Journal of Forestry, 2024, **122**, 31–53
https://doi.org/10.1093/jofore/fvad039
Advance access publication 1 November 2023
Research Article - geospatial technologies

Tree Condition and Analysis Program – Detecting Forest Disturbance at the Tree Level across the Contiguous United States with High Resolution Imagery

Sarah A. Wegmueller,[1,*] William B. Monahan,[2] and Philip A. Townsend[1]

[1]Department of Forest and Wildlife Ecology, University of Wisconsin, 1630 Linden Drive, Madison, WI, 53706, USA (wegmueller@wisc.edu, ptownsend@wisc.edu).
[2]Forest Health Assessment and Applied Sciences Team, USDA Forest Service, 2150A Centre Avenue, Suite 331, Fort Collins, CO, 80526, USA (william.monahan@usda.gov).
[*]Corresponding author email: wegmueller@wisc.edu

Abstract

Effective management of forest insects and diseases requires detection of abnormal mortality, particularly among a single species, sufficiently early to enable effective management. Remote detection of individual trees crowns requires a spatial resolution not available from satellites such as Landsat or Sentinel-2. In the United States, there are currently few operational systems capable of effectively and affordably detecting and mapping tree mortality over broad landscapes using high-resolution imagery. Here, we introduce the Tree Condition and Analysis Program (TreeCAP), an open-source system that uses freely available imagery from the National Agriculture Imagery Program (NAIP) to create maps of tree condition (healthy or damaged). We demonstrate the potential applications of TreeCAP in four study sites: (1) beetle-killed pines in California, (2) emerald ash borer progression in Wisconsin, (3) hemlock wooly adelgid mortality in Pennsylvania, and (4) drought damage in Texas. We achieved an average overall accuracy of 87% across all study sites.

Study Implications: TreeCAP is a software program, ready for operational use, intended to help manage forest health in the contiguous United States at the individual tree level. Using freely available high-resolution NAIP airborne imagery and LiDAR data, TreeCAP maps tree crown condition, highlighting areas that may warrant further attention to forest managers. We demonstrate the potential applications of TreeCAP in four study sites: (1) beetle-killed pines in California, (2) emerald ash borer progression in Wisconsin, (3) hemlock wooly adelgid mortality in Pennsylvania, and (4) drought damage in Texas. We achieved an average overall accuracy of 87% across all study sites.

Keywords: TreeCAP, NAIP, LiDAR, TCH, forest health, remote sensing

Detection of abnormal mortality from forest insects and diseases is important for assessing the impact and trajectory of outbreaks and mitigating potential damage. Many of the deadliest insects and disease are species-specific, meaning that the ability to detect mortality in individual tree crowns in a mixed forest is important. Aerial detection surveys have been used widely to document forest damage over large areas on an annual basis, but results are generally limited to understanding and reporting spatiotemporal patterns of forest conditions (Potter and Conkling 2017). Satellite-based remote sensing detection and mapping efforts have relied heavily on 250 m MODIS imagery, 30 m Landsat imagery, or 10 m Sentinel-2 imagery (Cohen et al. 2010; Housman et al. 2018; Miller et al. 2009; Norman et al. 2013). However, conditions of individual trees within the forest canopy often cannot be discerned at these spatial resolutions because the spectral signature of the few pixels covering damaged trees are mixed with the spectral signatures of the surrounding healthy vegetation. This typically limits applications of remotely sensed methods using these data sources to multi-tree, continuous disturbances (Cohen et al. 2017; Houborg and McCabe 2018; Wulder et al. 2008).

A range of methods have emerged that use high-spatial resolution imagery to map the condition of tree crowns with simple supervised classification. One recent approach, Tree Crown Health (TCH), uses imagery from the United States Department of Agriculture (USDA) National Aerial Imagery Program (NAIP) to evaluate the hue, saturation, and value (HSV) color space of training pixels selected over various tree crown condition classes (Monahan et al., 2022a). TCH classes of tree condition include green, indicating healthy trees; red, indicating damaged or morbid trees (foliage, especially in conifers, will often take on a reddish hue under extreme stress); gray, indicating damaged or dead trees (dead limbs will typically appear gray in color); and shadows, cast by the trees themselves or arising from topography, which are problematic in high-resolution imagery regardless of time of day.

The TCH series of equations—using only the red, green, and blue bands and tuned for each region—yielded a median overall accuracy of 90% and Kappa of 86% within treed areas across 122 different state and year combinations. However, TCH did not include a component for differentiating treed areas from nontree/bare ground areas (NT/BG),

Received: March 13, 2023. **Accepted:** August 23, 2023.

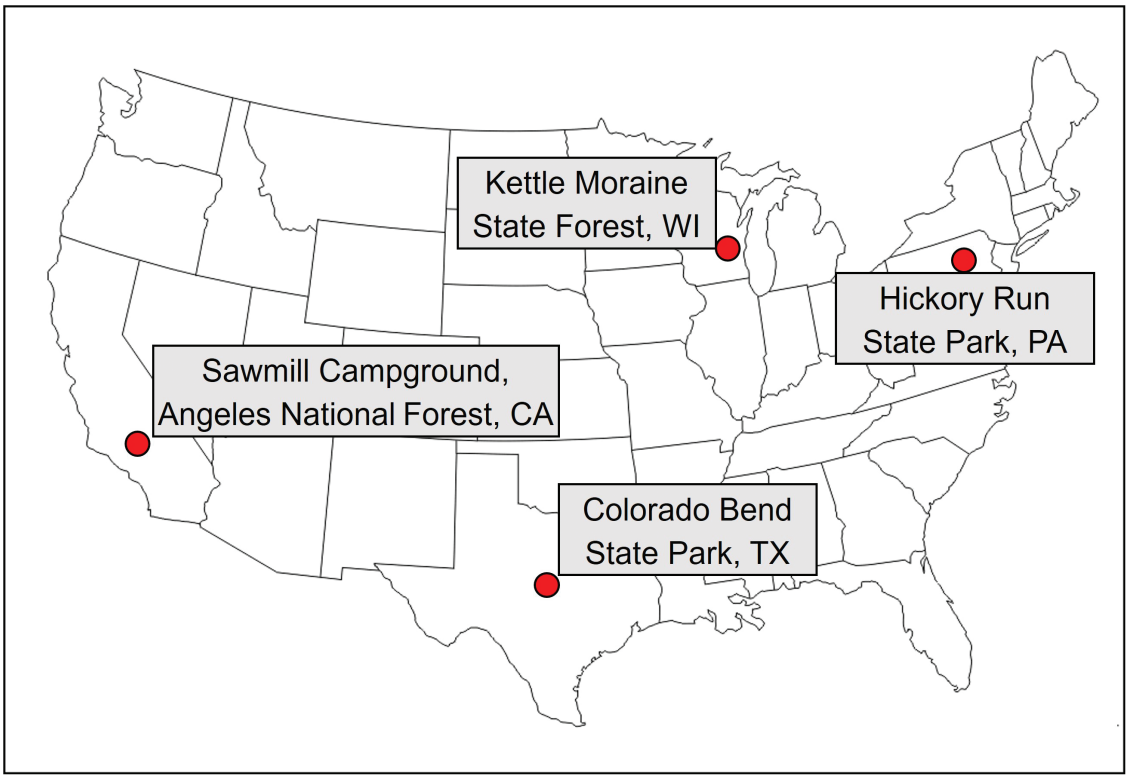


Figure 1. Study sites located in California, Texas, Wisconsin, and Pennsylvania in the USA. Each study site contains a different combination of forest and disturbance type.

and the variation in TCH accuracies across states and years suggest opportunities for alternative models to supplement and improve predictions. Here, we address these two needs by integrating information on NT/BG areas and providing an alternative machine learning model to complement the original TCH mathematical HSV model.

We introduce the Tree Condition and Analysis Program (TreeCAP), which provides a complementary and potentially more accurate machine learning model based on the TCH training data. TreeCAP's novel approach leverages the extreme gradient boosting algorithm (XGBoost) to create a single model that can be used seamlessly anywhere in the contiguous United States to classify NAIP imagery pixels or imagery-derived objects. We address the challenge of misclassification of NT/BG areas within the canopy by creating tree segments using light detection and ranging (LiDAR) data to exclude these areas.

The LiDAR-derived tree segments could enable users to track the extent of damage occurring in a particular tree species (and/or subsequent recovery) through time when combined with TreeCAP's XGBoost model and additional verification data. Although TreeCAP does not attribute a particular causal agent nor specific tree species being damaged, it can provide information regarding spatial patterns, rates, and timing of damage at a considerably more detailed and precise level than is available through other survey methodologies (e.g., aerial detection surveys). These metrics, combined with other information on causal agents and host species, could aid in the management of outbreaks.

We demonstrate the capability and flexibility of TreeCAP at four sites across the United States (figure 1). We show results spatially in two forms: (1) TreeCAP's raster (pixel-based) output, and (2) a vector map (i.e., shapefile) of change detection using the LiDAR-derived tree segments. These maps are intended to be used by forestry professionals to help manage large areas of continuous forests by mapping individual and small clusters of trees, highlighting abnormal or unexpected levels of damage that may warrant further attention.

With its use of freely available data and efficient computation, we believe TreeCAP has the potential to become a valuable tool to help inform management of forest health in the United States. Maps produced by TreeCAP could be used to cue continued monitoring and evaluation of potential issues, especially in cases where imagery or surveillance is available more frequently and with less latency than NAIP. There is also potential for the TreeCAP modeling framework to be applied to high spatial and temporal resolution spaceborne imagery, offering future capabilities for early detection of forest insects and diseases. Finally, TreeCAP could be integrated into an ensemble framework with other complementary models, including any future spaceborne models, which may offer even more accurate resulting maps by leveraging the strengths of multiple models and imagery sources.

Methods

Study Sites

We selected study sites in California, Wisconsin, Pennsylvania, and Texas to evaluate TreeCAP over four different combinations of damage causal agents, forest types, and years (Table 1). The Sawmill Campground area in Angeles National Forest in California exhibited widespread tree mortality in pines (*Pinus spp.)* from a combination of drought and bark beetle attack from 2016 through 2020. In the Kettle Moraine State Forest in Wisconsin, entire stands of ash trees (*Fraxinus*

spp.) were lost to emerald ash borer (*Agrilus planipennis)* from 2018 through 2020. In Pennsylvania's Hickory Run State Park, hemlock woolly adelgid (*Adelges tsugae*) caused mortality in eastern hemlock (*Tsuga canadensis*) from 2013 through 2015. Finally, in Colorado Bend State Park in Texas, drought led to extensive mortality in several species, including junipers (*Juniperus spp.*) and oaks (*Quercus spp.*), from 2010 through 2016.

Data

TCH Training Dataset and Models

The point data used to create the model in TreeCAP was provided by the USDA Forest Service (Forest Service) and is the same data used in TCH (Monahan et al., 2022b). These points were hand-digitized on NAIP imagery from 2012 to 2019 across all forty-eight contiguous states. Information included for each point are the image date, Bailey's ecoregion (Bailey 1995), coordinate, NAIP band values (red, green, blue, and near infrared), and tree condition class at the point. Tree class conditions are green (healthy), red (damaged or morbid), gray (damaged or dead), or shadow. These four tree condition classes were used throughout this study. Individuals were trained to visually inspect NAIP imagery and select representative pixels from each of the four crown color classes (green, red, gray, and shadow) in which to digitize points. These data are available from Monahan et al. (2022a).

In the TCH models, data points were used to solve for four model constants, corresponding to each of the four color or condition classes. The TCH models included for comparison in this study were based on optimal model constants, and the damage prediction threshold was omitted (used to help mitigate NT/BG areas). Imagery for TCH was masked using the same LiDAR-based tree segments used in the TreeCAP maps (details below) to offer a direct comparison of model performance.

NAIP Imagery

NAIP imagery was downloaded from the USGS Earth Explorer archive (USGS 2022a) (Table 2). NAIP is administered by the USDA Farm Service Agency (FSA) through the Aerial Photography Field Office (APFO). Imagery is resampled to 0.6–1 m ground sample distance and orthorectified prior to distribution. Individual images are mosaicked together into uniform tiles; as such, more than one flight line, with varying geometries, may be present in a single tile downloaded by the user. No further processing was done on the NAIP imagery in TreeCAP, as our methods were designed to account for variable scene characteristics within NAIP tiles. Because the TCH training data provided corresponding NAIP band values for each point, the NAIP imagery used in this study was used solely for projecting the XGBoost model (below).

Masking NT/BG Areas Using LiDAR Data

TreeCAP's raster map output benefits greatly from a NT/BG ground mask to reduce false positives by excluding pixels over areas that do not represent trees, such as soil, rivers, or roads (the spectral signature of these areas often resembles that of the red or gray classes). This masking can be accomplished using multiple data sets, including freely available land cover data and LiDAR data. In this study, we excluded urban areas and focused instead on forested areas. Although TreeCAP could be used for urban trees with appropriate masking, this capability was not investigated here. The TreeCAP software

Table 1. Characteristics of the four study sites used to test TreeCAP.

Study site	Area mapped (km^2)	Location	Bailey's Ecoregion Province (ECOCODE)	Primary mortality causal agent	Years mapped
Sawmill Campground in Angeles National Forest	3.5	California	California Costal Range Open Woodland-Shrub-Coniferous Forest-Meadow (M262B)	Drought, pine beetle	2016, 2018, 2020
Kettle Moraine State Forest	14	Wisconsin	Eastern Broadleaf Forest—Continental (-222K)	Emerald ash borer	2018, 2020
Hickory Run State Park	19	Pennsylvania	Central Appalachian Broadleaf Forest-Coniferous Forest-Meadow (M221A)	Hemlock woolly adelgid	2013, 2015
Colorado Bend State Park	22	Texas	Southwest Plateau and Plains Dry Steppe and Shrub (-315D)	Drought	2010, 2016, 2020

Table 2. NAIP imagery tile identifiers and dates used for each study site.

Study Site	NAIP Tiles	NAIP Collection Dates
Sawmill Campground in Angeles National Forest	3411820_NE_11 3411820_NW_11	20160512 20180722 20200501
Kettle Moraine State Forest	4308831_NE_16 4308831_NW_16	20180923 20200830
Hickory Run State Park	4107560_SW_18 4107559_SE_18	20130815/20130623 20150529
Colorado Bend State Park	3109860_SE_14 3109861_SW_14	20100731 20160803 20201011

Table 3. LiDAR data used in this study varied in tile area and density among our study sites.

