Remote Sensing of Conifer Encroachment in Marin County Oak Woodland

Patrick T. Jacobson

ABSTRACT

California oak woodlands provide ecological and cultural benefits through their acorns and the unique habitat they create for native California species. The fire tolerance of oak trees relative to conifer species allowed oak woodlands to thrive under the 12-year median fire return interval of the pre-settlement California fire regime. The advent of fire suppression in California during the 1870s opened opportunities for shade tolerant conifer species to establish in, pierce, overtop, and eventually shade out oak canopies. This conversion process, called conifer encroachment, has been observed throughout California in small, individual field studies but has not been quantified over large areas. In this study I used a remote sensing approach to investigate the extent, severity, and spatial pattern of conifer encroachment in Marin County oak woodland. I created a Marin County canopy shapefile that contained canopy type and height data by applying zonal statistics to a canopy shapefile derived from a 2019 Marin County canopy height model (CHM) and Random Forest (RF) classified 2020 NAIP imagery. I found that 59.02% of oak canopies in Marin County were experiencing conifer encroachment, and that 23.22% of these oak canopies were in the overtopping (severe) phase of conifer encroachment. However, the conifer encroachment and overtopping findings were derived from an image classification and zonal statistics canopy labeling process that could not accurately identify conifer and oak canopies. Therefore, these findings are not yet accurate enough to be used as a reliable management tool. However, this study serves as a starting point for future remote sensing studies of conifer encroachment in Marin County and California oak woodland.

KEYWORDS

National Agriculture Imagery Program, succession, tree crown delineation, land cover classification, random forest classifier

INTRODUCTION

Oak woodland is a valuable part of California's ecological and cultural heritage. Oak woodland makes up 8.9 million acres of California's forest land, making it the most extensive forest cover type in the state (Brodie and Palmer 2020). Oak woodland is defined by an oak-dominated overstory that has a tree density of 30% to 80% crown cover (Dey et al. 2017). These overstories are made up of one or more of the oak species native to California such as black oak (*Quercus kelloggii*), coast live oak (*Quercus agrifolia*), Oregon white oak (*Quercus garryana*), and a multitude of other oak species (Gaman and Firman 2006). The diverse overstory of California oak woodland supports a wide variety of plant and animal species. For instance, oak canopies have strong positive effects on the survival of California native grasses such as purple needlegrass (*Stipa pulchra*) and Sandberg bluegrass (*Poa secunda*) (Stahlheber and D'Antonio 2014), while black oak acorns support animals in California such as acorn woodpeckers (*Melanerpes formicivorus*), valley quails (*Callipepla californica*), and dusky-footed woodrats (*Neotoma fuscipes*) (Long et al. 2016). Equally important is the cultural value of acorns from oak woodland, which form the bedrock of important rituals, dances, and ceremonies practiced by indigenous Californians (Long et al. 2016). Native Californian prescribed burn practices likely contributed to the consistently short fire intervals in pre-colonial California oak woodlands (Finney and Martin 1992), which had a median fire return interval of 12 years (Van de Water and Safford 2011). The loss of this frequent fire regime due to the introduction of fire suppression in the 1870s (van Wagtendonk 2007) created opportunities for less fire-resistant tree species to encroach into California oak woodland.

Without consistently short fire intervals, native conifer species such as Douglas-fir (*Pseudotsuga menziesii*) encroach upon and succeed California oak woodland (Sugihara and Reed 1987, Barnhart et al. 1996). The oak species found in California are capable of resprouting after being burned (Holmes et al. 2008), meaning that repeated burns of California mixed conifer-oak forests shift canopy dominance towards oak species and away from fire intolerant conifer species such as Douglas-fir (Nemens et al. 2018). However, the recruitment and growth traits of certain conifer species give them advantages over oak species in unmaintained California oak woodland. For example, Douglas-fir's shade tolerance and high growth rate allow it to quickly establish and outgrow neighboring oaks when not suppressed by fire (Hunter and

Barbour 2001). Conifer encroachment progresses through five distinct phases: maintenance, establishment, piercing, overtopping, and decadent (Figure 1) (Cocking et al. 2015). Overtopping, during which the height of the conifer canopy exceeds that of the oak canopy, is critical for oak conservation because it is the last phase of conifer encroachment when conifer removal can prevent oak die off (Cocking et al. 2015). Despite the threat conifer encroachment poses to California's oak derived resources and benefits, the true extent of the problem is unknown due to limited field studies.

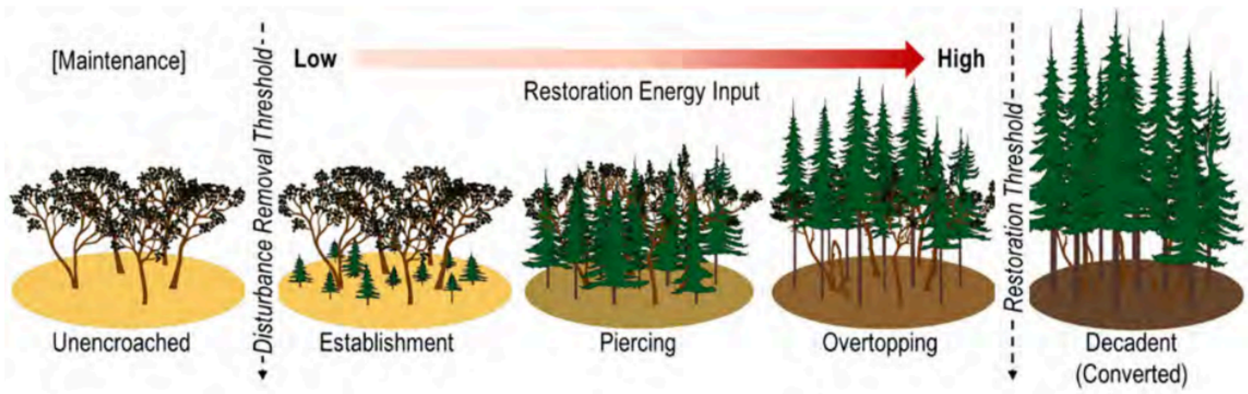

Figure 1. The phases of conifer encroachment in oak woodland. This diagram is originally from Cocking et al. 2015.

