INTERNATIONAL JOURNAL OF REMOTE SENSING
2024, VOL. 45, NO. 17, 5820–5844
https://doi.org/10.1080/01431161.2024.2380543

Taylor & Francis
Taylor & Francis Group

Check for updates

Machine learning approach to burned area mapping for Southern California

Chandler Ross[a], Douglas Stow[a], Daniel Sousa[a], Megan Jennings[a], Atsushi Nara[a] and Philip Riggan[b]

[a]Geography Department, San Diego State University, San Diego, CA, USA; [b]USDA, Forest Service Southwest Research Station, Riverside, CA, USA

ABSTRACT

Accurate representation of the location and amount of burned areas is vital to the understanding of patterns and impacts of fires. Some extant burned area maps appear to have high commission errors, which lead to an overrepresentation of burned area. The primary research objective of this study was to assess whether region-specific training data used for machine learning routines improve accuracy of burned area products for western San Diego County. We used training data derived from fine-scale aerial orthoimagery to create and compare three training sets, each with a different Landsat scale sub-pixel burn threshold: 20%, 50%, or 80%. Meaning either 20%, 50%, or 80% of a 30 m × 30 m pixel had to burn for the entire pixel to be classified as burned. High-resolution orthoimagery was used for the creation of the training/testing data as well as determining which sub-pixel threshold leads to more accurate burned area representation. These training data were input into a gradient-boosted regression model. We compared the burned area product from the region-specific gradient boosted model (L-GBRM) to the three products: Monitoring Trends in Burn Severity, Fire and Resource Assessment Program, and the Landsat Burned Area product. We found >20% sub-pixel burn threshold of a Landsat pixel yielded the most accurate classification results. We used a 50% sub-pixel burn threshold for the reference data to compare the results to since the burned area associated with it is closely aligned with the high-resolution orthoimagery determined burn area. The L-GBRM was the most accurate product while also mapping the smallest area burned, suggesting that the extant products have relatively high commission errors. Using region-specific training data achieved a higher accuracy than nationwide training data. Looking at sub-pixel burn thresholds for creating a burned area map could prove to make a more accurate map in terms of area burned represented.

ARTICLE HISTORY
Received 20 November 2023
Accepted 8 July 2024

KEYWORDS
Burned area; landsat; optical remote sensing

1. Introduction

The State of California, U.S.A., has seen an increase in frequency of large fires with nine of the top 20 largest ones (since the late 19th century) occurring since 2015 and 17 of

CONTACT Chandler Ross cross8046@sdsu.edu Geography Department, San Diego State University, 5500 Campanile Dr, San Diego, CA 92182, USA

© 2024 Informa UK Limited, trading as Taylor & Francis Group

the top 20 occurring since 2000 (CAL FIRE 2023). This has led to an increase in the importance of fire research and the development of tools and data products for effective fire management. An important information requirement for effective fire management and resource conservation is estimation of location and extent of burned areas over time.

Burned area extent and distribution are some of the least understood and inconsistent data types in fire science (Brown et al. 2002; Kolden and Weisberg 2007; Silva et al. 2003). Data and information on burned area amounts and distributions are used for understanding wildfire conditions and potential recovery (Bowman et al. 2009; Leland and Lutz 2011; Westerling et al. 2003).

By more accurately quantifying the size and number of wildfires, we can more reliably characterize wildfire regimes and changes in those regimes (Westerling et al. 2003). The amount of smoke and carbon entering the atmosphere from wildfires can be estimated from how much vegetation burned (Leland and Lutz 2011). Finally, with improved mapping of the amount and distribution of unburned patches, we can better understand and predict vegetation recovery, which is associated with the regeneration of fuels that influence the characteristics of future wildfires (Bowman et al. 2009; Kolden et al. 2012).

Fire plays an important role in influencing the composition and condition of vegetation communities. In California, some invasive species, particularly non-native grasses, flourish when native species in shrubland communities burn frequently, especially at short intervals (Syphard et al. 2019; Syphard, Brennan, and Keeley 2019). Plants in Southern California have adapted to fire, and for some, fire is a critical disturbance for their life cycles. However, short interval fires can cause damage to ecosystems, even for plants that rely on fires. Biologists can better understand the relationship between fire frequency and vegetation regrowth through more spatially and temporally precise burned area maps.

The increasing number and size of wildfires has resulted in a substantial increase in expenditure for fire suppression in California (CAL FIRE 2023). Burn extent is a very important metric for fire managers because it allows them to assess where fires have burned and if their suppression methods were effective. A better understanding of where a fire burned can help determine future fire management practices.

Remote sensing has been used to map burned areas since the early 1970s. With newer satellite sensors, spectral indices, and machine learning methods, burned area mapping accuracy has improved (Chuvieco et al. 2019). Different types of sensors have been used to map fires, but optical is the most prevalent type (Chuvieco et al. 2019). Synthetic aperture radar sensors have been used for burned area mapping; however, optical sensors lead to more accurate maps (Tanase et al. 2020). In this study, we utilize Landsat data because it is a moderate spatial resolution (30 m) satellite mission, with coverage going back to 1984 (for Landsat satellites 5–8). Landsat data provide a balance between spatial and temporal resolution. Higher spatial resolution data are often used for training, testing, and validation of satellite-derived products (Daldegan, Roberts, and Ribeiro 2019; Storey, Lee West, and Stow 2021; Tansey et al. 2008). Commonly used for reference data generation is imagery from the U.S. National Agriculture Imagery Program (NAIP) and similar aerial visible and near-infrared orthoimagery (Storey, Lee West, and Stow 2021). The benefit of using higher spatial resolution data to create reference data for Landsat-based burned area mapping is greater certainty of whether a pixel represents burned or unburned land surfaces. This finer spatial resolution allows the fraction of burned area to be quantified for the larger size of the ground resolution element associated with a Landsat pixel.

Machine learning (ML) is a powerful approach to automatically classifying imagery. A possible drawback to ML classifiers is they tend to require accurate and abundant training data to be effective. However, once ample training data are generated, ML can achieve accurate classifications with minimal human input. We specifically used the gradient boosted regression model (GBRM) because it has been used to accurately map burned areas using Landsat for the conterminous US (Hawbaker et al. 2020).

Gradient boosting regression models are similar to classification and regression tree (CART) models in that their basis is a CART model where each observation is used to better fit the model depending on the magnitude of residuals from the previous CART model. The larger the residual, the smaller the weight of the output is. The GBRM can yield more accurate results than a regular CART and even a random forest when it comes to binary decisions, plus it has extensive use in remote sensing applications (Hastie, Friedman, and Tibshirani 2009; Hawbaker et al. 2017). The algorithm does not assume independence among its predictors and can handle correlated predictors; though an increased number of predictors may not improve the accuracy (Hawbaker et al. 2020). Hawbaker et al. (2020) reduced computation time by eliminating ineffective predictors by selecting uncorrelated input features. They used various SVIs as predictors, as the indices are sensitive to presence of ash and charcoal and to changes in surface temperature from the multi-temporal thermal infrared features.

Several burned area mapping programs exist and thus, we compared products that we generated with those from (1) Monitoring Trends and Burn Severity (MTBS), a multiagency database between the US Forest Service and the US Geological Survey (USGS); (2) Fire and Resource Assessment Program (FRAP), a CAL FIRE database; and (3) Landsat Burned Area (LBA), a USGS product. BA maps from MTBS and FRAP were found by (Storey, Lee West, and Stow 2021) to overestimate burned area by roughly 44% due to inaccurate perimeters and the inclusion of internal unburned patches. LBA is a newer program and aims to map all small and large fires in the conterminous US. The geographic domain of LBA is the conterminous US and, therefore, their products are not specifically trained for the fuel, terrain and postfire conditions associated with predominantly chaparral fires of the Southern California ecoregion (Hawbaker et al. 2020).

MTBS is based on Landsat NBR, dNBR, and RdNBR inputs. Spectral indices based on NBR, though effective as inputs to image classifiers, can be inconsistent in terms of representing burn severity levels, especially in shrublands (Sparks et al. 2015). NBR is not an optimal index for detecting light burns under tree canopies, detecting very light homogeneous burns, and isn't optimal with severe burns in the same pixel as unburnable substances (e.g. rocks), differences in illumination angle between images, and quick regenerating areas (Kolden et al. 2012; Roy, Boschetti, and Trigg 2006). MTBS provides both burned extent and burn severity products. However, burn severity can be inconsistently represented from fire to fire due to illumination and vegetation differences between areas (Loboda, O'Neal, and Csiszar 2007). MTBS is produced for fire management purposes, but no field verification is conducted (Kolden, Smith, and Abatzoglou 2015). Furthermore, fires under 202 ha (500 acres) in the Eastern US and under 405 ha (1,000 acres) in the Western US are not mapped.

Unlike MTBS, FRAP uses a variety of image-derived inputs and approaches (i.e. not just NBR) to create fire perimeter maps, including manual delineation of burned area locations onto topographic maps, field surveys using global navigation satellite system equipment, and image classification (Storey, Lee West, and Stow 2021; Syphard and Keeley 2016). Similar to MTBS, FRAP uses a minimum fire size threshold (4.04 ha (10 acres) for timber fires, 12.14 ha (30 acres) for brush fires, and 121.41 ha (300 acres) for grass fires) before they will map a fire, so many small fires go unmapped (Syphard and Keeley 2016).