Study Site	Area per tile (km^2)	Tiles Used	Point Density (per m^2)	Year of Collection
Sawmill Campground in Angeles National Forest	2.5	2	11.51	2018
Kettle Moraine State Forest	1.9	12	4.92	2018
Hickory Run State Park	1	30	8.74	2019
Colorado Bend State Park	2.25	18	4.29	2018

was designed to be very flexible with masking inputs and can use data in both raster and vector formats, accepting multiple files for each as needed.

In this study, we used the National Land Cover Database (NLCD) available from the Multi-Resolution Land Characteristics Consortium (USGS 2022b). The NLCD is derived from Landsat data and is a thematic map of land cover types. Although masking with NLCD is more computationally efficient than processing LiDAR data (described below) and preferred for large NT/BG areas, NLCD is significantly more spatially coarse (30 m pixels) than LiDAR and may exclude areas that users prefer to map. However, as LiDAR data may not yet be available in some parts of the contiguous United States, NLCD could be used for masking, although this is best suited for areas with a mostly continuous canopy. In Hickory Run, we demonstrate how both types of masking (NLCD and LiDAR) can be used concurrently, with NLCD effectively masking large urban areas, roads, and water, and LiDAR masking smaller gaps within the canopy.

We used open-source LiDAR data from USGS's The National Map (TNM) (USGS 2022c) to mask NT/BG areas within a continuous canopy (i.e., areas too small for the NLCD to detect) (Table 3). LiDAR data were used to address two challenges, (1) mitigating classification errors due small patches of NT/BG areas within forests and (2) annotating changes in crown classes through time (addressed in the Using LiDAR-derived Tree Segments to Enable Change Detection section).

We used tree segmentation methods from the lidR R package (Roussel et al. 2020) on LiDAR data to create objects representing trees or parts of trees (Q. Chen et al. 2006). We first located treetops by analyzing the point return heights within a moving window (a small square subset of the data; size allowed to vary to accommodate trees of different heights) across the study area, looking for the maximum height within the window. Using the resulting treetop points, the watershed method was used to delineate individual trees. This method can be thought of as similar to draping a dust cloth over a series of chairs in a room; the cloth will conform to the chair shapes. A dividing line is assessed based on the shapes of the resulting domes. In this simplified analogy, the chair domes represent trees. This segmentation method is best suited for conifers because deciduous trees are more structurally complex.

To our knowledge, there is no LiDAR-based tree segmentation method that performs well across all forest types, and most struggle with deciduous stands (Ayrey et al. 2017; Jeronimo et al. 2018). Further, LiDAR data are collected at nadir, whereas NAIP imagery is generally off-nadir except at scene centers, with solar illumination further distorting apparent surface reflectance. Consequently, some of the resulting LiDAR-derived segments do not line up well with NAIP's spectral data of the same tree(s). Because NAIP imagery are delivered as mosaics with multiple flightlines per tile (and thus irregular viewing geometries within the tile), co-registration of the LiDAR and NAIP is problematic because the quality of the registration will vary within the tile. Because of these factors, registration was not attempted here.

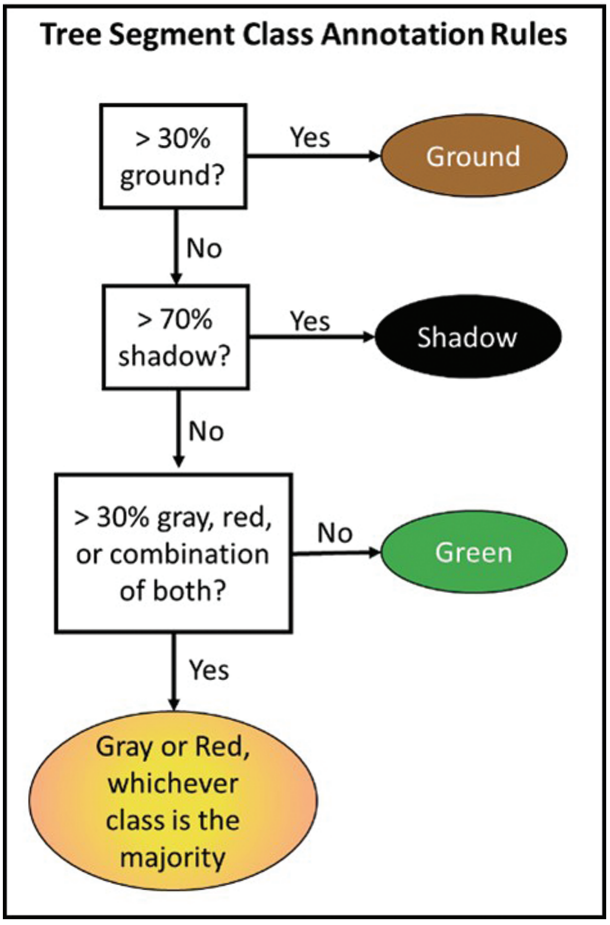


Figure 2. The process of determining a tree segment class after the XGBoost model had predicted the class of all pixels contained within that segment. The rule for gray and red classes is based on the Forest Service's thresholds for determining polygon damage severity using the Digital Mobile Sketch Mapping (DMSM) system.

Although in many cases the segmentation method does correctly delineate individual trees, particularly conifers, our goal was simply to create segments that approximated individual trees. Particularly in deciduous forests, this meant that some segments may contain small groups of trees (one to three small trees, for example). This was still suitable for

Table 4. The fuzzy accuracy level definitions used to evaluate TreeCAP accuracy (adapted from Gopal and Woodcock, 1994).

Level	Evaluation	Definition
5	Absolutely Right	No doubt about match. Perfect.
4	Good Answer	Would be happy to find this answer given the map. Very right. Segments where predicted value and truth value differ in number of underlying pixels by less than 10%. Also, segments which were predicted green (healthy) when the truth value was shadow.
3	Reasonable or Acceptable Answer	Maybe not the best possible answer, but it is acceptable; this answer does not pose a problem to the user if it is seen on the map. Right. Used for segments predicted to be red (morbid) when they were gray (dead), and vice versa.
2	Understandable but Wrong	Not a good answer. There is something about the segment that makes the answer understandable, but there is clearly a better answer. This answer would pose a problem for user of the map. Not right. Used for all other errors involving shadows.
1	Absolutely Wrong.	This answer is absolutely unacceptable. Very wrong. Used for segments predicted to be green (healthy) when they were gray (dead) or red (morbid), or vice versa.
0	Non-Tree/Bare Ground	Segments which occurred primarily over non-treed or bare ground areas. This error is a result of masking and not the classification algorithm.

Table 5. The confusion matrix and accuracy of the XGBoost model used in TreeCAP

		n	Predicted				Producer's accuracy %
			Green	Red	Gray	Shadows	
Reference data	Green	15665	15056	99	234	614	96
	Red	3512	33	3075	169	88	88
	Gray	5887	200	256	5092	294	86
	Shadows	10612	376	82	392	9616	91
User's accuracy %		35676	94	91	87	92	92

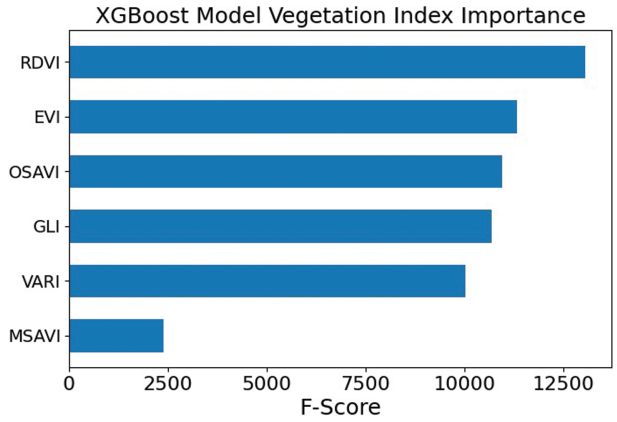


Figure 3. The ranking of feature, or variable, importance in the XGBoost model. Higher F-scores indicate more importance in determining the correct class. All vegetation indices are described in Table S6.

our purposes of locating damage at the tree level (discussed further in the Using LiDAR-derived Tree Segments to Enable Change Detection section), but we do not recommend segments be used for inventory purposes.

Because the segments conformed to the trees on the landscape better than data such as NLCD, they were more useful for excluding pixels in the image that represented NT/BG areas. By excluding pixels outside of the resulting LiDAR-derived tree segments, errors from NT/BG areas were significantly reduced as compared with using only NLCD to exclude these areas (Tables S1-4, Figures S1-4).

XGBoost Model: Inputs and Parameters

To maximize the potential of the categorical data available in the TCH training dataset while also minimizing the computational requirements for prediction, we use XGBoost in TreeCAP. XGBoost is a fast, scalable implementation of gradient-boosted regression trees, an ensemble machine learning classification technique that performs well with large, multivariable datasets, (Chen and Guestrin, 2016). XGBoost builds decision trees for classification in succession instead of independently. Each new tree is a result of what has been learned from classification errors in previous trees, producing better results with each successive tree (Chen and Guestrin 2016; Friedman 2001; Friedman et al. 2008).

Our goal was to create a model that performed well in the contiguous United States, was potentially expandable beyond this geographical range, and was designed to be testable with other image sources. To do this, we used 238,509 points from the TCH dataset with the XGBoost algorithm to test the predictive power of several vegetation indices (VIs) suitable for four-band imagery using only the spectral data from the TCH points. We started with a large selection of VIs and iteratively eliminated them based on their predictive power. We also ensured that VIs known to distinguish shadows well were included, regardless of their predictive power, to help avoid overfitting the model (Zhang et al. 2015). Our final model included six VIs: Renormalized Difference VI (RDVI), Enhanced VI (EVI), Optimized Soil Adjusted VI (OSAVI), Modified Soil Adjusted VI (MSAVI), Green Leaf Index (GLI), and Visible Atmospherically Resistant Index (VARI) (details

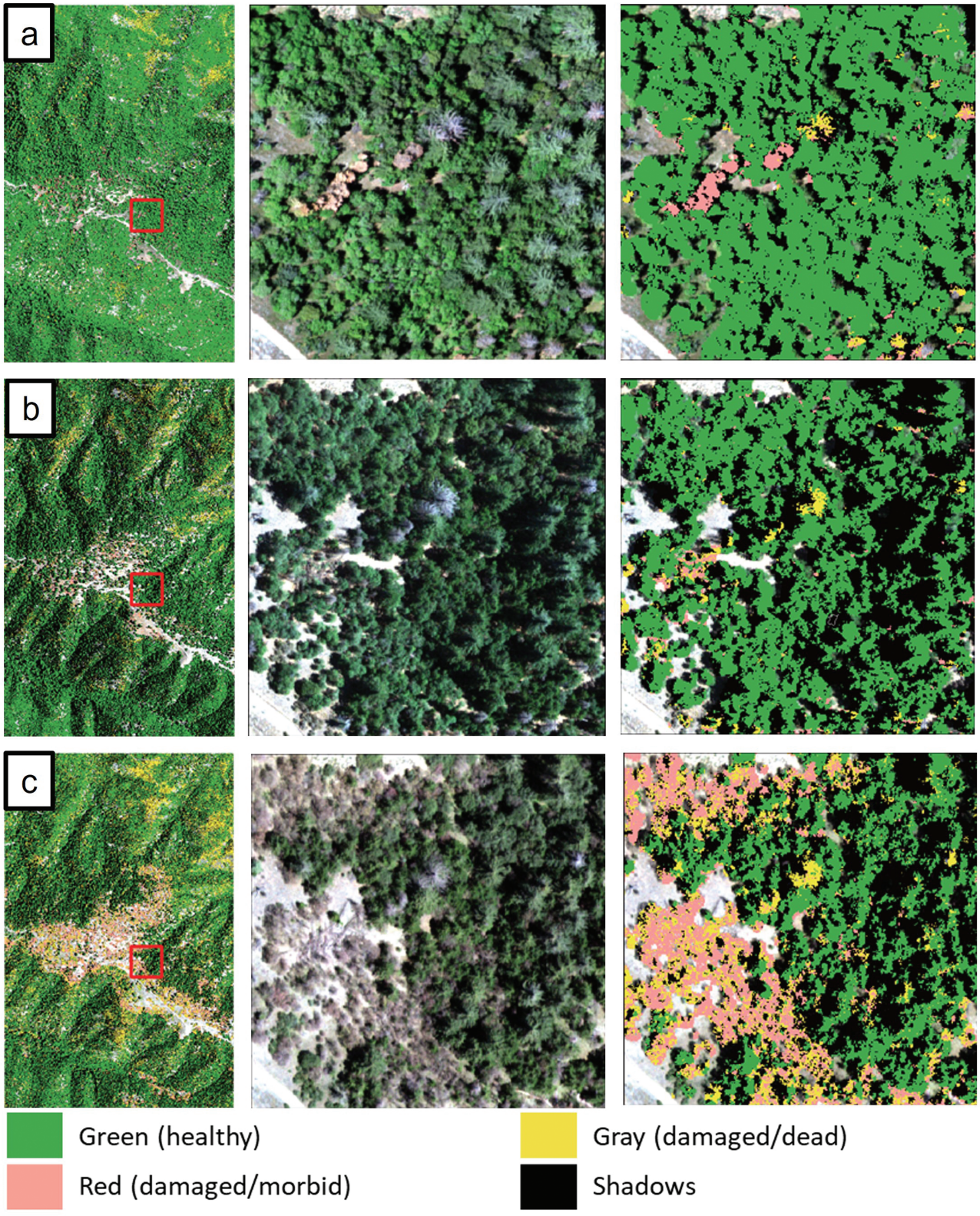


Figure 4. Sawmill Campground from 2016 to 2020. The first column is the entire study area as modeled by the TreeCAP raster output overlaid on the corresponding NAIP imagery. The red box is the 200 m subset area shown in the next two columns. The middle column is the original NAIP imagery with a 2% linear stretch applied and the right column is the TreeCAP raster output. Row (a) is 2016, row (b) is 2018, and row (c) is 2020. The drastic changes in the red (morbid) and gray (dead) classes (shown in pink and yellow colors, respectively) are due to drought and pine beetle infestation.

on computation are provided in Table S6). Parameters used for XGBoost model development (XGBoost Documentation) are provided in Table S5.

