map mortality levels, an approach that assumes normal data distribution and can be less effective with the complex, non-linear relationships inherent in ecological data [35]. Our use of the non-parametric RF algorithm has a more robust response to noise and does not rely on assumptions about data distribution, making it better suited for deconvolving the mixed spectral signals of heterogeneous forests. Our model leverages the time-series of the Normalized Burn Ratio (NBR) as predictive features, in contrast to the single-date or bi-temporal anomaly detection used in earlier work. Our work is distinct from those of other researchers, such as Bright et al. [36], who focused primarily on evaluating the accuracy of aerial survey data by applying existing methods to use legacy data [32]. In contrast, we develop a novel, direct predictive model that uses RF regression to estimate percent mortality from the Landsat time-series. This enables us to perform not only a robust binary classification (presence/absence) but also a regression analysis to estimate the percentage of tree mortality at the sub-pixel level. By training the RF model with high-resolution reference data, we can deconvolve the mixed spectral signal and provide a quantitative estimate of

mortality that is not readily available from ADS, LandTrendr, or BFAST. This represents a significant methodological advancement, providing a more nuanced and managerially relevant assessment of forest disturbance.

Our primary objective of this research is to develop, apply, and validate this robust and scalable methodology for modeling the spatiotemporal dynamics of MPBB infestations at a regional scale. To achieve this, the study is structured around two aims: 1) To model the annual progression of MPBB infestation, including its presence and severity (percent mortality), in the lodgepole pine-dominated forests of north-central Colorado for the epidemic period of 2005–2009. The resulting model outputs will be rigorously validated against high-resolution National Agriculture Imagery Program (NAIP) data and ground-referenced field observations. 2) To quantify the influence of environmental drivers on infestation patterns by evaluating the relative importance of topographic variables, forest structural characteristics, and climatic parameters in predicting the spatial extent and temporal rate of MPBB spread. By fulfilling these objectives, we aim to advance the scientific understanding of insect-driven forest dynamics and to provide an operational, data-driven tool to support proactive and adaptive forest management in an era of rapid environmental change.

2. Study area

We examine the effects of tree-killing bark beetles in subalpine forests of Colorado in the Upper Yampa Basin (Figure 1). This is a mountainous region with a complex topography (3299–2291 m) [37,38] covering 696.92 km^2. 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 −12.3 ˚C), 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 [39]. The common tree species in subalpine forests in the study area include lodgepole pine (*Pinus contorta* Dougl. ex. Loud.), five-needle pines (i.e., bristlecone pine–*Pinus aristata* Engelm., limber pine–*Pinus flexilis* James, and southwestern white pine–*Pinus strobiformis* Engelm.), Engelmann spruce (*Picea engelmannii* Parry ex Engelm.), subalpine fir (*Abies lasiocarpa* (Hook.) Nutt.), and quaking aspen (*Populus tremuloides* Michx.) are present, although the composition varies across the region [34,38]. In subalpine forests, the most important tree-killing bark beetles are the mountain pine bark beetle (MPBB; *Dendroctonus ponderosae* Hopkins) that attacks lodgepole pine and five-needle pines, spruce beetle (SB; *Dendroctonus rufipennis* (Kirby), which attacks Engelmann spruce, and western balsam bark beetle (WBBB; *Dryocoetes confusus* Swaine), which attacks subalpine fir.

3. Methodology

3.1. Random Forest (RF)

RF is a decision tree algorithm that randomly selects a sample of observations and a sample of variables many times (default value of 500) to produce many small classification trees that are 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 [40]. 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. The goal is

finding the optimal number of predictors to generate the result which yields the highest accuracy. Each decision 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. 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 [40,41]. 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 decision trees in the forest resembling the bootstrap aggregating idea [42].

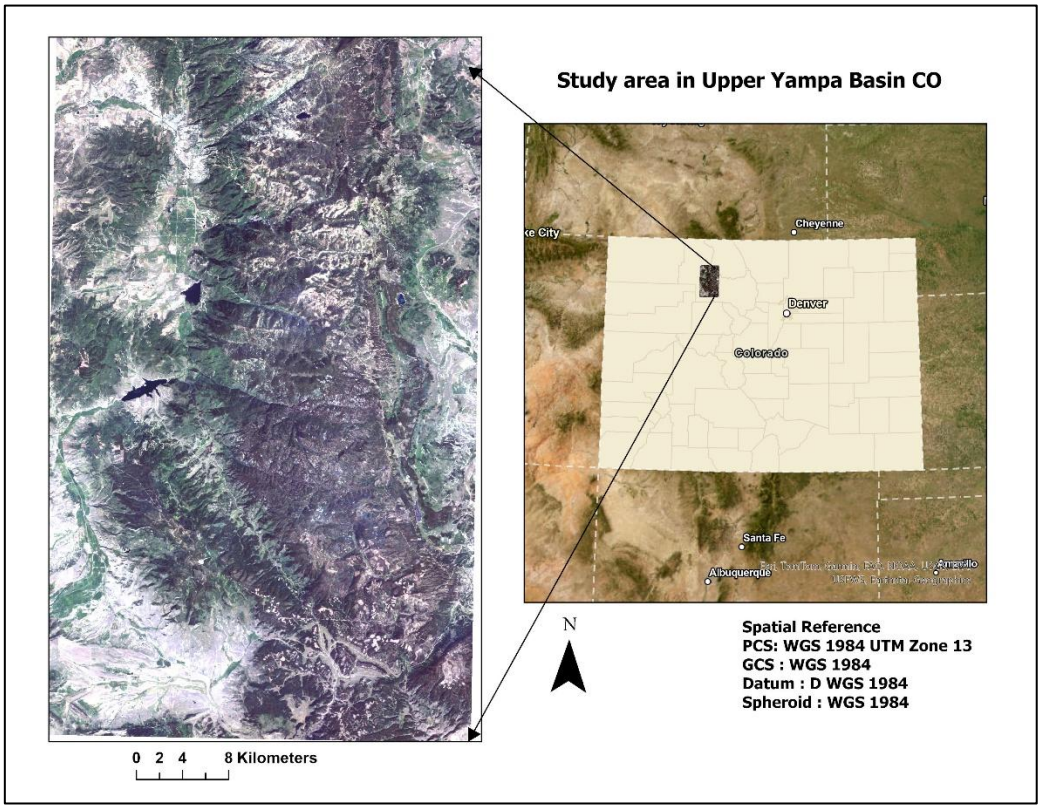

Figure 1. Study area.

The error rate of the forest can be measured using two values. The error rate depends on two calculations: The correlation between any two decision trees in the forest and the strength, or the error rate of each tree. For example, if we have n 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 n variables. RF also measures important variables 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 data from the number of correctly classified cases using non-permuted data will give the importance value for each n 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 [43].

3.2. Study plan

We followed a multi-stage workflow, as illustrated in the flowchart in Figure 2. The process began with the acquisition and preprocessing of satellite and topographic data, followed by the development of a robust reference dataset for model training and validation. The core of the analysis involved the development of two distinct RF models: A binary classification model to map the presence or absence of infestation, and a regression model to quantify the percentage of tree mortality within infested areas. The final stage involved model validation and the generation of disturbance maps for the study area.

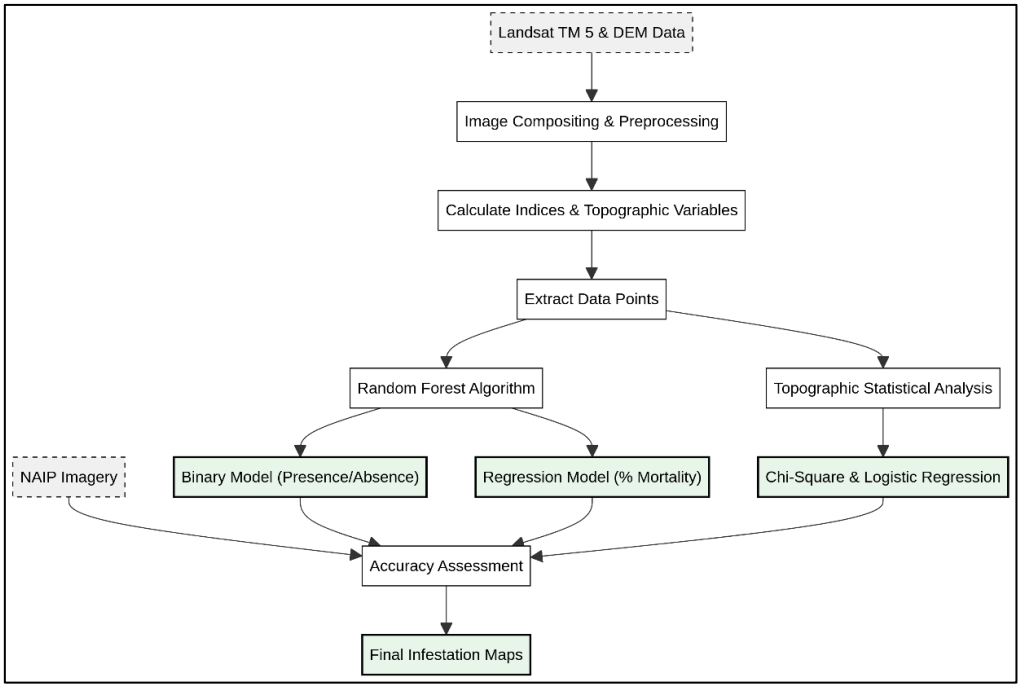

Figure 2. Flowchart of the methodology.

