Forest Ecology and Management 399 (2017) 9–23

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Mapping post-disturbance forest landscape composition with Landsat satellite imagery

Shannon L. Savage [a,*], Rick L. Lawrence [a], John R. Squires [b]

[a] Department of Land Resources & Environmental Sciences, Montana State University, PO Box 173120, Bozeman, MT 59717, United States
[b] Rocky Mountain Research Station, USDA Forest Service, 800 E. Beckwith, Missoula, MT 59801, United States

ARTICLE INFO

Article history:
Received 4 January 2017
Received in revised form 12 April 2017
Accepted 10 May 2017

Keywords:
Forest landscape composition
Spruce beetle
Disturbance ecology
Biotic legacy
Landsat satellite imagery
Percent canopy cover

ABSTRACT

Forests worldwide are impacted by a wide variety of disturbances that are happening more frequently with more intensity than in the past due to global climate change. Forest managers, therefore, need to identify new ways to quickly and accurately predict post-disturbance forest landscape composition. We suggest the use of Landsat satellite imagery and an image processing tool to map percent canopy cover (PCC) by species and sub-canopy species counts to be used in adaptive forest management strategies. We used zero-inflated models to successfully predict PCC and sub-canopy counts (number of regenerating trees per pixel, also called biotic legacies) for 4 tree species, along with overall PCC and percent mortality, for a large portion of the Rio Grande National Forest (RGNF) in 2013. The RGNF had recently been disturbed by spruce beetle (*Dendroctonus rufipennis*) infestation since the early 2000s and the West Fork Fire Complex in 2013. Our PCC models resulted in pseudo median differences between observed and predicted values of 0.2–6.5%, RMSE of 10.9–17.0%, and 95% confidence interval widths of 4.4–24.9%, depending on the species. The percent mortality model resulted in pseudo median differences between observed and predicted values of 1.1%, RMSE of 12.4%, and 95% confidence interval width of 4.6%. The sub-canopy PCC model resulted in a pseudo median differences between observed and predicted values of 1.3%, RMSE of 9.4%, and 95% confidence interval of 3.0%. The sub-canopy count models resulted in mean differences of 0.1–1.4 trees, RMSE of 3.0–13.4 trees, and 95% confidence interval widths of 1.1–5.0 trees, depending on species. By mapping PCC and sub-canopy counts, we have provided forest managers with knowledge of the current surviving forest (PCC) as well as the biotic legacies (sub-canopy counts) that can aid in forming hypotheses as to what the forest might become in the future, adding to the forest manager toolbox for forest management strategies. The methods described can be applied to a variety of issues within the field of disturbance ecology and, combined with change analyses, will provide forest managers with empirical evidence of current and future forest composition along with biological legacies that will impact forest regeneration.

1. Introduction

Large portions of forests worldwide have been disturbed over the past 50 years (Hansen et al., 2013; Hicke et al., 2016), and land managers face the daunting task of determining how these changes will affect their forests ecologically, economically, and historically. Land managers need a new way to quickly and accurately predict forest succession, since successional patterns, fire regimes, and pathogen resistance are evolving with the changes brought on by global climate change (Hicke et al., 2016). Adaptive management strategies that incorporate ecological principles and climate change-related effects are realistically the most appropriate approach for today's land managers (Crisafulli et al., 2005; Millar et al., 2007; Negrón et al., 2008; Schmid and Frye, 1977; Temperli et al., 2015).

Global climate change is a critical driving force in forest ecosystem balance (IPCC, 2014; Kennedy et al., 2014), resulting in severe disturbances in many locations across the globe. Bark beetles in particular have been observed to move to higher latitudes and elevations due to higher temperature suitability as well as depletion of host trees from drought and drought stress, and tree species at these higher locations have not yet adapted to insect outbreaks and are less likely to survive an outbreak (Bentz et al., 2010; Hart et al., 2014; Hicke et al., 2006, 2016; Jenkins et al., 2014; Jewett et al., 2011; Temperli et al., 2015).

* Corresponding author.
E-mail address: shannon.savage@msu.montana.edu (S.L. Savage).

http://dx.doi.org/10.1016/j.foreco.2017.05.017
0378-1127/

Land managers are limited in their ability to plan for economic, ecological, and historic impacts and protection without knowledge of the current species composition of their forests, particularly after a large disturbance. Land managers can make more informed decisions – and perhaps better decisions – when they approach their tasks through the lens of disturbance ecology. Traditional disturbance response/succession models tend to be long term and linear – i.e., moving from stand initiation through stem exclusion to understory re-initiation and finally arriving at an old growth stage with the process taking decades or centuries (Oliver, 1980) – leaving near-term (i.e., shortly after a disturbance) forest composition and structure uncertain, since successional patterns have changed due to global climate change (Derderian et al., 2016; Jewett et al., 2011; Temperli et al., 2015).

Disturbance ecology considers how the dynamic ecosystem might change over time (Bebi et al., 2003; Kennedy et al., 2014; Kulakowski et al., 2003; Lawrence and Ripple, 1999, 2000; McDowell et al., 2015; Walker et al., 2007). We have made great progress in understanding the driving forces and impacts of disturbances (Kennedy et al., 2014; Kulakowski et al., 2003). Biological legacies – vegetation that remains post-disturbance, including regeneration and colonizing vegetation – are among the most important variables influencing forest recovery after a disturbance (Crisafulli et al., 2005; Lawrence and Ripple, 2000; Sibold et al., 2007). Observation of forest vegetation patterns over time is essential for disturbance-related prevention, suppression, and restoration of forest ecosystems (Crisafulli et al., 2005; Fettig et al., 2007; Jenkins et al., 2014; Walker et al., 2007). Mortality of large trees in a forest stand from beetle-kill or stand-replacing fires, for example, could open the stand to faster and more dense regeneration (Derderian et al., 2016; Sibold et al., 2007) through reduction of competition for biological legacies. Efficient mapping of post-disturbance biological legacies using remote sensing has the potential to greatly assist in post-disturbance forest evaluation and planning.

Remote sensing scientists can take advantage of free Landsat satellite imagery to map many different types of ecosystems. Few studies have used Landsat imagery to predict species composition of a young forest after significant disturbance, or more specifically, through the lens of disturbance ecology. Lawrence and Ripple (2000) looked at the response of the Mount St. Helens landscape to the drastic changes caused by the volcanic eruption in 1981 and suggested that landscape-scale understanding of the distribution of biological legacies is critical to understanding post-disturbance vegetation responses. More recent studies have shown promise for mapping species composition and/or mortality as a percentage of each pixel rather than simply presence/absence. Savage et al. (2015) used zero-inflated modeling to accurately map percent canopy cover by species in northwestern Montana. Several recent studies have utilized Landsat imagery to detect tree mortality. Percent mortality within pixels was predicted using zero-inflated modeling for portions of the Helena National Forest in Montana with very accurate results (Long and Lawrence, 2016). A study in Texas was able to accurately detect drought-induced tree canopy loss using zero-or-one-inflated beta regression (Schwantes et al., 2016). In southwestern Colorado researchers successfully applied methods for classing mountain pine beetle-induced tree mortality and ecologically informed post-classification correction to detect spruce beetle-induced tree mortality (Hart and Veblen, 2015).

Our primary goal for this study was to characterize the structure and composition of a mixed conifer forest after major beetle-kill and large fires by using freely available Landsat imagery to evaluate within-pixel percent mortality and predict surviving tree composition by species for both dominant upper canopy species and sub-canopy regeneration. A secondary objective was to test and compare a series of prediction algorithms that have been previously applied to remotely sensed data in order to maximize predictive accuracy (as demonstrated in Savage et al. (2015)). We suggest that remote sensing applications can play a key role in monitoring ecosystems and assisting planning for future management (Nagendra et al., 2013). Our image analysis method is a powerful tool that allows land managers to predict and project compositions of future forests following disturbance (Long and Lawrence, 2016; Savage et al., 2015).

2. Methods

2.1. Study area

The Rio Grande National Forest (RGNF) covers 783,742 ha in Colorado and includes portions of the San Juan mountain range east of the continental divide. The study area covers approximately 347,000 ha, includes much of the RGNF Divide and Conejos Peak ranger districts (Fig. 1), and falls within one Landsat scene. It is a mountainous region ranging from 2550 to 4280 m in elevation with a variety of grassland, brushland, and forest types.

Rapid ecological changes have occurred in the high-elevation spruce-fir (*Picea engelmannii – Abies lasiocarpa*) zone within the RGNF due to a spruce beetle (*Dendroctonus rufipennis*) outbreak in the early 2000s that affected a large portion of the spruce-fir forests of the study area. Approximately 85% of the mature spruce-fir habitat on the RGNF had been influenced by the end of 2013 (Blakeman, 2013; RWEACT, 2016). The beetle-kill trees contributed to the dead fuel load of the forest, and the West Fork Fire Complex of 2013 burned approximately 44,515 ha of spruce-fir/aspen (*Populus tremuloides*) mix on the San Juan National Forest (SJNF) and RGNF (USFS, 2014). The fires, initially starting on the west (SJNF) side of the Continental Divide, spread to about 35,612 ha on the RGNF. Much of the burn occurred in spruce-fir cover types that already had significant rates of tree mortality due to the spruce beetle.