Once trained, the XGBoost model was then applied to classify each pixel within the NAIP images covering our study sites. This resulted in a raster output of only five values: 0,

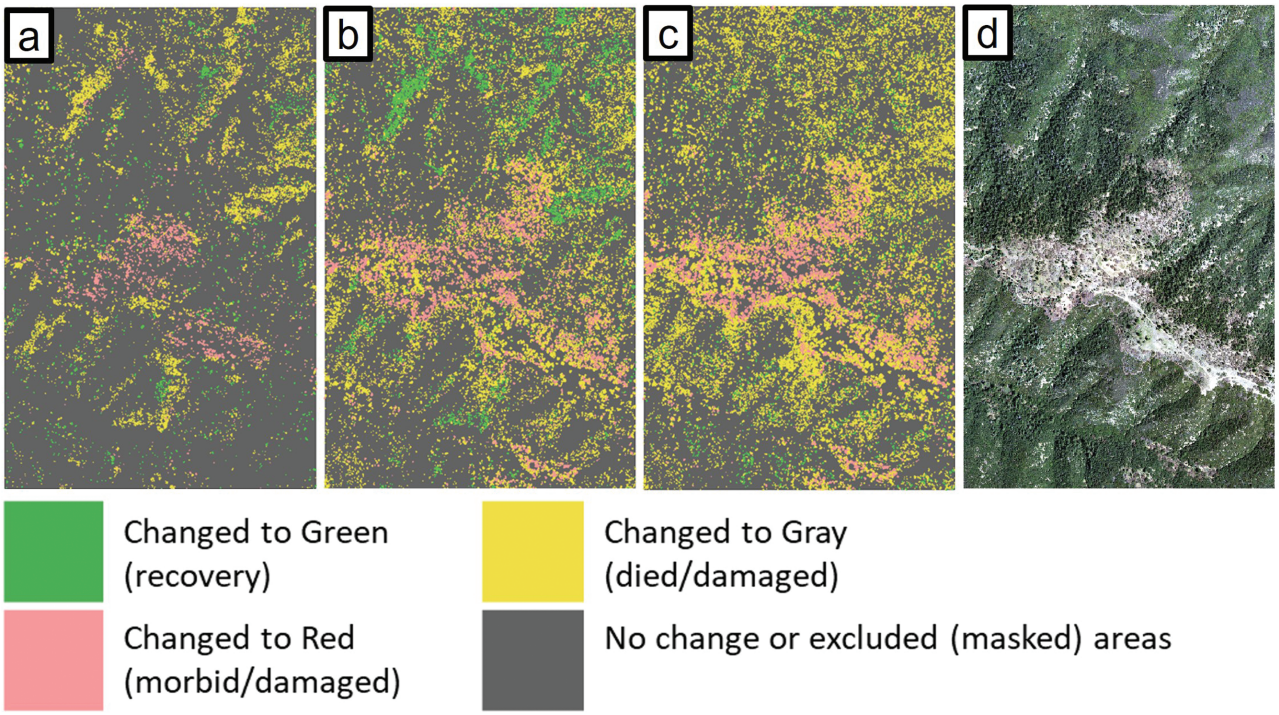


Figure 5. Changes highlighted by TreeCAP at Sawmill Campground from 2016 to 2020 using the tree segment vector layer to compare changes among years. (a) is the difference in tree condition between 2016 and 2018, (b) is the difference between 2018 and 2020, (c) is the difference between 2016 and 2020, and (d) is the 2020 NAIP imagery with a 2% linear stretch applied. The trend of high mortality (shown in pink and yellow) is extensive by 2018 and 2020.

Table 6. Quantification of the changes observed between sets of years in Sawmill Campground study area in hectares (ha) and percents.

	Change between images ha (% of area mapped)		
	2016–2018	2018–2020	2016–2020
Change to green	3 (1%)	8 (3%)	4 (2%)
Change to red	4 (2%)	11 (4%)	13 (5%)
Change to gray	15 (6%)	33 (13%)	48 (20%)
Change to shadow	61 (25%)	47 (19%)	31 (12%)
No change	164 (66%)	148 (60%)	151 (61%)

1, 2, 3, and 99, corresponding to shadows, green, red, gray, and no data (or masked areas), respectively. Prediction by the XGBoost model for each map took less than 5 minutes on an individual workstation (computational resources used are discussed below). The TreeCAP software has built-in batching capability that can be scaled depending on the computational power of the operating platform.

Using LiDAR-derived Tree Segments to Enable Change Detection

TreeCAP's methodology makes it possible to investigate change in crown condition among years. Solar azimuth and view angle variation in NAIP imagery present challenges when attempting a pixelwise comparison across years, so the LiDAR-derived tree segments (discussed above) provide a basis to compare maps of different years. For each segment, we annotated the best class, based on the underlying TreeCAP raster output for each year, as an attribute in the shapefile. These classes were determined based on rules that mimic those used by the Forest Service in their Digital Mobile Sketch Mapping system (USDA Forest Service 2019) and are intended to maximize the usability of these maps for forestry professionals (figure 2). In particular, segments containing at least 30% gray or red pixels were annotated as such (i.e., gray or red). If a segment contained pixels from both the gray and red class and they collectively represented greater than 30% of the total pixels in that segment, that segment was annotated with the class containing more pixels. For example, in a segment containing fifteen pixels, with four classified as red and six as gray, that segment will be annotated as gray. This feature of TreeCAP is particularly useful for quickly locating areas of recent damage, morbidity, or mortality on the landscape. However, this feature can also be used to map recovery (i.e., changes from red or gray to green).

Accuracy Assessment

Accuracy was assessed based on visual interpretation of the underlying NAIP imagery. The 0.6–1 m spatial resolution of NAIP makes it possible to identify individual trees to make assessments of crown condition. We use tree segments instead of individual pixels because pixels within all four model classes are all on continuous color gradients, and distinguishing among them can at times be extremely difficult for human vision and perception (i.e., is that canopy edge pixel more shadow or green?) (Erker et al. 2019). In contrast, within a segment containing a *group* of pixels, the appropriate class was usually clear to a human analyst.

Using tree segments as opposed to pixels is also a more meaningful way to assess the accuracy of the TreeCAP's output for operational uses. For example, if a tree contains a single large gray (damaged) branch covered by only one to

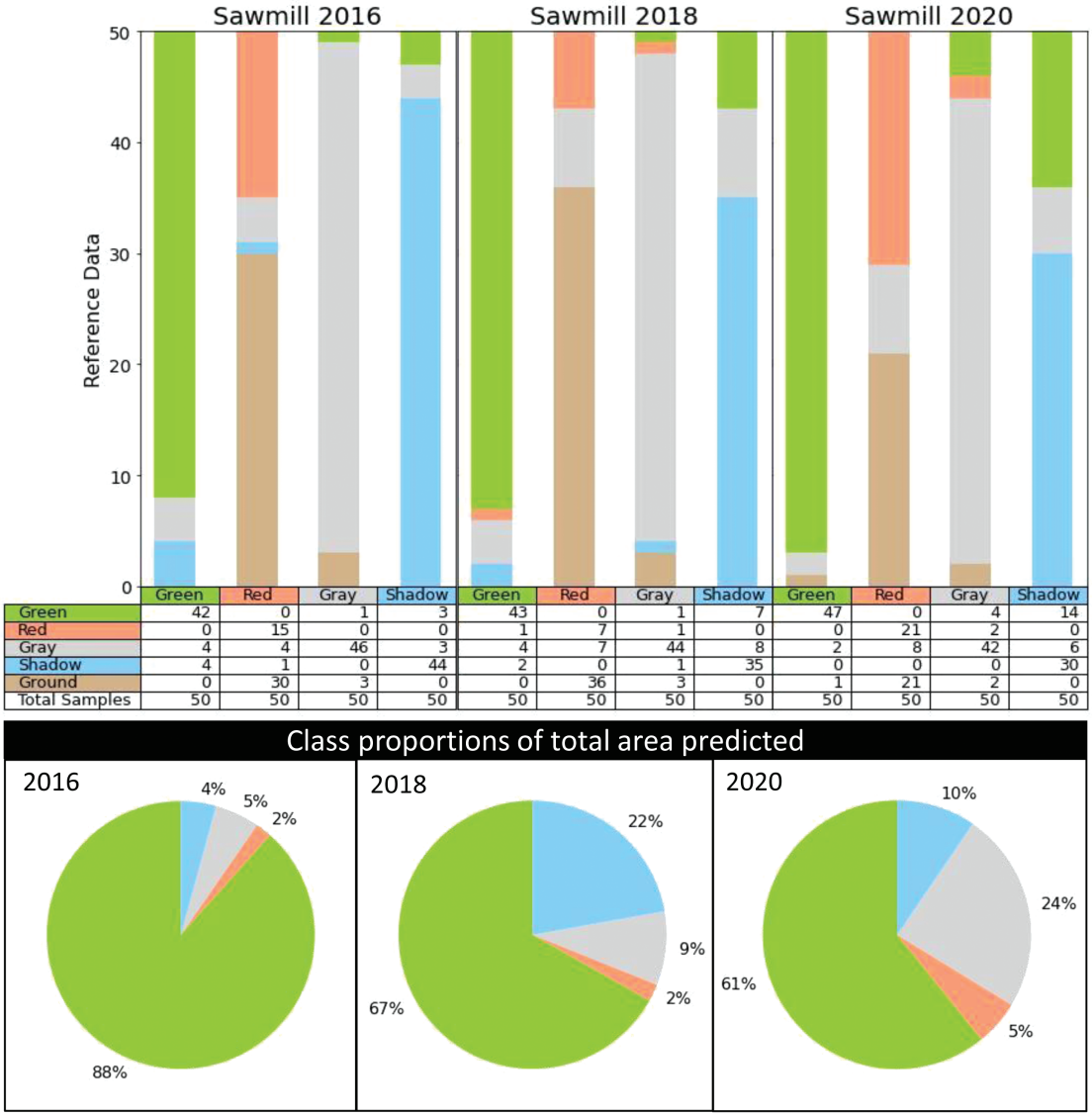
Table 7. The traditional and fuzzy accuracy data for the Sawmill Campground study site, 2016, 2018, and 2020. The left column provides accuracy values if masking errors (segments that were over nontree/bare ground areas) are excluded. The right column includes these errors. Producer's accuracy only assesses performance on the segments known to belong to these respective classes (i.e., it is the same for both columns because masking errors are already excluded in producer's accuracy). Masking errors often disproportionally affected the red class due to the scarcity of these segments and our sampling scheme.

		Model accuracy (excluding masking errors)			Model + masking accuracy (including masking errors)		
		n	Traditional (%)	Fuzzy (%)	n	Traditional (%)	Fuzzy (%)
2016							
User's Accuracy	Green	50	42 (84)	50 (100)	50	42 (84)	50 (100)
	Red	20	15 (75)	19 (95)	50	15 (30)	19 (38)
	Gray	47	46 (98)	47 (100)	50	46 (92)	47 (94)
	Shadow	50	44 (88)	47 (94)	50	44 (88)	47 (94)
Producer's Accuracy	Green	46	42 (91)	46 (100)	Does not change. (33 segments are masking errors)		
	Red	15	15 (100)	15 (100)			
	Gray	57	46 (81)	54 (95)			
	Shadow	49	44 (90)	48 (98)			
		Overall accuracy = 88% traditional, 98% fuzzy			**Overall accuracy = 74% traditional, 82% fuzzy**		
2018							
User's Accuracy	Green	50	43 (86)	50 (100)	50	43 (86)	50 (100)
	Red	14	7 (50)	14 (100)	50	7 (14)	14 (28)
	Gray	47	44 (94)	46 (98)	50	44 (88)	46 (92)
	Shadow	50	35 (70)	44 (88)	50	35 (70)	44 (88)
Producer's Accuracy	Green	51	43 (84)	49 (96)	Does not change. (39 segments are masking errors)		
	Red	9	7 (78)	9 (100)			
	Gray	63	44 (70)	59 (94)			
	Shadow	38	35 (92)	37 (97)			
		Overall accuracy = 80% traditional, 96% fuzzy			**Overall accuracy = 64% traditional, 77% fuzzy**		
2020							
User's Accuracy	Green	49	47 (96)	49 (100)	50	47 (94)	49 (98)
	Red	29	21 (72)	29 (100)	50	21 (42)	29 (58)
	Gray	48	42 (88)	48 (100)	50	42 (84)	48 (96)
	Shadow	50	30 (60)	42 (84)	50	30 (60)	42 (84)
Producer's Accuracy	Green	65	47 (72)	59 (91)	Does not change. (24 segments are masking errors)		
	Red	23	21 (91)	23 (100)			
	Gray	58	42 (72)	56 (97)			
	Shadow	30	30 (100)	30 (100)			
		Overall accuracy = 80% traditional, 95% fuzzy			Overall accuracy = 70% traditional, 84% fuzzy		

three pixels but is otherwise healthy, it is unlikely to be of concern to a forest manager. However, if that same tree has several branches that are damaged and many more pixels are classified as red or gray, it may be more concerning, especially if other similarly damaged trees are nearby. Using a segment-based approach means segments are only classified as damaged when a significant portion (30% or more) of the enclosed pixels are classified as damaged (the visual classification of each segment followed the same rule set defined in figure 2). This reduces "speckle" in the resulting map and draws attention to areas more likely to represent significant damage.

The accuracy assessments of each map used a random sample of fifty segments per class per map. Because the green class dominated most maps, a fully random or stratified random sample set would have returned a low number of samples for the red and gray classes, when in fact, these are the classes of greatest interest to users. In maps with classes containing fewer than fifty total segments in a particular class, the validation sample contained all available segments.

In addition to reporting traditional accuracy numbers, we also adapted the fuzzy accuracy assessment methods for thematic maps described by Gopal and Woodcock (1994). This



	Green	Red	Gray	Shadow	Green	Red	Gray	Shadow	Green	Red	Gray	Shadow
Green	42	0	1	3	43	0	1	7	47	0	4	14
Red	0	15	0	0	1	7	1	0	0	21	2	0
Gray	4	4	46	3	4	7	44	8	2	8	42	6
Shadow	4	1	0	44	2	0	1	35	0	0	0	30
Ground	0	30	3	0	0	36	3	0	1	21	2	0
Total Samples	50	50	50	50	50	50	50	50	50	50	50	50

Figure 6. These bar graphs provide insight on the origin of the errors in the assessment data for the Sawmill Campground study site. The pie charts show the proportion of that class for the entire map, regardless of total sample size; classes not seen in the pie charts mean the proportion was significantly less than 1%. Although some errors seem excessive (such as in the red class), they often represent a small portion of the map and are generally concentrated (such as in the campground area after tree removal) and do not obscure the observed trends.

method provides additional valuable information about the classification performance of an algorithm by quantifying the qualitative magnitude of the errors. For example, a pixel classified as gray when it was actually red—classes that both indicate trees are likely damaged—is not as grievous as the same pixel classified as green. The fuzzy set approach can also account for our use of segments for validation in which the quantities of different mapped classes may be mixed. We adapted the fuzzy set classification of Gopal and Woodcock (1994) to the TreeCAP evaluation with three levels of acceptable classifications, two incorrect levels due to XGBoost misclassification and one misclassification level due to masking errors (Table 4). Fuzzy accuracy is calculated for each class by grouping all three levels of acceptable classification (levels 3, 4, and 5) as "right" and unacceptable levels (levels 1 and 2) as "wrong."

