

FULL RESEARCH ARTICLE

Machine learning to understand patterns of burn severity from the SCU Lightning Complex Fires of August 2020

CHRISTOPHER POTTER^{1*} AND OLIVIA ALEXANDER²

¹ *Casa Systems 2100, 33 Tait Avenue, Los Gatos, CA 95030, USA*

² *San José State University, One Washington Square, San José, CA 95195, USA*

**Corresponding Author: christopher@casa2100.com*

The SCU Lightning Complex Fire started on 16 August 2020 and burned more than 395,000 acres of woodlands and grasslands in six California counties. Satellite images of pre-fire green vegetation biomass from both 2020 springtime (moist) and summertime (drier) periods, along with slope and aspect were used as predictors of burn severity patterns on the SCU Complex landscape using machine learning algorithms. The main finding from this analysis was that the overall burn severity patterns of the SCU Complex fires could be predicted from pre-fire vegetation biomass, slope, and aspect model input variables with high accuracies of between 50% and 80% using Random Forest machine learning techniques. The August and April biomass cover variables had the highest feature importance values. It can be concluded that the amount of dry biomass present at a given location was essential to predict how severely and completely the 2020 fires burned the vegetation cover and surface soils across this landscape.

Key words: burn severity, machine learning, NDVI, random forest, wildfire

The SCU Lightning Complex Fires started on 16 August 2020 as a result of hundreds of lightning strikes in the Diablo Range of northern California, and burned within six counties: Alameda, Contra Costa, Santa Clara, San Joaquin, Stanislaus, and Merced (CALFIRE 2020). A total of 222 structures were destroyed in these fires. The SCU Complex Fires were declared nearly 100% contained on 10 September 2020, after burning over an estimated 160,498 ha (396,600 acres; WERT 2020) and becoming the third-largest wildfire recorded in California's modern history.

In the wake of a disaster of this magnitude, resource managers require timely information about burn severity patterns, for purposes ranging from addressing immediate hazards such as landslides and tree falls, to monitoring runoff of chemicals in waterways, and managing long-term post-fire recovery of watersheds and woodland stands (WERT 2020). The use of multispectral (remotely sensed) burn severity metrics has become common across

North American forests (French et al. 2008). The normalized burn ratio (NBR; Key and Benson 2006) from satellite imagery was developed expressly to assess post-fire changes in reflectance of healthy vegetation, soils, and soil moisture (Potter 2016).

There have been several noteworthy modeling studies to predict burn severity levels from wildfires. For instance, Whitman et al. (2018) found that pre-fire stand structure and composition, topography, and fire weather at time of burning were the best predictors of burn severity from boreal forest fires. Wetlands burned less severely than uplands, and open stands with high basal areas showed lower burn severity than in upland vegetation stands. Burn severity has been shown to be a product of pre-fire vegetation conditions and fuel loads (Boucher et al. 2016; Lydersen et al. 2017) and topography (Krawchuk et al. 2016).

Topographic aspect can influence the amount of solar radiation and moisture availability on a hillslope, which in turn can directly influence fire behavior, as well as indirectly through the control over differences in vegetation composition and biomass fuel density (Estes et al. 2017). Steeper slopes may also lead to greater preheating of fuels and increased rate of spread when fire is moving upslope (Estes et al. 2017). Localized weather conditions related to topography, such as wind speeds and surface temperatures during the periods of intense burning, can strongly influence fire behavior and combustion rates (FCFDG 1992; Krawchuk et al. 2016). Along these lines, Potter (2017) reported that seasonal climate conditions (maximum air temperatures and low moisture) at the time of ignitions of large wildfires on the central and southern California coasts were significant controllers of the total area burned at high severity and the edge complexity of high severity burn patches on the fire landscape.