2.2. Data

2.2.1. Field data collection

We collected overstory and sub-canopy percent canopy cover (PCC) and sub-canopy tree count reference data in 2015 at 463 field locations randomly located within 500 m from roads and trails within the study area. Points were reviewed to be spatially homogenous and more than 40 m from the edge of a stand. Several points were excluded due to cloud cover in the imagery, providing 454 total reference points for use in modeling, exceeding the results of a Power test for proportion that suggested a minimum of 384 random sample points for our study (Chow et al., 2008).

We used two types of data collection methods: (1) 20-m × 20-m grid plot to estimate PCC and mortality (per Savage et al., 2015) and (2) line-intercept sampling to estimate sub-canopy count/-composition. Field crews established sample points on a 20-m by 20-m grid oriented to the north. Crews used a moosehorn, a tool for vegetation sampling (Fiala et al., 2006), to identify presence or absence of canopy cover every 5 m within the 20 × 20 grid (25 readings per field data point). Every tree observed as present with the moosehorn was identified (1) by species, (2) as live, red (dead or dying with red needles that have not fallen yet), or dead, and (3) whether upper canopy or sub-canopy. The total PCC for each species was calculated by adding together the number of times that species was listed as "live" in the upper canopy within the 25-point grid and multiplying that number by 4. Percent mortality was calculated by counting the number of times a tree's condition was listed as "dead" in the upper canopy within the 25-point grid and multiplying that number by 4 (regardless of species). The

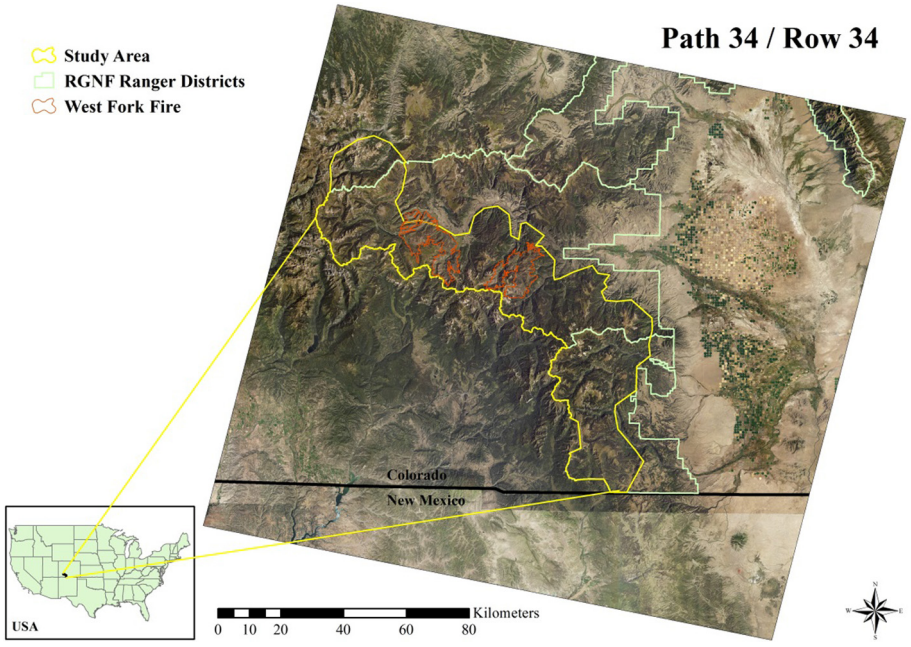

Fig. 1. Location map of the Colorado study area showing the Rio Grande National Forest (RGNF) ranger districts, West Fork Fire perimeters, and the boundary of the surrounding Landsat scene. The study area is displayed over Colorado and New Mexico 1-m NAIP (National Agriculture Imagery Program) imagery.

sub-canopy PCC was calculated by counting the number of times any tree was observed and listed as "live" in the sub-canopy within the 25-point grid and multiplying that number by 4 (regardless of species).

The line-intercept sampling was applied between each of the 25 grid points described above, where the presence of forest sub-canopy was recorded as a count of each sub-canopy species observed (live, red, or dead) on the line between points. The sub-canopy counts were calculated by adding the number of times a tree was recorded as "live" in the sub-canopy line intercept data. This was done for all instances to calculate total sub-canopy count as well as separately for each species observed.

We focused our analyses on four tree species that on average collectively represented approximately 94% of the total observed canopy cover and 97% of the live sub-canopy counts within our reference data: (1) subalpine fir (10% of total canopy cover), (2) Engelmann spruce (27% of total canopy cover), (3) quaking aspen (43% of total canopy cover), and (4) Douglas-fir (*Pseudotsuga menziesii*) (14% of total canopy cover). The remaining eight species were observed only rarely at these sites: (1) Rocky Mountain white fir (*Abies concolor*) (<1% of total canopy cover), (2) mountain alder (*Alnus tenuifolia*) (<1% of total canopy cover), (3) bristlecone pine (*Pinus aristata*) (3% of total canopy cover), (4) lodgepole pine (*P. contorta*) (0% of total canopy cover), (5) pinyon pine (*P. edulis*) (<1% of total canopy cover), (6) limber pine (*P. flexilis*) (1%of total canopy cover), (7) ponderosa pine (*P. ponderosa*) (<1% of total canopy cover), and (8) blue spruce (*Picea pungens*) (1% of total canopy cover).

2.2.2. Digital data acquisition and pre-processing

Airborne and satellite imagery are freely and easily accessible from a variety of outlets. Imagery from sensors that collect visible and infrared signatures is often used by image analysts to map vegetation communities (USGS, 2017; Xie et al., 2008). The most common bands of the electromagnetic spectrum used for distinguishing vegetation species include green, red, and near infrared (NIR), while the blue and short-wave infrared (SWIR) bands are helpful when considering vegetation moisture (i.e., deciduous versus conifer species) (USGS, 2017). Derived data such as tasseled cap (TC) and normalized difference vegetation index (NDVI) are especially useful in mapping vegetation with moderate resolution imagery. The TC transformation reduces the spectral information within the image (blue, green, red, NIR, SWIR) into three ecologically interpretable bands: (1) brightness (soil brightness or total reflectance), (2) greenness (relative amounts of leafy green vegetation), and (3) wetness (soil moisture status) (Crist and Cicone, 1984). The NDVI calculation uses the red and NIR bands of the electromagnetic spectrum to identify the presence of leafy green vegetation. These indices can be calculated with Landsat data and are commonly included in analyses for regional scale community-level vegetation mapping (Xie et al., 2008).

We downloaded the mostly cloud-free Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) images for Path 34/Row 34 from August 6, 2015 and June 16, 2014 from the USGS EROS Center (Table 1). The images were rectified by the USGS EROS Center in UTM coordinate system, Zone 13, WGS84 datum. There were no cloud-free Landsat 8 images acquired by the

Table 1
Data used in the predictions. Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) from Path 34, Row 34, August 6, 2015 with cloud-fill from June 16, 2014.

Number	Component name
1	August 2015 – OLI Band 1, Coastal Aerosol
2	August 2015 – OLI Band 2, Blue
3	August 2015 – OLI Band 3, Green
4	August 2015 – OLI Band 4, Red
5	August 2015 – OLI Band 5, NIR
6	August 2015 – OLI Band 6, SWIR1
7	August 2015 – OLI Band 7, SWIR2
8	August 2015 – OLI Band 10, TIR1
9	August 2015 – OLI Band 11, TIR2
10	DEM
11	Aspect (9 categories)
12	Slope (percent)
13	Tasseled Cap Brightness
14	Tasseled Cap Greenness
15	Tasseled Cap Wetness
16	NDVI
17	2015 NAIP texture mean
18	2015 NAIP texture minimum

satellite for the scene that covers our study area for either 2015 or 2014, therefore, the most cloud-free 2015 image was selected as the master image and the best 2014 Landsat 8 image was used to fill most of the data gaps in the 2015 image (after radiometric normalization through no-change regression normalization (Yuan and Elvidge, 1996)). Locations that were cloudy in both images were masked out and not used in this study.

Four-band (blue, green, red, near-infrared), one-meter aerial imagery from the National Agriculture Imagery Program (NAIP) for the growing season of 2015 was acquired for the study area from the USDA Farm Service Agency Aerial Photography Field Office (https://www.fsa.usda.gov/programs-and-services/aerial-photography/) and used to produce texture data (Brown and Barber, 2012) to be included in the analyses (Table 1). Texture information is helpful for distinguishing between species. We also used the NAIP imagery to check for errors in the field data such as bad GPS coordinates, cloud cover in the texture data, or incorrect identification of cover types. Topographic slope and aspect were derived from the digital elevation model (DEM) downloaded from the USGS National Elevation Dataset (ned.usgs.gov); and NDVI and three TC bands were calculated from the Landsat imagery (Table 1). The USFS provided boundary data for major disturbances within the study area, specifically for the West Fork Fire complex of 2013 and annual beetle infestation boundaries from 2000 to 2013.