Accuracy tables are reported with and without masking errors (fuzzy level 0) to separate errors due to masking from the performance of the XGBoost model on the spectral data from trees. We report these complementary accuracies to demonstrate the performance of the XGBoost model when masking is perfect (i.e., no NT/BG areas remain after masking). Masking quality is a result of user input and has the potential to be improved, such as with more concurrent LiDAR data, whereas the XGBoost model is built into TreeCAP.

Computational Resources

TreeCAP was developed and implemented on a laptop with the PopOS 21.10 operating system (Linux-based), 32 GB of RAM, and an Intel Core i7-9700k CPU with eight cores. TreeCAP is written in Python 3.9 using several libraries, but notably xgboost, numpy, rasterio, geopandas, and rsgislib (Clewley et al. 2014). A Docker image for TreeCAP has also been created and is available for download at https://hub.docker.com/repository/docker/sarahweg/treecap.

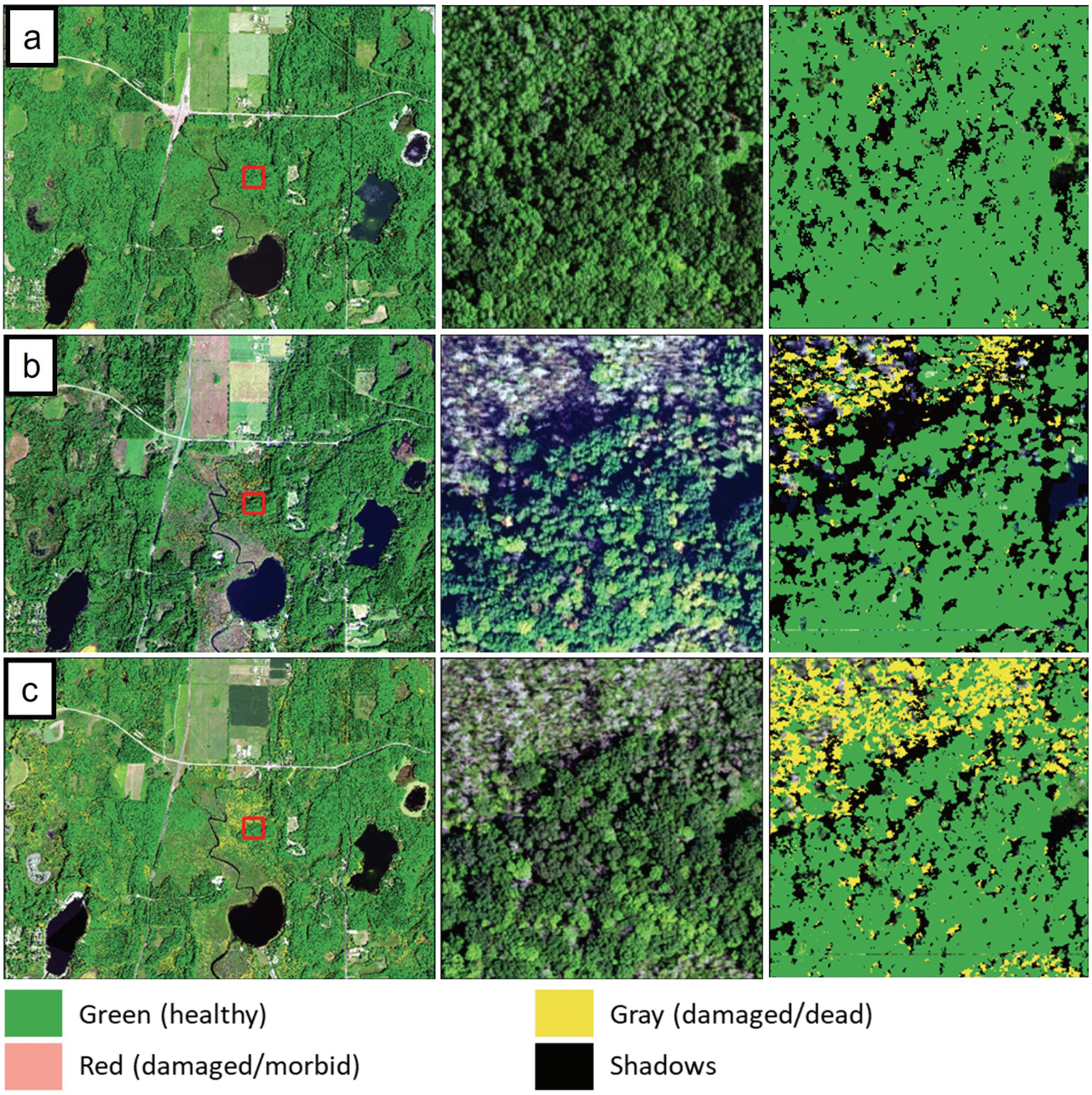


Figure 7. Kettle Moraine State Forest study site from 2016 to 2020. The first column are overviews on the entire study area as modeled by the TreeCAP raster output overlaid on the corresponding NAIP imagery. The red box is the 200 m subset area shown in the next two columns. The middle column is the original NAIP imagery with a 2% linear stretch applied, and the right column is the TreeCAP output. Row (a) is 2017, row (b) is 2018, and row (c) is 2020. The drastic increase in mortality (shown in yellow) is due to emerald ash borer infestation.

Results

XGBoost Model Performance

Our XGBoost model achieved an overall accuracy of 92% on the TCH data, with individual class accuracies all between 86% and 96% (Table 5). Figure 3 provides a summary of feature, or variable, importance in the final XGBoost model. During model development (including and excluding VIs), we found that the VIs most useful for improving distinction among vegetation classes were EVI, GLI, and VARI. VIs that improved distinction between the classes of gray and shadow were RDVI, OSAVI, and MSAVI (Table S6).

Performance by Site

For each study site, accuracy tables are given with both traditional and fuzzy accuracy estimates, with and without masking errors. In most cases, the difference between the traditional accuracy and fuzzy accuracy values are not drastic, indicating our definitions for the fuzzy category set described in Table 4 are appropriate. In most cases, any NT/BG errors remaining after masking affected only a minimal portion of the map, although this was disproportionately represented in the red class, likely due to its low occurrence and our nonproportionate sampling scheme. A minimal red class can occur in certain geographies when tree species do not tend to experience an extensive red phase during damage

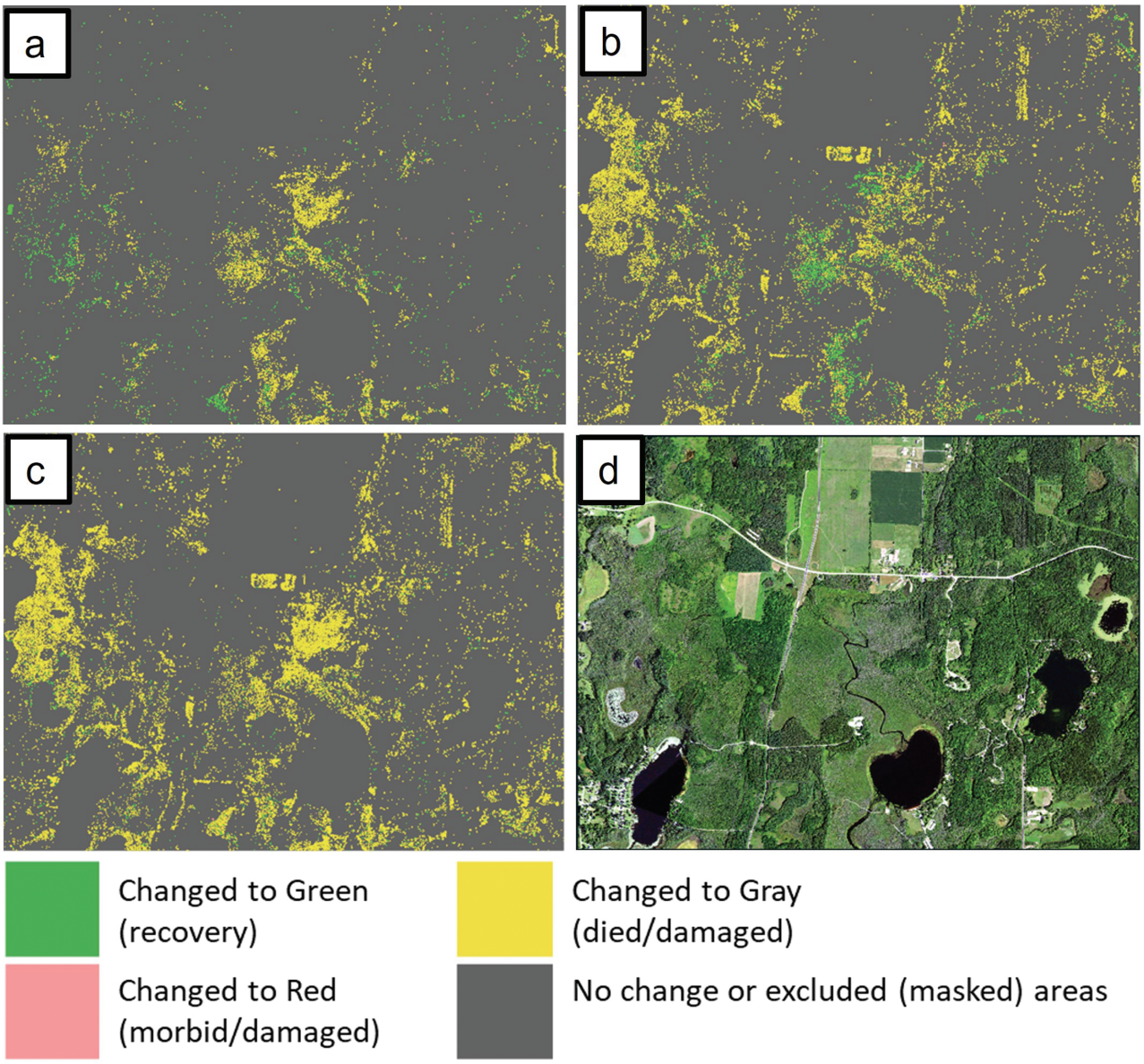


Figure 8. Changes highlighted by TreeCAP in the Kettle Moraine study site from 2017 to 2020. (a) is the difference in tree condition between 2017 and 2018, (b) is the difference between 2018 and 2020, (c) is the difference between 2017 and 2020, and (d) is the 2020 NAIP imagery with a 2% linear stretch applied. The spread of mortality due to emerald ash borer is clearly evident (shown in yellow).

Table 8. Quantification of the changes observed between sets of years in the Kettle Moraine study site in hectares (ha) and percents.

	Change between images ha (% of area mapped)		
	2017–2018	2018–2020	2017–2020
Change to green	6 (1%)	8 (1%)	3 (<1%)
Change to red	None	None	None
Change to gray	19 (3%)	60 (8%)	81 (11%)
Change to shadow	86 (12%)	78 (11%)	33 (5%)
No change	604 (84%)	569 (80%)	596 (83%)

(if at all) or when trees died in years prior to the imagery and remain standing dead/gray. For example, in Colorado Bend's 2020 map, only forty-three segments in total were classified as red and all were used for assessment purposes with only one segment representing canopy, as opposed to NT/BG. The disproportionate amount of NT/BG error in the red class make these errors appear more prominent than they actually are.

We compared accuracy of TreeCAP with TCH using the same tree segments (Table S7). We found that both models performed well in general at all sites but differed in the nature of their errors. For example, TreeCAP predicted the gray class more accurately in the Kettle Moraine and Hickory Run sites, whereas TCH obtained a higher accuracy in the same class in the Colorado Bend study site. We address these implications in the Discussion section.

Sawmill Campground, California

TreeCAP-produced maps show a clear progression visually of increasing mortality from 2016 through 2020 at Sawmill

Table 9. The traditional and fuzzy accuracy data for the Kettle Moraine study site, 2017, 2018, and 2020. The left column provides accuracy values if masking errors (segments that were over nontree/bare ground areas) are excluded. The right column includes these errors. Producer's accuracy only assesses performance on the segments known to belong to these respective classes (i.e., it is the same for both columns because masking errors are already excluded in producer's accuracy).