3.3. Landsat data collection and preparation

We utilized Landsat 5 Thematic Mapper (TM) Collection 2 Level 2 surface reflectance data, which were acquired via the Google Earth Engine (GEE) platform for the years 2000 to 2010. To minimize phenological variation and cloud contamination, annual, cloud-free image composites were generated for the peak growing season (May-September) of each year. This process was facilitated using functions from the LandTrendr algorithm framework within GEE, which selects the best available pixel from the time series based on a set of user-defined criteria. From these annual composites, we calculated a suite of spectral indices known to be sensitive to forest health, canopy structure, and moisture content (Table 1). To effectively detect change, it is crucial to establish a stable pre-disturbance baseline. Thus, we adapted the iterative method of Meddens et al. [34] to generate undisturbed mean NBR and B5/B4 images. This method iteratively identifies and masks out pixels that are spectrally anomalous (i.e., likely disturbed), recalculating the mean until a stable, representative baseline of the undisturbed forest state is achieved. Annual anomaly images were created by

subtracting this undisturbed mean from each annual image, providing a reliable input for detecting the timing and magnitude of forest disturbances.

Table 1. Vegetation Indices.

Index	Formula	Notes	reference
Normalized Burn Ratio (NBR)	(NIR-SWIR2)/(NIR+SWIR2	Moisture stress & needle structure	[35]
Normalized Difference Vegetation Index (NDVI)	(NIR-RED)/(NIR+RED)	Sensitive to conifer tree healthy	[44]
Band5/Band4 (B5/B4)	B5/B4	Sensitive to conifer tree healthy	[45]
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	[46]
Tasseled-Cap Greenness (TCG)	−0.1603 * B1 - 0.2819 - B2 * 0.4934 * B3 + 0.7940 * B4 - 0.0002 * B5 - 0.1446 * B7	Moisture stress	[46]
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	[46]
Normalized Difference Moisture Index (NDMI)	(NIR-SWIR1)/(NIR+SWIR1)	Sensitive to canopy water content	[47]

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

3.4. Reference and training data

3.4.1. Data sources and host species stratification

Our dataset was developed using a combination of high-resolution imagery and available field data. The primary source was 1-meter resolution aerial imagery from the National Agriculture Imagery Program (NAIP**)**, which provides sufficient detail to visually identify individual tree crowns and their health status (e.g., healthy green, infested red, and dead gray). This was supplemented with historical imagery from Google Earth Pro and plot-level data from the Insect and Disease Detection Surveys (IDDS).

To ensure our analysis targets mortality caused by the mountain pine bark beetle (*Dendroctonus ponderosae*), we first performed an ecological stratification. Using the LANDFIRE Existing Vegetation Type dataset, we created a forest mask that included only pixels classified as being dominated by lodgepole pine (*Pinus contorta*), the primary host for MPBB. This step is crucial for pest-specific attribution, as it minimizes the risk of confounding mortality from other pests that target different host species, such as the spruce beetle (*D. rufipennis*) or western balsam bark beetle (*Dryocoetes confusus*).

3.4.2. Sample selection and interpretation protocol

A total of 1,021 sample pixels were selected using a stratified random sampling approach to ensure adequate representation across land cover types and disturbance severities. While this stratification inherently resulted in spatial clustering (Nearest Neighbor Index = 0.13, $p < 0.001$), reflecting the patchy distribution of MPBB infestation and forest stands, spatial autocorrelation was mitigated by ensuring a minimum separation distance of 30 meters between sample centers. To ensure consistency, a single, trained analyst performed all image interpretations. Each 30x30 meter sample pixel was classified into one of five mutually exclusive categories based on a rigorous set of visual and temporal criteria (Table 2).

Table 2. Visual and Temporal Criteria for Reference Sample Classification.

Class	Primary Visual Indicators (from NAIP)	Temporal Signature (from Landsat)
Healthy Forest	Dense, contiguous green canopy; no significant gaps or discoloration.	Stable or slightly increasing NBR trajectory over time.
MPBB Infestation	Presence of red-needled trees (red-attack) or gray, needleless standing dead trees (gray-attack) in a "popcorn" or clustered pattern.	Gradual NBR decline over a 2–5-year period.
Fire	Complete or near-complete canopy loss; presence of blackened, charred soil and tree boles; sharp, distinct boundaries.	Abrupt, significant NBR drop in a single year.
Clear-cutting	Complete removal of all trees; visible stumps, skid trails, or logging roads; sharp, geometric boundaries.	Abrupt, significant NBR drop in a single year.
Regrowth	Dominated by small, young trees, shrubs, or dense herbaceous cover; lower canopy height than surrounding mature forest.	Gradually, sustained increase in NBR over multiple years.

3.4.3. Manual mortality estimation protocol

For the 162 pixels classified as infested, the percentage of tree mortality was manually estimated to provide the continuous dependent variable for our regression model. This was performed using an exacting dot-grid method in a GIS environment. A 10 ×10 grid (100 points) was overlaid on each 30 m pixel. At each of the 100 grid points, the underlying feature in the 1m NAIP imagery was classified as either “dead canopy” (red or gray foliage) or “other” (healthy canopy, shadow, understory, soil). The percentage mortality for the pixel was calculated as the number of points falling on dead canopy. To ensure the consistency and reliability of this manual process, a single analyst performed all estimations. Intra-observer reliability was assessed by having the same analyst re-evaluate 20% of the pixels (32 pixels) one month later. The resulting high level of agreement (Lin’s Concordance Correlation Coefficient = 0.94) confirmed the consistency of the manual estimation protocol.

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 × 1 m) was collected from National Agricultural Imagery Program (NAIP) (Aerial Photography Field Office (AFPO), 2012) for manual interpretation. NAIP images were repeated every 2 to 5 years, providing a continuous long-term record. There were 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 only as a general reference because there was not always a direct one-to-one correlation between reference data points and study data points due to 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 were 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). A total of 70% of total data points were selected as training data to determine the distribution of healthy versus unhealthy forest and degree of disturbance. The remaining 30 % of the data was used for validation.

3.5. Data selection for the region of interest

Ten points or regions of interest (ROI) were chosen based on NBR values and high-resolution images like NAIP to represent forested areas that were either healthy or had experienced events such as fire, clear cutting, or insect infestation (Table 3). These areas were selected so that they could be monitored over time using time series analyses. The assessment of health state and the occurrence of damaging events in each ROI was determined using Normalized Burn Ratio (NBR) values ranging from +1 to −1. These values were calculated for the years 2000, 2005, 2009, and 2010. High NBR values ranging from +1 to 0.4 indicated the presence of healthy trees, whereas low values ranging from 0.4 to −1 indicated areas that had been disturbed.

In addition to simply differentiating healthy vs. disturbed, the trajectories of time series analyses can offer insights into the nature of disturbances. Steep trajectories implied sudden events like clear cutting or fires, while gently sloped trajectories showed the progression of infection or regeneration following a fire or clear cutting. For each site and year in the time series (2000–2010), two satellite data bands (Band 4 and Band 7) (Table 3) were collected and analyzed. The NBR values were calculated using Equation (1).

$$NBR = \frac{(NIR-SWIR2)}{(NIR+SWIR2)} \quad \text{Eqe (1)}$$

Table 3. Spectral region of Landsat TM5 and ETM+.

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

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

The years 2005 and 2009 were selected since they enabled a comparison with high-resolution National Agricultural Image Program (NAIP) data for validation purposes. Four sites were chosen to exemplify historical events that resulted in distinct segmentation patterns.

The NBR thresholds used for classification were derived empirically from this ROI analysis (Table 4). ROIs that remained visually healthy (e.g., roi1 and roi4) consistently exhibited NBR values greater than 0.3. Infested ROIs (e.g., roi2 and roi7) showed a decline into the 0.1 to 0.3 range, while severely disturbed ROIs (e.g., roi9 and roi10) dropped to below 0.1. Moreover, values below 0.01 represented zones that were unvegetated, destroyed by fire, or clear-cut.