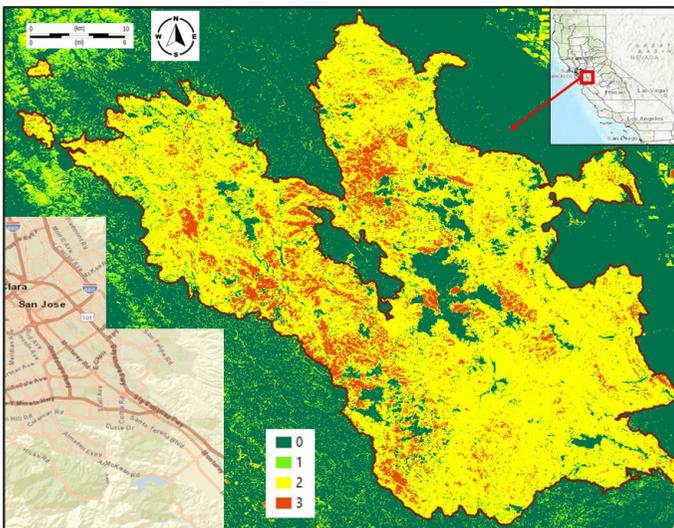


Figure 1. Map of Landsat dNBR burn severity classes resulting from the 2020 SCU Complex fires. Color legend of burn severity class labels is as follows: 0 = no burn, 1 = low burn, 2 = moderate burn severity, and 3 = high burn severity. Inset map of major highways in the Santa Clara Valley shows actual proximity of the fire boundary outline to San Jose urban areas.

The purpose of this study was to describe and explain the geographic variability in burn severity classes resulting from the 2020 SCU Complex Fires. Plant communities that burned in the SCU Complex wildfires included coast live oak (*Quercus agrifolia*), blue oak (*Q. douglasii*), valley oak (*Q. lobata*), and black oak (*Q. kelloggii*) woodlands, plus chamise (*Adenostoma fasciculatum*) shrublands, Diablan sage scrub, non-native annual grassland of brome grass, and native perennial grassland (White 1966; Fry 2008; Stahle et al. 2013). The main objective of this study was to characterize the relative importance of spatially

mapped landcover attributes, namely topography and vegetation cover density, as controls on burn severity in this extreme fire event. Satellite images of pre-fire green vegetation density in terms of relative biomass from both 2020 springtime (moist) and summertime (drier) periods, along with slope and aspect were used as predictors of August–September 2020 burn severity classes in a machine learning approach. For mapping of fire fuel in terms of vegetation biomass amounts prior to the 2020 fires in California, we have analyzed the Landsat normalized difference vegetation index (NDVI) as a surrogate for burnable biomass, as has been done in similar studies of wildfire mapping (Radočaj et al. 2021).

METHODS

Satellite Image Data

We calculated the SCU Complex NBR index from satellite image dates, both pre- and post-August of 2020, from the near infrared (NIR; 0.85–0.88 μm) and shortwave infrared (SWIR; 1.57–1.65 μm) bands of the Landsat 8 sensor Collection 2 images at 30-m pixel size, according to the equation:

$$NBR = (NIR - SWIR)/(NIR + SWIR)$$

We differenced pre-fire (24 July 2020) and post-fire (26 September 2020) NBR images to generate a dNBR map product for the SCU Complex Fires. Burn severity classes of low, moderate, and high levels can cover a dNBR value range of –500 to 12,000 over burned land surfaces. Positive dNBR values represent a decrease in vegetation cover and a higher burn severity class, while negative values represent an increase in live vegetation cover following the fire event.

We defined four classes of burn severity for this study as no burn (0-NB) at $dNBR < 500$, low burn severity (1-LBS) at $dNBR > 500$ and ≤ 1000 , moderate burn severity (2-MBS) at $dNBR > 1000$ and ≤ 5000 , and high burn severity (3-HBS) at $dNBR > 5000$ (Potter 2016). These classification levels generally followed the burn severity thresholds determined by Miller and Thode (2007) based on a composite burn index (CBI) for California forests. The CBI was developed to assess on-the-ground fire effects on plants and soils (i.e. burn severity) by sampling over strata of the vegetation remaining post-fire: litter, low shrubs, small trees, tall shrubs and sapling trees, intermediate trees, and tall trees.

The Landsat 8 Collection 2 normalized difference vegetation index (NDVI) provides consistent spatial and temporal profiles of relative vegetation canopy biomass (Verbesselt et al. 2010) according to the equation:

$$NDVI = (NIR - Red)/(NIR + Red)$$

resulting in values between –1.0 and 1.0 NDVI units. We multiplied NDVI values by 10,000 to preserve decimal places in integer file storage. Low values of NDVI (near 0.1) indicate barren land cover whereas high values of NDVI (above 0.8) indicate dense canopy cover. NDVI has been proven as an accurate index of herbaceous green cover in grasslands of California and can be converted into seasonal herbaceous biomass (g carbon/m²) each year (Potter 2014a). We obtained Landsat 8 images from both 3 April 3 and 9 August 2020 for cool season (April) and warm season (August) pre-fire NDVI map layers.