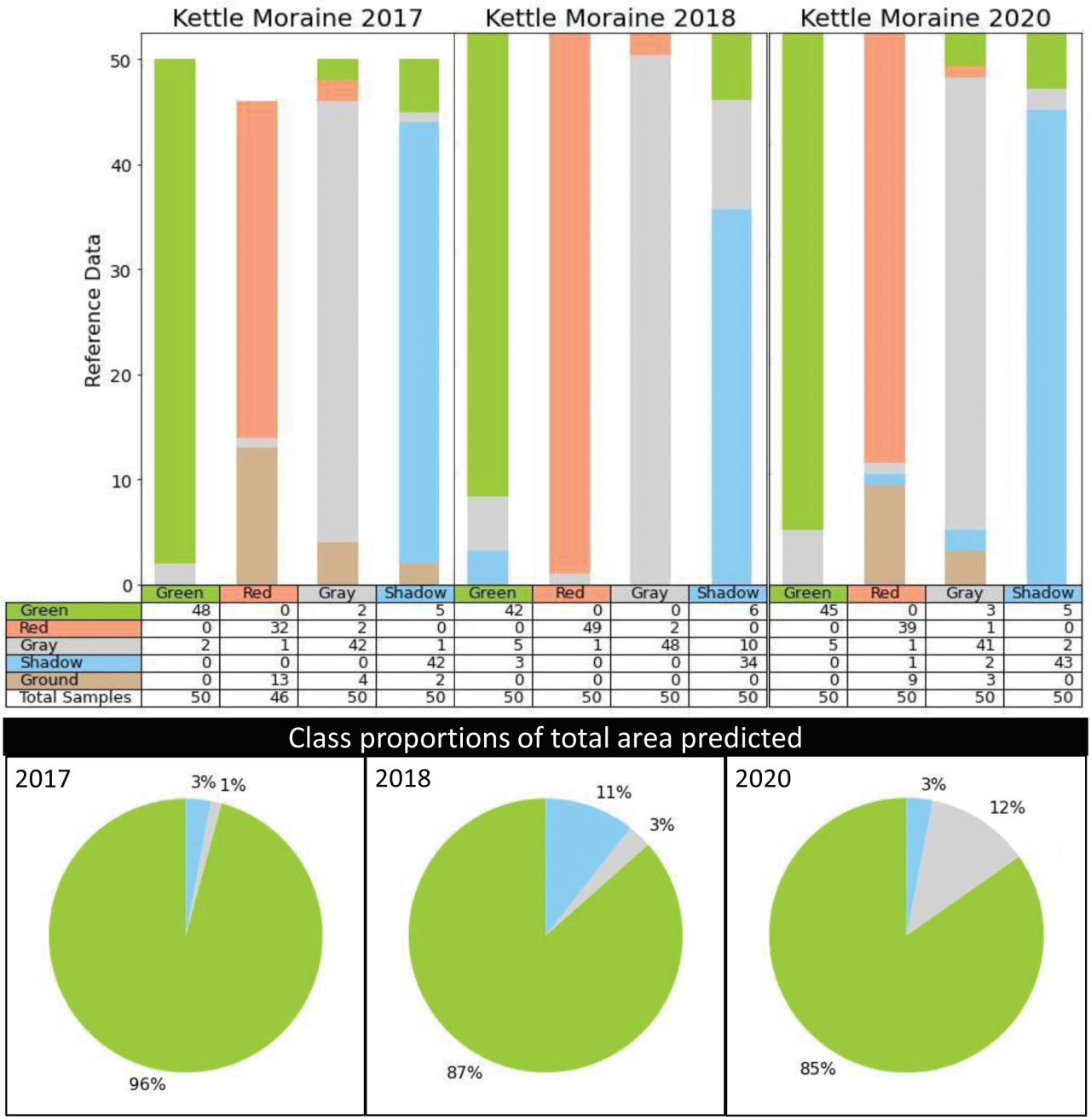
		Model accuracy (excluding masking errors)			Model + masking accuracy (including masking errors)		
		n	Traditional (%)	Fuzzy (%)	n	Traditional (%)	Fuzzy (%)
2017							
User's Accuracy	Green	50	48 (96)	50 (100)	50	48 (96)	50 (100)
	Red	33	32 (97)	33 (100)	46	32 (70)	33 (72)
	Gray	46	42 (91)	46 (100)	50	42 (84)	46 (92)
	Shadow	48	42 (88)	46 (96)	50	42 (84)	46 (92)
Producer's Accuracy	Green	55	48 (87)	53 (96)	Does not change. (4 segments are masking errors)		
	Red	34	32 (94)	34 (100)			
	Gray	46	42 (91)	46 (100)			
	Shadow	42	42 (100)	42 (100)			
		Overall accuracy = 93% traditional, 99% fuzzy			Overall accuracy = 84% traditional, 89% fuzzy		
2018							
User's Accuracy	Green	50	42 (84)	50 (100)	Does not change (no masking errors)		
	Red	50	49 (98)	50 (100)			
	Gray	50	48 (96)	50 (100)			
	Shadow	50	34 (68)	44 (88)			
Producer's Accuracy	Green	48	42 (88)	46 (96)			
	Red	51	49 (96)	51 (100)			
	Gray	64	48 (75)	60 (94)			
	Shadow	37	34 (92)	37 (100)			
		Overall accuracy = 86% traditional, 97% fuzzy					
2020							
User's Accuracy	Green	50	45 (90)	50 (100)	50	45 (90)	50 (100)
	Red	41	39 (95)	40 (98)	50	39 (78)	40 (80)
	Gray	47	41 (87)	45 (96)	50	41 (82)	45 (90)
	Shadow	50	43 (86)	49 (98)	50	43 (86)	49 (98)
Producer's Accuracy	Green	53	45 (85)	52 (98)	Does not change. (12 segments are masking errors)		
	Red	40	39 (98)	40 (100)			
	Gray	49	41 (84)	49 (100)			
	Shadow	46	43 (93)	43 (93)			
		Overall accuracy = 89% traditional, 98% fuzzy			Overall accuracy = 84% traditional, 92% fuzzy		

Campground area resulting from drought and pine beetle infestation (figure 4). The segment-based change detection maps illustrate how tree segments can be used to quickly identify areas of significant change between years (figure 5). The trend of high mortality is evidenced by the increasing number of segments that changed from the green class to either the red class (shown in pink) or gray class (shown in yellow) between 2016 and 2020 (Table 6).

Traditional accuracy for each map ranged from 80% to 88%, whereas fuzzy accuracy for each map ranged from 95% to 98% (Table 7). The most common errors were a result of confusion between red and gray. Errors as a result of NT/BG affected primarily the red class, as the soil in this area is often reddish in color (figure 6, Table 7). Of note, the proportion of NT/BG errors decreases as tree health declines. In contrast, segments classified as gray and red were more abundant in 2020, leading to a sample more representative of the images as a whole. However, there were also several trees removed from the campground between 2018 and 2020, after the LiDAR data was collected, resulting in additional errors concentrated in this area.

Kettle Moraine State Forest, Wisconsin

In the Kettle Moraine State Forest study site, mortality caused by emerald ash borer infestation is apparent between the maps of 2017 and 2020 (figures 7 and 8, Table 8). The



	Green	Red	Gray	Shadow	Green	Red	Gray	Shadow	Green	Red	Gray	Shadow
Green	48	0	2	5	42	0	0	6	45	0	3	5
Red	0	32	2	0	0	49	2	0	0	39	1	0
Gray	2	1	42	1	5	1	48	10	5	1	41	2
Shadow	0	0	0	42	3	0	0	34	0	1	2	43
Ground	0	13	4	2	0	0	0	0	0	9	3	0
Total Samples	50	46	50	50	50	50	50	50	50	50	50	50

Figure 9. These bar graphs provide insight into the origin of the errors in the assessment data for the Kettle Moraine study site. The pie charts show the proportion of that class for the entire map, regardless of total sample size; classes not seen in the pie charts mean the proportion was significantly less than 1%. Although some errors seem excessive (such as in the red class), they often represent a small portion of the map and do not obscure the observed trends.

traditional accuracy for all maps ranged from 84% to 92% whereas the fuzzy accuracy for all maps ranged from 97% to 99% (Table 9). However, in this study area, a large proportion of the red class was due to early fall senescence, particularly in maples (*Acer* spp.) (figure 9). Although the red class prediction on senescent trees is not incorrect given the XGBoost model's spectral training data, it highlights the importance of selecting NAIP imagery before this time period. Ground errors were minimal at this site, resulting in high accuracies in all classes (most over 80%) despite remaining errors (Table 9).

Hickory Run State Park, Pennsylvania

In Hickory Run State Park, mortality due to hemlock woolly adelgid infestation is apparent when comparing the 2015 map to the 2013 map (figures 10 and 11, Table 10). Traditional accuracies for the XGBoost model in these maps were 97% (2013) and 73% (2015), and fuzzy accuracies for these maps were 99% and 95%, respectively (Table 11). As in the Kettle Moraine study area, both maps had insignificant treed segments classified red (also gray in 2013), too few to provide meaningful accuracy metrics in this class for this map. Another error often exhibited in this site was due to

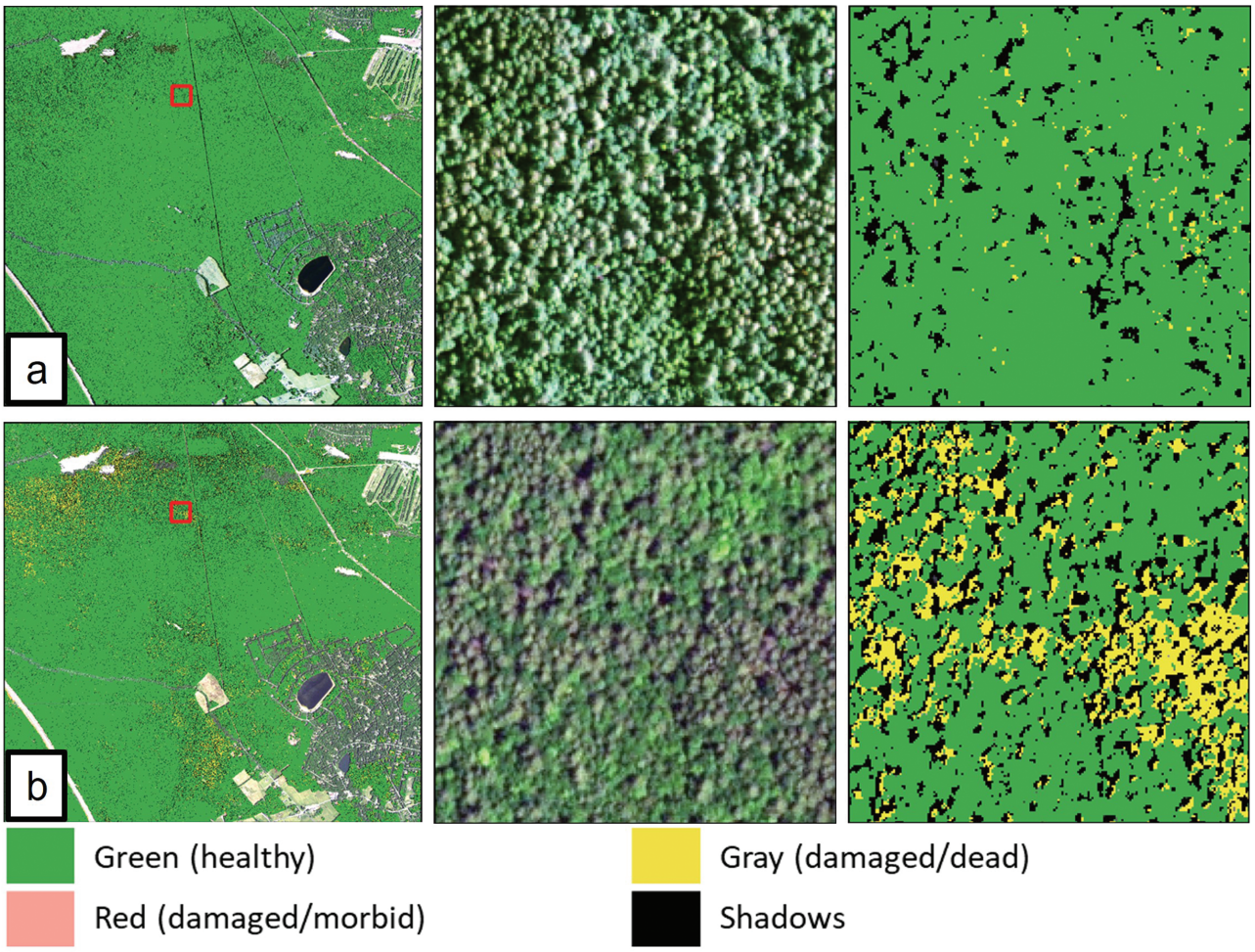


Figure 10. Hickory Run State Park study site in 2013 and 2015. The first column is overviews on the entire study area as modeled by the TreeCAP raster output overlaid on the corresponding NAIP imagery. The red box is the 200 m subset area shown in the next two columns. The middle column is the original NAIP imagery with a 2% linear stretch applied and the right column is the TreeCAP raster output. Row (a) is 2013 and row (b) is 2015. The increase in mortality, shown in yellow, is due to hemlock wooly adelgid infestation.

confusion between the gray class and shadows in 2015 (figure 12). However, despite the confusion with shadows, enough of the gray class was correctly mapped that the signature of increased mortality in general remains clear.

Colorado Bend State Park, Texas

In Colorado Bend State Park, the trend in mortality as a result of drought and some evidence of recovery can be seen in the maps of the study area from 2010, 2016, and 2020 (figures 13 and 14, Table 12). Model accuracies for these maps ranged from 86% and 93% for traditional accuracy, and 94% to 100% for fuzzy accuracy (Table 13, "Model Accuracy"). Despite good performance on this site from the XGBoost algorithm, the 2010 map suffered from false positives due to poor masking in comparison to maps from all other sites in this study (figure 13, Table 13 "Model + Masking Accuracy"). This is due to particularly disparate spatial differences in areas occupied by trees in the LiDAR data from 2018 and the 2010 NAIP imagery. The false positives occurring as a result of NT/BG areas, despite being obvious on crown edges, nonetheless obscure the real results over trees and shrubs (figure 13a). Segments classified red were extremely limited across all maps at this study site with only 11 and 43 segments of 45,184 total segments classified red in 2010 and 2020, respectively (figure 15). Of these, most were NT/BG errors, disproportionately affecting the accuracies of the red class.