Currently, no comprehensive conifer encroachment datasets exist for the state of California. Studies of conifer encroachment in California oak woodland have typically investigated encroachment progression and encroachment prevention (Hunter and Barbour 2001, Schriver et al. 2018, Nemens et al. 2018, van Mantgem et al. 2021). Therefore, these studies only gathered data over small scales, typically observing and experimenting on small numbers of small patches of public land using ground-based methods (Hunter and Barbour 2001, Schriver et al. 2018, Nemens et al. 2018, van Mantgem et al. 2021). The lack of data on conifer encroachment in California oak woodlands makes it difficult for forest managers to efficiently target interventions, giving encroachment more time to progress.

Marin County exemplifies this conifer encroachment data deficiency. The county is a potential hotspot for conifer encroachment on oak woodland, as it has vegetation alliances that include both oak species and Douglas-fir (Buck-Diaz et al. 2021), an efficient encroacher in oak canopy (Hunter and Barbour 2001), and a history suppressing native prescribed burns (Lavezzo et al. 2020). However, no dataset tracking the spatial distribution of the problem at the scale of individual trees exists. The lack of precise county and state data on conifer encroachment in oak woodland in California will slow interventions in a time sensitive problem.

For this study, I built a spatial dataset of conifer encroached oak woodland in Marin County. To achieve this goal, I described the extent and severity of conifer encroachment of oak woodland in Marin County using a remote sensing approach. I characterized encroachment extent by quantifying the proportion of conifer encroached oak woodland to total oak woodland in Marin County. I studied encroachment severity by quantifying the proportion of conifer encroached Marin County oak woodland that is in the overtopping phase of conifer encroachment. To study these aspects of conifer encroachment, I used high resolution aerial photography to distinguish conifers from oaks and a canopy height model to compare the heights of conifers and oaks. I expected to find that at least 29% of oak canopies in Marin county would be conifer encroached, given that 29% of oak woodland plots were found to be Douglas-fir dominated in the California North Coast Ranges (Schriver et al. 2018). I expected to find that about 50% of conifer encroached oak canopy would be in the overtopping phase of encroachment, given the 10-40% cover range of Douglas-fir in the emergent layer of oak woodland plots in the Angelo Coast Range Reserve (Hunter and Barbour 2001).

METHODS

Study site

The study area encompasses Marin County's land area which covers 1347.83 km^2 (U.S. Census Bureau 2024). Marin County has a Mediterranean/summer fog climate along its Pacific coast and a Mediterranean/cool summer climate inland (State of California et al. 2021). The county's elevation ranges from sea level (0 m) to 786 m at the peak of Mount Tamalpais, with flatter coastal areas and hillier inland areas (PRISM Climate Group 2024). Between 1991 and 2020, average annual precipitation levels ranged from 609.6 mm to 1524 mm, with the wettest conditions occurring in the inland areas with the highest elevations (PRISM Climate Group 2024). Between 1991 and 2020, maximum annual temperatures ranged from 13.33℃ to 23.33℃, while minimum annual temperatures ranged from 6.11℃ to 11.11℃ (PRISM Climate Group

2024). During the 1991 to 2020 timeframe, maximum annual temperatures tended to increase with increasing distance from the Pacific Ocean, while minimum annual temperatures tended to decrease with increasing distance from San Francisco Bay (PRISM Climate Group 2024).

Study organisms

There are eight species of oak present in Marin County (Table A1), which can occur in single species stands and in mixed oak species alliances (Buck-Diaz et al. 2021). With the exception of blue oak (*Quercus douglasii*), all of these species display at least some shade tolerance as saplings but become more shade intolerant as they mature (UC Agriculture and Natural Resources Cooperative Extension 2024). With the exception of valley oak (*Quercus lobata*), all of the Marin County oak species are at minimum tolerant of low intensity fires as mature trees due to a combination of traits such as vigorous resprouting and thick bark (UC Agriculture and Natural Resources Cooperative Extension 2024).

There are six species of conifer present in Marin County (Table A1) including coast redwood (*Sequoia sempervirens*) and Douglas-fir (*Pseudotsuga menziesii*). Douglas-fir is of particular concern because its saplings can establish and grow in the patchy shade of *Quercus* canopies (Hunter and Barbour 2001). Furthermore, the narrow crown and high potential height of Douglas-fir allow it to easily pierce and then rapidly overtop *Quercus* canopies (Hunter and Barbour 2001).

Detecting and quantifying conifer encroached area

To detect conifer encroachment in Marin County oak woodland, I used an algorithm to classify masked color-infrared (CIR) imagery using a classifier. For my CIR imagery I chose National Agriculture Imagery Program (NAIP) rasters which are high resolution (60 cm) aerial photographs with red, green, blue, and near infrared bands that are collected by the USDA every two to three years depending on the state. The high spatial resolution is needed to identify where oak and conifer canopies are mixing, while the NIR band makes vegetation easier to identify due to how reflective leaves are in the NIR bandwidth. To gather my NAIP imagery, I used Google Earth Engine (GEE) which is a cloud-based remote sensing platform that allows users to download and manipulate imagery. Using a shapefile of NAIP quarter quads cut to the boundaries of Marin County in ArcGIS Pro (version 3.2.2), I called 2020 NAIP imagery of Marin County from GEE's NAIP repository. I chose the 2020 NAIP collection because of its narrower collection timeframe (2020-04-15 to 2020-08-05) (NOAA 2023) rather than the 2018 NAIP collection (2018-06-29 to 2019-03-05) (NOAA 2024), which minimizes temporal variation in vegetation reflectance. I then merged the individual pieces of the NAIP imagery and selected a CIR visualization in GEE and exported the merged imagery to ArcGIS Pro. Because NAIP imagery uses the UTM system (USGS 2017) and because Marin County falls within UTM zone 10N, I reprojected all data used in this study into the NAD 1983 (2011) UTM zone 10N projection.