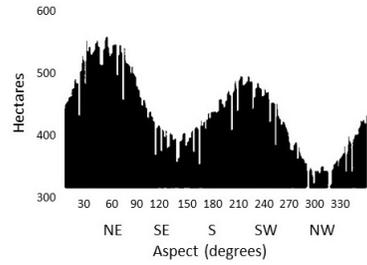
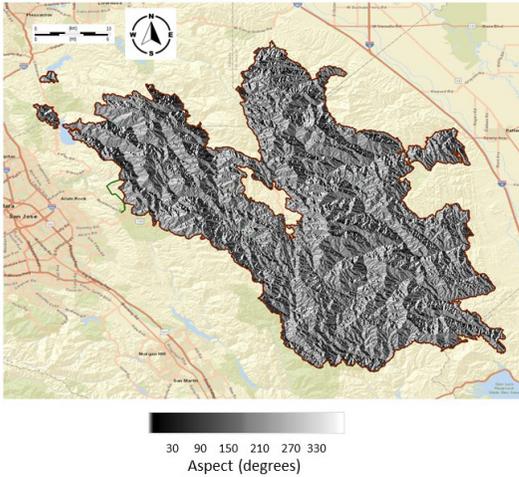


Figure 2. (a) Map of the aspect of the hillslopes and (b) Distribution of land area by hillslope aspect across the SCU Complex burned area.

Slope and Aspect Layers

Digital layers for slope and aspect for the SCU Complex burned area were determined at 30-m spatial resolution from the United States Geological Survey (USGS) National Elevation Dataset (NED) using the ArcGIS Spatial Analyst Toolbox (ESRI, 2021). This tool uses a 3 by 3 cell moving window to process the digital elevation data into continuous gridded slope and aspect values.

Machine Learning and Statistical Analysis

To predict dNBR burn severity classes for the SCU Complex fire area from NDVI, slope, and aspect spatial layers, we used the Scikit-learn machine learning library for the Python programming language (Pedregosa et al. 2011). Scikit-learn features various classification and regression algorithms including decision trees, support vector machines, Random Forest, and k-means nearest neighbor, all operating with the Python libraries NumPy and SciPy.

Among all the Scikit-learn machine learning methods, we selected the Random Forest method from Breiman (2001) for this analysis, because it has the ability to perform both classification and regression prediction. Random forests are an improved extension on classification and regression trees (CART) (Liaw and Weiner 2018). Moreover, Random Forest methods have the following advantages: handles categorical predictors naturally, computationally simple to fit, has no formal distributional assumptions, and performs automatic variable selection.

The Random Forest model operated as follows: first, the algorithms computationally “grow” a forest of *n* trees. For each tree from 1 to *n*, a sample of size *N* is taken from the dataset with replacement (bootstrap) to grow the tree. A selection of *m* variables, independently for each node tree, is made, and the tree is split at each node by determining which variable will create the highest proportion of homogenous classification using Gini impurity. Trees are grown until the nodes can no longer be split, unless otherwise specified

with a `max_depth` variable to prevent overfitting of the data. For classification, majority voting is used to generate aggregated predictions of the ntree trees. For model training, 70% of the data points are selected while the remaining 30% of data points are split to create the “testing” data, used to unbiasedly evaluate the model’s fit on the training dataset. The error rate of all the OOB predictions is the OOB error rate of the random forest result.

Random forest can also compute the importance of variables in two different ways. For this study and related classification problems, Gini criterion impurity can be used to measure variable importance (Pedregosa et al. 2011). For a given tree, the Gini variable importance for a particular variable of interest is the weighted average of the decrease in the Gini criteria impurity of the splits based on this variable. This is averaged over the ntree trees in the forest to get the Gini importance for the forest. The other variable importance calculation is called permutation importance, which is based on predictive accuracy. The testing error rate is computed from both a data set obtained from permuting the values of a particular variable of interest in the testing data and the original testing data. The difference between these two testing error rates gives the permutation variable importance.

Output statistics from the Random Forest model were generated as a classification matrix report including class prediction accuracies (as seen in Fig. 5), and as the F1 score for each predicted class, which can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its highest possible value at 1, indicating perfect precision and recall, and has a lowest value at 0 (Pedregosa et al. 2011). The relative contribution of precision and recall to the F1 score are equal. The F1 score is also known as the Sørensen–Dice coefficient. The feature importance of each predictor variable in the model is also captured to understand the weight of each variable in predicting the overall burn severity classes.