LBA products are generated by the USGS using a GBRM to classify burned area (Hawbaker et al. 2020). LBA consists of two products, continuous burn probability (BP) and burn classification (BC) maps. BP maps are created with a GBRM based on around 16 spectral index inputs. This product was validated both with high spatial resolution reference data and by visual analysis of Landsat images conducted by three independent image interpreters. The BC approach applies image segmentation to the BP product, then filters out segments smaller than 22 pixels. A burned probability with >96% is needed for a pixel to be considered burned prior to segmentation. Then, a random walker segmentation (Grady 2005) is used to grow the burned area with pixels with a >71% burned probability. Unlike MTBS and FRAP, no minimum fire size threshold is implemented, so more fires are mapped. The classification algorithm has been found to be robust for the conterminous US (Vanderhoof et al. 2017). However, no training data for Southern California fires were utilized, which likely reduces the accuracy and representativeness of the LBA products for this region.

The basis for determining and portraying fire history is important for understanding how the shape, area, and frequency of overlapping fires affect a landscape. Shape is defined here as the polygonal form of burned area perimeter. Shape can be important for conservation biologists because an accurately mapped burned polygon will influence the degree of overlap of historical fires. Overlapping burned area on overlain maps provides crucial information for studying multiple fires on a vegetation stand (Kolden et al. 2012). Lippitt et al. (2013) studied frequent fires to determine the vulnerability of chamise chaparral in San Diego County. They examined vegetation on land that was affected by fires with a 5-year or less return interval from 1977 to 2003 to determine if fire correlated with vegetation type conversion (Lippitt et al. 2013). They concluded that areas with two burns occurring in 5 years or less or three burns close in time were more likely to have type converted than areas with just a single burn. Other studies have examined the effects of different fire return intervals of 6 to 8 years on chaparral recovery and have even debated if type conversion from frequent fire occurs at all (Jacobsen, Davis, and Fabritius 2004; Meng et al. 2014).

Areal extent of burn is important for climate modellers to determine how much smoke/carbon from a fire enters the atmosphere (Leland and Lutz 2011). They are more concerned with the areal extent of burned land, than the exact location of where it burned. As the number of wildfires has increased in Southern California and many geographic locations, more emphasis has been placed on the effects of smoke (Black et al. 2017; Sicard et al. 2019). Area can be more accurately estimated by mapping the correct burned extent of a fire that excludes unburned patches, rather than simply using the burned perimeter.

The broad aim of this study is to increase understanding of fire history for Southern California with more accurate and precise burned area maps. The primary technical research objective is to assess whether region-specific training data used for image-based machine learning routines improve accuracy of burned area products for a study area in western San Diego County. To achieve this, we processed a time series of Landsat satellite image data using a gradient boosted model, trained and tested with reference data derived from higher spatial resolution orthoimagery from the US National Agricultural Imaging Program (NAIP). We created three different burned area maps based on three different percentage (of Landsat pixel) burn thresholds and tested their accuracy relative to the reference data. The three output burned area products were compared with each other and with three extant burned area products including Monitoring Trends in Burn Severity (MTBS), Fire Recovery Assessment Program (FRAP), and Landsat Burned Area (LBA) Products.

The study context is using Landsat imagery captured during summer and fall seasons to classify burned areas during the period 2003 to 2018 for a Southern Californian study area composed of chaparral and grasslands. We determined if burned area map time series from a regionally specific model yields a more accurate representation of burn history than a nationally trained model. Further, we assessed how changing the percentage burn threshold for a Landsat pixel-element, based on the number of burned higher resolution classified burned pixels within the Landsat pixel-element, affects the burned area map accuracy.

The following research questions were addressed, within the context stated above.

(1) How well do region-specific training data improve machine learning model (Gradient Boosted Classification) results when classifying Landsat multispectral data to generate burned area maps?
(2) How do the accuracies of burned area maps from the nationwide and regionally specific trained ML classifiers compare to more traditional burned area products from the FRAP & MTBS programs?
(3) How do the number, area, and spatial distribution of burned areas represented by these different burned area products differ?
(4) How does the accuracy of the burned area product derived from the region-specific trained ML model vary as a function of the percentage burned area threshold used to train and test the product?
(5) How do burned area products derived from the region-specific trained ML model for a time series of Landsat data change how fire history is represented compared to MTBS, FRAP, and the nationwide-trained ML burn area product?

2. Methods

2.1. Study area

The extent of the study area is delineated by the intersection of the USGS Level III Southern California/Northern Baja Coast, San Diego County, and the Landsat Analysis Ready Data (ARD) Horizontal 4 Vertical 13 footprint, as shown in Figure 1 (Griffith et al. 2016). The USGS divides the US into ecoregions based on differences in geology, physiography, vegetation, climate, soils, land use, wildlife, and hydrology (Griffith et al. 2016).

The study area encompasses the western section of San Diego County, within the footprint of the South Coast ecoregion. The landscape of the study area is topographically diverse, with low elevation coastal zone in the west to mountains in the east, encompassing an elevation range of approximately 3000 m (Omernik and Griffith 2014). The study period is from 2003 to 2020, roughly corresponding to when aerial orthoimage reference data for training classifiers are available.

Mediterranean-type climate and vegetation characterize the study area. On average, the region receives 355 mm of precipitation annually, with most occurring as rainfall from Pacific originated storms during the winter. A yearly drought occurs in the summer and early fall months when precipitation is negligible other than from occasional monsoon or tropical cyclone influences (Bachelet et al. 2016; Franklin 1998). Winters are mild and summers are warm to hot, with air temperatures following a coastal to inland temperature gradient.

Two general types of fire behaviour occur within the study area, slower spreading, lower intensity fuel (vegetation)-driven fires (June–September) and faster spreading, higher intensity wind-driven fires (September–December) (Jin et al. 2015). Santa Ana weather conditions result when high pressure forms in the Great Basin region, drying out the vegetation and are associated with strong northeasterly winds that rapidly spread wildfires (Jin et al. 2015). Fuel-driven fires tend to occur more frequently in the summer and are often easier to suppress. The total area burned by wind- and fuel-driven fires is about equal due to the large size of the few existing fall fires and the more frequent occurrence of fuel-driven fires in summer (Jin et al. 2015). Fires are predominantly ignited by humans, as lightning is rare for the area (Keeley and Syphard 2018).

2.2. Data sources

Multispectral imagery captured by Landsat 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper Plus (ETM+), and Landsat 8 Operational Land Imager (OLI) was used to map the distribution of burned area for the entire study area and period. Landsat 5 TM

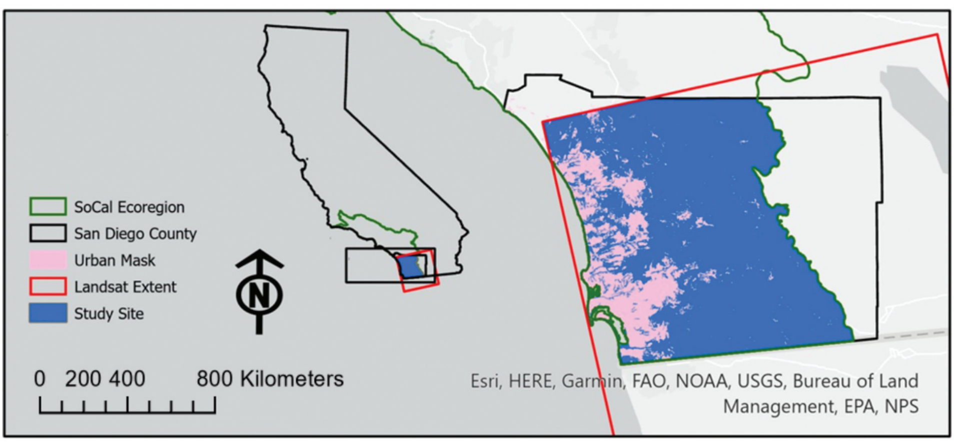

Figure 1. The extent of the study area encompasses the Southern California Ecoregion within San Diego County, excluding urban areas. Data were from the census, SanGIS, EPA, and USGS LUCAS.

data go back to 1984, and similar data continue to be collected with other Landsat sensors with a spatial resolution of 30 m, a radiometric resolution of 8 bits for Landsat 5 and 7 (12 bits for Landsat 8), and a temporal resolution of 16 days. The spatial and temporal resolutions of the Landsat series are sufficient for creating regional burned area products dating back to the mid-1980s.

Landsat Collection 2 Tier 1 surface reflectance data were downloaded from USGS Earth Explorer for time periods corresponding to the high spatial resolution aerial orthoimagery (i.e. periods within 30 days of wildfires) to overlay each other and used to generate the primary burned area products for subsequent analyses. Table 1 shows when the fires occurred and when the NAIP Imagery was acquired. Landsat images were also downloaded for the anniversary date of the post-fire scene for the previous 3 years to calculate a 3-year pre-fire average and standard deviation pixel value. From these data, we generated SVI images based on the formulae included in Table 2. To generate the SVI values are the predictors (input features) for the ML classifiers, we created a 3-year lag average for each of the SVIs which was used to derive prefire SVI, difference SVI, and standard deviation SVI. We ensured cloud cover was limited for each study period/image. Landsat images close to the orthoimage acquisition dates were utilized over those closer to the fire extinguishment date. Table 1 should be around here. Table 2 can follow by not far after.

We used the NAIP data as the source for generating reference data, through high spatial resolution classification of burned areas. We created reference data by classifying NAIP aerial orthoimage data as burned or unburned. NAIP is a program to capture, process, and provide aerial orthoimages of the US every 2–5 years. NAIP data from 2003 to 2008 have BGR spectral bands, from 2009 to present it has BGR and NIR spectral bands. The image spatial resolution varies between 0.6 m for imagery before 2016 to 1 m for imagery after 2016. We selected only NAIP data that were based on images captured in the months of April to August for fires that took place within 30 days of the image being taken. The immediate collection of NAIP imagery after a fire occurred means that the images were likely taken before ash and charcoal had been removed from the surface, making burned area mapping more accurate (Chuvieco et al. 2019).

Table 1. Dates for fires used for training and subsequent NAIP image dates.