Table 4. 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
roi10	−106.641	40.09607	0.35697	0.314189	−0.09	−0.19169

3.6. Time series trajectory segmentation and interpretation

Key factors to consider in understanding time series trajectories include the year when the change was first detected, the direction of slope of the change segment, the duration of the change, and the amount of the change. The NBR values were used to establish three thresholds, which distinguished

five distinct trajectories: stable healthy forest, insect infestation, destruction by fire, clear cutting, and regrowth (Figure 3). For instance, when referring to forested areas, a Normalized Burn Ratio (NBR) value greater than 0.3 is considered indicative of health. The NBR in an area impacted by insect infestation declined from a value above 0.3 to 0.05 between 2005 and 2010. The NBR in a region impacted by a forest fire experienced a decrease from a value above 0.3 to −0.45 between 2001 and 2002. The clear-cut area had a decline from 0.3 to −0.2 between 2008 and 2010. These NBR values were derived by doing calculations on the L5 and L7 data for Band 4 and Band 7, using Equation (1). The disturbance segments were categorized based on whether they were linked to a prolonged decrease (lasting five or more years) or a brief decrease (less than three years). The duration of a specific disruption was determined by identifying significant changes in the slopes of the trajectory, from the starting year to the ending year.

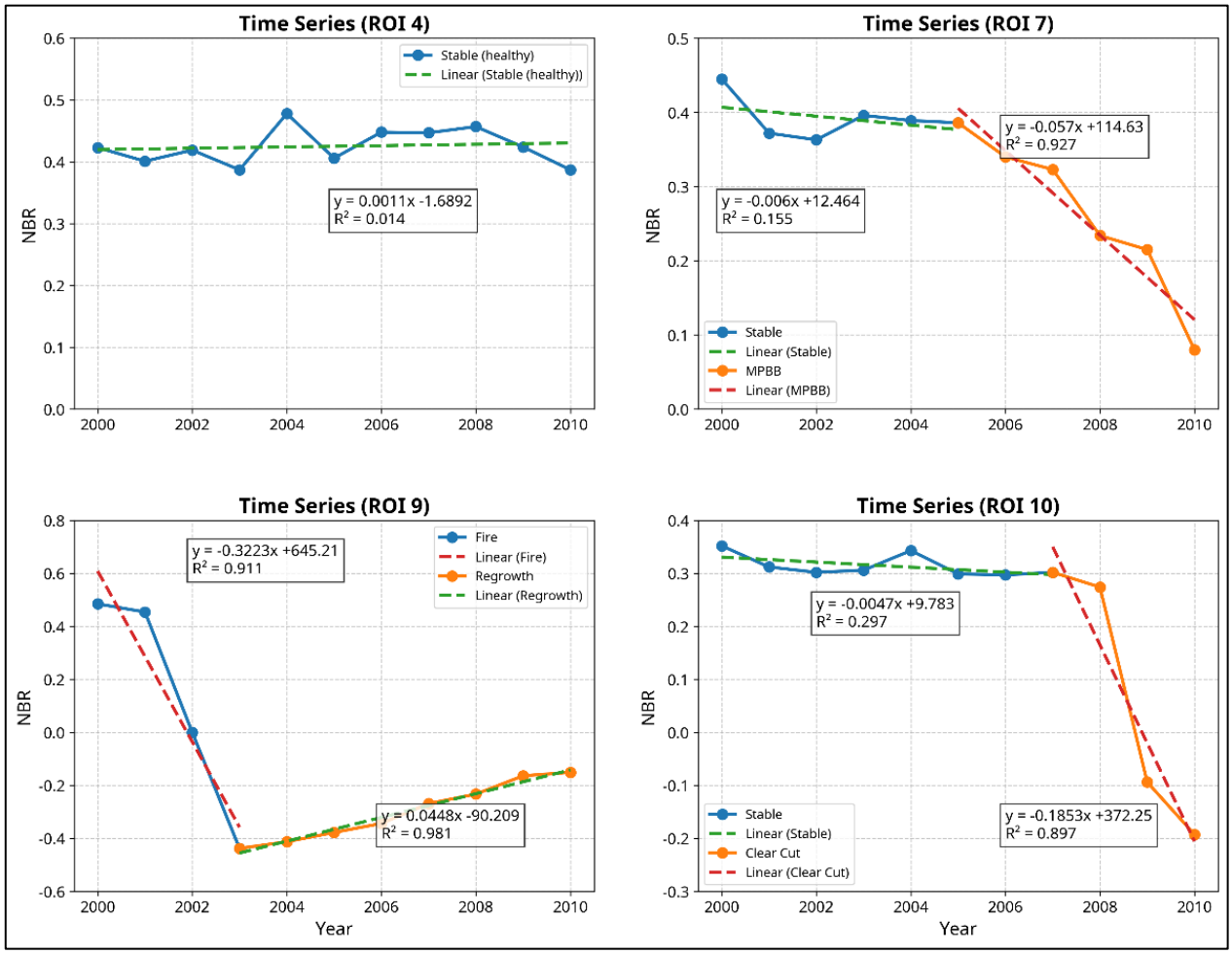

Figure 3. NBR time series analyses for 4 sampling locations.

Figure 3 shows changes in NBR values with time to show the patterns of change at 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). In Figure 3, ROI 9, c, there are signs of regrowth.

3.7. Developing a model to estimate the presence or absence of infestation and percent tree mortality

We developed two models using the RF algorithm. The 1,021 reference samples were split into a training set (70%) and a testing set (30%) for model development and validation [6,40,48] to avoid possible violations of normal linear regression [49].

3.7.1. Model 1: Binary classification (Infested vs. Non-infested)

This model was trained on all 1,021 samples to classify pixels as either infested or non-infested. The response variable was binary (1 or 0). In RF, the "rf.modelSel" function of the "rfUtilities" package [50] 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 that show correlation scores above important thresholds [51], using 10-fold cross-validation in the "caret" package to optimize hyperparameters [6,51]. In RF, all explanatory indices were used to determine the most important indices to accurately determine the 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 explained the most variations in the response variables using the fewest predictor variables was the most accurate (note: the response variable here denoted the presence or absence or percentage of infestation, while the predictor variables were NBR, B5/B4, TCG, band4, band 5, MeanB5/B4, DiffB5/B4, DEM, and AnomB5/B4) (Table 5).

Table 5. 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

Temporal anomalies represented 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 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 disturbed and undisturbed forest pixels. This model was designed to separate areas showing any signs of infestation from undisturbed areas.

3.7.2. Model 2: Regression (% Mortality)

This model was trained on the 162 infested pixels to predict the percentage of tree mortality. The response variable was the continuous percent mortality (0–100%) derived from our manual estimation protocol. Predictor variables included NBR, B5/B4, TCG, Mean B5/B4, Band5, and DEM. To give 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 corroborated with NBR and NDVI indices.

3.7.3. Hyperparameter tuning and accuracy assessment

For both models, hyperparameters (ntree and mtry) were tuned using 10-fold cross-validation to optimize performance. We tested all possible mtry values (1-6) and evaluated Out-of-Bag (OOB) error, test accuracy, and class-specific metrics. To successfully apply the RF algorithm, it was essential to configure two key parameters: The number of decision 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 for the program [48,52]. Liaw and Wiener [52] discussed that the inclusion of many trees in a model will yield a consistent estimation of variable importance. Furthermore, Breiman [40] noted that while employing too many trees may not improve results, it does not adversely impact the model. In addition, several authors [53,54] asserted that even using a minimum number of trees (ntree = 200), RF attained precise outcomes. Several authors discussed employing the default value mtry = $\sqrt{p}$ for the mtry parameter, where p represents the number of predictor variables [52,55]. In this study, an investigation was conducted to determine the most accurate RF model for classification [54]. To do this, a series of analyses, tests, and evaluations for several values were performed 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.

4. Results

Using 1,021 data samples (pixels) hand-selected and separated into disturbed (162) and undisturbed (859) categories, two 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, a programming toolbox and open-source software environment primarily designed for statistical computing and graphics [56]. Each model was analyzed by statistical methods to assess their accuracy.

4.1. Binary classification model

Data from two years in the time series were selected representing pre-infestation versus maximum infestation periods. Classification was done using R. R was used to analyze the two classes of data with the purpose of separating the study area into disturbed and undisturbed areas (Table 6), and to compare the predicted pixels with the actual data points (Refer Appendix A–R code).

Accuracy for this calculation was based on confusion matrix analysis [57]. 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 [57]. 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. The maximum-corrected binary model gave a simplified view of the presence or absence of infestations.

In Table 7, 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) [58].