I then removed spectrally irrelevant non-conifer and non-oak objects from the imagery such as roofs, short herbaceous vegetation, and non-oak and non-conifer trees like Pacific madrone. To accomplish this, I used both elevation and vegetation datasets. The first dataset was the 2019 Marin Canopy Height Model (CHM), which is a three-foot resolution canopy height raster derived from a LiDAR point cloud collected on 2019-12-19. The CHM has had all terrain feature and human-made structure heights set to zero so that only canopy heights are present in the dataset. The second dataset was the 2018 Marin County Fine Scale Vegetation Map (FSVM) which uses polygons labeled with vegetation type data to map out vegetation cover across all of Marin County. The FSVM was finished in 2018 using a combination of field work and classified aerial photography. Using ArcGIS Pro and the CHM, I removed all pixels with values <2 m, because pixels of this height are generally considered to be open areas, shrubs, shorter trees, or ground clutter (Kane et al. 2014, 2023). Then, I used ArcGIS Pro to create a layer from the FSVM that only included polygons that were labeled as conifer or oak vegetation types known to be present in Marin County (Table A1). I used this filtered FSVM layer to clip the ≥2 m CHM in ArcGIS Pro such that only conifer and oak containing areas of the ≥2 m CHM were left. In addition to this masked CHM layer, I also used ArcGIS Pro to create a mask where every pixel in the masked CHM <2 m was given a value of zero and every pixel ≥2 m was given a value of one. This binary mask would be used during the classification step in GEE.

I then exported the masked CHM to R (version 4.3.3) to begin canopy delineation. First, I split the masked CHM into 4000 m by 4000 m squares (tiles) using the terra package in R. Tiling splits canopies that straddle tile boundaries which can lead to single canopies being double counted during delineation. I compensated for this by first using a combination of the terra and raster packages to buffer each tile by 50 m, which prevented tile-edge canopies from being split temporarily. Then, I used the "locate_trees" algorithm from the lidR package (ws = 5) to identify individual treetops and their heights in each buffered tile. Finally, I removed the 50 m buffer and the points on it, which left the correctly located tile-edge canopy treetops that would have otherwise been pairs of points located on separate tiles. Using the correctly placed points, I delineated the tree canopies in each tile using the "dalponte2016" algorithm from the lidR package. I used the "dalponte2016" algorithm because of its ability to delineate canopies using CHMs, not just LiDAR point clouds. I converted the resulting canopy delineation rasters into shapefiles, which I merged into a single Marin County canopy shapefile. Then, using the sf and tidyverse packages, I gave each canopy a unique consecutive canopy ID starting with one, before exporting the canopy shapefile. My final masking step before classification was to apply the binary mask to the 2020 NAIP imagery to remove non-conifer, non-oak, and <2 m pixels that could confuse the classifier. The result of the mask application was 2020 NAIP imagery that had all non-conifer, non-oak, and <2 m pixels removed.

I used a random forest (RF) algorithm to classify my masked 2020 Marin County NAIP imagery. Random forest classifiers use a group (a forest) of trained decision trees to categorize data. RF classifiers are proven to be robust to noise in training data relative to other classification methods in land cover classification applications (Rodriguez-Galiano et al. 2012). RF classifiers have also been proven to be accurate in California vegetation classification studies in the past (Lydersen and Collins 2018, Fertel et al. 2023). Specifically, I used the "ee.Classifier.smileRandomForest" classifier in GEE because of its ability to draw on GEE cloud computing resources and its ability to use shapefiles as training samples. RF classifiers tend to see the largest drop in land cover classification error as the number of trees increases from one to 100, after which improvement decreases (Thanh Noi and Kappas 2018). I therefore decided to use 100 trees in my RF classifier to balance classification speed and accuracy.

I used four canopy classes when creating my classifier's training and validation datasets. "1" for oak canopies, "2" for conifer canopies, "3" for shade canopies or shadows, and "4" for other objects (Table 1). I chose these four classes because they encompassed my two groups of study trees and the two types of noise (shadow and other) I expected to have made it through my masking steps. To aid my manual canopy scoring, I set pixel threshold rules that determined whether a particular canopy would fit into a particular class. For example, if I observed a canopy that had >80% oak pixels within it, I classed it as "1", while a canopy that had ≥30% shadow pixels within its bounds was classed as "3" (Table 1). I chose lower thresholds for the shadow and other classes to compensate for the higher likelihood of observing discontinuous regions of shadow and other pixels relative to tree pixels. To create a training and a validation dataset for the RF classifier, I started by generating a list of random canopy IDs. I used the "runif" function from the stats package in R to create 1,000 canopy IDs that fell within bounds of the canopy IDs of my polygons. I then used the floor function to round all the random canopy IDs and exported the resulting list as a CSV file. I went down the random canopy ID list in the order the IDs appeared in the CSV file. For a given random canopy ID, I used ArcGIS Pro to zoom in on the associated canopy in the Marin County canopy shapefile and observed the 2020 NAIP pixels within the bounds of that canopy. Using ocular photo interpretation, I then assigned the random canopy one of four canopy classes (Table 1) and exported the canopy to a training and validation layer in ArcGIS Pro. I skipped canopies that did not fit into any of the four canopy classes or could not be classified using ocular photo interpretation. I then moved on to the next random canopy ID in the CSV file. I repeated this process until I had 90 canopies for each of the "1", "2", and "3" classes and 15 canopies for the "4" class. I selected an equal number of canopies for the first three classes to avoid creating bias in the classifier. However, I chose not to score 90 canopies in the "4" class because of the rarity of the class in the masked 2020 NAIP imagery.

I chose to select 80% of the 285 classified canopies to be used for training and 20% of the canopies to be used for validation. I chose the 80/20 training and validation dataset split because it is a widely used rule of thumb for training classifiers (Joseph 2022). I split the data by assigning each classified canopy a random number between zero and one in ArcGIS Pro. Canopies with random number values of ≤0.8 were exported in a training shapefile, while canopies with random number values of >0.8 were exported in a validation shapefile. I uploaded the training shapefile to GEE and ran the RF classification on the masked 2020 NAIP imagery. I downloaded the resulting classified NAIP imagery as a GeoTIFF and mosaiced it in ArcGIS Pro. I then used the exactextractr, terra, and sf packages in R to calculate the proportions of four classes within each canopy in the Marin County canopy shapefile and the validation shapefile. I exported the resulting zonal statistics tables to ArcGIS Pro where I used the calculate field tool and the classification pixel fraction thresholds in Table 1 to assign every canopy a predicted canopy class. I then created a table showing instances of canopy labeling confusion by comparing the true class column and predicted class columns in the validation dataset using the dplyr package in R. I calculated the unweighted Cohen's Kappa statistic for the validation dataset using the "kappa2" function from the irr package in R.