Number	Fire Start Date	Fire End Date	NAIP Acquisition Date	Date Difference (Days)
1	24 May 2009	25 May 2009	23 June 2009	30
2	22 May 2009	23 May 2009	27 May 2009	5
3	3 May 2014	3 May 2014	29 May 2014	7
4	19 May 2014	Unknown	3 June 2014	26
5	19 May 2014	Unknown	3 June 2014	15
6	19 May 2014	Unknown	3 June 2014	15
7	14 May 2014	19 May 2014	3 June 2014	15
8	14 May 2014	23 May 2014	3 June 2014	20
9	13 May 2014	30 May 2014	30 May 2014	20
10	14 May 2014	18 May 2014	3 June 2014	17
11	14 May 2014	19 May 2014	3 June 2014	20
12	18 June 2016	18 June 2016	10 July 2016	20
13	6 August 2018	6 August 2018	23 August 2018	22
14	28 July 2018	2 August 2018	23 August 2018	13
15	29 July 2018	18 December 2018	23 August 2018	7
16	25 October 2003	5 December 2003	N/A	N/A

Table 2. Spectral vegetation indices to be used in the burned area classifiers.

Vegetation Index Name	Abbreviation	Formula
Burned Area Index	BAI	$1/((0.1-(Red))^2+(0.06-(NIR))^2)$
Char Soil Index	CSI	$(NIR)/(SWIR2)$
Enhanced Vegetation Index	EVI	$2.5((NIR)-(Red))/((NIR)+(6.0(Red))-(7.5(Blue))+1.0)$
Global Environmental Monitoring Index	GEMI	$N(1.0-0.25\times N)-((Red)-0.125)/(1-(Red));\ N=(2\left((NIR)^2-(Red)^2\right)+(1.5(NIR))+(0.5(Red)))/((NIR)+(Red)+0.5)$
Mid Infrared Burn Index	MIRBI	$(10.0(SWIR2))-(9.8(SWIR1))+2.$
Normalized Burn Ratio	NBR	$((NIR)-(SWIR2))/((NIR)+(SWIR2))$
Normalized Burn Ratio 2	NBR2	$((SWIR1)-(SWIR2))/((SWIR1)+(SWIR2))$
Normalized Burn Ratio Thermal	NBRT1	$((NIR)-((SWIR2)\times(Thermal)))/((NIR)+((SWIR2)\times(Thermal)))$
Normalized Difference Moisture Index	NDMI	$((NIR)-(SWIR1))/((NIR)+(SWIR1))$
Normalized Difference Vegetation Index	NDVI	$((NIR)-(Red))/((NIR)+(Red))$
Normalized Difference Wetness Index	NDWI	$((Green)-(NIR))/((Green)+(NIR))$
Soil-Adjusted Vegetation Index	SAVI	$1.5((NIR)-(Red))/((NIR)+(Red)+0.5)$
Vegetation Index 6 Thermal	VI6T	$((NIR)-(Thermal))/((NIR)+(Thermal))$
NIR/Red Ratio	VI43	$(NIR)/(Red)$
NIR/SWIR1 Ratio	VI45	$(NIR)/(SWIR1)$
NIR/Thermal Ratio	VI46	$(NIR)/(Thermal)$
SWIR1/SWIR2 Ratio	VI57	$(SWIR1)/(SWIR2)$

NAIP imagery is primarily acquired in the summer season, so commercial high spatial resolution satellite imagery was used for fall fire data to minimize seasonal biases in the training data. Aerial imagery captured by a commercial aerial firm (ground sample distance = 0.5 m, four spectral bands, 12 bits spectral resolution) shortly after the October 2003 Cedar Fire was used for this purpose.

2.3. Reference data generation

To generate reliable reference data, we classified each airborne orthoimage using three different approaches: unsupervised, NDVI threshold, and ensemble classification, then selected the most accurate classified orthoimage for each burn site. An ISODATA unsupervised classification was conducted on a per-pixel basis using a subset of each orthoimage that encompasses the burned perimeter and its immediate surroundings. Forty clusters were generated and pixels were classified as one of the 40 cluster classes (Figure 2). Each of the 40 classes were manually labelled as either burned or not burned, based on the spectral signatures and spatial distribution of pixels of each cluster class. Cluster classes associated with pixels that were classified as both burned and not burned, based on visual image interpretation, were run through the classifier again. This step helped to ensure that cluster classes with both burned and unburned pixels were separated into burned and non-burned classes.

We used NDVI thresholding to classify the orthoimage subset by determining a threshold for each site where below the threshold is burned and above the threshold is unburned. Each threshold was manually set through visual interpretation.

An ensemble classification was performed by concatenating the two classified images (NDVI thresholded and unsupervised classified), yielding raster maps with values of zero (unburned both classification products), one (burned and unburned), and two (burned both). From this raster map, we created two new raster maps, Ensemble 1 represents the agreement in burned area between both maps (called the high threshold burn) maps. Ensemble 2 represents where only one of the two maps had to classify a pixel as burned for the final classification to be considered burned (low threshold burn).

To assess the accuracy of the reference burned area data, we visually interpreted 215 sample NAIP pixels randomly located within the same orthoimages used for image classification. That sample size was based on an expected accuracy of 85% and an allowable error of 5% (Fitzpatrick-Lins 1981):

$$N = \frac{Z^2(p)(q)}{E^2} \tag{1}$$

where N is the suggested sample size, Z is 2 from the standard normal deviate of 1.96 for the 95% two-sided confidence level, p is the expected percent accuracy, q is 100 minus p, and E is the allowable error. The burned/non-burned interpretation for the sampled pixels was compared to the corresponding pixels for each of the three NAIP image classification products. The classification approach yielding the highest agreement was used to generate reference data for each fire site (Fitzpatrick-Lins 1981). If the highest accuracy product for a site was below an 80% accuracy threshold derived from the 215

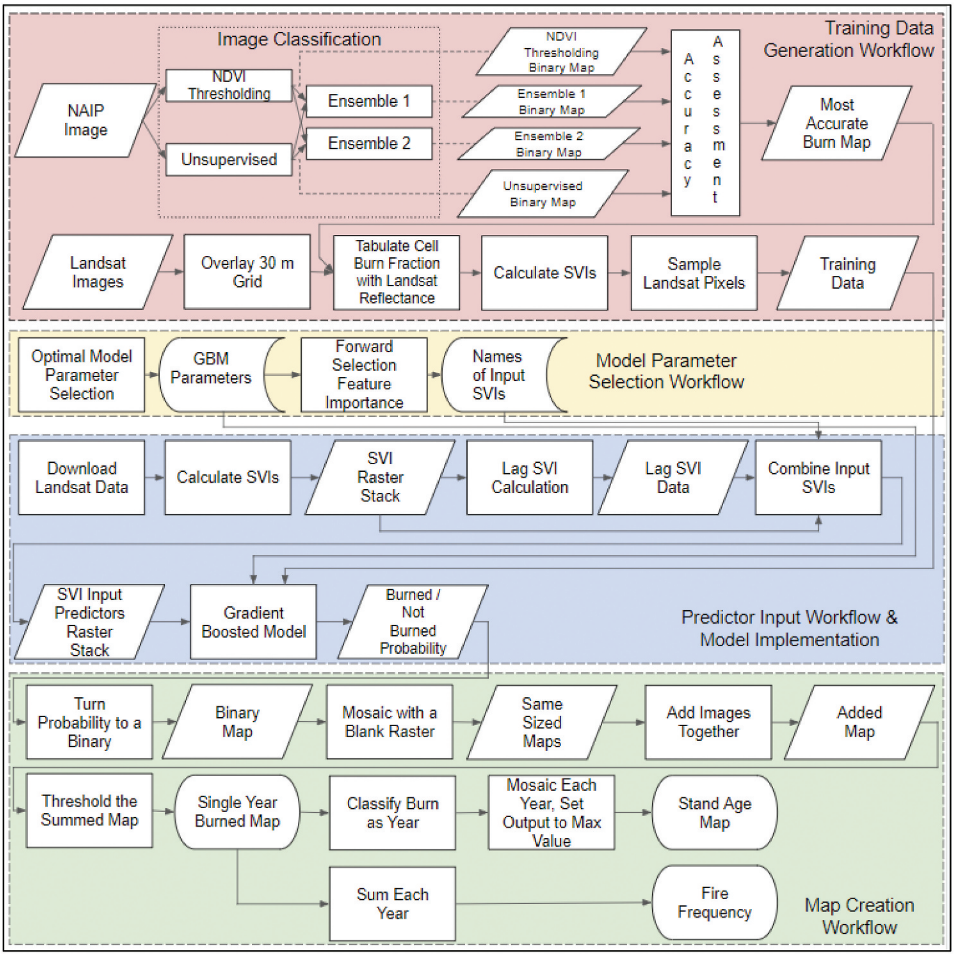

Figure 2. Flow chart of the process of creating burned area maps. Squares are processes, parallelograms are data, and rounded shapes are final products.

sample points, reference data for the site were not included for the training of the Landsat classifier. Figure 2 should be in the mid to later part of this section

The NAIP-derived burned area products and the Landsat data were projected to North American Datum 1983 Universal Transverse Mercator Zone 11 from a World Geodetic System 1984 datum and Albers Equal Area World Geodetic System 84 projection. Sampling grids with 30×30 m cells were overlain on the classified NAIP reference data in order to tabulate fractions of burned within each Landsat ground resolution element associated with Landsat pixels. From this sample grid, outputs were generated with Landsat sub-pixel burn thresholds of 20%, 50%, and 80% and compared in subsequent image classification and accuracy assessment stages. The sub-pixel burn thresholds were chosen to compare the accuracy of low, high, and evenly split sub-pixel burned thresholds in representing the area burned. After a visual inspection of the initial images were processed, we only continued with the 20% sub-pixel burned area threshold because the product appeared to represent the true burned area much more accurately than the products based on 50% and 80% thresholded criteria.