Table 6. 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%

Tables 7. 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 classifications is called the Kappa coefficient. The Kappa coefficient takes into account the agreement that could be expected by chance and the actual agreement observed [59]. The formula for calculating the Kappa coefficient is:

2) $K = (Pa - Pe)/(1 - Pe)$

Here, Pa is the overall accuracy and Pe is the probability of agreement by chance [60]. The Kappa statistics serve 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 by an amount proportional to the probability of chance). The Kappa coefficient can range from −1 to 1. A value of 1 indicates perfect agreement. A value of 0 indicates agreement that is 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 was 0.84; and the Kappa value for 2009 was 0.92.

Output results of the binary model are represented in Figure 4, comparing the distribution of infested vs. non-infested sections of the study area for the years 2005 vs. 2009.

From this diagram, infestation spread significantly from 2005 to 2009, matching the trend seen in the time series analysis discussed in Section 3.6. After looking at topographic information, most of the spread was on southeast and southwest-facing slopes. Although there is some debate about this, topographic factors, including aspect, slope, and elevation, may also have major effects on local variability in temperature and moisture [61], which in turn affect a tree's ability to muster its defenses.

Nelson et al. [62] reported that warmer and drier slopes on southerly aspects at low elevation might decrease tree defenses against a beetle attack, while colder and wetter slopes on northerly aspects at high elevation increase tree defenses.

The binary classification model for 2005 and 2009 shows the spread of infestation.

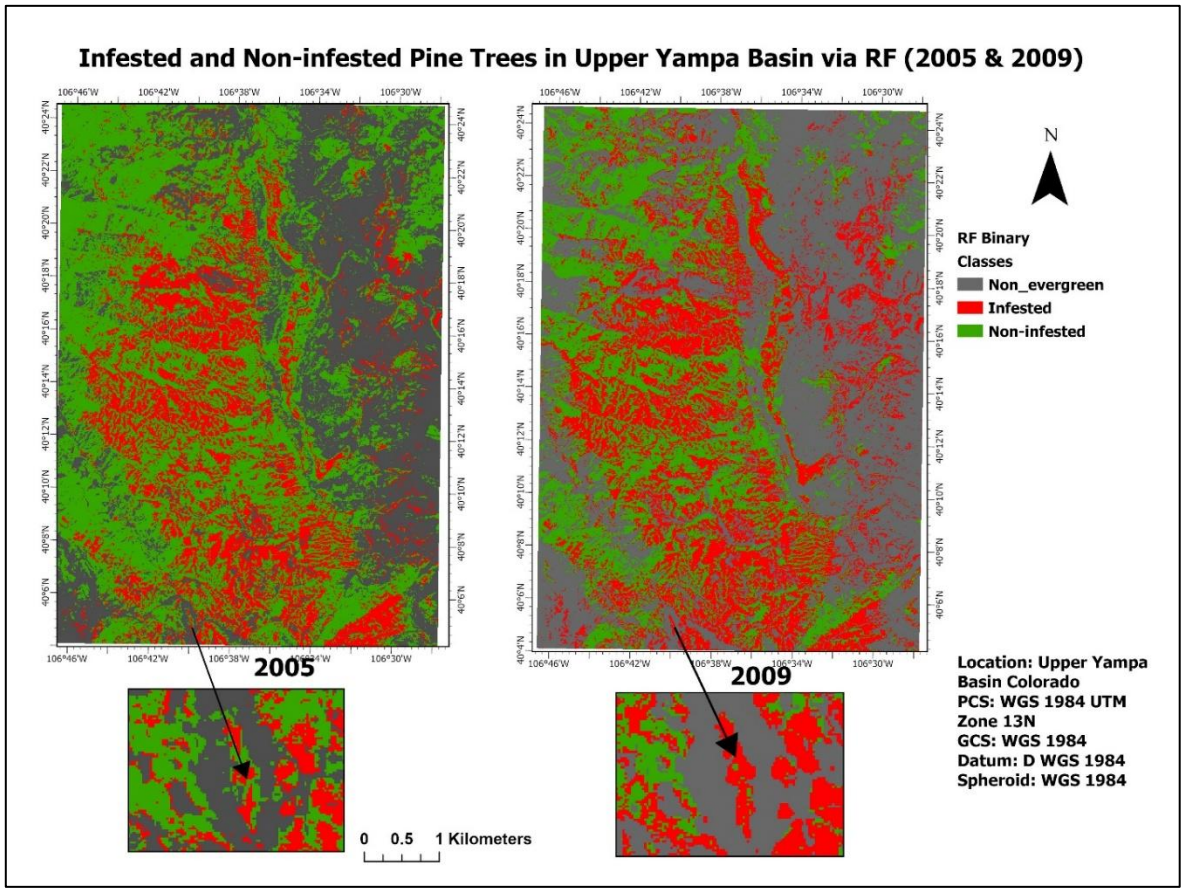

Figure 4. Binary classification.

4.2. Regression analysis

To predict the percentage of damage in any given area, a regression analysis is the most useful method [63]. Regression analysis is a statistical method used for modeling and analyzing the relationships between variables. It is 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 [63].

The "rf.modelSel" function in RF 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 [50]. In this study, ten predictor variables were chosen and correlated with one another, and, in particular, with NBR. In the next round of regression analysis, variables that correlated too strongly with the response variable were removed from the model because too many variables with high correlations will result in RF reducing the number of decision trees and limiting the accuracy of the results [55]. A 10-fold cross-validation in the "caret" package of R was used to assess the RF model performance in predicting new predictor variables [51].

A final RF model predicting percent mortality was applied to the data sets for 2005 and 2009. Predictor variables that were used in this model were: NBR, B5/B4, TCG, MeanB5/B4, Band5, and DEM. In the model, 500 decision trees used two data points at a time for analysis of the data set. The result showed 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 was the sum of the squared differences between each data point and the overall mean of the data set. Higher values of "% var" indicated that the model was capturing the underlying patterns in the data.

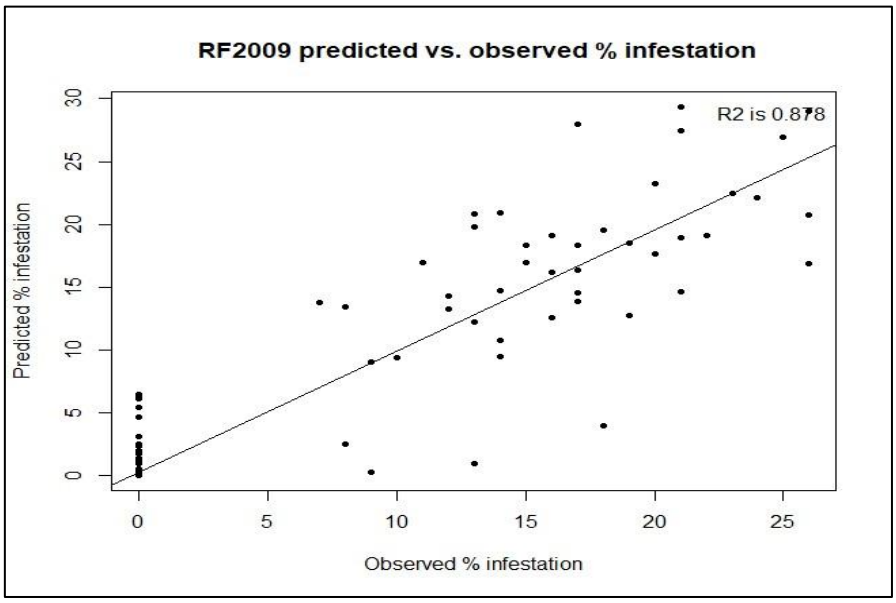

Figure 5. Scatter plot between predicted and observed percent of infestation. Note that the R^2 value for this scatter plot is 0.878 and RMSE is 0.1425.

A second RF model was run on just the testing data for 2009 (Figure 5). The predictor variables were the same; however, the 500 decision trees used 4 mtry data points at a time for each run. A lower RMSE indicated that the model's predictions were closer to the actual values, which meant the model was performing better. In this case, an R-squared value of 87.8% suggested that the model explained approximately 87.8% of the variability in the target variable (Figure 5), and the low RMSE of 0.1425 agreed that the model was performing very well. These data indicate that the model was capturing a substantial portion of the underlying patterns in the data.

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; and 1 indicates that the model perfectly explains all 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 of the percentage of tree mortality (Figure 6). The map on the left shows the distribution and percentage of damage across the study area; the map on the right is a close-up image showing the percentage damage map of a small area.