Table 1. Canopy Classification Criteria. The rows highlighted in gray were the only canopy labels/thresholds used for generating training data and running the RF classifier. All canopy labels/thresholds were used during the final canopy labeling process in ArcGIS Pro. Class 5 is considered to be conifer encroached oak canopy.

Canopy Class	Classification Pixel Fraction Thresholds
1 - Oak	>80% oak
2 - Conifer	>80% conifer
3 - Shade	≥30% shade
4 - Other	≥30% dead tree, roof, asphalt, concrete, etc.
5 - Oak-conifer (conifer encroached oak canopy)	>50% oak and >20% conifer
6 - Oak-shade	>50% oak and <30% shade and ≥20% shade
7 - Oak-other	>50% oak and <30% other and ≥20% other
8 - Conifer-oak	>50% conifer and >20% oak
9 - Conifer-shade	>50% conifer and <30% shade and ≥20% shade
10 - Conifer-other	>50% conifer and <30% other and ≥20% other
11 - Uncategorized	A pixel fraction not accounted for in classes 1 to 10

Subsetting conifer encroached area in the overtopping phase

To subset conifer encroached oak woodland in the overtopping phase, I used the "polygon neighbors" tool in ArcGIS Pro on the classed Marin County canopy shapefile produced in the previous step. The "polygon neighbors" tool records the traits of a polygon's neighboring polygons in a table. The tool does this for every polygon in a given dataset and the user can select the traits that get reported in the final table. I set the input feature of the tool as the canopy shapefile and set the reporting fields as canopy ID, canopy height, and canopy class. This produced a table where each neighbor relationship was represented with a row, meaning a canopy with three neighbors would be represented with three separate rows. I then exported the table to R where I used the dplyr package to filter the table such that only class "5" canopies that had class "2", "8", "9", or "10" neighbors remained. I then used dplyr to remove all duplicate canopy IDs. The result was a table that contained all encroached oak canopies that were experiencing overtopping in Marin County.

RESULTS

Conifer encroached area

I found that the overall accuracy of the masked NAIP image classification and zonal statistics canopy labeling process was 39.34%. The unweighted Cohen's Kappa statistic of the canopy labeling process was 0.26. I found that the process identified pure oak canopies (class "1") with an accuracy rate of 7.14% (Table 2). I calculated that the process mislabeled pure oak canopies as conifer encroached oak canopies (class "5") 28.57% of the time and as uncategorized canopies (class "11") 42.86% of the time (Table 2). I found that the process identified pure conifer canopies (class "2") with an accuracy rate of 15.00% (Table 2). I calculated that the process mislabeled pure conifer canopies as shade (class "3") 15.00% of the time, as conifer-oak canopies (class "8") 20.00% of the time, and as uncategorized canopies 35.00% of the time (Table 2). I found that the process identified pure shadow canopies (class "3") with an accuracy rate of 79.17% (Table 2). After calculating canopy makeup by using zonal statistics on the classified NAIP image and applying the pixel thresholds from Table 1, I found that 59.02% of all oak canopies (classes "1", "5", "6", and "7") in Marin County were experiencing conifer encroachment (class "5") (Figure 2).

Table 2. Accuracy assessment for the predicted canopy class in the validation dataset. The rows highlighted in gray are correct classifications.

True/Referenced Canopy Class	Predicted Canopy Class	Number of Occurrences
1 (Oak)	1	1
1	3	1
1	5	4
1	8	2
1	11	6
2 (Conifer)	1	1
2	2	3
2	3	3
2	8	4
2	9	2
2	11	7
3 (Shade)	3	19
3	6	1
3	8	1
3	9	1
3	11	2
4 (Other)	1	1
4	4	1
4	11	1

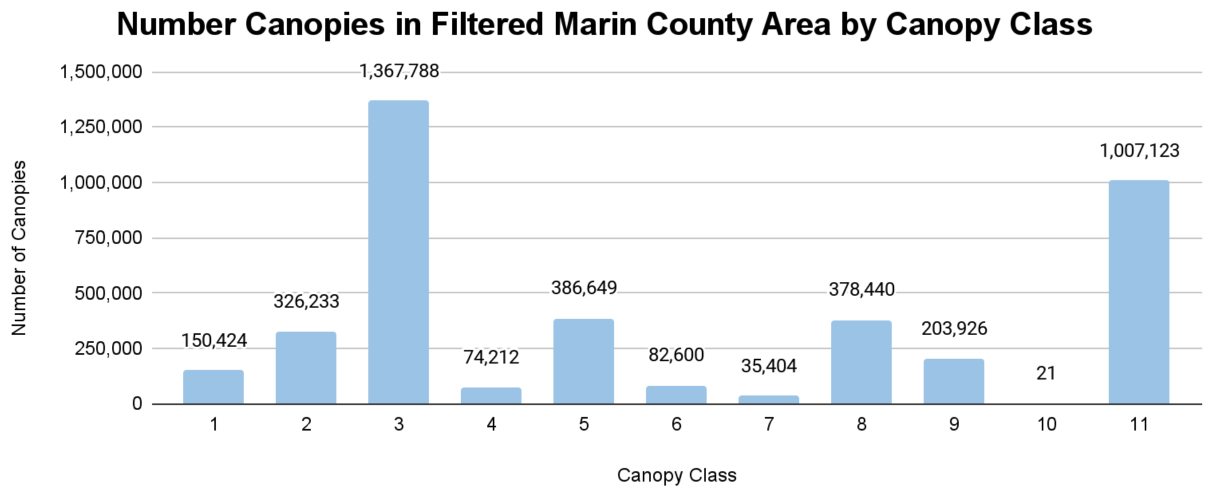

Figure 2. Number of canopies in filtered Marin County area by canopy class. Marin County conifer encroachment proportions. See Table 1 for a canopy class key.