2.3. Zero-inflated modeling with built-in 10-fold validation

Most of the datasets were zero-inflated, i.e., a large proportion of the data points within the dataset had values of zero, however, mortality and the combined species sub-canopy data were not (Table 2). We therefore approached the different datasets with different methodologies.

Zero-inflated data were analyzed using the methodology developed for predicting relative species composition in mixed conifer forests (Savage et al., 2015). We altered the approach slightly in that we separated the process into three distinct steps: (1) a binary prediction of presence/absence of the category of interest, (2) a continuous prediction of data that were predicted as present in step 1, and (3) combining the best models from steps 1 and 2. We were able, using this approach, to test 17 different statistical models (Table 3) without testing every possible 2-step combination of these methods (see Fig. 3 in Savage et al., 2015). The Caret Package in R enabled parameter tuning for each of the methods, which allowed for improved model fitting compared to accepting default model parameters.

Two sets of predicted maps were produced for our final zero-inflated results, one of presence/absence with an overall accuracy and kappa statistic and one of continuous values with a *p*-value, RMSE, 95% confidence interval (CI) width, and pseudo median or mean differences. The Wilcoxon's signed rank test was used for all percent canopy cover and mortality predictions since they are proportions (pseudo median). A standard *t*-test was used for the sub-canopy count predictions since they have a normal distribution (mean differences). These statistics were calculated through a 10-fold validation process where each model was executed 10 times while withholding a random 25% of the data each time. The results of the 10 models were averaged to provide the final validation statistics for each predicted map. We chose the models with the smallest average width as the final models with which to predict canopy cover (as long as the *p*-value was greater than 0.05, indicating the difference was not statistically significantly different from zero). The best binary model and the best continuous model for each category were then combined to produce one complete map of that category for the study area. All values of zero (absence) in the binary model were included, while the values from the continuous model were used where presence was predicted in the binary model.

Table 2
Number of zeroes observed in the reference data by category modeled – out of 463 total reference points. A high number of observed zeroes indicates that the data are zero-inflated and zero-inflated modeling should be utilized (Savage et al., 2015) if the Power test indicates there are enough observations to model. A low number of observed zeroes indicates that the data should be modeled using traditional methods. Percent canopy cover (PCC) and sub-canopy count were predicted using zero-inflated methods. Percent mortality and total sub-canopy PCC were predicted using traditional methods. Dead PCC and dead sub-canopy count were not modeled, however, the values are included here for species composition comparison. Dead is defined as a tree that is either fully dead (grey or no needles/leaves) or dying (red needles or unhealthy leaves) and NOT a snag (a dead tree that cannot be identified as a particular species).

Species/category	Number of zeroes	Number of non-zeroes
Live subalpine fir PCC	323	140
Live Engelmann spruce PCC	195	268
Live quaking aspen PCC	283	180
Live Douglas-fir PCC	386	77
Mortality[a]	80	383
Sub-canopy PCC[a]	89	374
Live subalpine fir sub-canopy count	263	200
Live Engelmann spruce sub-canopy count	145	318
Live quaking aspen sub-canopy count	295	168
Live Douglas-fir sub-canopy count	402	61
Dead subalpine fir PCC	414	49
Dead Engelmann spruce PCC	205	258
Dead quaking aspen PCC	394	69
Dead Douglas-fir PCC	437	26
Dead subalpine fir sub-canopy count	423	40
Dead Engelmann spruce sub-canopy count	337	126
Dead quaking aspen sub-canopy count	395	68
Dead Douglas-fir sub-canopy count	446	17

[a] Not zero-inflated.

Table 3
Statistical models/functions tested in this study (R Code for these models can be found in the Caret Package at http://topepo.github.io/caret/List.html). Several models were used for both the binary and continuous predictions, while others were used only for binary predictions ([b]) or only for continuous predictions ([c]).

Model	Function	R package
CART	rpart	rpart
Conditional Inference Random Forest[c]	cforest	party
Cubist[c]	cubist	Cubist
C5.0[b]	C5.0	C5.0, plyr
Generalized Linear with Step AIC Feature Selection[c]	glmStepAIC	MASS
Linear Discriminant Analysis[b]	lda	MASS
Multivariate Adaptive Regression Splines[c]	earth	earth
Naive Bayes[b]	nb	klaR
K Nearest Neighbors	kknn	kknn
Feed-Forward Neural Net	nnet	nnet
Partial Least Squares[c]	pls	pls
Random Forest	rf	randomForest
Ridge Regression with Variable Selection[c]	foba	foba
Stacked AutoEncoder Deep Neural Network[c]	dnn	deepnet
Support Vector Machine with Linear Kernel	svmLinear	kernlab
Support Vector Machine with Polynomial Kernel	svmPoly	kernlab
Support Vector Machine with Radial Kernel	svmRadial	kernlab

2.4. Continuous modeling with built-in 10-fold validation

We used a single-step process where the data were not zero-inflated (percent mortality and sub-canopy PCC; Table 2): continuous prediction using all of the reference data (rather than just those with a value greater than zero). A *p*-value, RMSE, 95% CI width, and pseudo median (Wilcoxon's signed rank test for PCC and mortality) or mean differences (standard *t*-test for sub-canopy counts) were calculated by performing a 10-fold validation process as described

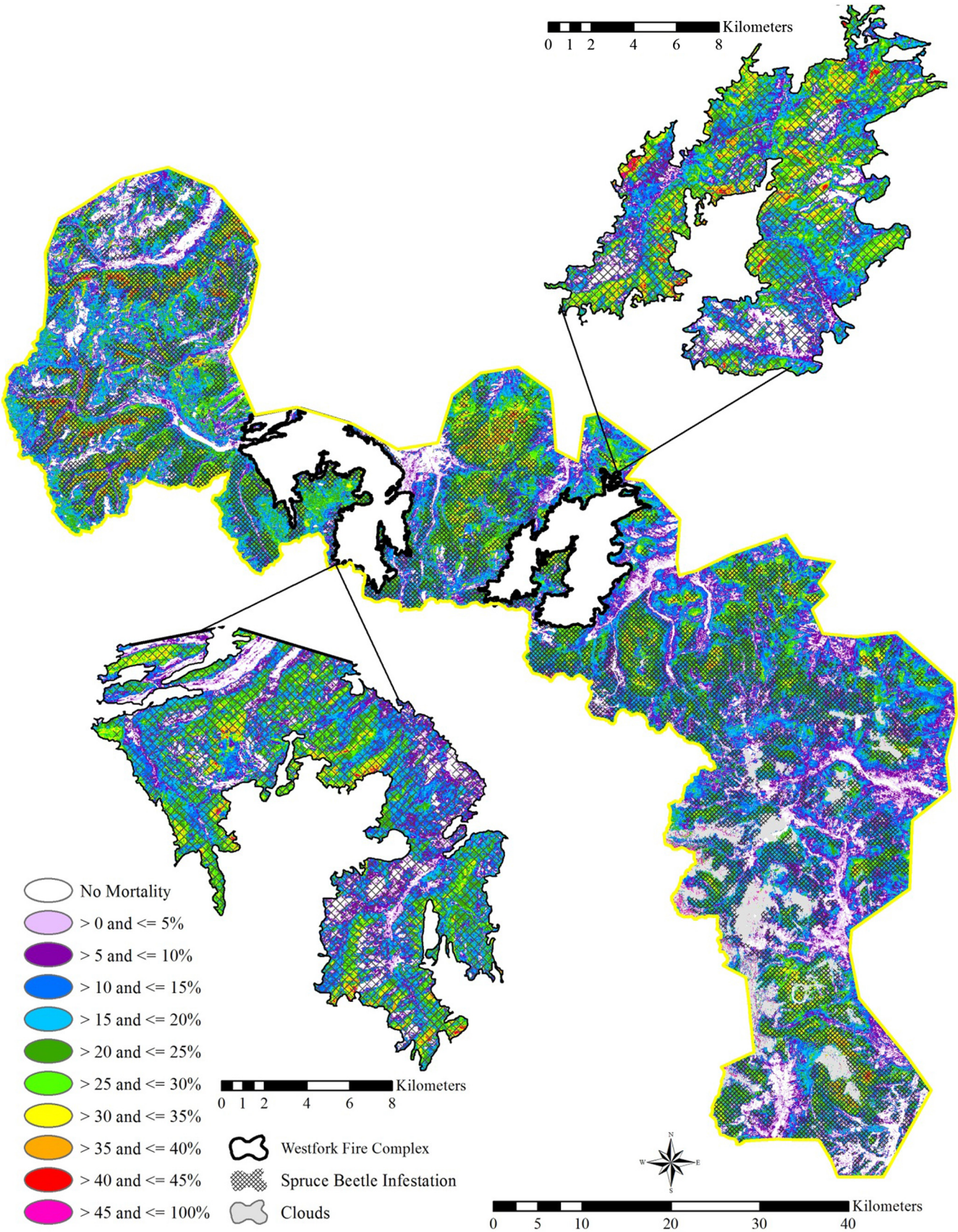

Fig. 2. Percent Mortality as of 2013 for the study area with the Westfork Fire Complex enlarged.