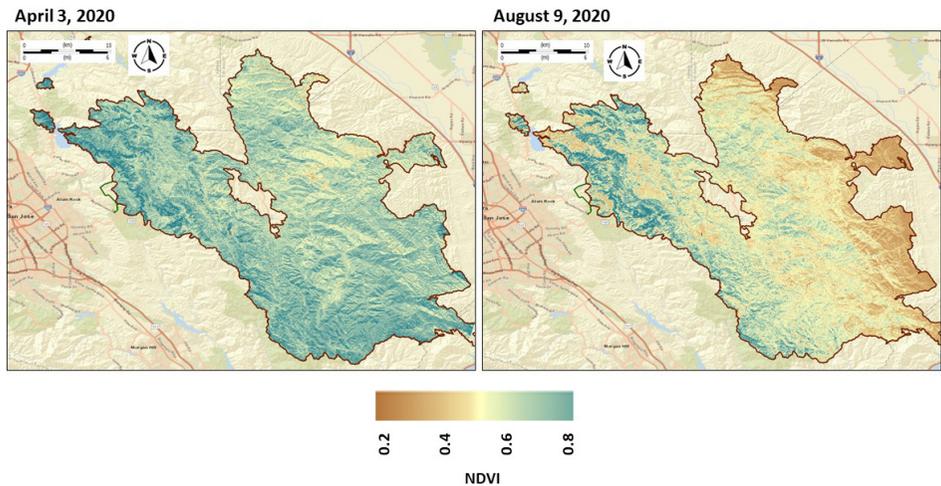


Figure 3. Pre-fire maps of NDVI in April and August of 2020 across the SCU Complex burned area.

As a large image data set with multiple variables, the entire SCU Complex burned area proved to be too large (at 1.7 million row entries) to run all at one in Scikit-learn. Therefore, we first tested a random sampling approach into smaller image subsets, about one-third the size of the entire burned area, which would still allow one to make strong statistical inferences about the entire dataset. Several down-sampling methods were tested to compare their performance to the random subsets sampling approach. Subsequently,

we tested down-sampling methods, including Near Miss and Edited Nearest Neighbor, to compare their performance to the random subsets sampling approach.

In cases such the SCU Complex fires with a skewed burn severity distribution among classes, data sampling methods can be used to compensate for a large class imbalance. Random down-sampling (over or under) methods generally show improved overall results in machine learning applications (Leevy et al. 2018). As a result of this type of down-sampling, the majority burn severity class should not take over the other classes during the training process, and all classes will be well-represented by the decision function.

The Near Miss undersampling method selects all data from the minority class and then focuses on sampling from the larger class(es). The algorithm computes the distance between all data in the majority class to the data in the minority class then selects pixel datapoints of the majority class with the smallest distance to the minority class(es). In this case, the burn severity class with the smallest number of pixels in the SCU Complex area, low burn severity (LBS 1), was used to set the maximum number of pixels for sampling of all the other burn severity classes for new Random Forest runs. Therefore, all data for class 1 is selected and burn severity classes 0, 2, and 3 are sampled equally using the Near Miss method.

Edited Nearest Neighbors (ENN) was also tested as an undersampling method. This technique focuses on removing noisy and ambiguous data on the class boundaries to address the class imbalance and also make the distinction between classes clearer. The data in the majority class that are misclassified as the minority class are removed and those correctly classified are selected. In addition, the data in the minority class that are misclassified have their nearest neighbors from the majority class deleted to reduce classification ambiguity. This method, unlike the Near Miss undersampling method, does not create an equal amount of pixel data across each class, but rather attempts to select the least ambiguous dataset to increase prediction accuracy.

RESULTS

Landsat Burned Severity Patterns

The SCU Complex burned severity classes mapped using 2020 Landsat NBR images (Fig. 1) were bounded to the north by the Livermore Valley, to the east by the San Joaquin Valley, to the south by the Pacheco-Pass Highway, and to the west by the Santa Clara Valley and the City of San Jose. A total of 1,454 km² (359,220 acres) was estimated by the Landsat dNBR to have burned in low to high severity classes during this lightning complex fire. The distribution of land area among burned severity classes was low severity 5%, moderate severity 83%, and high severity 12%. Taken as a whole, the dNBR data set for SCU Complex burned area classes can be described as highly imbalanced, because the four burn severity classes were unequally represented across the study landscape.