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

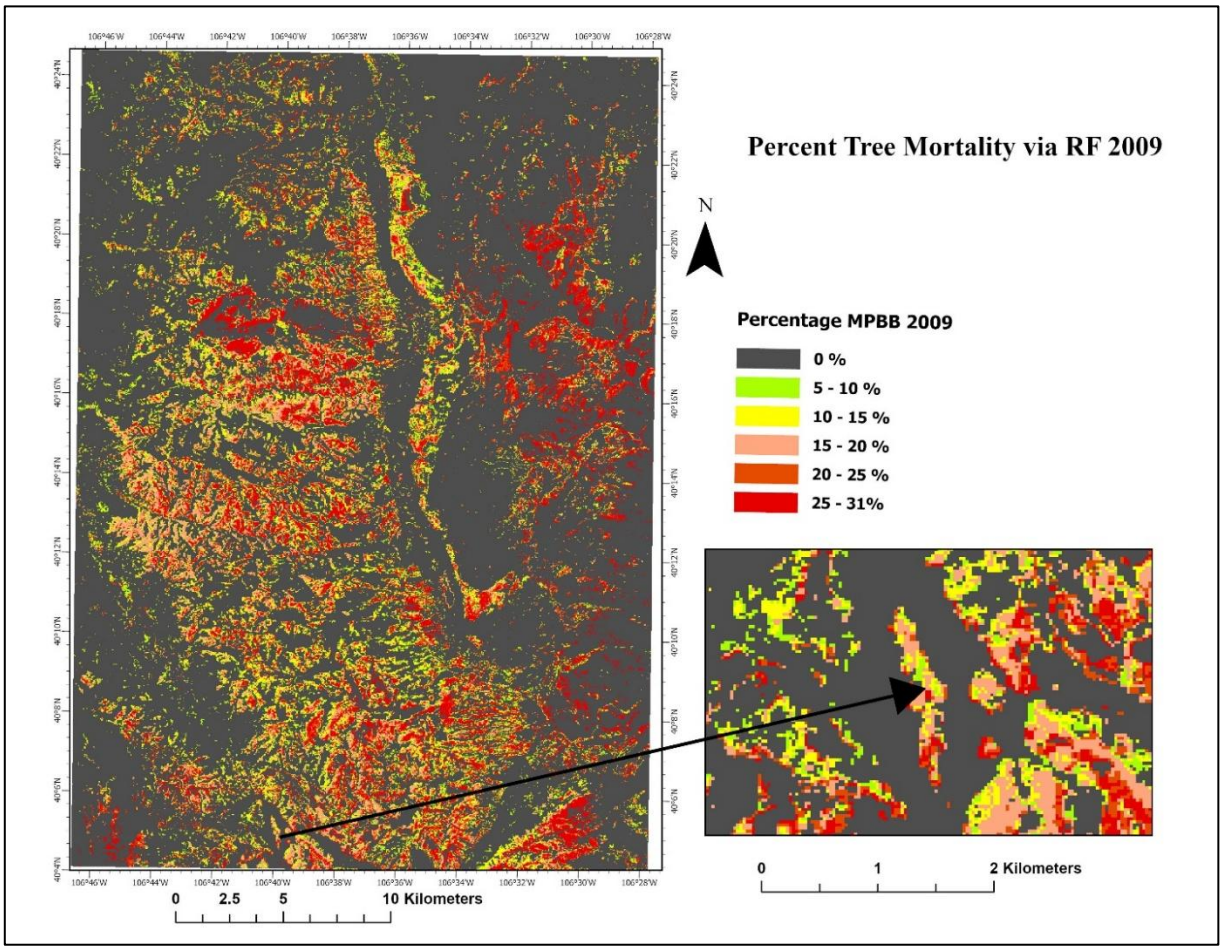

Figure 6. Percentage tree mortality.

4.3. Topographic influence on infestation spread

As noted in the binary classification model (Section 4.1), initial observations suggested that infestation spread was most prominent on south- and west-facing slopes. To statistically validate this observation and quantify the influence of topography on infestation patterns, a detailed analysis was conducted. These results aligned with research indicating that topographic factors such as aspect, slope, and elevation can significantly affect local microclimates (e.g., temperature and moisture), which mediate a tree's defensive capabilities and influence bark beetle population dynamics [9,61].

4.3.1. Chi-Square analysis of aspect

To test the hypothesis that infestation was not randomly distributed by aspect, a chi-square test of independence was performed. The analysis revealed a highly significant association between cardinal aspect and the presence of MPBB infestation (χ^2 (7, N = 1021) = 89.16, $p < 0.001$). This result provided strong statistical evidence that the spatial distribution of beetle-induced tree mortality was dependent on slope direction.

A detailed breakdown of infestation rates across aspect categories (Table 8) confirmed a distinct preference for southerly-exposed slopes. The highest rates were observed on the Southeast (29.0%), South (26.9%), and Southwest (19.1%) facing slopes. This pattern was consistent with the findings of Nelson et al. [62], who reported that warmer and drier slopes on southerly aspects can decrease tree defenses against beetle attack. The increased solar radiation on these slopes creates more favorable thermal conditions, which can accelerate beetle development and improve overwinter survival [64,65]. In contrast, northerly and westerly slopes in this study exhibited substantially lower infestation rates.

Table 8. Distribution and Rate of MPB Infestation by Aspect Category.

Aspect Category	Total Samples	Infested Samples	Non-Infested Samples	Infestation Rate (%)
North	360	18	342	5.0
Northeast	51	3	48	5.9
East	49	6	43	12.2
Southeast	31	9	22	29.0
South	446	120	326	26.9
Southwest	21	4	17	19.1
West	33	0	33	0.0

4.3.2. Logistic regression of topographic predictors

To further quantify the influence of topographic variables, a logistic regression analysis was performed to model the probability of infestation. The final model predicted the log-odds of infestation using elevation (DEM), south-facing aspect, and southeast-facing aspect as predictors. The model was statistically significant (χ^2 (3, N = 1021) = 118.0, $p < 0.001$) and explained approximately 13.2% of the variance in infestation status (Pseudo $R^2 = 0.132$).

Table 9. Logistic Regression Results for Topographic Predictors of MPB Infestation.

Predictor	Coefficient (β)	Std. Error	Odds Ratio (OR)	95% CI for OR	*p*-value
Intercept	−14.251	2.161	0.000	(0.000, 0.000)	<0.001
DEM (Elevation)	0.004	0.001	1.004	(1.002, 1.005)	<0.001
South Aspect	2.197	0.236	8.996	(5.667, 14.280)	<0.001
Southeast Aspect	1.923	0.455	6.844	(2.805, 16.697)	<0.001

The odds ratios (Table 9) confirmed the critical role of aspect and provided additional insight into the effect of elevation.

The analysis indicated that south-facing slopes had approximately 9 times higher odds of being infested compared to all other aspects, holding elevation constant. This strong effect was consistent with studies that identified south-facing slopes as having higher solar radiation loads, which creates thermally favorable conditions for beetle populations [64,66]. Furthermore, for every 100-meter increase in elevation, the odds of infestation increased by a factor of approximately 1.46, suggesting that higher-elevation sites were more susceptible within the study area. This finding aligns with research showing that warming trends are making higher-elevation habitats more thermally suitable for MPBB population growth and range expansion [67,68].

These results provide statistical support for the initial observation that topography, through its influence on microclimate, is a significant driver of MPBB infestation patterns in the study area. The warmer, sun-exposed microclimates found on southerly slopes, particularly at higher elevations, appear to be a critical predisposing factor for beetle-induced tree mortality [9,69].

5. Discussion

Our purpose of this study was to develop a method by integrated field and remotely sensed products to determine the presence and severity of insect-caused tree mortality in a 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 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 2005 and 2009, 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 in 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 regression analysis provided critical insights into the severity of tree mortality. The RF regression model achieved a high coefficient of determination ($R^2 = 0.878$), indicating that the model explained nearly 88% of the variance in the observed mortality data. The Root Mean Square Error (RMSE) was 0.1425, which underscored the model's precision in predicting the percentage of tree mortality. These metrics confirm that the combination of spectral indices (particularly NBR) and topographic variables can reliably quantify the gradient of infestation severity, not just its presence or absence.

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. [36]) 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 [70,71]. 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 minimize the danger of inappropriate extrapolation outside calibration areas.

Questions remain about the reliability of this analysis method and possible remedies. Researchers such as Meddens, et al. [34] have cautioned about the usefulness of using Landsat data for early stages of infestation or slowly progressing disturbances. Meddens et al. 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. They 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. [32] 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 fundamentally different, vector-based representations 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.