The reference data were divided into two groups, testing and training using 30% of the over quarter million pixels for each of the NAIP derived fires for training data. We ensured that the 30% testing and training data consisted of half burned pixels and half unburned pixels. The training data were further split 50/50 for testing and validation, following the Hawbaker et al. (2020) study. The training data were used to calibrate the classifiers and as testing data to assess the accuracy of multiple burn/no-burn classification products.

The testing data were used for the accuracy assessment of the model. A 50% burned threshold was used for the final accuracy assessment because the Landsat scale reference data with a 50% burned threshold matched the amount of burned area estimated with the NAIP-derived fine-scale reference data much closer than the 20% pixel burn threshold generalized Landsat scale reference data. We found that the 50% burned threshold data matched the NAIP-derived fine-scale reference data much closer than did the 20% pixel burn threshold generalized Landsat scale reference data.

2.4. Image classification

We implemented the GBRM ML model with Landsat data using a Python routine and the Scikit-Learn library to classify and map burned areas. The GBRM is a modified version of one developed by Hawbaker et al. (2020). We used training data for the San Diego study region, whereas Hawbaker et al. (2020) used training data sampled from across the conterminous US with no data from within our study region.

The GBRM requires many parameter inputs to control the structure of the model, three of which we manipulated. Following Hawbaker et al. (2020), we tested different learning rates (0.01, 0.05, and 0.1), splits per tree (one, three, five, and seven), and numbers of trees (2,000, 2,500, 3,000, 3,500, 3,500, 4,000, and 4,500) using a parameter grid where all the combinations were compared. Also replicating the methods of Hawbaker et al. (2020), we used 50% of the 266,664 reference data plots for model training and 50% for model evaluation. The Scikit-Learn library was used for splitting the training data and validating the training data with the testing data, which is the same library Hawbaker et al. (2020) used. A subset of the predictor SVIs (Table 2) was used for the final model. Using the Python Scikit-Learn library, we determined the most important predictors used in the final model using a forward feature input selection and removed the least significant predictors (Figure 2). This reduced computation time and yielded similar results as if all the predictors were used.

We applied the GBRM to classify a stack of Landsat ARD images covering the study area from 2003 to 2020 using the Python Scikit-Learn library. Training pixels were labelled as burned or not burned. We converted the Landsat ARD stack into a stack of SVIs; then a 3-year burn history was created for the selected SVIs. We chose the 3-year lag SVIs through an SVI forward feature selection process detailed in the previous paragraph. The postfire SVIs and the 3-year lag average SVIs were combined into a single raster stack that was inputted into the GBRM model and classified as burned or not burned. The output was a matrix of burned probabilities. To create a binary, burned/not burned raster we compared raster cell burn probability values of 0.5, 0.7, 0.75, and 0.8. We found 0.75 to be the optimal threshold through a visual assessment comparing the thresholded classification to a postfire Landsat scene using the NIR, Red, and Green bands. The binary rasters produced with the different thresholds were similar; however, the higher probability thresholds led to greater omission errors with a minimal decrease in commission errors.

Once the images were classified, they were mosaicked to a common extent. We performed a temporal composite of the classified images for 3 months after a fire took place to account for clouds, cloud shadow, and classification errors. Compositing for 3-month postfire was found to yield the highest classification accuracies based on a visual assessment compared to Landsat data using NIR, Red, and Green bands. For each pixel of the composited image, four classified Landsat images of the 3 months of Landsat images had to be classified as burned yield an output as 'burned' for periods when two Landsat systems were operating – i.e. if Landsat 7 EMT+ and Landsat 8 OLI systems provided imagery. Two of the multi-temporal burn products had to be classified as burned to be output as burned, for periods when one Landsat system was operating – i.e. in 2012 when only Landsat 7 provided data. Those values yielded the lowest commission and omission errors.

MTBS fire boundary vectors with a 250 m buffer were used to mask the extent of each composite burn classification map, to ensure only fires above 404.69 ha were generated (the MTBS minimum fire size). Finally, urban/developed areas and surface water were masked based on urban areas delineated from the California Department of Conservation's Farmland Mapping and Monitoring Program and the surface water information from the National Land Cover Database 2016 map.

An annual composite map of burned areas were generated from stacks of burned area maps for a given year. The map displays pixels as burned or not burned throughout a year. We assembled these annual composited maps to fire maps created by other programs (LBA, FRAP, MTBS) for each year and constructed a fire history map.

2.5. Accuracy assessment

Accuracy of the various burned area products was assessed using the NAIP high spatial resolution reference data. For each of the 30×30 m sample points that represent a Landsat pixel, we assessed whether a hard classification of burn/not burn fractional cover from the testing reference data agreed with that of each Landsat classification product. The threshold used to determine burn/not burned for the accuracy assessment is 50% because the fires generated with this threshold best match the fine-scale reference data in terms of area burned.

Accuracies of the product we created using region-specific training data (referred to hereafter as localized gradient boosted regression model (L-GBRM), FRAP, MTBS, and LBA products) were assessed based on a sample design following Fitzpatrick-Lins (1981). We assumed an *a priori* accuracy of 75%, which is below the lowest agreement percentage between the area burned for the L-GBRM product and the area burned for the reference data (Table 3). The sample size formula yielded 300 sample points for all products with the Landsat scale reference data. To reduce errors from spatial autocorrelation, we created a buffer of 121 m for each of the sample points, which ensured there would be, at minimum, four pixels between sampled pixels (an extra metre was added to reduce possible errors in the program that would cause only three pixels to

Table 3. Area burned for the reference data, L-GBRM, FRAP, MTBS, and LBA products.

Reference fire	Fine-scale Reference (ha)	Generalized Reference (ha)	FRAP (ha)	MTBS (ha)	LBA BC (ha)	LBA BP (ha)	L-GBRM (ha)
Tomahawk Complex	1,406.7	1,530.2	1,916.8	1,760	1,825.2	1,915.7	1,608.3
Pulgas Complex	3,571.9	3,546.2	5,202.4	4,280.1	3,838.7	3,921.8	3,497.3
Cocos	668.9	647.64	738.6	607.1	725.1	848	622.5
Bernardo	435	448.3	466.3	479.3	523.9	564.5	499.1
Cedar	86,875.3	88,238.1	104,160.8	92,773.7	78,839.4	77,423.3	73,700.1

be the minimum). Finally, masking out the urban and water areas, in addition to areas used for training, enabled us to omit pixels in these areas from the accuracy assessment.

To understand the differences in areal size, we also compared burned area extents represented by the four burned area products. This was done for five fires with our fine-scale data above the minimum fire size of 404.69 ha (1000 acres) and based on the total area burned each year. These five fires are the Tomahawk Complex Fire (ignited 5/19/2014), the Pulgas Complex Fire (ignited 5/19/2014), the Cocos Fire (ignited 5/14/2014), the Bernardo Fire (ignited 5/13/2014), and the Cedar Fire (ignited 10/25/2003).

2.6. *Fire history product generation*

Yearly burn maps generated from the L-GBRM product were compared with those created from the extant burned area products (MTBS, FRAP, LBA BC, and LBA BP) to explore their reliability for representing fire histories.

To create a fire history map for the LBA BC data, we downloaded 3 months of post-fire burned classification products for fires with sizes larger than 404.69 ha (1,000 acres) (the MTBS minimum fire size). We projected the LBA BC scenes into the NAD 1983 UTM Zone 11 projection. Then we summed the burn (1) and non-burned (0) pixel values for all images and classified the aggregated image as a binary burn/not burn if a pixel had at least a value of three. The summation value of three burned pixels out of the 3 months of satellite data was found to be the optimal threshold to reduce omission and commission errors.

The LBA BP-based fire history maps were made using the same process as the L-GBRM product. The burn percent product was made into a binary product by applying a 75% burn threshold to each pixel, a threshold chosen through visual interpretation. We used 3 months of the binary burn products to create the composite LBA BP product. A threshold of four burned pixels was used if data from two satellites were available and a threshold of two was used if only one satellite was available.

To create the MTBS burn history maps, we first converted the data to binary burned/not burned to match with the other products. Low, medium, and high severity burns were reclassified as burned, and all other classes as unburned. Water and urban areas were masked.

The FRAP fire history map was created by clipping the map to the study site and removing all fires smaller than 404.69 ha or out of the study period. Then, we applied a vectorized version of the urban and water mask.

3. Results

3.1. GBRM classification

Results from testing different model parameters indicated that when trained with region-specific reference data, the gradient boosted model (GBRM) fit was maximized using a learning rate of 0.05, 2500 trees, and five splits per tree. The model fit with AUC was 0.97. These parameters were then used to select predictor variables. The final predictor variables chosen by having a relative importance threshold of above 0.05 from the predictors, listed in Table 2, are NBR2, dNBR2, VI57, dVI57, SAVI, and NDWI (descriptions found in Table 2 and relative importance in Figure 3). These selected SVI predictors were based on green, red, NIR, SWIR1, and SWIR2 spectral reflectance values. Blue and thermal bands were not used in the final model. More than half of the chosen predictors relied upon SWIR1 and SWIR2 wavelengths (NBR2, dNBR2, VI57, dVI57).

Burned area extent data for the L-GBRM product and for other burned area products are presented in Table 3. The L-GBRM product consistently under-represented burned area by approximately 1.1% relative to the NAIP-derived reference data that was based on the 50% of burn pixels within a 30×30 grid cell. Of the five reference fires that were evaluated, the L-GBRM product had the greatest agreement (98.6%) with the reference data for the Pulgas Complex Fire. The greatest difference in burned area was found for the Cedar Fire, where the L-GBRM predicted 83.5% of the pixels identified as burned from the reference data. The average percent area mapped of the L-GBRM product compared to the reference is 101.0%. Though the perimeter of the Cedar Fire in the L-GBRM product was similar to that of the reference map, portrayal of the interior unburned patches was different (Figure 4). The L-GBRM product represented more and larger unburned patches within the Cedar Fire perimeter than the reference data, as may be expected since the spatial resolution of the Landsat data is much coarser than for the NAIP imagery used to generate the reference map.