Overtopping phase conifer encroached area

After performing the polygon neighbor analysis for all encroached oak canopies (class "5"), I found that 23.22% of oak canopies in Marin County experiencing conifer encroachment were in the overtopping phase of conifer encroachment (Figure 3).

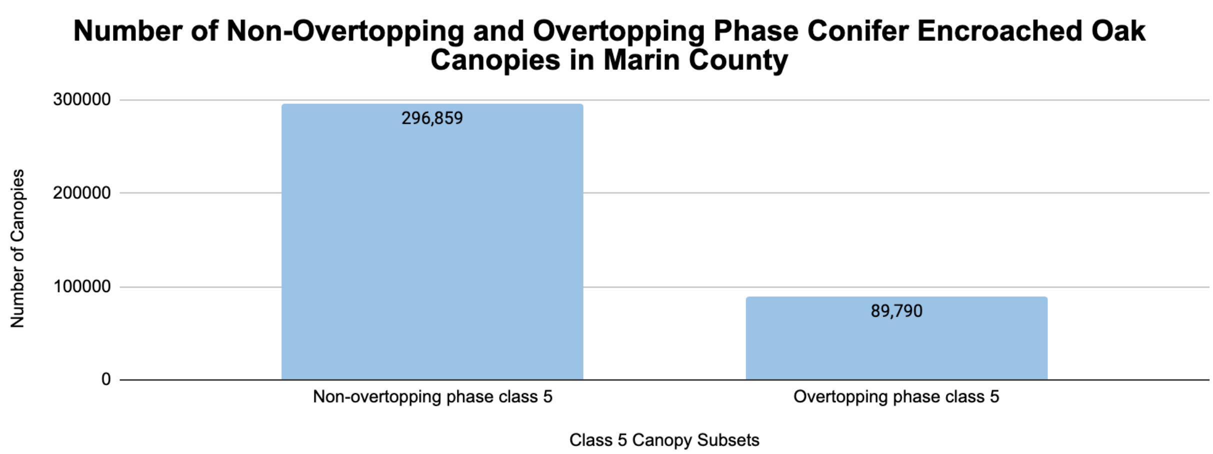

Figure 3. Number of non-overtopping and overtopping phase conifer encroached oak canopies in Marin County. Marin County overtopping phase conifer encroachment proportions. See Table 1 for a canopy class key.

DISCUSSION

I found that 59.02% oak canopies in Marin County were experiencing conifer encroachment. Of these encroached canopies, 23.22% were found to be in the overtopping phase of conifer encroachment. However, the conifer encroachment and overtopping and findings were derived from an image classification and zonal statistics canopy labeling process that could not accurately identify conifer and oak canopies. Therefore, these findings are not yet accurate enough to be used as a reliable management tool.

Quantification of encroachment

My finding that 59.02% of oak canopies in Marin County are conifer encroached was derived from a canopy labeling process with low accuracy. Specifically, my image classification and zonal statistics canopy labeling process tended to mislabel pure oak canopies as conifer encroached oak canopies. This tendency may have led to an overestimate of the number of conifer encroached oak canopies and an underestimate of the number of pure oak canopies in Marin County. Furthermore, my canopy labeling process tended to mislabel pure oak canopies as uncategorized canopies. This tendency may have led to a further underestimation of the number of pure oak canopies in Marin County. The combination of the underestimation of pure oak canopies and the overestimation of conifer encroached oak canopies means that my finding that 59.02% of oak canopies in Marin County are conifer encroached is likely higher than the true proportion.

The inaccuracy of my canopy labeling process makes understanding the true extent of conifer encroachment in Marin County difficult. Although I was able to describe the process' tendency towards conifer encroached oak canopy false positives, I was not able to ascertain the magnitude of the problem. Deriving an accurate measurement of the extent of conifer encroachment in Marin County oak woodland from the current dataset would likely be time consuming and difficult. Therefore, a new dataset created with an improved canopy labeling process will be needed to calculate the true extent and spatial distribution of conifer encroachment in Marin County oak woodland. This is especially important for forest management applications, which require spatially accurate datasets to make effective interventions in conifer encroachment.

Subsetting overtopping phase encroachment

My finding that 23.22% of conifer encroached oak canopies in Marin County are in the overtopping phase of conifer encroachment was derived from an inaccurate canopy labeling process. My image classification and zonal statistics canopy labeling process overestimated the number of conifer encroached oak canopies in Marin County. This tendency likely led to a general overestimation of the number conifer encroached oak canopies experiencing overtopping. In addition, my canopy labeling process underestimated the number of pure conifer canopies in the county. Specifically, my canopy labeling process tended to confuse pure conifer canopies with shade and uncategorized canopies. This likely led to the "polygon neighbor" tool labeling fewer conifer encroached oak canopies as overtopped than what was actually true. The combination of the overestimation of the number of conifer encroached oak canopies and the underestimation of conifer encroached oak canopies experiencing overtopping undermines the accuracy of my overtopping findings.

The combination of conifer encroached oak canopy overestimation and pure conifer canopy underestimation means that I was not able to ascertain if the "polygon neighbor" tool tended towards false positives or false negatives. This makes estimating the true number of conifer encroached oak canopies that are experiencing overtopping in Marin County near impossible using my current dataset. Here again I recommend the creation of a new dataset with an improved canopy labeling process. The reliance of the overtopping subsetting step on an accurate canopy labeling process makes this new dataset a necessity.

Synthesis

The inaccuracy of my image classification and zonal statistics canopy labeling process meant I could not ascertain the extent and severity of conifer encroachment in Marin County oak woodland. I found that the canopy labeling process overestimated the number of conifer encroached oak canopies in Marin County which undermined my ability to accurately calculate the proportion of conifer encroached oak canopies to total oak canopies and the proportion of conifer encroached oak canopies experiencing overtopping to total conifer encroached oak canopies. I believe that creating a new Marin County conifer encroachment dataset with a more accurate canopy labeling process will address the conifer encroached oak canopy false positive issue.