Advances in multi-sensor fusion offer promising solutions to these resolution challenges. The integration of Landsat with Sentinel-2 imagery, which offers higher spatial (10–20 m) and spectral resolution, has been shown to significantly improve the detection of mountain pine bark beetle infestations, particularly in the early stages [72,73]. Sentinel-2's inclusion of Red Edge bands provides enhanced sensitivity to the physiological stress signals, such as chlorophyll fluorescence reduction, that preceded visible needle fading [74]. The use of Harmonized Landsat-Sentinel (HLS) datasets enables the construction of dense, gap-free time series, enabling the retrieval of fine-scale phenological metrics [75]. These metrics can reveal subtle shifts in growing season timing or “browning” trends that serve as early warning indicators of forest stress before mortality becomes spectrally dominant [76]. For sub-pixel mortality estimation, the deployment of Unmanned Aerial Vehicles (UAVs) represents a critical leap forward. UAV-mounted sensors can capture tree-level data that bridges the gap between ground surveys and satellite observations [77], accurately detecting "hidden symptoms" of infestation weeks before they are visible to the naked eye [78].

The combination of RF models with NAIP images proved useful in identifying other forest disturbances. Although the predominant disturbance in the study area was mountain pine bark 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-cutting and fires had taken place, and the presence of these areas was confirmed through NAIP image examination so that they could be removed from areal beetle-kill calculation [35]. When using LTS and other remotely sensed data, it is challenging to distinguish areas that are unaffected and those with low-severity tree mortality [70,71,79]. For example, the occurrence and timing of spruce beetle outbreaks are difficult to determine when less than 35% of Landsat pixels are impacted [6]. Additionally, the role of mixed host composition in influencing spectral detection accuracy remains an underexplored area. While we focused on lodgepole pine (*Pinus contorta*), the presence of other susceptible five-needle pines, such as limber pine (*Pinus flexilis*), introduces variability. Different pine species possess distinct physiological defense mechanisms, such as varying terpene profiles and resin duct structures, which can influence the timing and intensity of their response to beetle attacks [80,81]. Consequently, the spectral trajectory of a dying limber pine may differ from that of a lodgepole pine due to differences in needle retention and canopy architecture [82]. In future research, researchers should investigate species-specific spectral signatures to refine detection algorithms in mixed-conifer landscapes [22].

For the 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 attacks was an important determinant of pine mortality [83]. Koll et al., [84] showed that pine forests under water stressed conditions are less able to use their normal defense systems against mountain pine bark beetle attacks. Mountain pine 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 does not. Mountain pine bark beetle attacks also induce an increase in resin flow for unstressed trees but not for water-stressed trees. Thus, mortality is highest for beetle-attacked and water-stressed trees [85]. The extensive bark beetle-caused tree mortality in forests throughout the SRM raises concerns about the trajectories of these ecosystems in a warmer future. The potential for numerous forest stands to revert to a similar normal structural state within a few decades is indicated by the abundance of advance regeneration and the establishment of new seedlings following a bark beetle outbreak [86–88]. Still, the density of post-outbreak stands, and the composition, are highly changeable because of outbreak severity, weather conditions, local microclimate, and previous disturbances due to fire or blowdown that impact regeneration dynamics [15,89–91]. While warming climate conditions by rising temperatures have the potential to increase mountain pine bark beetle activity in the future, these effects will be tempered by the reduction of appropriate host trees, changes in species composition, and differing recovery rates across the landscape [6,92].

6. Conclusions and limitations

Climate change is leading to increased tree mortality from mountain pine bark beetle outbreaks, wildfires, and drought-related forest disturbances. Understanding the consequences of disturbances on vegetation structure, composition, and spatial patterns is crucial for predicting future forest ecosystems. Remote sensing and field inventories are key techniques for analyzing and understanding ecosystem changes at large spatial scales. Our purpose of this study was 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 RF classifier algorithm to determine the presence or absence of MPBB infestations, and to predict the percentage of attacked trees in each infested area.

RF was successful in performing these analyses with high accuracy (96% for 2005 and 97% for 2009), permitting the construction of regional maps showing the distribution of infested areas and changes in that distribution over time. Our topographic analysis further revealed that infestation risk is non-randomly distributed, with south- and southeast-facing slopes exhibiting significantly higher odds of attack (OR = 8.99 and 6.84, respectively) compared to north-facing slopes. This suggests that solar insolation and associated microclimatic water stress are critical drivers of MPBB susceptibility in this region.

However, this study has several limitations. First, the stratified random sampling design resulted in statistically significant spatial clustering of sample points (Nearest Neighbor Index = 0.13), which, while reflective of the patchy nature of beetle outbreaks, may introduce spatial autocorrelation effects. Second, the 30-meter resolution of Landsat imagery limits the detection of low-severity mortality (under 25% canopy loss) and early-stage “green attack” phases. Furthermore, researchers should integrate higher-resolution sensors (e.g., Sentinel-2, UAVs) and phenological metrics from Harmonized Landsat-Sentinel

data to improve sub-pixel detection. Finally, while we focused on lodgepole pine, the spectral response of other host species (e.g., limber pine) to infestation remains underexplored. Investigating species-specific spectral trajectories would refine mortality estimates in mixed conifer stands. Despite these limitations, the study demonstrates that RF and Landsat time series can provide the critical, scalable information needed by forest managers to monitor and mitigate these expanding infestations.

Author contributions

Hamza Taleb: conceptualization, Methodology, Software, Formal data analysis, Writing original draft, Visualization. Melinda Latturi: Conceptualization, Methodology, Writing-review and editing, supervision.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

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. 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. Lastly, a special mention goes to my dear friend Dr. Sharon Stonecipher and CSU Geospatial Centroid for supporting and reviewing this work.

Conflicts of interest

The authors declare that there is no conflict of interest.

Appendix: Supplementary data

Appendix A: Data Collection

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

Appendix B: Time Series

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

Appendix C: 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.

References

1. Hicke JA, Meddens AJ, Kolden CA (2016) Recent tree mortality in the western United States from bark beetles and forest fires. *For Sci* 62: 141–153. https://doi.org/10.5849/forsci.15-086
2. Hicke JA, Xu B, Meddens AJ, et al. (2020) Characterizing recent bark beetle-caused tree mortality in the western United States from aerial surveys. *For Ecol Manage* 475: 118402. https://doi.org/10.1016/j.foreco.2020.118402
3. Janousek WM, Hicke JA, Meddens AJ, et al. (2019) The effects of mountain pine beetle outbreaks on avian communities in lodgepole pine forests across the greater Rocky Mountain region. *For Ecol Manage* 444: 374–381. https://doi.org/10.1016/j.foreco.2019.04.047
4. Jenkins MJ, Hebertson E, Page W, et al. (2008) Bark beetles, fuels, fires and implications for forest management in the Intermountain West. *For Ecol Manage* 254: 16–34. https://doi.org/10.1016/j.foreco.2007.09.045
5. Morris JL, Cottrell S, Fettig CJ, et al. (2018) Bark beetles as agents of change in social-ecological systems. *Front Ecol Environ* 16: S34–S43. https://doi.org/10.1002/fee.1754
6. Rodman KC, Andrus RA, Butkiewicz CL, et al. (2021) Effects of Bark Beetle Outbreaks on Forest Landscape Pattern in the Southern Rocky Mountains, USA. *Remote Sens* 13: 1089. https://doi.org/10.3390/rs13061089
7. Safranyik L, Carroll AL (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. Available from: https://publications.gc.ca/collections/collection_2012/rncan-nrcan/Fo144-4-2006-eng.pdf.
8. Jenkins MJ, Runyon JB, Fettig CJ, et al. (2014) Interactions among the mountain pine beetle, fires, and fuels. *For Sci* 5: 21–27. https://doi.org/10.3390/f5010021
9. Raffa KF, Aukema BH, Bentz BJ, et al. (2008) Cross-scale drivers of natural disturbances prone to anthropogenic amplification: the dynamics of bark beetle eruptions. *Bioscience* 58: 501–517. https://doi.org/10.1641/b580607
10. Weed AS, Ayres MP, Hicke JA (2013) Consequences of climate change for biotic disturbances in North American forests. *Ecol Monogr* 83: 441–470. https://doi.org/10.1890/13-0160.1
11. Buonanduci MS, Morris JE, Agne MC, et al. (2020) Neighborhood context mediates probability of host tree mortality in a severe bark beetle outbreak. *Ecosphere* 11: e03236. https://doi.org/10.1002/ecs2.3236
12. Koontz MJ, Latimer AM, Mortenson LA, et al. (2021) Cross-scale interaction of host tree size and climatic water deficit governs bark beetle-induced tree mortality. *Nat Commun* 12: 129. https://doi.org/10.1038/s41467-020-20455-y
13. Andrus RA, Chai RK, Harvey BJ, et al. (2021) Increasing rates of subalpine tree mortality linked to warmer and drier summers. *J Ecol* 109: 2203–2218. https://doi.org/10.1111/1365-2745.13634
14. Hart SJ, Veblen TT, Schneider D, et al. (2017) Summer and winter drought drive the initiation and spread of spruce beetle outbreak. *Ecology* 98: 2698–2707. https://doi.org/10.1002/ecy.1963
15. Harvey BJ, Andrus RA, Battaglia MA, et al. (2021) Droughty times in mesic places: factors associated with forest mortality vary by scale in a temperate subalpine region. *Ecosphere* 12: e03318. https://doi.org/10.1002/ecs2.3318