The burned area extent represented by the L-GBRM product for the Pulgas Complex Fire was the closest to the reference data out of the five fires (Figure 5). The fire boundary of the product closely aligns with that of the reference map, while the interior unburned

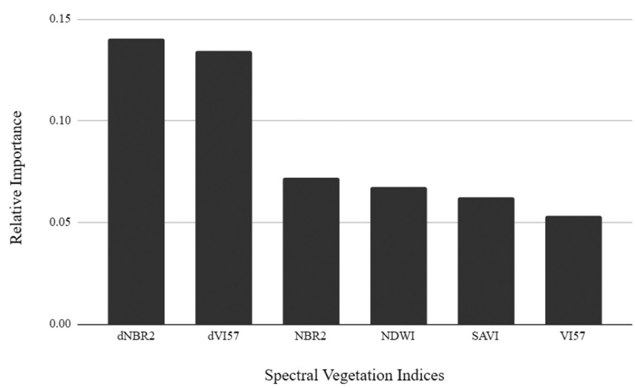

Figure 3. The relative importance of the SVI feature inputs used in the final L-GBRM product. The relative importance values are a percent value of how useful each SVI is to the model if all SVIs were included.

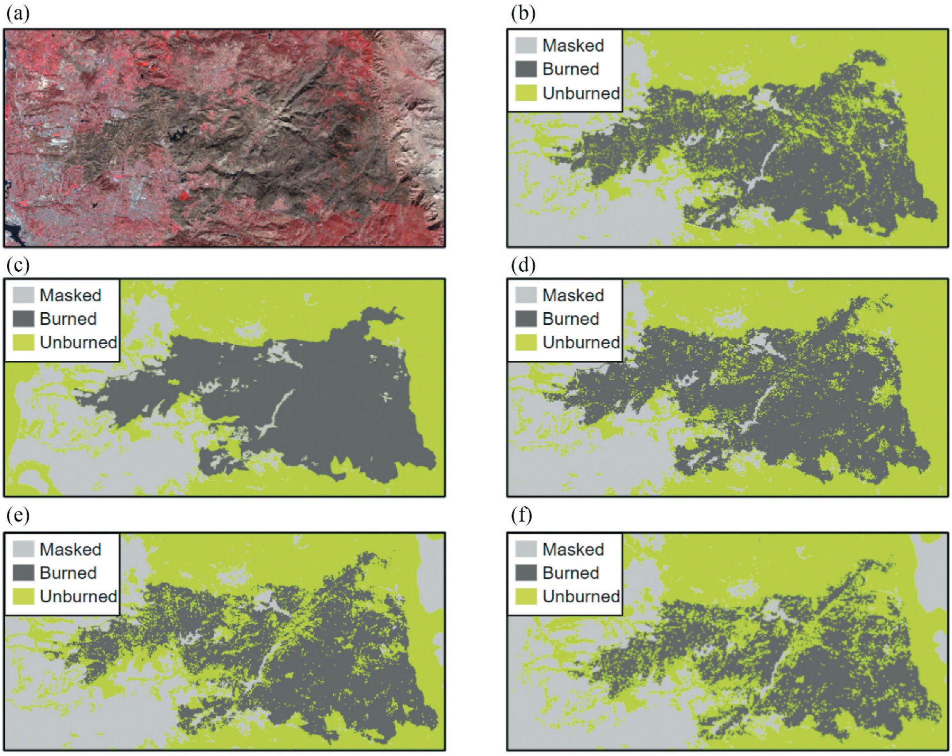

Figure 4. Burned area maps comparison for the cedar fire which burned from 10/25/2003 to 11/5/2003. a. Landsat 5 false-color infrared (4-3-2) image taken on 11/19/2003; b. 30 m reference where 50% or more of the cells burned; c. LBA; D. L-GBRM; e. CAL FIRE; f. MTBS.

patches differed slightly (Figure 5). Figure 4 illustrates the differences in burned area mapped across all burned area products for the Cedar Fire, the least accurate map in terms of area mapped between the L-GBRM product and the reference data, while Figure 5 represents the product for the Pulgas Complex fire.

The L-GBRM products were highly accurate for the five reference fires, with a mean accuracy of 92% (Table 53), based on comparison with the Landsat scale reference data for 300 randomly selected sampling units. The most accurate product is for the Cocos Fire with an accuracy of 97% and the least accurate product is for the Cedar Fire with an accuracy of 81.3%. Since the L-GBRM product tended to underestimate the area burned, the omission errors are higher than the commission errors. The average omission error is 5% whereas the average commission error is 2.9%.

When creating the L-GBRM product, we combined the stack of binary burn images 3 months after a fire, which helped to reduce errors associated with single-date products. This map compositing was the same process used for LBA BP production, so that we could isolate the influence of using region specific vs. global training data. Temporal compositing reduced classification errors stemming from the Landsat 7 scan-line corrector off (SLC-off) issue by combining different images with different locations of missing data from SLC-off, thus mapping areas that burned as burned if the SLC-off was present in a few scenes. However, for some fires, especially before Landsat 8 data were available, L-GBRM products still

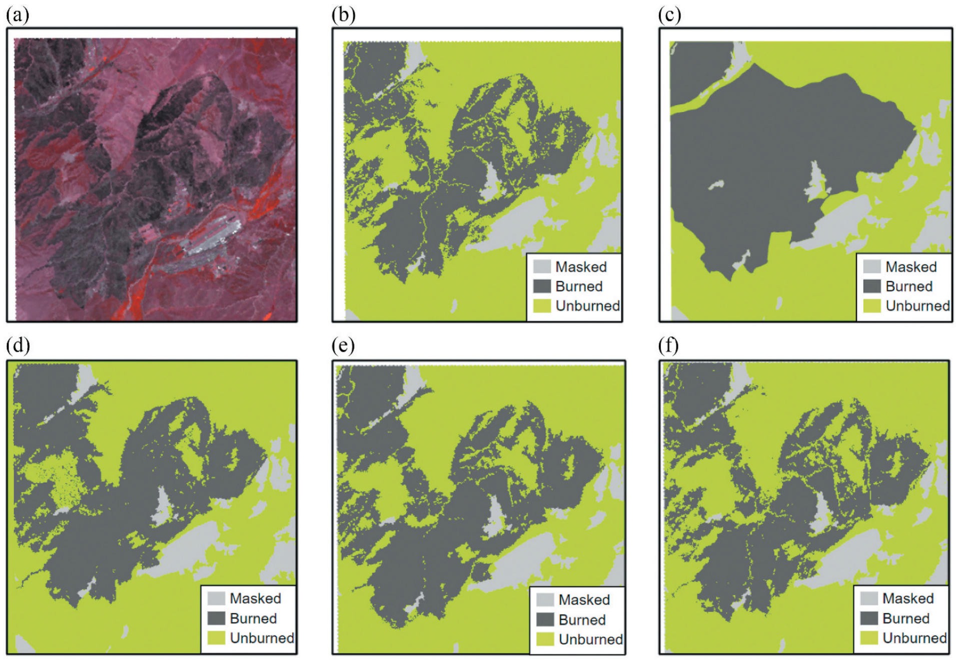

Figure 5. Burned area maps comparison for the pulgas complex fire which burned from 5/14/2014 to 5/21/2014. a. Landsat 8 false-color infrared (5-4-3) image taken on 5/25/2014; b. 30 m reference where 50% or more of the cells burned; c. CAL FIRE; D. MTBS; e. L-GBRM; f. LBA.

contained some striping associated with the SLC-off. Other Landsat-based products (LBA and MTBS) also contain Landsat 7 SLC-off errors, while the FRAP products do not. This could be a factor for why some of the Landsat-based products portray a lower amount of burned area compared to the vector-based FRAP product.

3.2. Comparison of burned area maps

The L-GBRM product is the most accurate (92% overall accuracy) when compared to the reference data derived by aggregation with the 50% threshold. The LBA BC and MTBS products exhibit accuracies of 90.3% and 89.9%, respectively (Table 4). FRAP has the lowest accuracy at 82.7% (Table 4). The MTBS product has the lowest average omission error at 2.6% and the L-GBRM product has the highest at 5%. The L-GBRM product exhibits the lowest average commission error at 2.9% and the FRAP product the highest average commission error rate at 13.6% (Table 4). The FRAP product is the only product to have a higher commission error rate than the omission error rate for a majority of the reference fires.

The LBA BC, which is the more refined product, has a higher average overall accuracy (90.3%) than the LBA BP product (87.3%), which is similar to the L-GBRM in terms of processing (Table 5). The average omission error for the LBA BC is lower at 6.1% than the LBA BP at 8.7% (Table 5). The average commission error of the LBA BC is 3.5% and for the LBA BP is 3.9% (Table 5). The LBA BC product exhibits more burned area than the LBA BC

Table 4. Accuracy metrics for the different burn map products compared to the generalized reference data.