Limitations

The final classified map incorporated datasets that each had some level of inherent error as is typical of remote sensing investigations. The 2018 Marin County FSVM had an overall accuracy of 77%, leading me to think that some conifers and oaks were removed during the masking while some non-conifer and non-oak trees were left in. I also believe that the CHM introduced some level of error into the overtopping analysis due to the LiDAR point cloud to raster conversion process used in its creation. I think that error was also introduced due to temporal differences within and between datasets. 2020 California NAIP imagery was collected between 2020-4-15 and 2020-08-04 meaning that late-spring and summer vegetation spectra were collected. I don't know for certain whether this temporal difference was present in the Marin County 2020 NAIP imagery specifically. Differences between dataset collection dates were also present, with the NAIP imagery being from 2020, the CHM being from 2019, and the FSVM being from 2018. I think that changes such as treefall or canopy death that occurred between the collection times of these datasets introduced temporal noise into my final dataset.

The classification in GEE relied on a relatively small training sample size. Out of the 4,012,820 canopies identified in the canopy delineation process, I manually scored and used 224 to create the training dataset. My use of a small training dataset to classify the county-spanning area of the masked 2020 NAIP imagery likely led to misclassifications.

A lack of computing power and memory necessitated the tiling of the CHM during the treetop identification and tree crown delineation process. I used a buffer and clip process to avoid the double counting of treetops, but this did not remove edge effects. Ultimately, I accepted the inaccuracies in canopy delineation due to edge effects because of the difficulties involved in removing these effects and the cost of acquiring computers powerful enough to avoid tiling. I also accepted inaccuracies in tree crown delineation due to my search window setting choice for the "locate_trees" algorithm due to the impossibility of finding an optimal search window size for the range of tree crown diameters found across the county.

Future directions

Creating a more accurate canopy labeling process is the most immediate next step. I found that the tendency of my canopy labeling process to overestimate conifer encroached oak canopies made my final data set too inaccurate to ascertain the true extent and severity of conifer encroachment in Marin County oak woodland. I believe there are several ways to improve canopy labeling accuracy. I recommend using ground truthing instead of photo interpreted NAIP imagery to label canopies for the RF classifier's training and validation datasets. Using ground truthed tree data would decrease the likelihood that the classifier is being fed mislabeled training canopies, increasing accuracy. I recommend increasing the number of canopies in the training and validation datasets. Increasing the number of training canopies will give the RF classifier a more comprehensive sample of the diverse array of canopy spectra in Marin County, decreasing mislabeling. I recommend rerunning the analysis on more temporally recent and more temporally synced data. Given the fast growth rate of conifer species like Douglas-fir, the 2018-2020 data will increasingly differ from the on-the-ground reality in conifer encroached Marin County oak woodland. A rerun of the analysis should also seek to minimize temporal noise by using datasets that have collection dates that match as closely as possible. If NAIP data is going to be used in future analyses, I recommend that Marin County synchronize its LiDAR data collection dates with NAIP imagery collection dates (i.e., on even years). Finally, I recommend employing more computing power and memory in this analysis, which will reduce the need for tiling and thus canopy delineation edge effects.

Broader implications

This study is an important first step in characterizing the extent and severity of conifer encroachment in California oak woodland. The study used remote sensing techniques to move beyond the time, resource, and coverage constraints of plot-level studies to characterize conifer encroachment in oak woodland at a county-wide scale. While the canopy labeling process ultimately had a variety of inaccuracies, I believe that there are multiple opportunities to fix these inaccuracies and that creating an improved labeling process will be worth the effort. This is because an accurate spatial dataset of conifer encroached oak canopies at a county or statewide scale would decrease the time forest managers need to spend performing field studies on conifer encroachment while increasing the time they spend on removing encroaching conifers. An accurate spatial dataset of conifer encroached oak canopies could also allow for the study of the spatial covariates of conifer encroachment at a county or statewide scale, something I was unable to accomplish in this paper due to the inaccuracy of my canopy labeling process.

ACKNOWLEDGEMENTS

Thank you to Saba Saberi for comprehensively mentoring and supporting me throughout my senior thesis journey, your dedication has been incredible. I am grateful to Patina Mendez and Melissa von Mayrhauser for providing continuous feedback on every part of my senior thesis. Dr. Jeffery Chambers, Xihan Yao, and Yanlei Feng provided technical assistance for my classification pipeline. Dr. Lynn Huntsinger inspired me to investigate conifer encroachment in California oak woodland with her excellent lectures on the subject. I greatly benefited from the comradeship and peer review support provided by my incredible Environmental Sciences cohort: Tessa Auer, Mina Burns, Moses Castillo, Angela Kou, Brandon Lin, Akshay Patel, Mark Sun, and Ethan Xie. Finally, thank you to Craig, Hui, and Katherine Jacobson for supporting me throughout my thesis journey and beyond.

REFERENCES

Barnhart, S. J., J. R. McBride, and P. Warner. 1996. Invasion of Northern Oak Woodlands by Pseudotsuga Menziesii (mirb.) Franco in the Sonoma Mountains of California. Madroño 43:28–45.

Brodie, L. C., and M. Palmer. 2020. California's forest resources, 2006–2015: Ten-Year Forest Inventory and Analysis report. General Technical Report, USDA, Forest Service, Pacific Northwest Research Station.

Buck-Diaz, J., K. Sikes, and J. M. Evens. 2021. Appendix D Vegetation Descriptions.

Cocking, M. I., J. M. Varner, and E. A. Engber. 2015. Conifer encroachment in California oak woodlands. Gen. Tech. Rep. PSW-GTR-251. Berkeley, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station: 505-514 251:505–514.

Dey, D. C., J. M. Kabrick, and C. J. Schweitzer. 2017. Silviculture to Restore Oak Savannas and Woodlands. Journal of Forestry 115:202–211.

Fertel, H. M., B. M. Collins, J. M. Lydersen, and S. L. Stephens. 2023. Vegetation type change in California's Northern Bay Area: A comparison of contemporary and historical aerial imagery. Forest Ecology and Management 542:121102.

Finney, M. A., and R. E. Martin. 1992. Short Fire Intervals Recorded by Redwoods at Annadel State Park, California. Madroño 39:251–262.

Gaman, T., and J. Firman. 2006. Oaks 2040: The Status and Future of Oaks in California.

Holmes, K., K. Veblen, T. Young, and A. Berry. 2008. California Oaks and Fire: A Review and Case Study. USDA Forest Service - General Technical Report PNW PSW-GTR-217:551–565.