16. Senf C, Seidl R, Hostert P (2017) Remote sensing of forest insect disturbances: Current state and future directions. *Int J Appl Earth Obs Geoinf* 60: 49–60. https://doi.org/10.1016/j.jag.2017.04.004
17. Trumbore S, Brando P, Hartmann H (2015) Forest health and global change. *Science* 349: 814–818. https://doi.org/10.1126/science.aac6759
18. Meddens AJ, Hicke JA, Vierling LA (2011) Evaluating the potential of multispectral imagery to map multiple stages of tree mortality. *Remote Sens Environ* 115: 1632–1642. https://doi.org/10.1016/j.rse.2011.02.018
19. Meigs GW, Kennedy RE, Gray AN, et al. (2015) Spatiotemporal dynamics of recent mountain pine beetle and western spruce budworm outbreaks across the Pacific Northwest Region, USA. *For Ecol Manage* 339: 71–86. https://doi.org/10.1016/j.foreco.2014.11.030
20. Williams DW, Birdsey RA (2003) *Historical patterns of spruce budworm defoliation and bark beetle outbreaks in North American conifer forests : an atlas and description of digital maps*, US Department of Agriculture, Forest Service, Northeastern Research Station.
21. Bode ET, Lawrence RL, Powell SL, et al. (2018) Time-series approach for mapping mountain pine beetle infestation extent and severity in the US Central Rocky Mountains. *J Appl Remote Sens* 12: 046030. https://doi.org/10.1117/1.JRS.12.046030
22. Coops NC, Johnson M, Wulder MA, et al. (2006) Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation. *Remote Sens Environ* 103: 67–80. https://doi.org/10.1016/j.rse.2006.03.012
23. Vogelmann JE, Xian G, Homer C, et al. (2012) Monitoring gradual ecosystem change using Landsat time series analyses: Case studies in selected forest and rangeland ecosystems. *Remote Sens Environ* 122: 92–105. https://doi.org/10.1016/j.rse.2011.06.027
24. Wulder MA, Dymond CC, White JC, et al. (2006) Surveying mountain pine beetle damage of forests: A review of remote sensing opportunities. *For Ecol Manage* 221: 27–41. https://doi.org/10.1016/j.foreco.2005.09.021
25. Ahern F (1988) The effects of bark beetle stress on the foliar spectral reflectance of lodgepole pine. *Int J Remote Sens* 9: 1451–1468. https://doi.org/10.1080/01431168808954952
26. Cheng T, Rivard B, Sánchez-Azofeifa G, et al. (2010) Continuous wavelet analysis for the detection of green attack damage due to mountain pine beetle infestation. *Remote Sens Environ* 114: 899–910. https://doi.org/10.1016/j.rse.2009.12.005
27. Goodwin NR, Coops NC, Wulder MA, et al. (2008) Estimation of insect infestation dynamics using a temporal sequence of Landsat data. *Remote Sens Environ* 112: 3680–3689. https://doi.org/10.1016/j.rse.2008.05.005
28. Bentz B, Endreson D (2004) Evaluating satellite imagery for estimating mountain pine beetle-caused lodgepole pine mortality: current status, *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. 154–163. https://research.fs.usda.gov/treesearch/43497.
29. Jewett JT, Lawrence RL, Marshall LA, et al. (2011) Spatiotemporal relationships between climate and whitebark pine mortality in the Greater Yellowstone Ecosystem. *For Sci* 57: 320–335. https://doi.org/10.1093/forestscience/57.4.320

30. Franklin S, Wulder M, Skakun R, et al. (2003) Mountain pine beetle red-attack forest damage classification using stratified Landsat TM data in British Columbia, Canada. *Photogramm Eng Remote Sens* 69: 283–288. https://doi.org/10.14358/PERS.69.3.283
31. Long JA, Lawrence RL (2016) Mapping percent tree mortality due to mountain pine beetle damage. *For Sci* 62: 392–402. https://doi.org/10.5849/forsci.15-046
32. Bright BC, Hudak AT, Egan JM, et al. (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: 529. https://doi.org/10.3390/f11050529
33. Wulder MA, Masek JG, Cohen WB, et al. (2012) Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens Environ* 122: 2–10. https://doi.org/10.1016/j.rse.2012.01.010
34. Meddens AJ, Hicke JA, Vierling LA, et al. (2013) Evaluating methods to detect bark beetle-caused tree mortality using single-date and multi-date Landsat imagery. *Remote Sens Environ* 132: 49–58. https://doi.org/10.1016/j.rse.2013.01.002
35. Meigs GW, Kennedy RE, Cohen WB (2011) A Landsat time series approach to characterize bark beetle and defoliator impacts on tree mortality and surface fuels in conifer forests. *Remote Sens Environ* 115: 3707–3718. https://doi.org/10.1016/j.rse.2011.09.009
36. Bright BC, Hudak AT, Meddens AJ, et al. (2020) Mapping multiple insect outbreaks across large regions annually using Landsat time series data. *Remote Sens* 12: 1655. https://doi.org/10.3390/rs12101655
37. Romme WH, Floyd ML, Hanna D, et al. (2009) Historical range of variability and current landscape condition analysis: South Central Highlands section, Southwestern Colorado and Northwestern New Mexico: Colorado Forest Restoration Institute Fort Collins, Colorado, USA. https://hdl.handle.net/10217/240808
38. Wilson BT, Lister AJ, Riemann RI, et al. (2013) Live tree species basal area of the contiguous United States (2000–2009).
39. Group PC (2018) PRISM Climate Data. Oregon State University. https://prism.oregonstate.edu.
40. Breiman L (2001) Random forests. *Mach Learn* 45: 5–32. https://doi.org/10.1023/A:1010933404324
41. Peters J, De Baets B, Verhoest NE, et al. (2007) Random forests as a tool for ecohydrological distribution modelling. *Ecol Modell* 207: 304–318. https://doi.org/10.1016/j.ecolmodel.2007.05.011
42. Kulkarni AD, Lowe B (2016) Random forest algorithm for land cover classification. *Int J Recent Innov Trends Comput Commun* 4: 58–63.
43. Cutler DR, Edwards Jr TC, Beard KH, et al. (2007) Random forests for classification in ecology. *Ecology* 88: 2783–2792. https://doi.org/10.1890/07-0539.1
44. Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens Environ* 8: 127–150. https://doi.org/10.1016/0034-4257(79)90013-0
45. Vogelmann J (1990) Comparison between two vegetation indices for measuring different types of forest damage in the north-eastern United States. *Int J Remote Sens* 11: 2281–2297. https://doi.org/10.1080/01431169008955175
46. Crist EP (1985) A TM tasseled cap equivalent transformation for reflectance factor data. *Remote Sens Environ* 17: 301–306. https://doi.org/10.1016/0034-4257(85)90102-6