Discussion

TreeCAP accurately maps tree crown condition at sites across the United States at a fine spatial scale (1 m or less), often at subtree level, using a single model that can be used seamlessly on NAIP imagery in all forty-eight contiguous United States and possibly elsewhere. The efficiency and replicability of our approach are conducive to implementation on an operational level, providing a means to detect unexpected damage soon after NAIP imagery becomes available. Maps resulting from this methodology could offer valuable information about forest health in areas that are normally difficult to access or not routinely monitored. In particular, TreeCAP could be used to help track abnormal mortality occurring in a single tree species (as demonstrated here with emerald ash borer and hemlock wooly adelgid) because of its ability to map mortality in individual trees in mixed forests, especially when using the raster output. Although independent verification would be needed to determine the affected species, maps produced by TreeCAP could provide information regarding the extent

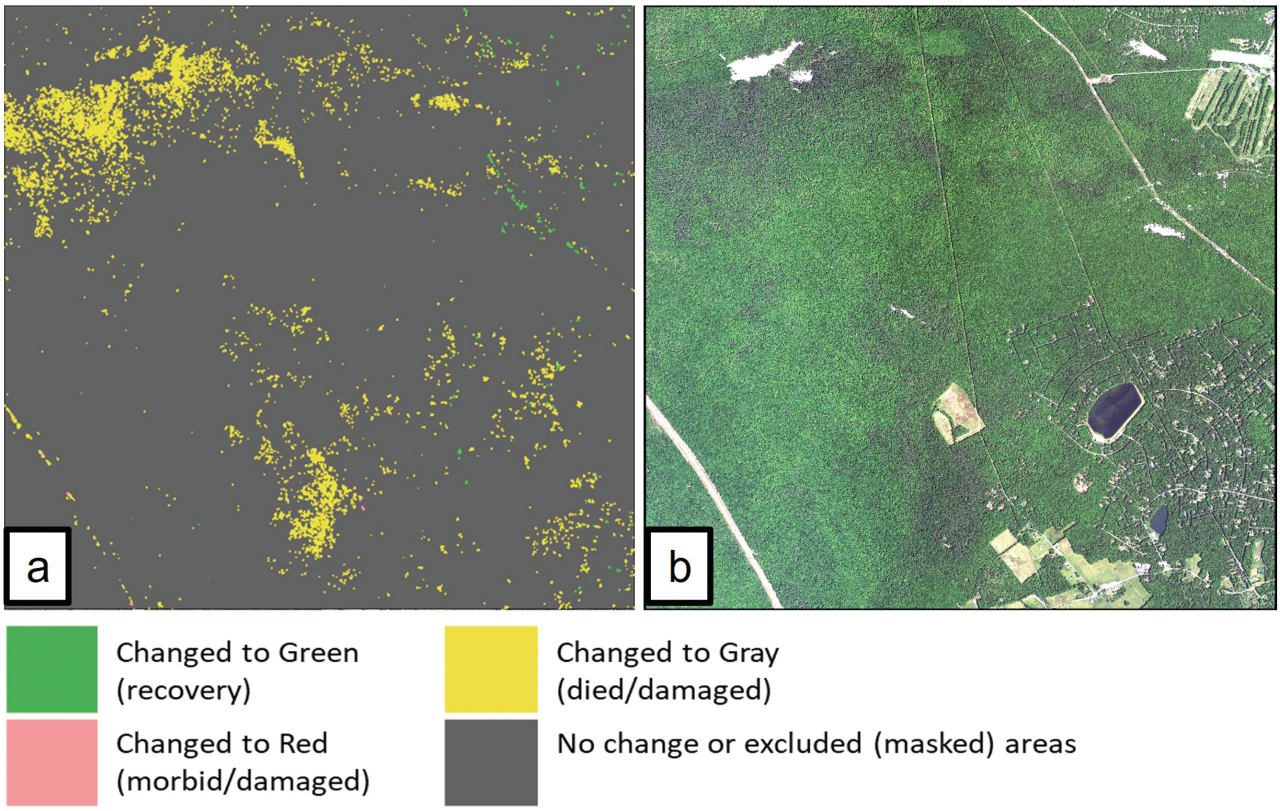


Figure 11. Changes highlighted by TreeCAP in the Hickory Run State Park study site between 2013 and 2015 (a). (b) is the 2015 NAIP imagery with a 2% linear stretch applied. The trend in increase in mortality due to hemlock wooly adelgid is clearly evident in yellow.

Table 10. Quantification of the changes observed between sets of years in the Hickory Run study site in hectares (ha) and percents.

	Change between images ha (% of area mapped)
	2013–2015
Change to green	2 (<1%)
Change to red	None
Change to gray	72 (5%)
Change to shadow	1 (<1%)
No change	1371 (95%)

and spatial patterns of damage for individual trees and small clusters of trees.

The inclusion of fuzzy accuracy highlights the quality of TreeCAP's XGBoost model by offering insight into the nature of the classification errors. Although the average traditional model accuracy across all maps was 87% (minimum 73%), the average fuzzy accuracy was 97% (minimum 94%). This indicates that the errors arise primarily from the understandable confusion between the red and gray classes or a slight difference in the class quantity of pixels predicted by the algorithm in comparison with human visual interpretation. From this, we infer that there are benefits to using both the original raster output from TreeCAP and the annotated tree segments shapefile. The shapefile offers the ability to quickly locate areas that may be of concern to the user, especially across multiple years. After these areas are located, it may be more prudent to use the raster for a greater level of detail.

Notably, the training data used by TreeCAP spanned 2012 to 2019, yet here we show examples of how the XGBoost model also performed well when projected to nontraining years (2010 in Texas, 2020 in California and Wisconsin). We conclude that TreeCAP's fitted XGBoost model can be applied across multiple years of NAIP imagery, including future years. This means it is possible to use the output from TreeCAP in conjunction with follow-on analyses to investigate continuing trends or detect new infestations. Progression of forest insect and disease outbreaks could potentially be tracked to assess metrics such as spread rate and mortality area and intensity. Although such studies would likely need additional observations to verify the causal agent, TreeCAP's output has great potential to streamline field work by providing spatially precise maps of tree crown damage. These types of evaluation monitoring studies could be invaluable for mitigating damage from emerging forest insects and diseases.

A significant concern with NAIP imagery is the latency between collection and public release, which can be longer than a year. However, NAIP imagery can be accessed in near-real time by federal agencies that contribute funding to the program. Although these agencies cannot release actual imagery, they are able to apply models and share screen captures and analyses derived from the imagery. This means that adoption of TreeCAP by any of the federal agencies that contribute to NAIP could produce and share TreeCAP results in near-real time on lands they support or manage. An example of this is shown for a subset of the Kettle Moraine study site, based on NAIP imagery collected in 2022, which was

Table 11. The traditional and fuzzy accuracy data for the Hickory Run study site, 2013 and 2015. The left column provides accuracy values if masking errors (segments that were over nontree/bare ground areas) are excluded. The right column includes these errors. Producer's accuracy only assesses performance on the segments known to belong to these respective classes (i.e., it is the same for both columns because masking errors are already excluded in producer's accuracy). Masking errors disproportionally affected the red class due to the scarcity of these segments and our sampling scheme (also in the gray class in 2013).

		Model accuracy (excluding masking errors)			Model + masking accuracy (including masking errors)		
		n	Traditional (%)	Fuzzy (%)	n	Traditional (%)	Fuzzy (%)
2013							
User's Accuracy	Green	50	50 (100)	50 (100)	50	50 (100)	50 (100)
	Red	3	2 (67)	3 (100)	26	2 (8)	3 (12)
	Gray	10	9 (90)	10 (100)	50	9 (18)	10 (20)
	Shadow	47	46 (98)	46 (98)	50	46 (92)	46 (92)
Producer's Accuracy	Green	51	50 (98)	51 (100)	Does not change (66 segments are masking errors)		
	Red	4	2 (50)	3 (75)			
	Gray	9	9 (100)	9 (100)			
	Shadow	46	46 (100)	46 (100)			
		Overall accuracy = 97% traditional, 99% fuzzy			Overall accuracy = 61% traditional, 62% fuzzy		
2015							
User's Accuracy	Green	50	45 (90)	50 (100)	50	45 (90)	50 (100)
	Red	No treed red segments			10	0 (0)	0 (0)
	Gray	47	40 (85)	46 (98)	50	40 (80)	46 (92)
	Shadow	50	23 (46)	44 (88)	50	23 (46)	44 (88)
Producer's Accuracy	Green	51	45 (74)	60 (98)	Does not change (13 segments are masking errors)		
	Red	No treed red segments					
	Gray	47	40 (65)	57 (92)			
	Shadow	50	23 (96)	23 (96)			
		Overall accuracy = 73% traditional, 95% fuzzy			Overall accuracy = 68% traditional, 88% fuzzy		

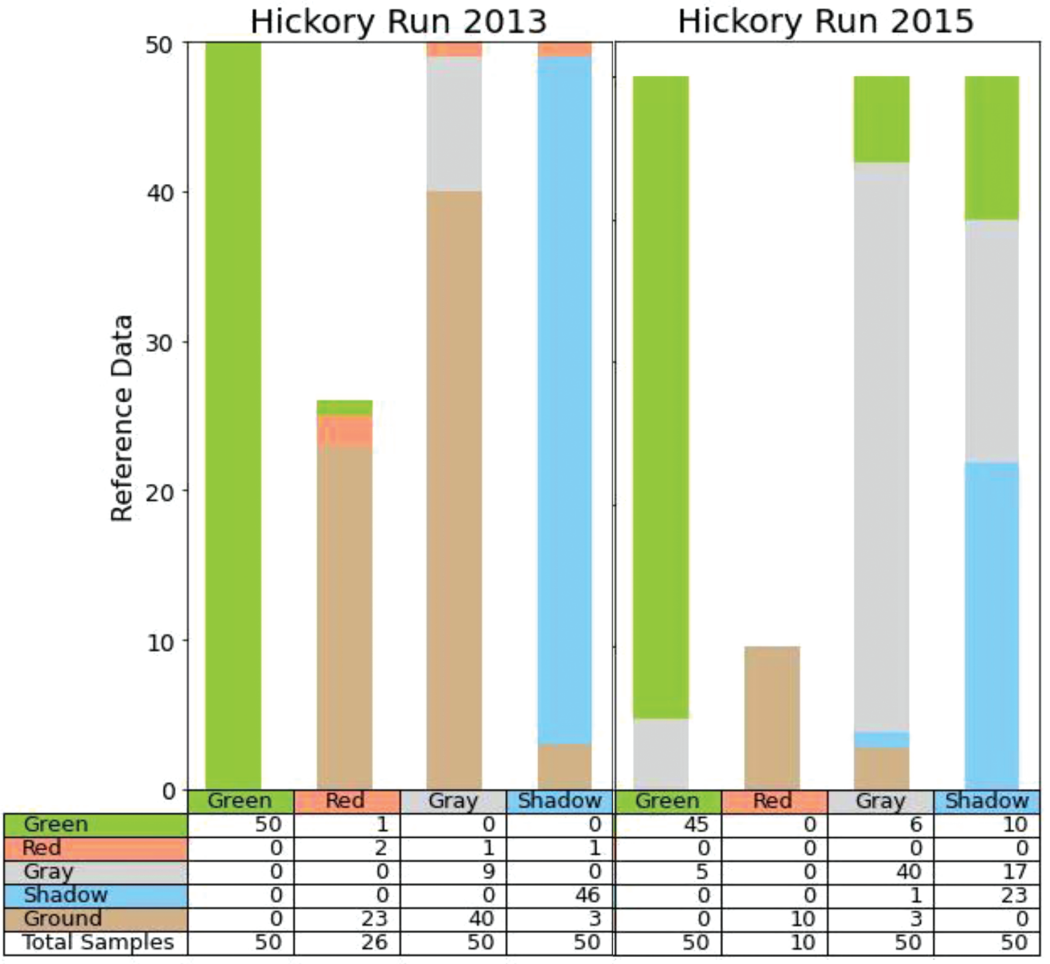
mapped with TreeCAP while the imagery was still being collected (figure 16).

Although we developed and tested TreeCAP on NAIP imagery, the XGBoost model is not limited to NAIP and can be applied to any imagery with similar spectral characteristics. The high accuracy values frequently achieved TreeCAP's XGBoost model (Tables 7, 9, 11, and 13, Model accuracy columns) indicates that there is great potential to use TreeCAP to aid in the early detection of species-specific damage. However, this would require timely suitable imagery and LiDAR to be collected temporally near the imagery. Use of alternate NAIP-like imagery would enable TreeCAP to be leveraged with much higher temporal frequency than the current NAIP collection revisit interval. It may also be possible to conduct band-matching activities to apply the model across different sensors.

Important next steps include cross-walking the TreeCAP methodology to high temporal-resolution satellite imagery such as Maxar's Worldview and/or Planet's Dove or SkySat. Although the spatial resolutions of these sources are often lower than that of NAIP imagery (1–3 m versus 0.6–1 m), they still may capture damage occurring in individual tree crowns or small clusters of trees and can potentially be used globally in forested ecoregions (i.e., because the imagery is collected globally). For demonstration, we tested 3 m, band-matched Dove imagery at the Kettle Moraine study site (Figure S5). Although Dove imagery spatial resolution is coarser, the revisit time is near daily. In comparison to the NAIP imagery, application to Dove imagery resulted in less mortality being mapped, but the map does correctly highlight larger areas of contiguous mortality. The ability to leverage high-resolution satellite imagery sources such as Dove will enable much higher temporal frequency and reduce errors resulting from sun-sensor-target geometry, including shadows. Recreating the training data for these sources will be the most significant challenge, especially as sensor characteristics change.

Our use of LiDAR segmentation effectively reduced errors of commission for NT/BG ground areas (Figures S1–4, Tables S1–4). However, the characteristics of the available LiDAR data varied widely (Table 3). Specifically, point density, time of collection, and coverage all influence the effectiveness of the LiDAR. For instance, most of the LiDAR used was leaf-off, collected for terrain mapping, and therefore tree identification can be compromised in areas with sparse tree cover. Because most areas do have sufficient coverage and point density even under leaf-off acquisitions, the ideal solution to this is to have LiDAR data collected the same year as NAIP to better account for changes such as forest harvests, stand thinning, and hazard tree removal.