Fire	Product	Overall Accuracy (%)	Commission Error (%)	Omission Error (%)
Tomahawk	FRAP	88.7	10	1.3
	MTBS	91.3	6.3	2.3
	LBA BC	90.3	8.7	1
	LBA BP	88.3	10.7	1
	L-GBRM	95	2.7	2.3
Pulgas	FRAP	73	24.3	2.3
	MTBS	87.7	11.7	0.3
	LBA BC	89.3	5.7	4.7
	LBA BP	89.3	7.3	3
	L-GBRM	90.3	4	5.3
Cocos	FRAP	93.7	6	0.3
	MTBS	91.3	3.3	5.3
	LBA BC	94.7	4.3	1
	LBA BP	89.3	9.3	1.3
	L-GBRM	97	0.7	2.3
Bernardo	FRAP	90	5	5
	MTBS	93	3.7	3.3
	LBA BC	92.7	5.3	2
	LBA BP	90.7	7.3	2
	L-GBRM	96.3	1	2.7
Cedar	FRAP	68.3	22.7	9
	MTBS	86	12	2
	LBA BC	84.3	6.7	9
	LBA BP	79	8.7	12.3
	L-GBRM	81.3	6.3	12.3

Table 5. Comparison of burned area of the mapped fires relative to the fine-scale NAIP reference data. Percentage of agreement between the generalized reference data and different products with the fine-scale reference data.

Fire	Generalized Reference area of NAIP area (%)	FRAP area of NAIP area (%)	MTBS area of NAIP area (%)	LBA BC area of NAIP area (%)	LBA BP area of	L-GBRM area of NAIP area (%)
Tomahawk	108.8	136.3	125.1	129.8	136.2	114.3
Pulgas	99.3	145.6	119.8	107.5	109.8	97.9
Cocos	96.8	110.4	90.8	108.4	126.8	93.1
Bernardo	103.1	107.2	110.2	120.4	129.8	114.7
Cedar	101.6	119.9	106.8	90.8	89.1	84.8
Average	**101.9**	**123.9**	**110.5**	**111.4**	**118.3**	**101.0**

product (Table 5). The L-GBRM is closer to the LBA BC product in terms of area burned and overall accuracy than the LBA BP.

In terms of representing the area burned of the reference fires, the L-GBRM product is the most accurate, representing 101.0% of the area burned (Table 5). The accuracies of the MTBS and LBA BC products are similar, with an average percentage area burned of the reference fires of 110.5% and 111.4%, respectively (Table 5). The LBA BP and FRAP products most overrepresent the area burned, at 118.3% and 123.9%, respectively (Table 5).

Generalized reference data are the 50% fine-scale burn threshold for the 30 m resolution reference data.

The burned areal extents for the different burned area products are shown in Table 6. The L-GBRM product has the lowest total burned area at 165,752.4 ha burned. The LBA product exhibits the next lowest ha burned at 214,983.99. The MTBS product is higher

Table 6. Area burned by product for each year; fires above 404.69 ha (1,000 acres).

Year	FRAP (ha)	MTBS (ha)	LBA (ha)	L-GBRM (ha)
2003	152,136.6	135,391.6	109,215	98,369.1
2004	3,543.2	4,311.2	3,093.3	2,435
2005	1,176.7	1,530.8	989.5	762.8
2006	6,732.2	5,632.7	6,650.3	6,051.2
2007	124,511.1	80,592.9	67,031.6	38,1
2008	0	1,767.5	2,165.9	2,411.5
2009	0	0	0	0
2010	385.5	1,616.9	783.5	702.9
2011	2,418.3	3,865.4	1,492.1	667.8
2012	3,152.7	7,408.6	2,787.3	1,021.9
2013	3,766.7	4,378.9	3,223.2	719
2014	8,324.2	7,329.2	6,912.9	6,177.2
2015	0	0	0	0
2016	3,073.6	1,811.5	2,729.3	1,732.9
2017	2,261.6	2,193.6	1,302.3	585.2
2018	0	0	0	0
2019	0	0	0	0
2020	6,970.5	7,085.3	6,607.9	5,995.9
Average	7,405.9	6,160.8	4,999.6	3,854.7
Sum	318,452.8	264,916.1	214,984	165,752.4

Table 7. Percent of the study area* by number of burns.

Product	0 Burns (%)	1 Burn (%)	2 Burns (%)	3 Burns (%)
FRAP	55.3	37.5	7	0.2
MTBS	62.5	34.2	3.2	0.1
LBA BC	66.3	32.4	1.3	0.1
L-GBRM	73.8	25.4	0.9	0.01

*Doesn't include urban or water land cover types within the study area.

with 264,916.1 ha burned. The FRAP product portrays the largest area of burned area mapped at 318,452.8 ha.

For the total area burned, the FRAP product mapped the most, followed by the MTBS, LBA, and L-GBRM products. The same trend of the most to least area burned continued for areas with successive burns. All the products identify more than a quarter of the study area (excluding water and urban land cover types) as having burned once, and less than 7% of the study area as containing successive fires. The sequence of L-GBRM products portrayed the least number of successive fires, whereas the combination of FRAP products portrayed substantially more area having multiple fires (Table 7).

4. Discussion

The USGS LBA product generated with the gradient boosted classifier that was trained with reference data from across the CONUS is fairly accurate in portraying burned vs non-burned areas. By using region-specific training data, we produced a more accurate burned area map for the study area relative to the LBA product. The L-GBRM product had a 92% average accuracy for the five reference fires. The LBA BC product had an average accuracy of 90% for the five reference fires, while the LBA BP had an 86% average accuracy. The nationwide training data used to generate LBA products did not include reference data for Southern California, likely a factor for the lower accuracy compared to the L-GBRM product.

Further, we experimented with three different percentages of sub-pixel burned thresholds (percentage of a Landsat pixel needed to burn for the entire pixel to be classified as burned) and found 20% to be the most accurate sub-pixel burned threshold compared to 50% and 80%. Before the study, we hypothesized that 50% would be the optimal threshold. Storey, Lee West, and Stow (2021) similarly found that 10% is the optimal burned threshold to consider a pixel burned. Using a lower pixel burned threshold (20%) yielded a product that matches well with the NAIP scale data in terms of area burned and site-specific accuracy. Machine learning models that are trained on training sets with a small variance and/or a small number of training points underfit the data, which leads to inaccurate predictions (Hastie, Friedman, and Tibshirani 2009). The regionally specific training data with the 20% burn pixel threshold had a larger spectral variance and more training points than the 50% and 80% pixel burn threshold datasets. The higher variance and number of training points for the 20% thresholded data most likely lead to the higher accuracy for the model.

We used the same SVIs as input to the GBRM as Hawbaker et al. (2020), which led to 64 feature inputs in total being incorporated into the model. To reduce the number of variables, we determined the relative importance for each variable from the Python Scikit-Learn library and we chose predictor variables with a relative importance metric of above 0.05 to reduce the number of predictors used. By using fewer predictors, we ran the model with less computational cost. The selected SVIs (NBR2, dNBR2, VI57, dVI57, SAVI, and NDWI) mostly relied upon the SWIR1 and SWIR2 bands, along with red, green, and NIR bands. Hawbaker et al. (2020) chose SVIs that mostly relied upon the thermal infrared band, a band not selected for any of our model runs. Since the Hawbaker et al. (2020) study incorporated training data from throughout the CONUS, a diverse set of vegetation was represented, whereas our study area is primarily composed of chaparral, coastal sage scrub, grassland, and oak woodland vegetation. This led to the different SVIs being selected, given that SVIs based on different spectral bands have a varied level of suitability towards detecting specific vegetative conditions (Alonso-Canas and Chuvieco 2015).

The other major difference between the predictors selected for our models and those from the Hawbaker et al. (2020) study is that we only tested scene-level predictors and differenced SVIs, whereas the Hawbaker study used no differenced SVIs, but included scene-level predictors, 3-year lagged mean SVI, and the lag mean standard deviation SVI. Unlike this study, Hawbaker et al. (2020) biased their selection against non-scene level predictors to save computation time, which explains why we ended up with different SVI features being selected.

The MTBS and FRAP programs utilize different methodologies for creating burned area maps, which affect the accuracy of the products differently. MTBS uses Landsat imagery and applies dNBR and RdNBR, thresholding in order to map burn severity for a fire (Eidenshink et al. 2007). The FRAP product was made by visually delineating a perimeter and considering the interior as completely burned. This leads to high commission errors. Such delineation seems to be conducted without consistency between fires (Syphard and Keeley 2016). Both programs map fires primarily by an analyst conducting visual image interpretation and each fire is mapped separately. In contrast, the mapping process for the L- GBRM product is automated and every Landsat scene was fed into the model. One exception to the automation is the selection of the 3 months of Landsat scenes after a fire occurs, which currently uses an analyst to select the dates. However, this procedure is systematic and can be automated in the future. The same cannot be said for the MTBS and FRAP products. An MTBS analyst can select Landsat images that are cloud-free or do not have Landsat 7 SLC-off striping. Conversely, L-GBRM products are based on Landsat images of varying quality, but without the high cost of manual user inputs. The FRAP product leads to inaccuracies for mapping fires due to its creation process. Essentially, a perimeter was drawn around a fire and this leads to over mapping of fires.

The MTBS program maps different severities for a fire. The lowest severity is unburned/low. Other classifications are masked and enhanced greenness. In this study, we considered these classifications to all be unburned. Some of the areas in these classifications may have burned, thus making our MTBS burn area estimate slightly lower than it could be.

Constructing a locally tested and validated burned area model (L-GBRM) resulted in more accurate mapping of unburned patches. The L-GBRM product contains the greatest number and areal extent of and largest unburned patches, followed by the LBA BC, MTBS, and FRAP products. This is likely because a major difference among each of the burned area map products we compared is how they map unburned patches. The FRAP product rarely includes unburned patches, but for a few fires, such as the Pulgas Complex Fire (Figure 5), a single unburned area was represented within the fire perimeter. The MTBS, LBA BC, and L-GBRM products tend to have similar number of unburned patches, though the MTBS product often had fewer than the other two, depending on the fire. These trends can be seen in the Pulgas Complex Fire (Figure 5) and the Cedar Fire (Figure 4). For the Pulgas Complex Fire (Figure 5), the MTBS, LBA BC, and L-GBRM products have a similar number of unburned patches, and the L-GBRM had bigger unburned patches. In the Cedar Fire (Figure 4), the L-GBRM product had similar number of unburned patches but they tend to be larger than the LBA BC product and in greater quantity than the MTBS product. The L-GBRM product has the lowest average commission error and highest average omission error for the five reference fires than any other product.