47. Wilson EH, Sader SA (2002) Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sens Environ* 80: 385–396. https://doi.org/10.1016/s0034-4257(01)00318-2
48. Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* 2: 18–22. https://journal.r-project.org/articles/RN-2002-022/
49. Schwantes AM, Swenson JJ, Jackson RB (2016) Quantifying drought-induced tree mortality in the open canopy woodlands of central Texas. *Remote Sens Environ* 181: 54–64. https://doi.org/10.1016/j.rse.2016.03.027
50. Evans JS, Murphy MA (2019) Package 'rfUtilities'. R Core Team: Vienna, Austria. Available from: https://cran.r-project.org/src/contrib/Archive/rfUtilities/.
51. Kuhn M (2015) Caret: classification and regression training. *Astrophysics Source Code Library* ascl-1505. https://ascl.net/1505.003.
52. Duro DC, Franklin SE, Dubé MG (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 Sens Environ* 118: 259–272. https://doi.org/10.1016/j.rse.2011.11.020
53. Laabs BH, Westenberger A, König IR (2024) Identification of representative trees in random forests based on a new tree-based distance measure. *Adv Data Anal Classif* 18: 363–380. https://doi.org/10.1007/s11634-023-00537-7
54. 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 Geosci Remote Sens Lett* 12: 1650–1654. https://doi.org/10.1109/LGRS.2015.2418159
55. 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: 18. https://doi.org/10.3390/s18010018
56. Wickham H, Grolemund G (2017) Tidy, transform, visualize, and model data, *R for data science*, https://r4ds.had.co.nz.
57. Sammut C, Webb GI (2011) *Encyclopedia of machine learning*, Springer New York, NY. https://doi.org/10.1007/978-0-387-30164-8
58. Zerrouki N, Bouchaffra D. Pixel-based or object-based: Which approach is more appropriate for remote sensing image classification? *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, San Diego, CA, USA, 864–869. https://doi.org/10.1109/SMC.2014.6974020
59. Cohen J (1960) A coefficient of agreement for nominal scales. *Photogramm Eng Remote Sens* 20: 37–46. https://doi.org/10.1177/001316446002000104
60. Stehman S (1996) Estimating the kappa coefficient and its variance under stratified random sampling. *Photogramm Eng Remote Sens* 62: 401–407.
61. Oke TR (2002) Boundary layer climates, Routledge.
62. Nelson TA, Boots B, Wulder MA, et al. (2007) Environmental characteristics of mountain pine beetle infestation hot spots. *J Ecosys Manag* 8. https://doi.org/10.22230/jem.2007v8n1a367
63. Freedman DA (2009) *Statistical models: theory and practice*, Cambridge University Press.
64. Bentz BJ, Logan JA, Amman GD (1991) Temperature-dependent development of the mountain pine beetle (Coleoptera: Scolytidae) and simulation of its phenology. *Can Entomol* 123: 1083–1094. https://doi.org/10.4039/Ent1231083-5
65. Mezei P, Potterf M, Škvarenina J, et al. (2019) Potential solar radiation as a driver for bark beetle infestation on a landscape scale. *Forests* 10: 604. https://doi.org/10.3390/f10070604

66. Bolstad PV, Bentz BJ, Logan JA (1997) Modelling micro-habitat temperature for Dendroctonus ponderosae (Coleoptera: Scolytidae). *Ecol Modell* 94: 287–297. https://doi.org/10.1016/S0304-3800(96)00021-X
67. Bentz BJ, Duncan JP, Powell JA (2016) Elevational shifts in thermal suitability for mountain pine beetle population growth in a changing climate. *Forestry* 89: 271–283. https://doi.org/10.1093/forestry/cpv054
68. Raffa KF, Powell EN, Townsend PA (2013) Temperature-driven range expansion of an irruptive insect heightened by weakly coevolved plant defenses. *Proc Natl Acad Sci U S A* 110: 2193–2198. https://doi.org/10.1073/pnas.1216666110
69. Logan JA, Bentz BJ (1999) Model analysis of mountain pine beetle (Coleoptera: Scolytidae) seasonality. *Environ Entomol* 28: 924–934. https://doi.org/10.1093/ee/28.6.924
70. Cohen WB, Healey SP, Yang Z, et al. (2017) How similar are forest disturbance maps derived from different Landsat time series algorithms? *Forests* 8: 98. https://doi.org/10.3390/f8040098
71. Kennedy RE, Yang Z, Gorelick N, et al. (2018) Implementation of the LandTrendr algorithm on google earth engine. *Remote Sens* 10: 691. https://doi.org/10.3390/rs10050691
72. Kluczek M, Zagajewski B (2025) Mapping spatiotemporal mortality patterns in spruce mountain forests using Sentinel-2 data and environmental factors. *Ecol Inform* 86: 103074. https://doi.org/10.1016/j.ecoinf.2025.103074
73. König S, Thonfeld F, Förster M, et al. (2023) Assessing combinations of Landsat, Sentinel-2 and Sentinel-1 time series for detecting bark beetle infestations. *GISci Remote Sens* 60: 2226515. https://doi.org/10.1080/15481603.2023.2226515
74. Bárta V, Lukeš P, Homolová L (2021) Early detection of bark beetle infestation in Norway spruce forests of Central Europe using Sentinel-2. *Int J Appl Earth Obs Geoinf* 100: 102335. https://doi.org/10.1016/j.jag.2021.102335
75. Tran KH, Zhang X, Ye Y, et al. (2023) HP-LSP: A reference of land surface phenology from fused Harmonized Landsat and Sentinel-2 with PhenoCam data. *Sci Data* 10: 691. https://doi.org/10.1038/s41597-023-02605-1
76. Grabska EM, Sulikowska A (2025) What can Sentinel-2 time series tell us about the phenology of dominant deciduous tree species in Poland? *Forestry* 99: cpaf064. https://doi.org/10.1093/forestry/cpaf064
77. Klouček T, Komárek J, Surový P, et al. (2019) The use of UAV mounted sensors for precise detection of bark beetle infestation. *Remote Sens* 11: 1561. https://doi.org/10.3390/rs11131561
78. Bozzini A, Huo LN, Brugnaro S, et al. (2025) Multispectral drone images for the early detection of bark beetle infestations: assessment over large forest areas in the Italian South-Eastern Alps. *Front For Glob Change* 8: 1532954. https://doi.org/10.3389/ffgc.2025.1532954
79. Coops NC, Shang C, Wulder MA, et al. (2020) Change in forest condition: Characterizing non-stand replacing disturbances using time series satellite imagery. *For Ecol Manage* 474: 118370. https://doi.org/10.1016/j.foreco.2020.118370
80. Erbilgin N, Ma C, Whitehouse C, et al. (2014) Chemical similarity between historical and novel host plants promotes range and host expansion of the mountain pine beetle in a naïve host ecosystem. *New Phytol* 201: 940–950. https://doi.org/10.1111/nph.12573

81. Kichas NE, Trowbridge AM, Raffa KF, et al. (2021) Growth and defense characteristics of whitebark pine (Pinus albicaulis) and lodgepole pine (Pinus contorta var latifolia) in a high-elevation, disturbance-prone mixed-conifer forest in northwestern Montana, USA. *For Ecol Manage* 493: 119286. https://doi.org/10.1016/j.foreco.2021.119286
82. Gibson K, Skov K, Kegley S, et al. (2008) Mountain pine beetle impacts in high-elevation five-needle pines: current trends and challenges. US Department of Agriculture Forest Service, Northern Region, Missoula, Montana R1-08-020, 1–32. Available from: https://www.govinfo.gov/app/details/GOVPUB-A13-PURL-LPS107316.
83. Restaino C, Young DJ, Estes B, et al. (2019) Forest structure and climate mediate drought-induced tree mortality in forests of the Sierra Nevada, USA. *Ecol Appl* 29: e01902. https://doi.org/10.1002/eap.1902
84. Knutzen F, Averbeck P, Barrasso C, et al. (2023) Impacts and damages of the European multi-year drought and heat event 2018–2022 on forests, a review. *EGUsphere* 2023: 1–56. https://doi.org/10.5194/egusphere-2023-1463
85. Kolb T, Keefover-Ring K, Burr SJ, et al. (2019) Drought-mediated changes in tree physiological processes weaken tree defenses to bark beetle attack. *J Chem Ecol* 45: 888–900. https://doi.org/10.1007/s10886-019-01105-0
86. Collins BJ, Rhoades CC, Hubbard RM, et al. (2011) Tree regeneration and future stand development after bark beetle infestation and harvesting in Colorado lodgepole pine stands. *For Ecol Manage* 261: 2168–2175. https://doi.org/10.1016/j.foreco.2011.03.016
87. Romme WH, Knight DH, Yavitt JB (1986) Mountain pine beetle outbreaks in the Rocky Mountains: regulators of primary productivity? *Am Nat* 127: 484–494. https://doi.org/10.1086/284497
88. Veblen TT, Hadley KS, Reid MS, et al. (1991) The response of subalpine forests to spruce beetle outbreak in Colorado. *Ecology* 72: 213–231. https://doi.org/10.2307/1938916
89. Pettit JM, Burton JI, DeRose RJ, et al. (2019) Epidemic spruce beetle outbreak changes drivers of Engelmann spruce regeneration. *Ecosphere* 10: e02912. https://doi.org/10.1002/ecs2.2912
90. Andrus RA, Hart SJ, Veblen TT (2020) Forest recovery following synchronous outbreaks of spruce and western balsam bark beetle is slowed by ungulate browsing. *Ecology* 101: e02998. https://doi.org/10.1002/ecy.2998
91. Gill NS, Jarvis D, Rogan J, et al. (2020) Disturbance history modulates how litter and herbaceous cover influence conifer regeneration after fire. *Int J Wildland Fire* 29: 519–529. https://doi.org/10.1071/WF19028
92. Temperli C, Veblen TT, Hart SJ, et al. (2015) Interactions among spruce beetle disturbance, climate change and forest dynamics captured by a forest landscape model. *Ecosphere* 6: 1–20. https://doi.org/10.1890/ES15-00394.1