Hunter, J. C., and M. G. Barbour. 2001. Through-Growth by Pseudotsuga menziesii: A Mechanism for Change in Forest Composition without Canopy Gaps. Journal of Vegetation Science 12:445–452.

Joseph, V. R. 2022. Optimal ratio for data splitting. Statistical Analysis and Data Mining: The ASA Data Science Journal 15:531–538.

Kane, V. R., B. N. Bartl-Geller, G. R. Cova, C. P. Chamberlain, L. Van Wagtendonk, and M. P. North. 2023. Where are the large trees? A census of Sierra Nevada large trees to determine their frequency and spatial distribution across three large landscapes. Forest Ecology and Management 546:121351.

Kane, V. R., M. P. North, J. A. Lutz, D. J. Churchill, S. L. Roberts, D. F. Smith, R. J. McGaughey, J. T. Kane, and M. L. Brooks. 2014. Assessing fire effects on forest spatial structure using a fusion of Landsat and airborne LiDAR data in Yosemite National Park. Remote Sensing of Environment 151:89–101.

Lavezzo, T., B. Penfold, S. Huang, and C. Scarborough. 2020. Community Wildfire Protection Plan.

Long, J. W., M. K. Anderson, L. Quinn-Davidson, R. W. Goode, F. K. Lake, and C. N. Skinner. 2016. Restoring California black oak ecosystems to promote tribal values and wildlife. Gen. Tech. Rep. PSW GTR-252. Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station. 110 p. 252.

Lydersen, J. M., and B. M. Collins. 2018. Change in Vegetation Patterns Over a Large Forested

Landscape Based on Historical and Contemporary Aerial Photography. Ecosystems 21:1348–1363.

van Mantgem, P. J., M. C. Wright, and E. A. Engber. 2021. Patterns of conifer invasion following prescribed fire in grasslands and oak woodlands of Redwood National Park, California. Restoration Ecology 29:e13366.

Nemens, D. G., J. M. Varner, K. R. Kidd, and B. Wing. 2018. Do repeated wildfires promote restoration of oak woodlands in mixed-conifer landscapes? Forest Ecology and Management 427:143–151.

NOAA. 2023, May 30. 2020 California NAIP 4-Band 8 Bit Imagery | InPort. https://www.fisheries.noaa.gov/inport/item/68159.

NOAA. 2024, January 11. 2018 California NAIP 4-Band 8 Bit Imagery | InPort. https://www.fisheries.noaa.gov/inport/item/60705.

PRISM Climate Group. 2024. PRISM Climate Group at Oregon State University. https://prism.oregonstate.edu/.

Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 67:93–104.

Schriver, M., R. L. Sherriff, J. M. Varner, L. Quinn-Davidson, and Y. Valachovic. 2018. Age and stand structure of oak woodlands along a gradient of conifer encroachment in northwestern California. Ecosphere 9:e02446.

Stahlheber, K. A., and C. M. D'Antonio. 2014. Do Tree Canopies Enhance Perennial Grass Restoration in California Oak Savannas? Restoration Ecology 22:574–581.

State of California, California Natural Resources Agency, and California Department of Fish and Wildlife. 2021. Atlas of the Biodiversity of California. Second edition. California Department of Fish and Wildlife.

Sugihara, N. G., and L. J. Reed. 1987. VEGETATION OF THE BALD HILLS OAK WOODLANDS, REDWOOD NATIONAL PARK, CALIFORNIA.

Thanh Noi, P., and M. Kappas. 2018. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors 18:18.

UC Agriculture and Natural Resources Cooperative Extension. 2024. Oak Tree Species ID & Ecology. https://oaks.cnr.berkeley.edu/oak-tree-species-id-ecology/.

U.S. Census Bureau. 2024. Marin County, California - Census Bureau Profile.

https://data.census.gov/profile/Marin_County,_California?g=050XX00US06041.

USGS. 2017. National Agriculture Imagery Program (NAIP). U.S. Geological Survey.

Van de Water, K. M., and H. D. Safford. 2011. A Summary of Fire Frequency Estimates for California Vegetation before Euro-American Settlement. Fire Ecology 7:26–58.

van Wagtendonk, J. W. 2007. The History and Evolution of Wildland Fire Use. Fire Ecology 3:3–17.

APPENDIX A: Fine Scale Vegetation Map (2018) Vegetation Classes

Table A1. A list of all the vegetation types found in the 2018 Marin County Fine Scale Vegetation Map. The vegetation types highlighted in yellow are those that were considered as conifer and oak areas during the masking step of the study.

Fine Scale Map Class in '18	Common Name
1. Acacia spp. – Grevillea spp. – Leptospermum laevigatum Semi-Natural Alliance	Acacia - spider flower - coast tea tree
2. Acer macrophyllum – Alnus rubra Alliance	Bigleaf maple - red alder
3. Acer macrophyllum Association	Bigleaf maple
4. Acer negundo / (Rubus ursinus) Association	Boxelder maple - pacific blackberry
5. Adenostoma fasciculatum Alliance	Chamise
6. Aesculus californica Alliance	California buckeye
7. Alnus rhombifolia Alliance	White alder
8. Ammophila arenaria Semi-Natural Alliance	European marram grass
9. Annual Cropland	NA
10. Aquaculture	NA
11. Arbutus menziesii Alliance	Pacific madrone
12. Arctostaphylos (bakeri, montana) Alliance	Manzanitas (various types)
13. Arctostaphylos (canescens, manzanita, stanfordiana) Alliance	Manzanitas (various types)
14. Arctostaphylos (nummularia, sensitiva) – Chrysolepis chrysophylla Alliance	Manzanitas (various types) - Golden chinquapin
15. Arctostaphylos glandulosa Alliance	Manzanitas
16. Arid West Freshwater Marsh Group	NA
17. Artemisia californica – (Salvia leucophylla) Alliance	California sagebrush
18. Artemisia pycnocephala Association	Coast sagewort
19. Atriplex prostrata – Cotula coronopifolia Semi-Natural Alliance	Spear-leaved orache - Brass buttons
20. Baccharis pilularis Alliance	Coyote brush
21. Barren and Sparsely Vegetated	NA
22. Bolboschoenus maritimus Alliance	Sea clubrush
23. Calamagrostis nutkaensis Alliance	Pacific reedgrass
24. Californian Annual & Perennial Grassland Mapping Unit	NA
25. Californian Cliff, Scree & Rock Vegetation Group	NA
26. Californian Vernal Pool / Swale Bottomland Group	NA
27. Carthamus lanatus Invasive Mapping Unit	Wooly distaff thistle
28. Ceanothus cuneatus Alliance	Buckbrush
29. Ceanothus thyrsiflorus Alliance	Blueblossom
30. Channel	NA
31. Conifer (Urban Window)	NA
32. Conium maculatum – Foeniculum vulgare Semi-Natural Alliance	Hemlock - Fennel
33. Cortaderia (jubata, selloana) Semi-Natural Alliance	Pampas grasses (various types)
34. Corylus cornuta / Polystichum munitum Association	Beaked hazelnut/Western sword fern
35. Cotoneaster (lacteus, pannosus) Provisional Semi-Natural Association	Cotoneaster shrub (various types)
36. Cytisus scoparius Provisional Semi-Natural Association	Scotch broom
37. Deciduous Hardwood (Urban Window)	NA
38. Developed	NA