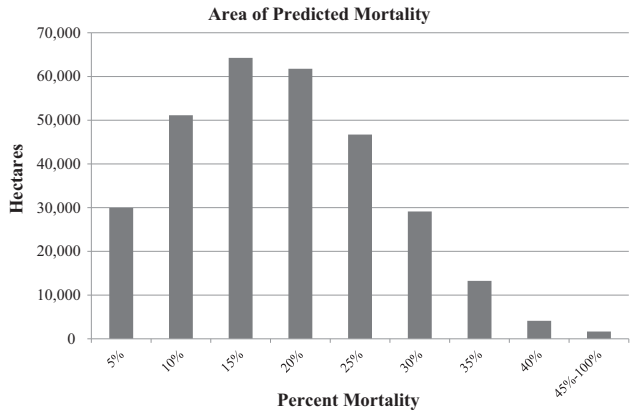

Fig. 3. Hectares of predicted percent mortality as of 2013 for the study area (~302,000 ha out of ~347,000 ha total).

above and choosing the model with the smallest average CI width as the final model to predict for each species/category.

3. Results

3.1. Field data

Most of the species that were recorded in the 463 reference site data were zero-inflated – i.e., a large portion of reference sites contained observations of zero for said species (as noted in Table 2). The minimum PCC values were zero for all species, while the maximum values ranged from 48% for subalpine fir to 100% for aspen, with the maximum percent mortality of 64% (Table 4). Average live PCC values ranged from 10.9% for subalpine fir to 38.6% for aspen, with the average percent mortality of 21.0% (Table 4). Maximum live sub-canopy counts ranged from 34 trees per plot for Douglas-fir to 151 trees for aspen (Table 4). Average live sub-canopy count values ranged from 5 trees for Douglas-fir to 16.6 trees for aspen (Table 4). The dominant species of each site was identified as the species that had the highest PCC or count for each data point. For live trees, Engelmann spruce was the most dominant for both the PCC (33% of the reference sites were dominated by Engelmann spruce) and the sub-canopy counts (41% of the reference sites were dominated by Engelmann spruce) (Table 4).

Engelmann spruce was observed dead or red more often than the other species within the reference data (Table 2), both in the PCC data (258 of the 463 sites had dead Engelmann spruce PCC) and the sub-canopy count data (126 of the 463 sites had dead Engelmann spruce sub-canopy counts). Dead Engelmann spruce had the highest maximum and mean PCC values (64% and 22.0%, respectively), while dead quaking aspen had the highest maximum and mean sub-canopy count values (42 trees and 4.9 trees, respectively; Table 4).

3.2. Percent canopy cover

A total of six predicted PCC maps were created in this process. The four PCC-by-species maps were modeled with zero-inflated methods (Table 5). All four PCC-by-species models had *p*-values

Table 4
Statistics from the 463 reference sites. Percent dominant represents the percent of the 463 points where that species was observed more often than the other species.

Species/category	Maximum (%)	Mean (%)	Standard deviation (%)	Dominant (%)
Live subalpine fir PCC	48	10.9	7.0	10
Live Engelmann spruce PCC	72	15.9	13.4	33
Live quaking aspen PCC	100	38.6	26.0	26
Live Douglas-fir PCC	84	29.7	13.6	11
Percent Mortality	64	21.0	15.1	
Sub-canopy PCC	76	12.2	9.3	
Dead subalpine fir PCC	24	8.0	3.0	
Dead Engelmann spruce PCC	64	22.0	14.9	
Dead quaking aspen PCC	40	8.0	3.7	
Dead Douglas-fir PCC	20	7.2	2.0	
	Maximum (# of trees)	Mean (# of trees)	Standard deviation (# of trees)	Dominant (%)
Subalpine fir sub-canopy count	86	8.5	7.8	17
Engelmann spruce sub-canopy count	43	7.5	6.2	41
Quaking aspen sub-canopy count	151	16.6	15.2	25
Douglas-fir sub-canopy count	34	5.0	2.8	6
Dead subalpine fir sub-canopy count	15	2.9	1.3	
Dead Engelmann spruce sub-canopy count	15	2.2	1.5	
Dead quaking aspen sub-canopy count	42	4.9	3.4	
Dead Douglas-fir sub-canopy count	11	2.6	0.7	

Table 5
Statistical results for percent canopy cover (PCC) predictions for individual species, overall mortality, and sub-canopy total. The models used in the multi-step zero-inflated process are shown (results of the models that were tested but not chosen to be used for the final predictions are listed in the Appendix A). The results of these models were combined to create the final predicted maps. RMSE, CI width and pseudo median from Wilcoxon's signed rank test, maximum, and mean values are displayed. The minimum value for all predictions was zero. Mortality and sub-canopy PCC were not treated as zero-inflated (see Table 2). Dominant represents the percent of all predicted pixels where that species had a larger PCC than the other species. All values are in percent (%). Model functions are from Table 3.

Species/category	Binary model function	Continuous model function	RMSE	CI width	Pseudo median	Maximum value	Mean value	Dominant (percent of total predicted map)
Subalpine Fir	svmLinear	svmLinear	10.9	4.4	2.5	12.8	4.3	1
Engelmann Spruce	svmLinear	cubist	14.2	6.5	−0.2	100.0	9.7	73
Quaking Aspen	svmPoly	rf	17.0	13.1	0.2	92.9	35.3	21
Douglas-Fir	rf	rf	12.7	24.9	−6.5	66.6	22.3	1
Mortality	–	svmPoly	12.4	4.6	−1.1	100.0	16.6	
Sub-canopy	–	CART	9.4	3.0	1.3	18.1	7.0	

greater than 0.05 and pseudo median values ranging from 0.2 to 6.5, showing none were significantly different between observed and predicted. RMSE values ranged from 10.9% for subalpine fir to 17.0% for aspen (Table 5). PCC-by-species 95% confidence interval widths ranged from 4.4% for subalpine fir to 24.9 for Douglas-fir (Table 5). PCC-by-species values ranged from 0 to 100%, with averages ranging from 4.3% to 35.1% depending on species (Table 5). The mortality and sub-canopy PCC maps were modeled using only the best continuous methods rather than zero-inflated methods, because the reference data were not zero-inflated (see Table 2). Percent mortality resulted in a *p*-value greater than 0.05, a pseudo median of −1.1, an RMSE of 12.4%, and a 95% confidence interval width of 4.6% (Table 5). Percent mortality values ranged from 0 to 100% with an average of 16.6% (Table 5). Total sub-canopy PCC resulted in a *p*-value greater than 0.05, a pseudo median of 1.3, an RMSE of 9.4%, and a 95% confidence interval width of 3.0% (Table 5). Sub-canopy PCC values ranged from 0 to 18.1% with an average of 7.0% (Table 5).

3.3. Sub-canopy counts

A total of four predicted count maps were created in this process. All species count maps were modeled with zero-inflated methods. All four models had *p*-values greater than 0.05 and mean differences less than 1.4 trees. RMSE values ranged from 3.0 trees for Douglas-fir to 13.4 trees for aspen (Table 6). 95% confidence interval widths ranged from 1.1 for Douglas-fir to 5.0 for aspen (Table 6). Sub-canopy counts ranged from 0 to 42 trees per pixel with averages ranging from 1 to 7 trees per pixel, depending on species (Table 6).

4. Discussion

We developed a Landsat-based approach to predict the composition of a mixed conifer forest after major disturbances. Through this approach we tested and compared a series of prediction algorithms with the aim of maximizing our predictive accuracy. Our Landsat-based zero-inflated models successfully predicted PCC-by-species and sub-canopy species counts out of the zero-rich reference data (at an alpha of 0.05, the predicted versus observed is not significantly different from zero), similar to results from Savage et al. (2015) (Tables 5 and 6). The non-zero-inflated models were also successful in predicting percent mortality (similar to Long and Lawrence (2016)) and sub-canopy PCC (Table 5). The methods applied in this study can provide land managers with a powerful tool that can help them predict present and future forest composition.

We believe that all of the bands, indices, and ancillary data incorporated into our methods (Table 1) are important for mapping PCC and sub-canopy counts within our study area. While some of the algorithms used in our models have built-in variable importance reporting, most of them do not, so we were unable to explicitly identify the most important variables that were used for predicting every PCC and sub-canopy count map. We suggest that variable importance in predicting forest composition is a topic for further study, especially to determine if specific datasets might not be needed and thus need not be included in data acquisition and pre-processing. Nearly the entire study area (~302,000 ha or 87%) showed some amount of mortality in the percent mortality map (Figs. 2 and 3), as expected after viewing the study area from the ground and from the air (large swaths of the study area had high levels of observed tree mortality and the majority of the reference data (383 sites out of 463) included observed mortality). The majority of the mortality pixels were predicted to have values from 11% to 20% with a mean value of 16.6% (Table 5), while relatively few pixels were predicted with 45% or more mortality (Fig. 3), though a maximum value of 100% was predicted for 402 ha (0.13% of the entire map) (Table 5).