Although we were unable to create a single programming script capable of effectively creating tree segments regardless



	Green	Red	Gray	Shadow	Green	Red	Gray	Shadow
Green	50	1	0	0	45	0	6	10
Red	0	2	1	1	0	0	0	0
Gray	0	0	9	0	5	0	40	17
Shadow	0	0	0	46	0	0	1	23
Ground	0	23	40	3	0	10	3	0
Total Samples	50	26	50	50	50	10	50	50

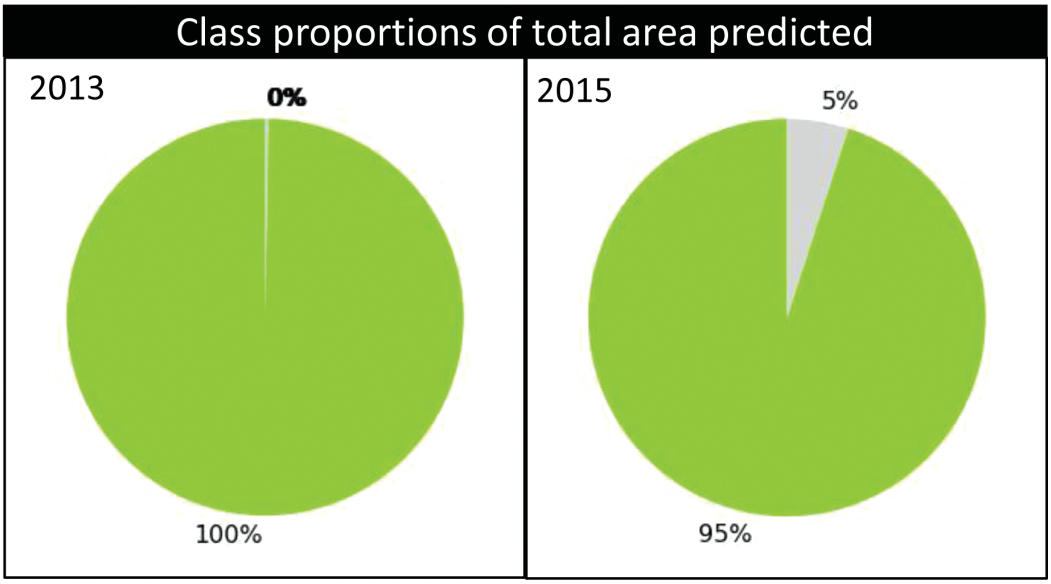


Figure 12. These bar graphs provide insight on the origin of the errors in the assessment data for the Hickory Run study site. The pie charts show the proportion of that class for the entire map, regardless of total sample size; classes not seen in the pie charts mean the proportion was significantly less than 1%. The poor accuracy of gray in 2013 was due to a limitation of gray segments in the image—less than 50. Because of this, erroneous segments over ground were disproportionately selected. This is also true in the red class for both maps. Additionally, in 2015, many of the errors in the gray class were due to confusion with shadows.

of input without some parameter tuning, the creation of tree segments was not arduous. As repeat LiDAR becomes more widely available and techniques to analyze it improve, we expect there may even be ready-made tree segmentation datasets available for use in the future for applications such as TreeCAP. For example, a nationwide canopy height model

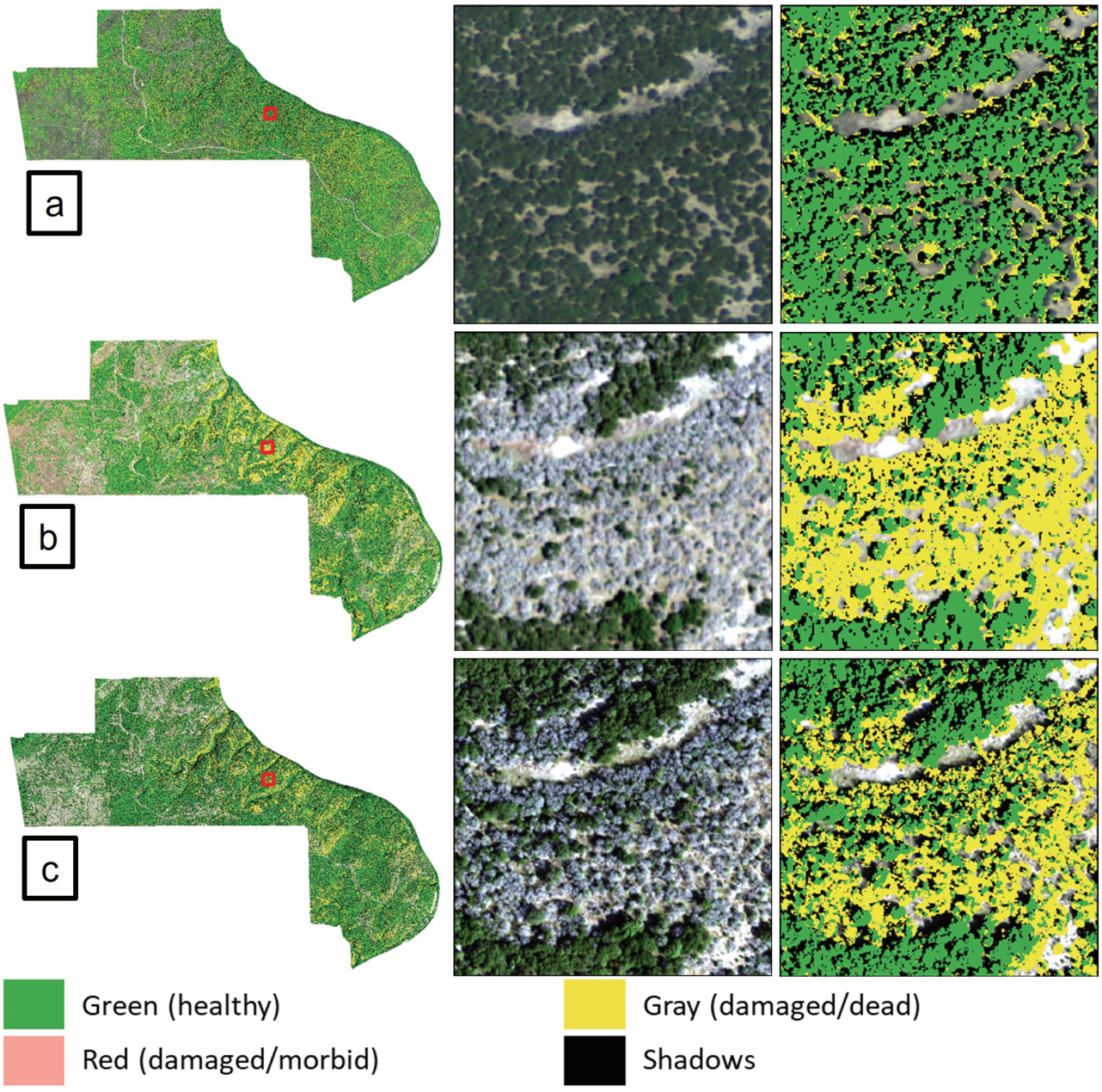


Figure 13. Colorado Bend State Park study site from 2010 to 2020. The first column are overviews on the entire study area as modeled by the TreeCAP raster output overlaid on the corresponding NAIP imagery. The red box is the 200 m subset area shown in the next two columns. The middle column is the original NAIP imagery with a 2% linear stretch applied, and the right column is the TreeCAP raster output. Row (a) is 2017, row (b) is 2018, and row (c) is 2020. Drought-induced mortality was heavy in 2016 but some recovery can be seen by 2020.

could potentially be used instead of tree segments for masking purposes (thresholding at certain height, depending location, to exclude NT/BG areas). If this were to be done, a grid system could be implemented to detect change through time instead of using the segments. Although the methodology for annotation of change outlined in figure 2 may have to be adjusted to accommodate these modifications, this combination could offer an even more efficient and expandable way to use TreeCAP nationwide.

Despite the use of LiDAR, NT/BG errors remained in our maps, but as only a small fraction of area and largely in areas where other factors influenced misclassification. For example, in Sawmill Campground 2020 (figure 4c and figure 6), the number of ground errors appears high, particularly in the red class, but most were disproportionately located in the area of the campground where hazard trees that had been present in previous imagery were apparently removed. Because the LiDAR for this area was collected in 2018, the tree segments for these recently cut trees caused the algorithm to predict over bare ground. TreeCAP's output is intended to be used by local forest managers and health specialists, and these errors would be readily apparent to such an expert of the area. The same is true for the errors in maps with very low incidents of red or gray. Unless these areas of gray or red are concentrated in an unexpected area of continuous canopy, the NT/BG errors will be either be readily apparent (i.e., a known recent harvest area or a newly cut road) or so spread out that they do not cause concern.

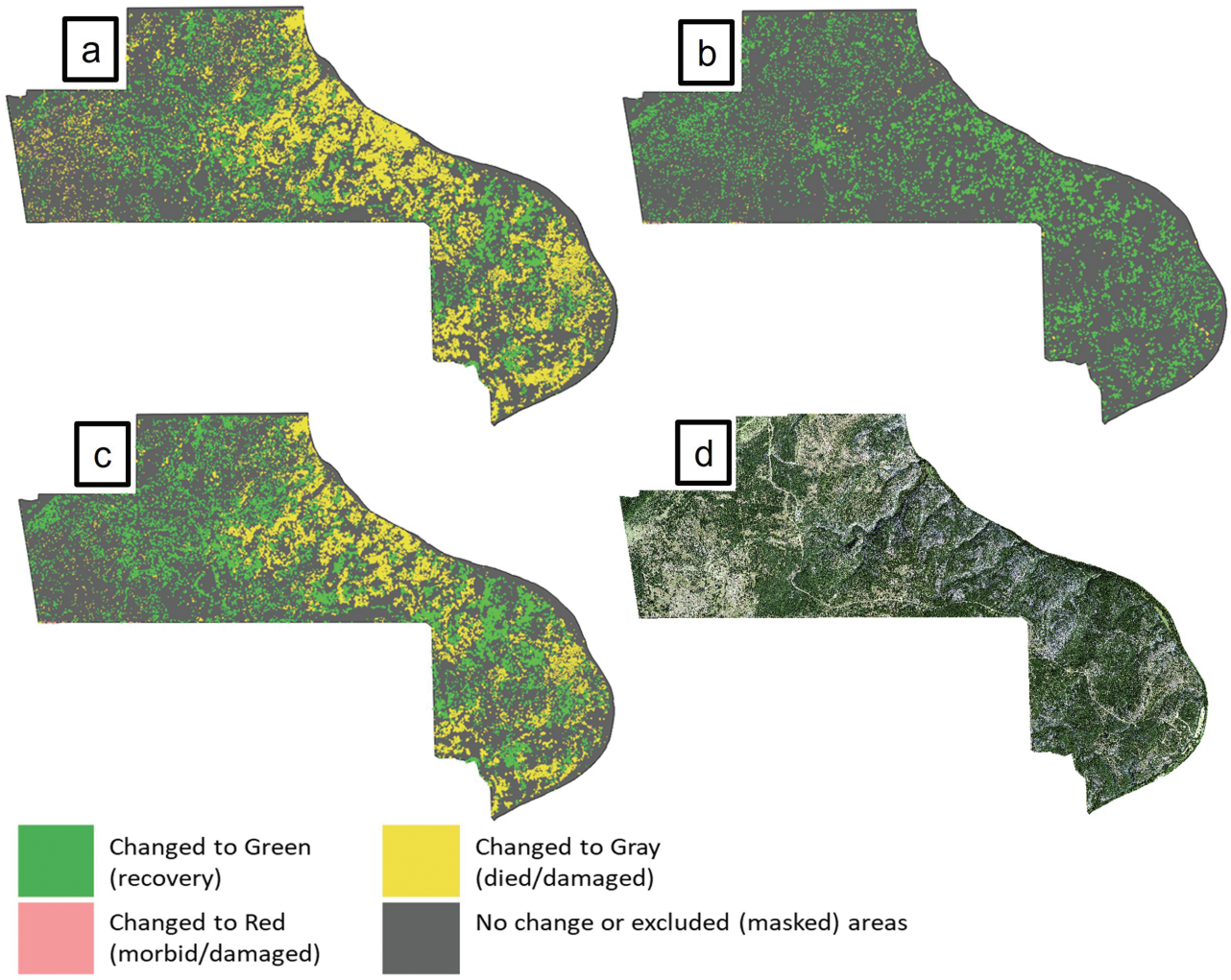


Figure 14. Changes highlighted by TreeCAP in the Colorado Bend study site from 2010 to 2020. (a) is the difference in tree condition between 2010 and 2016, (b) is the difference between 2016 and 2020, (c) is the difference between 2016 and 2020, and (d) is the 2020 NAIP imagery with a 2% linear stretch applied. Although drought-induced mortality caused extensive damage in this park, some recovery can be seen by 2020.

Table 12. Quantification of the changes observed between sets of years in the Colorado Bend study site in hectares (ha) and percents.

	Change between images ha (% of area mapped)		
	2010–2016	2016–2020	2010–2020
Change to green	133 (10%)	163 (12%)	196 (14%)
Change to red	2 (<1%)	None	None
Change to Gray	277 (20%)	4 (<1%)	181 (13%)
Change to shadow	1 (<1%)	3 (<1%)	3 (<1%)
No change	955 (70%)	1197 (88%)	986 (72%)

Natural morbidity and mortality occur in forests, and it is typically only when unexpectedly high levels are exhibited in a concentrated location that professionals may decide the area warrants further investigation.

Although TreeCAP was built to be computationally efficient, implementation on an operational scale—such as processing of entire states—would still be better deployed using larger computational capacity. Specifically, we believe implementing TreeCAP in Google Earth Engine would provide a reliable, expandable, and flexible means of making TreeCAP-produced maps freely available to forest managers and forest health specialists. The XGBoost model developed for TreeCAP can be loaded into other types of computer code, making it possible to implement TreeCAP in applications built in R and Google Earth Engine, among others.

Conclusion

In this article, we introduce TreeCAP, a system to detect abnormal mortality at the individual tree level or small clusters of trees in forests using high spatial resolution imagery and machine learning. TreeCAP was built to be fast and efficient to enable scaling up to operational levels and facilitate implementation on different computing platforms as needed. In addition, we demonstrate the use of LiDAR-derived tree segments to mask NT/BG areas and better allow comparison among years while mitigating the effects of shadows that are prevalent in high-spatial resolution imagery. Finally, TreeCAP's 4-class output is intended to be easily interpreted by forest managers and forest health specialists, informing and aiding their management of their forests.

Table 13. The traditional and fuzzy accuracy data for the Colorado Bend study site, 2010, 2016, and 2020. The left column provides accuracy values if masking errors (segments that were over nontree/bare ground areas) are excluded. The right column includes these errors. Producer's accuracy only assesses performance on the segments known to belong to these respective classes (i.e., it is the same for both columns because masking errors are already excluded in producer's accuracy). The red class was limited in 2010, and insignificant in 2016 and 2020. Masking errors disproportionally affected the red class due to the scarcity of these segments and our sampling scheme.

		Model accuracy (excluding masking errors)			Model + masking accuracy (including masking errors)		
		n	Traditional (%)	Fuzzy (%)	n	Traditional (%)	Fuzzy (%)
2010							
User's Accuracy	Green	50	50 (100)	50 (100)	50	50 (100)	50 (100)
	Red	3	3 (100)	3 (100)	11	3 (27)	3 (27)
	Gray	8	5 (62)	8 (100)	50	5 (10)	5 (10)
	Shadow	49	37 (76)	42 (86)	50	37 (74)	42 (84)
Producer's Accuracy	Green	46	42 (91)	46 (100)	Does not change (49 segments are masking errors)		
	Red	15	15 (100)	15 (100)			
	Gray	57	46 (81)	54 (95)			
	Shadow	49	44 (90)	48 (98)			
		Overall accuracy = 95% traditional, 94% fuzzy			Overall accuracy = 59% traditional, 64% fuzzy		
2016							
User's Accuracy	Green	48	47 (98)	48 (100)	50	47 (94)	48 (96)
	Red	No treed red segments			50	0 (0)	0 (0)
	Gray	38	36 (95)	38 (100)	50	36 (72)	38 (76)
	Shadow	49	39 (80)	49 (100)	50	39 (78)	49 (98)
Producer's Accuracy	Green	52	47 (90)	52 (100)	Does not change. (65 segments are masking errors)		
	Red	No treed red segments					
	Gray	44	36 (82)	44 (100)			
	Shadow	39	39 (100)	39 (100)			
		Overall accuracy = 90% traditional, 100% fuzzy			Overall accuracy = 61% traditional, 68% fuzzy		
2020							
User's Accuracy	Green	49	46 (94)	49 (100)	50	46 (92)	49 (98)
	Red	1	1 (100)	1 (100)	43	1 (2)	1 (2)
	Gray	43	41 (95)	43 (100)	50	41 (82)	43 (86)
	Shadow	50	45 (90)	50 (100)	50	45 (90)	50 (100)
Producer's Accuracy	Green	49	46 (94)	49 (100)	Does not change (7 segments are masking errors)		
	Red	2	1 (50)	2 (100)			
	Gray	47	41 (87)	47 (100)			
	Shadow	45	45 (100)	45 (100)			
		Overall accuracy = 93% traditional, 100% fuzzy			Overall accuracy = 69% traditional, 74% fuzzy		

Supplementary Material

Supplementary data are available at *Journal of Forestry* online.