39. Distichlis spicata Alliance	Desert saltgrass
40. Eriophyllum staechadifolium – Erigeron glaucus – Eriogonum latifolium Alliance	Beach plant alliance
41. Eucalyptus (globulus, camaldulensis) Provisional Semi-Natural Assocation	Eucalyptus (various types)
42. Evergreen Hardwood (Urban Window)	NA
43. Forest Fragment	NA
44. Frangula californica ssp. californica – Baccharis pilularis / Scrophularia californica Association	Shrub alliance
45. Fraxinus latifolia Alliance	Oregon ash
46. Garrya elliptica Provisional Association	Silk tassel bush
47. Gaultheria shallon – Rubus (ursinus) Alliance	Shrub - Pacific blackberry
48. Genista monspessulana Semi-Natural Association	French broom
49. Grindelia stricta Provisional Association	Oregon gumweed
50. Hesperocyparis macrocarpa Ruderal Provisional Semi-Natural Association	Monterey cypress
51. Hesperocyparis sargentii / Ceanothus jepsonii – Arctostaphylos spp. Association	Sargent's cypress/musk brush
52. Hesperocyparis sargentii Association	Sargent's cypress
53. Intensively Managed Hayfield	NA
54. Irrigated Pasture	NA
55. Lepidium latifolium – (Lactuca serriola) Semi-Natural Alliance	Perennial pepperweed - milk thistle
56. Lotus scoparius – Lupinus albifrons – Eriodictyon spp. Alliance	Deerweed - silver lupine - yerba santa
57. Lupinus arboreus Alliance	Yellow bush lupine
58. Lupinus chamissonis – Ericameria ericoides Alliance	Chamisso bush lupine - california goldenbush
59. Major Road	NA
60. Mesembryanthemum spp. – Carpobrotus spp. Semi-Natural Alliance	Iceplant - Pigface
61. Mudflat/Dry Pond Bottom Mapping Unit	NA
62. Non-native Forest	NA
63. Non-native Herbaceous	NA
64. Non-native Shrub	NA
65. Notholithocarpus densiflorus Alliance	Tanoak
66. Nursery or Ornamental Horticulture Area	NA
67. Orchard or Grove	NA
68. Pacific Coastal Beach & Dune Macrogroup	NA
69. Perennial Cropland	NA
70. Pinus muricata – Pinus radiata Alliance	Bishop pine - Monterey pine
71. Pinus radiata Plantation Provisional Semi-Natural Association	Monterey pine
72. Populus fremontii – Fraxinus velutina – Salix gooddingii Alliance	Fremont cottonwood - Velvet ash - Goodding's willow
73. Pseudotsuga menziesii – (Notholithocarpus densiflorus – Arbutus menziesii) Alliance	Douglas-fir - Tanoak - Pacific madrone
74. Quercus (agrifolia, douglasii, garryana, kelloggii, lobata, wislizeni) Alliance	Oak alliance
75. Quercus agrifolia Alliance	Coast live oak
76. Quercus chrysolepis Alliance	Canyon live oak
77. Quercus douglasii Alliance	Blue oak
78. Quercus durata Alliance	Leather oak
79. Quercus garryana Alliance	Oregon white oak
80. Quercus kelloggii Alliance	California black oak

81. Quercus lobata Alliance	Valley oak
82. Quercus wislizeni – Quercus chrysolepis (shrub) Alliance	Interior live oak - shrub canyon live oak
83. Rhododendron columbianum - Gaultheria shallon / Carex obnupta Association	Western labrador tea - shrub/grass
84. Rubus armeniacus Semi-Natural Association	Himalayan blackberry
85. Rubus spectabilis – Morella californica Alliance	Salmonberry - Pacific wax myrtle
86. Salix exigua Alliance	Narrowleaf willow
87. Salix gooddingii – Salix laevigata Alliance	Willow alliance
88. Salix hookeriana – Salix sitchensis – Spiraea douglasii Alliance	Willow alliance
89. Salix lasiolepis Alliance	Willow
90. Salix lucida ssp. lasiandra Association	Willow
91. Sarcocornia pacifica (Salicornia depressa) Alliance	
92. Sequoia sempervirens Alliance	Coast redwood
93. Shrub (Urban Window)	NA
94. Shrub Fragment	NA
95. Spartina foliosa Association	California cordgrass
96. Tidal Salt Marsh (Out of County)	NA
97. Toxicodendron diversilobum – (Baccharis pilularis) Association	Pacific poison oak - coyote brush
98. Triglochin maritima Association	Seaside arrowgrass
99. Ulex europaeus Provisional Semi-Natural Association	Gorse
100. Umbellularia californica Alliance	California bay
101. Vancouverian Freshwater Wet Meadow & Marsh Group	NA
102. Vancouverian Lowland Marsh, Wet Meadow & Shrubland Macrogroup	NA
103. Vineyard	NA
104. Water	NA
105. Western North American Freshwater Aquatic Vegetation Macrogroup	NA
106. Zostera (marina, pacifica) Pacific Aquatic Alliance	Eelgrass (various types)