We believe that the higher error for aspen PCC is due to confusion with shrub species, since aspen is the only deciduous species we analyzed for this study. The zero-inflated Douglas-fir PCC map showed relatively poor results compared to the other three species we modeled, though it still had a low RMSE. The results for the Douglas-fir PCC indicate that there were possibly too few reference data points with values greater than zero (77 out of 463 reference points; Table 2) to be able to predict a continuous PCC value for the entire study area with as high confidence as other species. On the other hand, the presence/absence map for Douglas-fir PCC is highly accurate at 89.7% overall accuracy.

Based on traditional successional models as well as discussion and observation in the field, we expected the sub-canopy to be mostly fir while the overstory would have mostly dead spruce and live fir. The results of the predictions for spruce do not support these expectations. Despite the observed large-scale spruce beetle kill and disturbances from the 2013 West Fork fire, there is medium to high regeneration of spruce where there are recorded disturbances (particularly within the spruce beetle kill areas) (Fig. 4b). Even some areas that have been disturbed show high values of overstory canopy cover, most likely due to the canopy being opened up for rapid regrowth of the sub-canopy species that in this case are often the same species as the upper canopy (Fig. 4). Additionally, when comparing the dominant species of both the PCC and sub-canopy counts, we observed that 72% of the study area has the same dominant species in both categories (or has no dominant species in both categories), and Engelmann spruce is the dominant species in both categories for much of the study area (Fig. 5a). The majority (62%) of the unmatched dominant species falls within the undisturbed areas of the study area, and here we see that Engelmann spruce and aspen are the dominant sub-canopy species in much of the area where the dominants do not match (38% and 17% of the areas that do not match, respectively) (Fig. 5b).

Table 6
Statistical results for sub-canopy count data for individual species. The models used in the multi-step zero-inflated process are shown (results of the models that were tested but not chosen to be used for the final predictions are listed in the Appendix A). The results of these models were combined to create the final predicted maps. RMSE, CI width and mean of the differences from a standard *t*-test, maximum, and mean values are displayed. The minimum value for all predictions was zero. "Dominant" represents the percent of all predicted pixels where that species was observed more often than the other species. Values are number of trees per pixel unless otherwise noted. Model functions are from Table 3.

Species/category	Binary model function	Continuous model function	RMSE	CI width	Mean differences	Maximum value	Mean value	Dominant (percent of total predicted map)
Subalpine Fir	rf	cubist	9.0	3.4	−0.9	15.3	3.4	5
Engelmann Spruce	svmLinear	foba	5.6	2.1	0.1	15.4	6.7	76
Quaking Aspen	svmLinear	svmPoly	13.4	5.0	−1.4	41.9	5.3	18
Douglas-Fir	lda	svmPoly	3.0	1.1	−0.5	9.3	1.2	<1

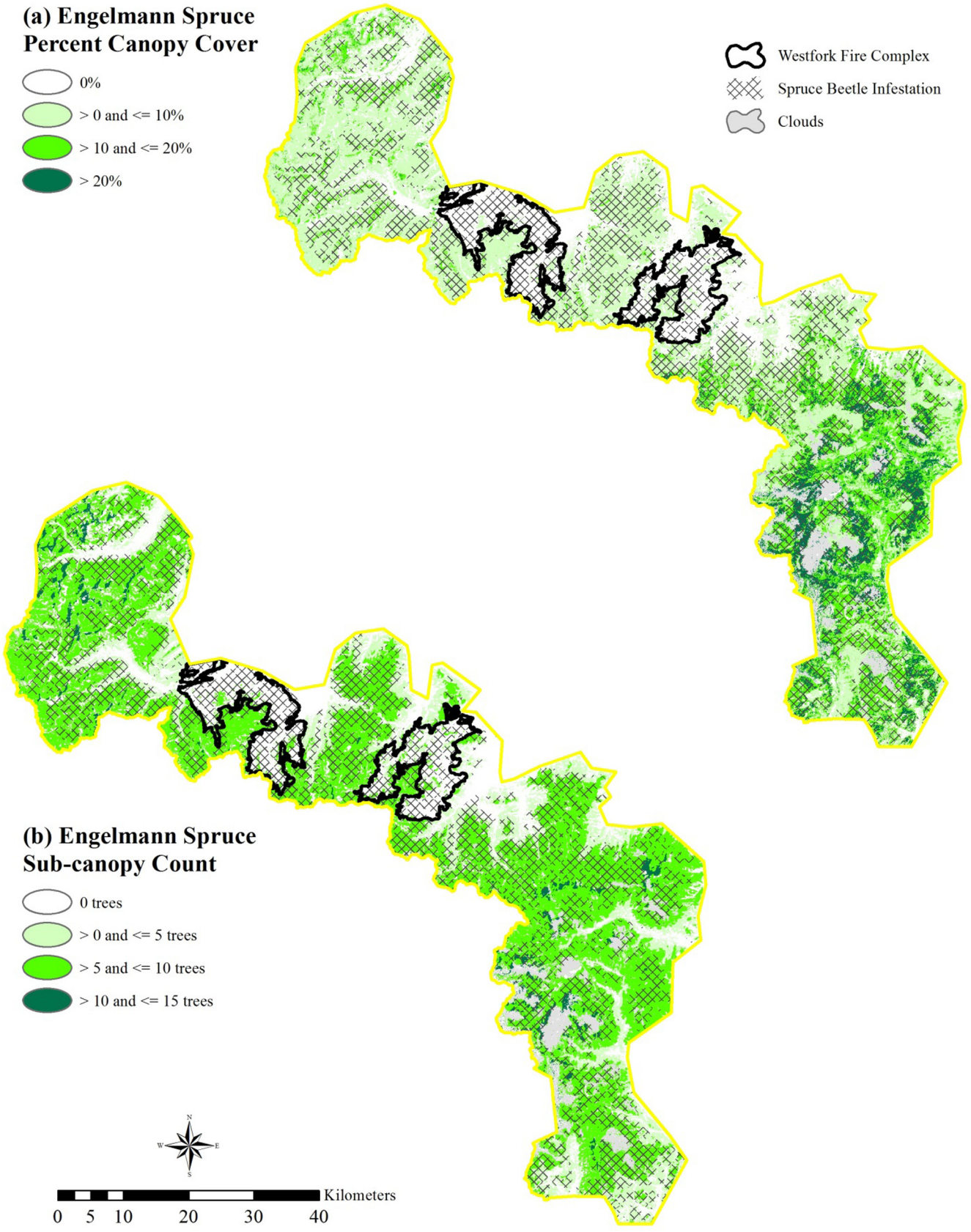

Fig. 4. Percent canopy cover (a) and sub-canopy count (b) for Engelmann spruce.