Land-cover classes are similar but not equally represented for the training data used in the L-GBRM. Fifty-two per cent of the training data represented shrubland cover, 28% grassland, 16% on urban/barren land, and 4% for the combined classes of trees, open water, and crops according to 2013 National Land Cover Database (NLCD) data. After masking out urban and water areas, the study area is comprised of 61% scrub, 22% grassland, 9% built/barren, and 7% trees, water, and crops.

We examined whether the misclassified pixels tended to be associated with particular land-cover types and found that a higher proportion of error occurred in grassland areas of the reference fires. The misclassified pixels within grasslands tended to occur more as commission rather than omission errors. Fewer misclassified pixels occurred shrubland and built areas and were mixed between omission and commission errors. The inaccuracies for the grassland could have been caused by the senescent grasses having a similar spectral signature to bare ground and burned area.

Misclassified pixels tend to particularly occur near the edges of the reference fires, as determined through a visual assessment of the false colour infrared NAIP and Landsat imagery. Most of the edges are on the perimeter of the fire, but some include interior unburned patches. Misclassification near the edges of the fire could have been due to the aggregation of burned/not burned pixels near the edge of the fire scar. A notable exception is the Cedar Fire, where the errors appear to be more randomly distributed.

The different size and shape of the burned area products affect the representation of fire history, often by over-representing burned areas. These errors include the exclusion of interior unburned patches and the inclusion of missing burned areas, often stemming from the Landsat 7 SLC-off issue.

The Landsat 7 SLC-off issue may lead to an underestimation of the true area burned for a product. For example, in years with SLC-off striping for the L-GBRM product, such as in 2007, the area burned is substantially less than for the other products, whereas in years with minimal striping, such as in 2014, the area burned is similar to the other products (Table 6). The MTBS product usually includes burned maps without signs of the Landsat 7 scan line corrector-off (SLC-off) striping but uses images with striping if no others are available. The LBA product is similar to the L-GBRM where several images are added to reduce the effects from striping. Along with the SLC-off issue, the representation of interior unburned patches is another significant reason why the L-GBRM product portrays less area burned than the other burned area products. The extant products do not map the unburned patches as extensively as the L-GBRM product does.

Differences in the burned area product maps occur both in the amount of area burned, and the location of burn patches. The spatial-temporal distribution of burned areas determines which locations have had multiple burns, especially in short succession. For this study, the L-GBRM product had the least amount of overlapping burned area. Excluding water and urban areas, only 0.9% of the study area burned twice, and 0.01% burned three times in the same location between 2003 and 2020 (Table 7). The FRAP product had the most overlapping burns with 7.2% and 0.2% for two and three fires in the same location, respectively (Table 7). Though only a small percentage of the total area burned multiple times (for each of the maps), when studying these areas with multiple burned area overlaps, the exact location of unburned patches can affect the results of a study. Successive fires in the same area affect how various plants and animals repopulate an area after a fire. For example, the time period needed for multiple burns to cause a type conversion from chaparral to grass is commonly debated (Jacobsen, Davis, and Fabritius 2004; Lippitt et al. 2013; Meng et al. 2014). Inaccurate mapping of burned areas causes issues for how we understand the effects of fire.

The L-GBRM product can provide a more refined approach to improving these studies. Mapping with a localized training set leads to a higher accuracy than a globally created model, but it can only be used in the region it was trained on. In order for a state agency to utilize a similar approach, they would need either a single model trained for the entire state or perhaps separate models tuned to each bioregion. Current fire map products that exist are good at determining the approximate location where a fire took place but may have some errors with the exact areas burned. If an agency would like a more refined product, they may have to train one or more models on their areas of interest.

5. Conclusion

The main objective of this study was to explore procedures that might improve upon an extant burned area products for Southern California. This was primarily achieved by evaluating different Landsat burn percentage thresholds used for training a machine learning classifier and by using region specific rather than national-level training data. We used freely available, high spatial resolution NAIP imagery to generate reference data for training and to determine an appropriate percentage of pixel burn threshold for classifying pixels as burned or unburned. Our research builds upon the robust methodology established by Hawbaker et al. (2020) but used regional, fine-scale training data rather than analyst classified Landsat pixels. The very high spatial resolution NAIP imagery also enabled us to create more accurate reference data to which to compare with Landsat-derived burned area products. We found that 20% was the most accurate percentage of pixel burn threshold for selecting training data, while 50% was the optimal pixel burn threshold for generating reference data since its burned area most closely matched the burn area of the NAIP scale data. Since the NAIP orthoimagery is free to the public, this process proves to be affordable to researchers and practitioners. The NAIP data can also be coupled with commercial high spatial resolution satellite imagery to increase the amount and coverage of training data.

Four key results stemming from this study are as follows: (1) the L-GBRM product was the most accurate product for the five reference fires evaluated, (2) region-specific training data achieved a higher accuracy than the nationwide model for the study area, (3) a percentage of Landsat pixel burned threshold of 20% yielded the most accurate prediction of burn vs. non-burn compared to 50% and 80% thresholds, and (4) the L-GBRM product shows less area burned than extant products while exhibiting higher accuracy, suggesting that extant products have high commission errors for burned area in the study area. We found the L-GBRM product to be the most accurate for the five reference fires, and it depicts the least amount of area burned. The L-GBRM product avoids high commission errors primarily by reliably mapping unburned patches within the interior of fire perimeters. As the results from this study demonstrate, the current LBA algorithm described in Hawbaker et al. (2020) can be modified with region-specific training data to produce more accurate burned area maps for Southern California. A low pixel burn threshold was found to lead to an accurate burned area; future studies can build on this finding and compare other pixel burn thresholds around 20% to further refine the model.

Several limitations associated with availability of imagery and inability to control for all aspects of burn area mapping procedures associated with different products influenced the results of this study. One limitation is the amount of available training data is an issue caused by a relatively small quantity of fire events and an even smaller number of large fires within the study region and period. There are only 15 images of fires collected from the NAIP program for the study area. This was limited in part because we chose to only collect data for fires that were captured within 21 days of a fire event to reduce the probability of ash and char blowing away, which was an issue for grass-fuelled fires. Of the 15 NAIP collected fires, only four were larger than 405 ha, the minimum fire size of MTBS, which has the largest minimum fire size of the products we compared. The lack of large fires may lead to an overtraining on smaller fires, which possibly have different fire effects from the larger ones. Also, smaller sized fires have fewer pixels from which to select for training and machine learning models require large amounts of data. The lack of large fires led to only four acceptable NAIP image sets being available, and all four fires occurred within the same month of the same year. The temporal bias was both seasonal and yearly. The NAIP program collects imagery in the study period every 2 to 3 years during the months of April to September, so the data were primarily from late spring through summer, resulting in a lack of seasonal diversity. Fall fires are not included, but they burn roughly the same amount of area and are generally much larger than summer and spring fires (Jin et al. 2015).

Future research could emphasize testing other sources of training data and possibly expanding the training region to greater Southern California, to see if the accuracy of the burned area product can be improved. Further, Hawbaker et al. (2020) created two data products, the LBA Burn Probability (BP) and the LBA Burn Classification (BC). The LBA BC was more accurate than the LBA BP in terms of areal representation and in the accuracy assessment, compared to the Landsat scale data. However, the L-GBRM is more accurate than each for the study site and had similar processing steps as the LBA BP product.

Acknowledgements

USFS and CWC grants; We would like to thank Dr. Todd Hawbaker for meeting with us and providing scripts to help us understand how the USGS model works. His advice and knowledge was invaluable for help in this research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the California Climate Investments Initiative [CCRP0061]; U.S. Forest Service [G00012222, G00014700].

ORCID

Douglas Stow http://orcid.org/0000-0001-5246-7073
Atsushi Nara http://orcid.org/0000-0003-4173-7773

References

Alonso-Canas, I., and E. Chuvieco. 2015. "Global Burned Area Mapping from ENVISAT- MERIS and MODIS Active Fire Data." *Remote Sensing of Environment* 163:140–152. https://doi.org/10.1016/j.rse.2015.03.011.

Bachelet, D., K. Ferschweiler, T. Sheehan, and J. Strittholt. 2016. "Climate Change Effects on Southern California Deserts." *Journal of Arid Environments* 127:17–29. https://doi.org/10.1016/j.jaridenv.2015.10.003.

Black, C., Y. Tesfaigzi, J. A. Bassein, and L. A. Miller. 2017. "Wildfire Smoke Exposure and Human Health: Significant Gaps in Research for a Growing Public Health Issue." *Environmental Toxicology and Pharmacology* 55:186–195. https://doi.org/10.1016/j.etap.2017.08.022.

Bowman, D. M. J. S., J. K. Balch, P. Artaxo, W. J. Bond, J. M. Carlson, M. A. Cochrane, C. M. D'Antonio, et al. 2009. "Fire in the Earth System." *Science* 324 (5926): 481–484. https://doi.org/10.1126/science.1163886.

Brown, T. J., B. L. Hall, C. R. Mohrle, and H. J. Reinbold. 2002. "Coarse Assessment of Federal Wildland Fire Occurrence Data." *Program for Climate, Ecosystem and Fire Applications* 35.

CAL FIRE. 2023. Stats & Events [Gov] CA. https://www.fire.ca.gov/stats-events/.

Chuvieco, E., F. Mouillot, G. R. van der Werf, J. San Miguel, M. Tanase, N. Koutsias, M. García, et al. 2019. "Historical Background and Current Developments for Mapping Burned Area from Satellite Earth Observation." *Remote Sensing of Environment* 225:45–64. https://doi.org/10.1016/j.rse.2019.02.013.