Acknowledgments

This research was supported by USDA Forest Service agreement 20-CA-11094200-196 to PT and SW. Photo interpretation and digitizing support for TCH was provided in part through agreement 18-CR-11221676-160 between the USDA Forest Service and Michigan Technological University. Colton Arnspiger and David Gudex-Cross assisted with developing canopy height models. Ting Zheng and Kyle Kovach provided guidance in working with LiDAR point cloud data. Numerous forest managers and forest health specialists with the Pennsylvania Department of Conservation and Natural Resources, the Wisconsin Department of Natural Resources, the Texas A&M Forest Service, and Texas Parks and Wildlife Service provided critical on-the-ground information and feedback for our study sites and results. Finally, we extend our

	Green	Red	Gray	Shadow	Green	Red	Gray	Shadow	Green	Red	Gray	Shadow
Green	50	0	1	12	47	0	2	3	46	0	2	1
Red	0	3	2	0	0	0	0	0	1	1	0	0
Gray	0	0	5	0	1	0	36	7	2	0	41	4
Shadow	0	0	0	37	0	0	0	39	0	0	0	45
Ground	0	8	42	1	2	50	12	1	1	42	7	0
Total Samples	50	11	50	50	50	50	50	50	50	43	50	50

Figure 15. These bar graphs provide insight on the origin of the errors in the assessment data for the Colorado Bend study site. The pie charts show the proportion of that class for the entire map, regardless of total sample size; classes not seen in the pie charts mean the proportion was significantly less than 1%. Poor alignment between the LiDAR data and NAIP caused a greater proportion of ground to be misclassified in 2010. In the red class, very few segments of this classification—less than 50—were available for evolution, making this class insignificant in all maps for this study site.

gratitude to Annemarie Schneider, Glen Stanosz, and three anonymous reviewers for providing feedback that greatly improved this article.

Contribution Statement

Conceptualization: Wegmueller, Monahan, Townsend. Methodology: Wegmueller, Monahan, Townsend. Software: Wegmueller. Validation: Wegmueller, Monahan, Townsend. Formal analysis: Wegmueller, Monahan, Townsend. Investigation: Wegmueller, Monahan, Townsend. Data curation: Monahan, Wegmueller. Writing-Original draft: Wegmueller. Writing-Review & Editing: Monahan, Townsend, Wegmueller. Visualization: Wegmueller, Monahan, Townsend. Project administration: Wegmueller, Monahan, Townsend. Funding acquisition: Townsend, Wegmueller.

Funding

This research was funded by USDA Forest Service agreement 20-CA-11094200-196 to Philip Townsend and Sarah Wegmueller.

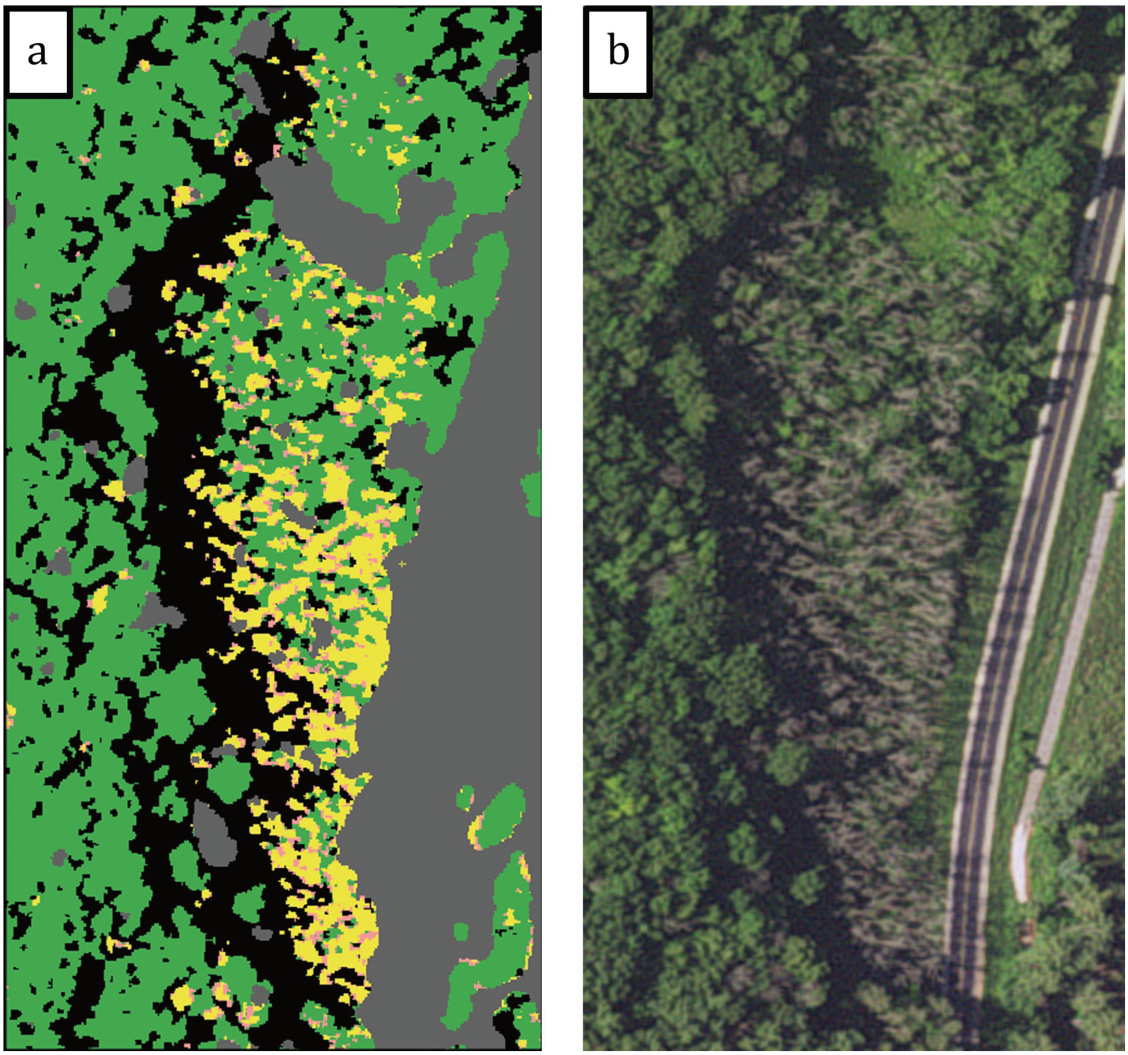


Figure 16. TreeCAP output (a) derived from 2022 NAIP imagery (b) in an area of tree damage in the Kettle Moraine Study site.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Literature Cited

Ayrey, Elias, Shawn Fraver, John A. Kershaw, Laura S. Kenefic, Daniel Hayes, Aaron R. Weiskittel, and Brian E. Roth. 2017. "Layer Stacking: A Novel Algorithm for Individual Forest Tree Segmentation from LiDAR Point Clouds." *Canadian Journal of Remote Sensing* 43 (1): 16–27. https://doi.org/10.1080/07038992.2017.1252907.

Bailey, Robert G. 1995. *Description of the Ecoregions of the United States*, second edition. Washington, DC: US Department of Agriculture, Forest Service.

Chen, Qi, Dennis Baldocchi, Peng Gong, and Maggi Kelly. 2006. "Isolating Individual Trees in a Savanna Woodland Using Small Footprint Lidar Data." *Photogrammetric Engineering & Remote Sensing* 72 (8): 923–932. https://doi.org/10.14358/PERS.72.8.923.

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. San Francisco: ACM. https://doi.org/10.1145/2939672.2939785.

Clewley, Daniel, Peter Bunting, James Shepherd, Sam Gillingham, Neil Flood, John Dymond, Richard Lucas, et al. 2014. "A Python-Based Open Source System for Geographic Object-Based Image Analysis (GEOBIA) Utilizing Raster Attribute Tables." *Remote Sensing* 6: 6111–6135. https://doi.org/10.3390/rs6076111.

Cohen, Warren B., Sean P. Healey, Zhiqiang Yang, Stephen V. Stehman, C. Kenneth Brewer, Evan B. Brooks, Noel Gorelick, et al. 2017. "How Similar are Forest Disturbance Maps Derived from Different Landsat Time Series Algorithms?." *Forests* 8 (98): 1–19. https://doi.org/10.3390/f8040098.

Cohen, Warren B., Zhiqiang Yang, and Robert Kennedy. 2010. "Detecting Trends in Forest Disturbance and Recovery Using Yearly Landsat Time Series: 2. TimeSync - Tools for Calibration and Validation." *Remote Sensing of Environment* 114 (12): 2911–2924. https://doi.org/10.1016/j.rse.2010.07.010.

Erker, Tedward, Lei Wang, Laura Lorentz, Andrew Stoltman, and Philip A. Townsend. 2019. "A Statewide Urban Tree Canopy Mapping Method." *Remote Sensing of Environment* 229 (June 2018): 148–158. https://doi.org/10.1016/j.rse.2019.03.037.

Friedman, Jerome. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *The Annals of Statistics* 29 (5): 1189–1232. https://doi.org/10.1214/009053606000000795.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2008. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York: Springer.

Gopal, S., and C. Woodcock. 1994. "Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets." *Photogrammetric Engineering & Remote Sensing* 60 (2): 181–188.

Houborg, Rasmus, and Matthew F. McCabe. 2018. "A Cubesat Enabled Spatio-Temporal Enhancement Method (CESTEM) Utilizing Planet, Landsat and MODIS Data." *Remote Sensing of Environment* 209 (January): 211–226. https://doi.org/10.1016/j.rse.2018.02.067.

Housman, Ian, Robert Chastain, and Mark Finco. 2018. "An Evaluation of Forest Health Insect and Disease Survey Data and Satellite-Based Remote Sensing Forest Change Detection Methods: Case Studies in the United States." *Remote Sensing* 10 (8): 1184. https://doi.org/10.3390/rs10081184.

Jeronimo, Sean M A, Van R. Kane, Derek J. Churchill, Robert J. McGaughey, and Jerry F. Franklin. 2018. "Applying LiDAR Individual Tree Detection to Management of Structurally Diverse Forest Landscapes." *Journal of Forestry* 116 (4): 336–346. https://doi.org/10.1093/jofore/fvy023.

Miller, Jay D, Eric E. Knapp, Carl H. Key, Carl N. Skinner, Clint J. Isbell, R. Max Creasy, and Joseph W. Sherlock. 2009. "Calibration and Validation of the Relative Differenced Normalized Burn Ratio (RdNBR) to Three Measures of Fire Severity in the Sierra Nevada and Klamath Mountains, California, USA." *Remote Sensing of Environment* 113 (3): 645–656. https://doi.org/10.1016/j.rse.2008.11.009.

Monahan, William B., Colton E Arnspiger, Parth Bhatt, Zhongming An, Frank Krist Jr., and Tao Liu. 2022a. "A Spectral Three-Dimensional Color Space Model of Tree Crown Health." *Dryad Digital Repository* 2022. https://doi.org/10.5061/dryad.wm37pvmpp.

Monahan, William B, Colton E Arnspiger, Parth Bhatt, Zhongming An, Frank J. Krist, Tao Liu, Robert P Richard, et al. 2022b. "A Spectral Three-Dimensional Color Space Model of Tree Crown Health." *PLoS One* 17 (10): e0272360. https://doi.org/10.1371/journal.pone.0272360.

Norman, Steven P., William W. Hargrove, Joseph P. Spruce, William M. Christie, and Sean W. Schroeder. 2013. "Highlights of Satellite-Based Forest Using the ForWarn System Change Recognition and Tracking." *General Technical Report SRS-180*. Asheville: US Department of Agriculture, Forest Service, Southern Reserach Station.

Potter, Kevin M., and Barbara L. Conkling. 2017. "Forest Health Monitoring: National Status, Trends, and Analysis 2015." *General Technical Report SRS-222*. US Department of Agriculture, Forest Service, Southern Research Station.

Roussel, Jean-Romain, David Auty, Nicholas C. Coops, Piotr Tompalski, Tristan R.H. Goodbody, Andrew Sánchez Meador, Jean-François Bourdon, et al. 2020. "LidR: An R Package for Analysis of Airborne Laser Scanning (ALS) Data." *Remote Sensing of Environment* 251 (August): 112061. https://doi.org/10.1016/j.rse.2020.112061.

USDA Forest Service. 2019. "Digital Mobile Sketch Mapping (DMSM) Updates to Forest Health Survey with DMSM." *United States Department of Agriculture, Forest Service*.

USGS. 2022a. "Earth Explorer." 2022. https://earthexplorer.usgs.gov.

USGS. 2022b. "MRLC." 2022. https://www.mrlc.gov.

USGS. 2022c. "TNM Download." 2022. https://apps.nationalmap.gov/downloader/.

Wulder, Michael A, Joanne C. White, Nicholas C. Coops, and Christopher R. Butson. 2008. "Multi-Temporal Analysis of High Spatial Resolution Imagery for Disturbance Monitoring." *Remote Sensing of Environment* 112 (6): 2729–2740. https://doi.org/10.1016/j.rse.2008.01.010.

Zhang, Lifu, Xuejian Sun, Taixia Wu, and Hongming Zhang. 2015. "An Analysis of Shadow Effects on Spectral Vegetation Indexes Using a Ground-Based Imaging Spectrometer." *IEEE Geoscience and Remote Sensing Letters* 12 (11): 2188–2192. https://doi.org/10.1109/LGRS.2015.2450218.