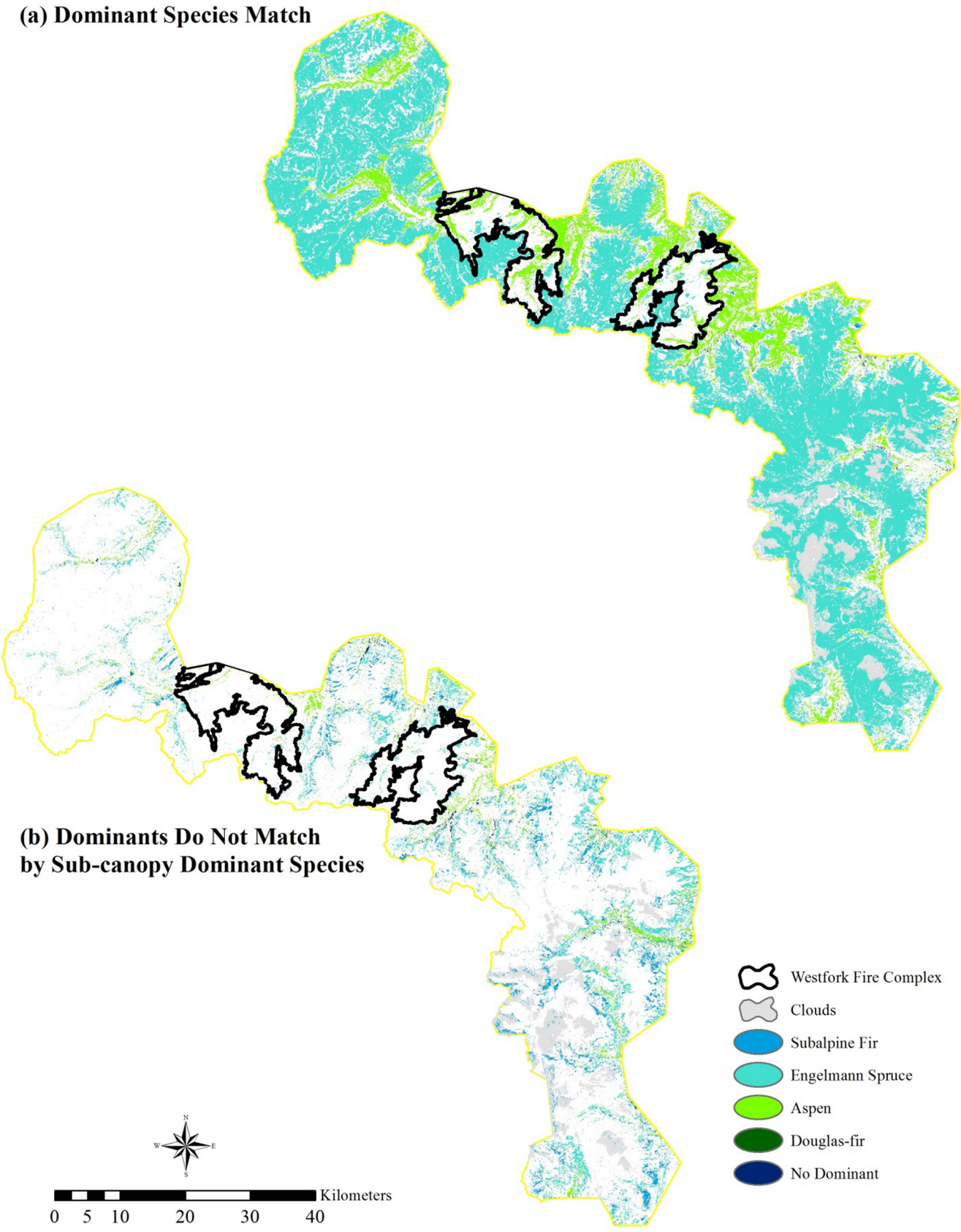

Fig. 5. Where dominant species are (a) the same and (b) different for upper canopy percent canopy cover (PCC) and sub-canopy count. Sub-canopy dominant is displayed where dominants do not match (b). The dominant species is identified as the species that had the highest PCC or sub-canopy count for each pixel.

Our results indicate that post-disturbance forest composition is likely to be much like pre-disturbance composition, rather than a shift or successional change, with some exceptions. Mismatches where the dominant species is Engelmann spruce (Fig. 5b) are probably a result of substantially or entirely dead overstory Engelmann spruce, suggesting the sub-canopy has returned to its previous state. Where the dominant sub-canopy is subalpine fir or aspen, we likely are looking at changes in forest composition post-disturbance, although this is rare (e.g., the sub-canopy aspen east of the fire perimeters in Fig. 5b). Though rare, several studies have addressed using Landsat imagery to map PCC and percent mortality (Hart and Veblen, 2015; Long and Lawrence, 2016; Savage et al., 2015; Schwantes et al., 2016), however, even more rare are studies that are able to distinguish between upper canopy and sub-canopy with high accuracy. Recently, for the first time, researchers in northern Colorado were able to use snow to mask out understory vegetation to better detect upper canopy vegetation (Baker et al., 2017). Application of this method to our PCC-by-species mapping might improve our results. Another recent study utilized LiDAR to predict basal area and tree density at species-level in northern Idaho and through comparison of many different machine learning algorithms concluded that random forest produced the best overall results (Hudak et al., 2008). Our method of testing several algorithms to maximize predictive accuracy also found that random forest often produces the best results (Tables 5 and 6), however, LiDAR data are cost-prohibitive when mapping large areas, so we do not expect to attempt this method to improve our PCC and sub-canopy count predictions.

Finally, we know that major disturbances such as stand-replacing fires and large insect outbreaks are happening at a more rapid pace in recent years and can have a variety of effects on vegetation and forest composition, including by altering the successional patterns within in a forest, particularly for the biotic legacies that will establish early after a disturbance (Hicke et al., 2016; Sibold et al., 2007). We expect to see forest composition and structure altered in distribution, both in increasing latitude and altitude, and in increasing extents (Hicke et al., 2006, 2016; Jewett et al., 2011; Temperli et al., 2015). Establishment, regeneration, competition, and the ability of trees to respond to major disturbances are all changing due to global climate change. Land managers, in order to protect their resources, must be able to adapt to these drastic changes as they happen. Using freely accessed Landsat imagery and the methods described in this paper will not only save time and money for these land managers, but will also quickly and easily provide information with which to make highly informed management decisions and pursue additional analyses of forest composition and succession. From a management standpoint, knowledge of the dominant species establishing after a disturbance (i.e., biotic legacies) is very meaningful for planning.

5. Conclusion

We successfully used free and easily accessible Landsat satellite imagery and remote sensing tools to accurately predict PCC and sub-canopy counts of four different tree species in southwest Colorado. Intense and detailed field data collection, along with the application of zero-inflated prediction methods, was crucial for the resulting accuracies. Empirical knowledge, in addition to local knowledge, of the current surviving forest (PCC) combined with the sub-canopy/regeneration/biotic legacies (counts) will allow land managers to form an evidence-based hypothesis as to what the forest will become in the future. Predicting what the "next" forest will be – using empirical data – is an important tool for management/planning regarding many forested landscapes worldwide and the habitats therein.

The methods described in this paper can be applied to many different disturbance ecology questions. Forest managers will be able to provide empirical evidence of current and future forest composition (as well as past composition with change analyses using the Landsat satellite imagery archive (data available as far back as 1972)) and the biological legacies that will affect the regeneration and successional patterns that prevail. The knowledge provided by these methods will allow managers to make more informed and better decisions when planning for or reacting to ecological, economic, and historical impacts within their regions.

Acknowledgements

Funding for this study was provided by the Rio Grande National Forest, Monte Vista, Colorado and the U. S. Forest Service Region 2. We thank D. Gomez and R. Ghormley, Rio Grande National Forest, for their considerable logistical support. We acknowledge Colorado Parks and Wildlife for added logistical support. We thank the many field technicians for collecting the reference data and L. Olson for her analytical and field support.

Appendix A

Part 1: Statistical results for percent canopy cover (PCC) predictions for individual species, overall mortality, and sub-canopy total.

BINARY – Subalpine Fir PCC	Overall Accuracy
Random Forest	70.36%
Support Vector Machine with Linear Kernel	70.98%
Support Vector Machine with Radial Kernel	69.20%
Support Vector Machine with Polynomial Kernel	70.54%
CART	69.64%
C5.0	68.39%
Feed-Forward Neural Net	65.09%
K Nearest Neighbors	66.70%
Naïve Bayes	66.34%
Linear Discriminant Analysis	69.55%

CONTINUOUS – Subalpine Fir PCC	RMSE	p-value	CI Width	Pseudo Median
Stacked AutoEncoder Deep Neural Network	10.10	0.21	6.20	1.34
Partial Least Squares	9.60	0.32	6.34	2.10
Ridge Regression with Variable Selection	9.91	0.34	5.78	2.20
Cubist	10.63	0.55	6.96	−1.12
Conditional Inference Random Forest	9.71	0.40	6.28	1.54
Generalized Linear with Step AIC Feature Selection	10.46	0.50	6.19	1.31
Random Forest	9.62	0.25	6.32	2.32
Support Vector Machine with Linear Kernel	9.51	0.37	5.59	0.18
Support Vector Machine with Radial Kernel	9.47	0.44	5.76	−0.46
Support Vector Machine with Polynomial Kernel	9.82	0.40	6.01	−1.00
CART	9.93	0.31	6.20	2.05
Feed-Forward Neural Net	15.19	0.00	6.20	−10.00
K Nearest Neighbors	10.24	0.17	5.95	3.15
Multivariate Adaptive Regression Splines	9.57	0.33	6.32	1.72

BINARY – Engelmann Spruce PCC	Overall Accuracy
Random Forest	72.95%
Support Vector Machine with Linear Kernel	73.75%
Support Vector Machine with Radial Kernel	70.18%
Support Vector Machine with Polynomial Kernel	71.79%
CART	69.11%
C5.0	74.11%
Feed-Forward Neural Net	66.07%
K Nearest Neighbors	64.64%
Naïve Bayes	71.07%
Linear Discriminant Analysis	73.75%

CONTINUOUS – Engelmann Spruce PCC	RMSE	p-value	CI Width	Pseudo Median
Stacked AutoEncoder Deep Neural Network	15.26	0.31	7.40	0.94
Partial Least Squares	14.22	0.57	7.17	0.56
Ridge Regression with Variable Selection	14.23	0.60	7.20	0.39
Cubist	14.57	0.37	6.71	−2.09
Conditional Inference Random Forest	14.02	0.54	6.95	−0.54
Generalized Linear with Step AIC Feature Selection	14.61	0.53	6.99	0.26
Random Forest	13.88	0.54	6.87	0.94
Support Vector Machine with Linear Kernel	14.94	0.19	6.95	−2.59
Support Vector Machine with Radial Kernel	14.53	0.20	6.83	−2.47
Support Vector Machine with Polynomial Kernel	14.61	0.20	6.96	−2.76
CART	15.16	0.61	7.60	0.39
Feed-Forward Neural Net	22.96	0.00	7.40	−16.00
K Nearest Neighbors	15.09	0.60	7.01	0.38
Multivariate Adaptive Regression Splines	14.73	0.59	7.51	0.59