Daldegan, G. A., D. A. Roberts, and F. D. F. Ribeiro. 2019. "Spectral Mixture Analysis in Google Earth Engine to Model and Delineate Fire Scars Over a Large Extent and a Long Time-Series in a Rainforest-Savanna Transition Zone." *Remote Sensing of Environment* 232:111340. https://doi.org/10.1016/j.rse.2019.111340.

Eidenshink, J., B. Schwind, K. Brewer, Z.-L. Zhu, B. Quayle, and S. Howard. 2007. "A Project for Monitoring Trends in Burn Severity." *Fire Ecology* 3 (1): 3–21. https://doi.org/10.4996/fireecology.0301003.

Fitzpatrick-Lins, K. 1981. "Comparison of Sampling Procedures and Data Analysis for a Land-Use and Land-Cover Map." *Photogrammetric Engineering* 47 (3): 343–351.

Franklin, J. 1998. "Predicting the Distribution of Shrub Species in Southern California from Climate and Terrain-Derived Variables." *Journal of Vegetation Science* 9 (5): 733–748. https://doi.org/10.2307/3237291.

Grady, L. 2005. "Multilabel Random Walker Image Segmentation Using Prior Models." *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 763–770, *1*. https://doi.org/10.1109/CVPR.2005.239.

Griffith, G. E., J. M. Omernik, D. W. Smith, T. D. Cook, E. Tallyn, K. Moseley, and C. B. Johnson. 2016. "Ecoregions of California [USGS Numbered Series]." In *Ecoregions of California*, Vol. 2016–1021. U.S. Geological Survey. https://doi.org/10.3133/ofr20161021.

Hastie, T., J. Friedman, and R. Tibshirani. 2009. *The Elements of Statistical Learning: Data Mining, Inference, Prediction*. 2nd ed. New York, NY: Springer.

Hawbaker, T. J., M. K. Vanderhoof, Y.-J. Beal, J. D. Takacs, G. L. Schmidt, J. T. Falgout, B. Williams, et al. 2017. "Mapping Burned Areas Using Dense Time-Series of Landsat Data." *Remote Sensing of Environment* 198:504–522. https://doi.org/10.1016/j.rse.2017.06.027.

Hawbaker, T. J., M. K. Vanderhoof, G. L. Schmidt, Y.-J. Beal, J. J. Picotte, J. D. Takacs, J. T. Falgout, and J. L. Dwyer. 2020. "The Landsat Burned Area Algorithm and Products for the Conterminous United States." *Remote Sensing of Environment* 244:111801. https://doi.org/10.1016/j.rse.2020.111801.

Jacobsen, A. L., S. D. Davis, and S. L. Fabritius. 2004. "Fire Frequency Impacts Non- Sprouting Chaparral Shrubs in the Santa Monica Mountains of Southern California." *Proceedings 10th MEDECOS Conference*, April 25–May 1, 2004, 6.

Jin, Y., M. L. Goulden, N. Faivre, S. Veraverbeke, F. Sun, A. Hall, M. S. Hand, S. Hook, and J. T. Randerson. 2015. "Identification of Two Distinct Fire Regimes in Southern California: Implications for Economic Impact and Future Change." *Environmental Research Letters* 10 (9): 094005. https://doi.org/10.1088/1748-9326/10/9/094005.

Keeley, J. E., and A. D. Syphard. 2018. "Historical Patterns of Wildfire Ignition Sources in California Ecosystems." *International Journal of Wildland Fire* 27 (12): 781. https://doi.org/10.1071/WF18026.

Kolden, C. A., J. A. Lutz, C. H. Key, J. T. Kane, and J. W. van Wagtendonk. 2012. "Mapped Versus Actual Burned Area within Wildfire Perimeters: Characterizing the Unburned." *Forest Ecology & Management* 286:38–47. https://doi.org/10.1016/j.foreco.2012.08.020.

Kolden, C. A., A. M. S. Smith, and J. T. Abatzoglou. 2015. "Limitations and Utilisation of Monitoring Trends in Burn Severity Products for Assessing Wildfire Severity in the USA." *International Journal of Wildland Fire* 24 (7): 1023. https://doi.org/10.1071/WF15082.

Kolden, C. A., and P. J. Weisberg. 2007. "Assessing Accuracy of Manually-Mapped Wildfire Perimeters in Topographically Dissected Areas." *Fire Ecology* 3 (1): 22–31. https://doi.org/10.4996/fireecology.0301022.

Leland, T., and J. Lutz. 2011. "Sustainable Fire: Preserving Carbon Stocks and Protecting Air Quality As Sierra Nevada Forests Warm." *Park Science* 28 (1): 76.

Lippitt, C. L., D. A. Stow, J. F. O'Leary, and J. Franklin. 2013. "Influence of Short-Interval Fire Occurrence on Post-Fire Recovery of Fire-Prone Shrublands in California, USA." *International Journal of Wildland Fire* 22 (2): 184. https://doi.org/10.1071/WF10099.

Loboda, T., K. J. O'Neal, and I. Csiszar. 2007. "Regionally Adaptable dNBR-Based Algorithm for Burned Area Mapping from MODIS Data." *Remote Sensing of Environment* 109 (4): 429–442. https://doi.org/10.1016/j.rse.2007.01.017.

Meng, R., P. E. Dennison, C. M. D'Antonio, M. A. Moritz, and B. Bond-Lamberty. 2014. "Remote Sensing Analysis of Vegetation Recovery Following Short-Interval Fires in Southern California Shrublands." *PLOS ONE* 9 (10): e110637. https://doi.org/10.1371/journal.pone.0110637.

Omernik, J. M., and G. E. Griffith. 2014. "Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework." *Environmental Management* 54 (6): 1249–1266. https://doi.org/10.1007/s00267-014-0364-1.

Roy, D. P., L. Boschetti, and S. N. Trigg. 2006. "Remote Sensing of Fire Severity: Assessing the Performance of the Normalized Burn Ratio." *IEEE Geoscience & Remote Sensing Letters* 3 (1): 112–116. https://doi.org/10.1109/LGRS.2005.858485.

Sicard, P., Y. O. Khaniabadi, S. Perez, M. Gualtieri, and A. De Marco. 2019. "Effect of O3, PM10 and PM2.5 on Cardiovascular and Respiratory Diseases in Cities of France, Iran and Italy." *Environmental Science and Pollution Research* 26 (31): 32645–32665. https://doi.org/10.1007/s11356-019-06445-8.

Silva, J. M. N., J. M. C. Pereira, A. I. Cabral, A. C. L. Sá, M. J. P. Vasconcelos, B. Mota, and J.-M. Grégoire. 2003. "An Estimate of the Area Burned in Southern Africa During the 2000 Dry Season Using SPOT-VEGETATION Satellite Data: AREA BURNED in SOUTHERN AFRICA 2000." *Journal of Geophysical Research Atmospheres* 108 (D13): n/a–n/a. https://doi.org/10.1029/2002JD002320.

Sparks, A. M., L. Boschetti, A. M. S. Smith, W. T. Tinkham, K. O. Lannom, and B. A. Newingham. 2015. "An Accuracy Assessment of the MTBS Burned Area Product for Shrub–Steppe Fires in the Northern Great Basin, United States." *International Journal of Wildland Fire* 24 (1): 70. https://doi.org/10.1071/WF14131.

Storey, E. A., K. R. Lee West, and D. A. Stow. 2021. "Utility and Optimization of LANDSAT-Derived Burned Area Maps for Southern California." *International Journal of Remote Sensing* 42 (2): 486–505. https://doi.org/10.1080/01431161.2020.1809741.

Syphard, A. D., T. J. Brennan, and J. E. Keeley. 2019. "Extent and Drivers of Vegetation Type Conversion in Southern California Chaparral." *Ecosphere* 10 (7). https://doi.org/10.1002/ecs2.2796.

Syphard, A. D., T. J. Brennan, J. E. Keeley, and T. Vaclavik. 2019. "Drivers of Chaparral Type Conversion to Herbaceous Vegetation in Coastal Southern California." *Diversity & Distributions* 25 (1): 90–101. https://doi.org/10.1111/ddi.12827.

Syphard, A. D., and J. E. Keeley. 2016. "Historical Reconstructions of California Wildfires Vary by Data Source." *International Journal of Wildland Fire* 25 (12): 1221. https://doi.org/10.1071/WF16050.

Tanase, M. A., M. A. Belenguer-Plomer, E. Roteta, A. Bastarrika, J. Wheeler, Á. Fernández- Carrillo, K. Tansey, et al. 2020. "Burned Area Detection and Mapping: Intercomparison of Sentinel-1 and Sentinel-2 Based Algorithms Over Tropical Africa." *Remote Sensing* 12 (2): 334. https://doi.org/10.3390/rs12020334.

Tansey, K., J.-M. Gregoire, P. Defourny, R. Leigh, J.-F. Pekel, E. van Bogaert, and E. Bartholome. 2008. "A New Global Multi-Annual 2000-2007 Burnt Area Product at 1 Km.Pdf." *Geophysical Research Letters* 35 (L01401): 6. https://doi.org/10.1029/2007GL031567.

Vanderhoof, M. K., N. Fairaux, Y.-J. G. Beal, and T. J. Hawbaker. 2017. "Validation of the USGS Landsat Burned Area Essential Climate Variable (BAECV) Across the Conterminous United States." *Remote Sensing of Environment* 198:393–406. https://doi.org/10.1016/j.rse.2017.06.025.

Westerling, A. L., A. Gershunov, T. J. Brown, D. R. Cayan, and M. D. Dettinger. 2003. "Climate and Wildfire in the Western United States." *Bulletin of the American Meteorological Society* 84 (5): 595–604. https://doi.org/10.1175/BAMS-84-5-595.