BINARY – Aspen PCC	Overall Accuracy
Random Forest	81.07%
Support Vector Machine with Linear Kernel	79.73%
Support Vector Machine with Radial Kernel	81.70%
Support Vector Machine with Polynomial Kernel	82.23%
CART	78.39%
C5.0	80.18%
Feed-Forward Neural Net	72.14%
K Nearest Neighbors	69.38%
Naïve Bayes	76.43%
Linear Discriminant Analysis	80.45%

CONTINUOUS – Aspen PCC	RMSE	p-value	CI Width	Pseudo Median
Stacked AutoEncoder Deep Neural Network	30.87	0.30	20.80	−1.11
Partial Least Squares	20.12	0.50	13.20	1.07
Ridge Regression with Variable Selection	19.52	0.47	12.80	1.05
Cubist	18.88	0.51	11.87	1.22
Conditional Inference Random Forest	18.94	0.45	11.96	1.31
Generalized Linear with Step AIC Feature Selection	19.85	0.55	12.75	1.38
Random Forest	18.40	0.44	11.76	1.54
Support Vector Machine with Linear Kernel	20.31	0.60	13.05	0.36
Support Vector Machine with Radial Kernel	19.65	0.56	13.27	−0.06
Support Vector Machine with Polynomial Kernel	19.69	0.55	13.12	−0.63
CART	21.58	0.30	13.35	2.69
Feed-Forward Neural Net	47.36	0.00	20.80	−36.40
K Nearest Neighbors	21.92	0.46	14.21	2.22
Multivariate Adaptive Regression Splines	23.46	0.38	15.44	2.40

BINARY – Douglas-fir PCC	Overall Accuracy
Random Forest	89.73%
Support Vector Machine with Linear Kernel	87.68%
Support Vector Machine with Radial Kernel	87.59%
Support Vector Machine with Polynomial Kernel	87.41%
CART	86.70%
C5.0	87.95%
Feed-Forward Neural Net	87.77%
K Nearest Neighbors	85.27%
Naïve Bayes	86.70%
Linear Discriminant Analysis	89.11%

CONTINUOUS – Douglas-fir PCC	RMSE	p-value	CI Width	Pseudo Median
Stacked AutoEncoder Deep Neural Network	23.20	0.48	23.20	−2.68
Partial Least Squares	21.19	0.53	21.35	−2.68
Ridge Regression with Variable Selection	21.76	0.45	21.78	−2.59
Cubist	20.90	0.41	21.07	−2.55
Conditional Inference Random Forest	22.31	0.40	22.19	−3.79
Generalized Linear with Step AIC Feature Selection	25.33	0.44	24.88	−4.47
Random Forest	21.02	0.50	20.67	−2.12
Support Vector Machine with Linear Kernel	23.34	0.42	23.62	−1.03
Support Vector Machine with Radial Kernel	21.54	0.33	21.02	−4.82
Support Vector Machine with Polynomial Kernel	22.37	0.33	21.84	−5.47
CART	22.93	0.42	23.20	−3.58
Feed-Forward Neural Net	40.88	0.00	23.20	−33.40
K Nearest Neighbors	21.86	0.48	22.04	−1.08
Multivariate Adaptive Regression Splines	22.11	0.46	22.42	−2.51

CONTINUOUS – Percent Mortality	RMSE	p-value	CI Width	Pseudo Median
Stacked AutoEncoder Deep Neural Network	15.18	0.29	6.20	1.88
Partial Least Squares	13.78	0.35	5.50	1.06
Ridge Regression with Variable Selection	12.53	0.38	4.65	0.79
Cubist	12.59	0.45	4.77	−0.29
Conditional Inference Random Forest	12.45	0.36	4.91	0.74
Generalized Linear with Step AIC Feature Selection	12.42	0.35	4.64	0.77
Random Forest	12.58	0.35	4.82	1.16
Support Vector Machine with Linear Kernel	12.66	0.42	4.68	−0.87
Support Vector Machine with Radial Kernel	12.40	0.33	4.62	−1.34
Support Vector Machine with Polynomial Kernel	12.43	0.36	4.60	−1.08
CART	14.36	0.52	5.57	0.73
Feed-Forward Neural Net	22.46	0.00	6.20	−15.00
K Nearest Neighbors	13.66	0.23	5.10	2.32
Multivariate Adaptive Regression Splines	12.60	0.50	4.85	0.50

CONTINUOUS – Sub-canopy PCC	RMSE	p-value	CI Width	Pseudo Median
Stacked AutoEncoder Deep Neural Network	9.48	0.34	3.00	0.95
Partial Least Squares	8.97	0.33	3.24	1.01
Ridge Regression with Variable Selection	8.98	0.32	3.12	1.01
Cubist	9.11	0.63	3.21	−0.46
Conditional Inference Random Forest	8.99	0.34	3.19	0.93
Generalized Linear with Step AIC Feature Selection	9.07	0.40	3.21	0.86
Random Forest	9.16	0.21	3.26	1.24
Support Vector Machine with Linear Kernel	9.12	0.57	3.17	−0.29
Support Vector Machine with Radial Kernel	9.00	0.42	3.12	−0.65
Support Vector Machine with Polynomial Kernel	9.12	0.39	3.13	−0.66
CART	9.42	0.59	3.00	1.30
Feed-Forward Neural Net	12.96	0.00	3.00	−7.60
K Nearest Neighbors	9.58	0.33	3.27	0.98
Multivariate Adaptive Regression Splines	9.28	0.27	3.30	1.18

Part 2: Statistical results for sub-canopy count data for individual species.

BINARY – Subalpine Fir Sub-canopy Count	Overall Accuracy
Random Forest	75.09%
Support Vector Machine with Linear Kernel	74.73%
Support Vector Machine with Radial Kernel	73.21%
Support Vector Machine with Polynomial Kernel	72.95%
C5.0	74.20%
K Nearest Neighbors	68.75%
Naïve Bayes	68.66%
Linear Discriminant Analysis	74.46%

CONTINUOUS – Subalpine Fir Sub-canopy Count	RMSE	p-value	CI Width	Mean Differences
Partial Least Squares	9.46	0.41	5.28	−0.03
Ridge Regression with Variable Selection	9.84	0.37	5.50	0.03
Cubist	9.59	0.25	5.27	−2.11
Conditional Inference Random Forest	9.59	0.34	5.36	−0.43
Generalized Linear with Step AIC Feature Selection	10.13	0.38	5.67	−0.09
Random Forest	9.98	0.45	5.54	0.41
Support Vector Machine with Linear Kernel	9.82	0.32	5.41	−2.02
Support Vector Machine with Radial Kernel	9.72	0.19	5.31	−2.38
Support Vector Machine with Polynomial Kernel	9.70	0.18	5.29	−2.43
K Nearest Neighbors	10.31	0.48	5.79	0.29
Multivariate Adaptive Regression Splines	9.75	0.38	5.45	−0.20

BINARY – Engelmann Spruce Sub-canopy Count	Overall Accuracy
Random Forest	78.93%
Support Vector Machine with Linear Kernel	79.64%
Support Vector Machine with Radial Kernel	76.52%
Support Vector Machine with Polynomial Kernel	78.93%
C5.0	77.50%
K Nearest Neighbors	73.57%
Naïve Bayes	72.32%
Linear Discriminant Analysis	79.20%

CONTINUOUS – Engelmann Spruce Sub-canopy Count	RMSE	p-value	CI Width	Mean Differences
Partial Least Squares	9.46	0.41	5.28	−0.03
Ridge Regression with Variable Selection	9.84	0.37	5.50	0.03
Cubist	9.59	0.25	5.27	−2.11
Conditional Inference Random Forest	9.59	0.34	5.36	−0.43

(*continued on next page*)

Appendix A (*continued*)

CONTINUOUS – Engelmann Spruce Sub-canopy Count	RMSE	p-value	CI Width	Mean Differences
Generalized Linear with Step AIC Feature Selection	10.13	0.38	5.67	−0.09
Random Forest	9.98	0.45	5.54	0.41
Support Vector Machine with Linear Kernel	9.82	0.32	5.41	−2.02
Support Vector Machine with Radial Kernel	9.72	0.19	5.31	−2.38
Support Vector Machine with Polynomial Kernel	9.70	0.18	5.29	−2.43
K Nearest Neighbors	10.31	0.48	5.79	0.29
Multivariate Adaptive Regression Splines	9.75	0.38	5.45	−0.20

BINARY – Aspen Sub-canopy Count	Overall Accuracy
Random Forest	75.71%
Support Vector Machine with Linear Kernel	79.55%
Support Vector Machine with Radial Kernel	77.86%
Support Vector Machine with Polynomial Kernel	79.38%
C5.0	76.61%
K Nearest Neighbors	71.07%
Naïve Bayes	73.57%
Linear Discriminant Analysis	78.39%

CONTINUOUS – Aspen Sub-canopy Count	RMSE	p-value	CI Width	Mean Differences
Partial Least Squares	20.17	0.36	12.45	1.06
Ridge Regression with Variable Selection	20.08	0.41	12.52	0.79
Cubist	21.77	0.29	13.40	−3.20
Conditional Inference Random Forest	20.61	0.37	12.76	−0.04
Generalized Linear with Step AIC Feature Selection	20.86	0.47	13.26	0.38
Random Forest	20.16	0.39	12.41	1.59
Support Vector Machine with Linear Kernel	19.74	0.14	12.26	−4.56
Support Vector Machine with Radial Kernel	19.95	0.14	12.39	−4.67
Support Vector Machine with Polynomial Kernel	19.58	0.15	12.14	−4.71
K Nearest Neighbors	21.89	0.49	13.79	1.66
Multivariate Adaptive Regression Splines	21.32	0.34	13.23	0.60

BINARY – Douglas-fir Sub-canopy Count	Overall Accuracy
Random Forest	87.05%
Support Vector Machine with Linear Kernel	87.68%
Support Vector Machine with Radial Kernel	85.18%
Support Vector Machine with Polynomial Kernel	85.27%
C5.0	85.89%
K Nearest Neighbors	85.98%
Naïve Bayes	86.61%
Linear Discriminant Analysis	87.86%

CONTINUOUS – Douglas-fir Sub-canopy Count
These data would not run in the continuous mode, so we ran the zero-inflated process with Linear Discriminant Analysis for the binary prediction combined with every continuous option. The support vector machine with polynomial kernel produced the best results and was used for the final prediction of Douglas-fir sub-canopy counts.

References

Baker, E.H., Painter, T.H., Schneider, D., Meddens, A.J., Hicke, J.A., Molotch, N.P., 2017. Quantifying insect-related forest mortality with the remote sensing of snow. Remote Sens. Environ. 188, 26–36.

Bebi, P., Kulakowski, D., Veblen, T.T., 2003. Interactions between fire and spruce beetles in a subalpine Rocky Mountain forest landscape. Ecology 84 (2), 362–371.

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.

Blakeman, M., 2013. Fact Sheet: Results from the 2013 Forest Health Aerial Survey. Rio Grande National Forest, 2p.

Brown, S., Barber, J., 2012. The Region 1 Existing Vegetation Mapping Program (VMap) Flathead National Forest Overview; version 12. Region One Vegetation Classification, Mapping, Inventory and Analysis Report 12–34. USDA Forest Service, Missoula, MT, 6p.

Chow, S., Shao, J., Wang, H., 2008. Sample Size Calculations in Clinical Research. Chapman & Hall/CRC Biostatistics Series.

Crisafulli, C.M., Dale, V.H., Swanson, F.J., 2005. Ecological Responses to the 1980 Eruption of Mount St. Springer, Helens.
Crist, E.P., Cicone, R.C., 1984. A physically-based transformation of Thematic Mapper data--The TM Tasseled Cap. IEEE Trans. Geosci. Remote Sens. 3, 256–263.
Derderian, D.P., Dang, H., Aplet, G.H., Binkley, D., 2016. Bark beetle effects on a seven-century chronosequence of Engelmann spruce and subalpine fir in Colorado, USA. For. Ecol. Manage. 361, 154–162.
Fettig, C.J., Klepzig, K.D., Billings, R.F., Munson, A.S., Nebeker, T.E., Negrón, J.F., Nowak, J.T., 2007. The effectiveness of vegetation management practices for prevention and control of bark beetle infestations in coniferous forests of the western and southern United States. For. Ecol. Manage. 238 (1), 24–53.
Fiala, A.C.S., Garman, S.L., Gray, A.N., 2006. Comparison of five canopy cover estimation techniques in the western Oregon Cascades. For. Ecol. Manage. 232, 188–197.
Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., 2013. High-resolution global maps of 21st-century forest cover change. Science 342 (6160), 850–853.
Hart, S.J., Veblen, T.T., 2015. Detection of spruce beetle-induced tree mortality using high-and medium-resolution remotely sensed imagery. Remote Sens. Environ. 168, 134–145.
Hart, S.J., Veblen, T.T., Kulakowski, D., 2014. Do tree and stand-level attributes determine susceptibility of spruce-fir forests to spruce beetle outbreaks in the early 21st century? For. Ecol. Manage. 318, 44–53.
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. J. Geophys. Res.-Biogeosci. 111 (G2).
Hicke, J.A., Meddens, A.J., Kolden, C.A., 2016. Recent tree mortality in the western United States from bark beetles and forest fires. For. Sci. 62 (2), 141–153.
Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E., Falkowski, M.J., 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sens. Environ. 112 (5), 2232–2245.
IPCC, 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In: Pachauri, R.K., Meyer, L.A. (Eds.), Core Writing Team. IPCC, Geneva, Switzerland, 151p.
Jenkins, M.J., Hebertson, E.G., Munson, A.S., 2014. Spruce beetle biology, ecology and management in the Rocky Mountains: an addendum to spruce beetle in the Rockies. Forests 5 (1), 21–71.
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. For. Sci. 57 (4), 320–335.
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., Meigs, G.W., 2014. Bringing an ecological view of change to Landsat-based remote sensing. Front. Ecol. Environ. 12 (6), 339–346.
Kulakowski, D., Veblen, T.T., Bebi, P., 2003. Effects of fire and spruce beetle outbreak legacies on the disturbance regime of a subalpine forest in Colorado. J. Biogeogr. 30 (9), 1445–1456.
Lawrence, R.L., Ripple, W.J., 1999. Calculating change curves for multitemporal satellite imagery: Mount St. Helens 1980–1995. Remote Sens. Environ. 67 (3), 309–319.
Lawrence, R.L., Ripple, W.J., 2000. Fifteen years of revegetation of Mount St. Helens: a landscape-scale analysis. Ecology 81 (10), 2742–2752.
Long, J.A., Lawrence, R.L., 2016. Mapping percent tree mortality due to mountain pine beetle damage. For. Sci. 62 (4), 392–402.
McDowell, N.G., Coops, N.C., Beck, P.S., Chambers, J.Q., Gangodagamage, C., Hicke, J. A., Huang, C.Y., Kennedy, R., Krofcheck, D.J., Litvak, M., Meddens, A.J., 2015. Global satellite monitoring of climate-induced vegetation disturbances. Trends Plant Sci. 20 (2), 114–123.
Millar, C.I., Stephenson, N.L., Stephens, S.L., 2007. Climate change and forests of the future: managing in the face of uncertainty. Ecol. Appl. 17 (8), 2145–2151.
Nagendra, H., Lucas, R., Honrado, J.P., Jongman, R.H., Tarantino, C., Adamo, M., Mairota, P., 2013. Remote sensing for conservation monitoring: assessing protected areas, habitat extent, habitat condition, species diversity, and threats. Ecol. Ind. 33, 45–59.
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., Seybold, S.J., 2008. US Forest Service bark beetle research in the western United States: looking toward the future. J. For. 106 (6), 325–331.
Oliver, C.D., 1980. Forest development in North America following major disturbances. For. Ecol. Manage. 3, 153–168.
RWEACT, 2016. Summation of Forestry Studies for Rio Grande Basin: Watershed Health Forest Biomass Opportunities Evaluation, Watershed and Community Health and Resilience Assessment and Action Plan. Spruce Sawlog Quality Changes due to Spruce Bark Beetle Mortality, p. 10.
Savage, S.L., Lawrence, R.L., Squires, J.R., 2015. Predicting relative species composition within mixed conifer forest pixels using zero-inflated models and Landsat imagery. Remote Sens. Environ. 171, 326–336.
Schmid, J.M., Frye, R.H., 1977. Spruce beetle in the Rockies. USDA Forest Service General Technical Report RM-49.
Schwantes, A.M., Swenson, J.J., Jackson, R.B., 2016. Quantifying drought-induced tree mortality in the open canopy woodlands of central Texas. Remote Sens. Environ. 181, 54–64.
Sibold, J.S., Veblen, T.T., Chipko, K., Lawson, L., Mathis, E., Scott, J., 2007. Influences of secondary disturbance on lodgepole pine stand development in Rocky Mountain National Park. Ecol. Appl. 17 (6), 1638–1655.
Temperli, C., Veblen, T.T., Hart, S.J., Kulakowski, D., Tepley, A.J., 2015. Interactions among spruce beetle disturbance, climate change and forest dynamics captured by a forest landscape model. Ecosphere 6 (11), 1–20.
USFS, 2014. West Fork Fire Complex Fire Programmatic/Cost Fire Review. San Juan and Rio Grand National Forest. National Incident Management Organization, p. 21.
USGS, 2017. What are the best spectral bands to use for my study? <https://landsat.usgs.gov/what-are-best-spectral-bands-use-my-study> (last accessed 11 April 2017).
Walker, L.R., Walker, J., Hobbs, R.J., 2007. Linking Restoration and Ecological Succession (No. Sirsi). Springer, London (i9780387353029).
Xie, Y., Sha, Z., Yu, M., 2008. Remote sensing imagery in vegetation mapping: a review. J. Plan Ecol. 1 (1), 9–23.
Yuan, D., Elvidge, C.D., 1996. Comparison of relative radiometric normalization techniques. ISPRS J. Photogrammetry Remote Sensing 51, 117–126.