Most of the SCU Complex burned area was mapped on steep terrain with an average slope gradient of over 35% (WERT 2020). The burned area exhibited 1242 m (4,076 ft) of vertical relief, ranging from about 90 m (300 ft) above mean sea level (amsl) in deep, lower elevation canyons, up into the mountainous eastern sections of the burned area at 1334 m (4,376 feet) elevation. Aspect of the hillslopes across SCU Complex burned area was skewed slightly to more northeastern-facing slopes than to southeastern- and southwestern-facing slopes (Fig. 2).

Maps of pre-fire NDVI in 2020 across the SCU Complex burned area showed the

patterns in the density of green plant cover during the relatively cool season (April) and again during the warmer season (August), including areas where evergreen oak woodland and shrubland cover predominated (Fig. 3). These oak woodland and shrub-covered watersheds were most extensive in the northwestern portions of the SCU Complex burned area. Locations where annual grassland plant cover predominated are identified by high NDVI (> 0.4) in April and lower NDVI (< 0.4) in August. These herbaceous plant-covered watersheds were most extensive in the eastern margins of the SCU Complex burned area. Judging from the cool-to-warm season transition in NDVI shown in Fig. 3, the majority of vegetation cover that burned in late August of 2020 had dried out and turned from green to brown at the time of ignition.

Correlation Matrix

The correlation matrix results for the four predictor variables and the predicted burn severity classes (Fig. 4) showed that the only significant (linear) correlation detected was between NDVI in April and in August. This NDVI correlation at $R = 0.57$ was not unexpected, because areas with evergreen woodland and shrub cover do not change in live canopy cover as much as grass-covered areas and in grazed rangelands of the study landscape. However, slope, aspect, and NDVI (in either April or August) were not strongly correlated across the 2020 burned area in any other one-to-one comparison of these predictor variables.

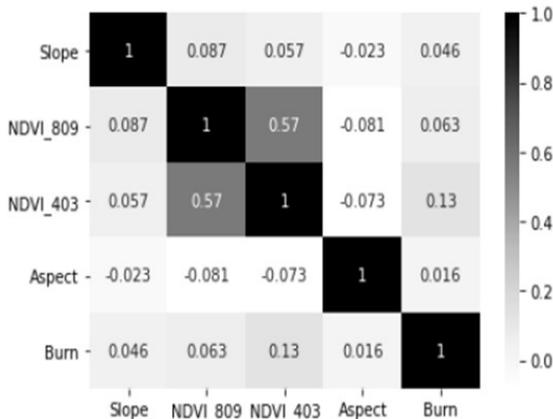


Figure 4. Correlation matrix for predictor (model input) layers and predicted burn severity classes in the SCU Complex fires.

Machine Learning Results

Running Random Forest on the randomly sampled (one-third) subsets of the dataset resulted in a 75% prediction accuracy overall. However, as seen in its normalized confusion matrix (Fig. 5a), this model mainly resulted in correctly classifying the most unbalanced (majority) class, namely the moderate burn severity (MBS 2), at a 95% prediction accuracy, while the high burn severity class (HBS 3) result had only a 23% prediction accuracy. For the purposes of this study, better prediction accuracies across all the burn classes are necessary and would be preferred over a high accuracy dominated by the majority burn class area. While the majority class, in this case MBS 2, makes up most of the burned area dataset, any of the minority burn classes may be considered to be of at least as great of interest.

A usefully predictive model should be able to generalize its learnings for new datasets. Generalization in part can be achieved by not overfitting the model to the training data. In the case of this study, because of the class-imbalance for MBS class 2, the model began to overfit this class. We alleviated this imbalance to reduce overfitting by performing under-sampling of the training data to have an equally distributed amount of data in each class.

Applying the Near Miss undersampling method, with the smallest burn class to set the sampling level being the low burn severity (LBS 1; N = 79,269 pixels) class, Random Forest results produced a significant overall prediction accuracy of 54% for the four burn severity classes. The normalized confusion matrix (Fig. 5b) showed that this sampling method resulted in the moderate burn severity (MBS 2) with a 61% prediction accuracy, while the high burn severity class (HBS 3) had a 71% prediction accuracy, and the low burn severity (LBS 1) had a 49% prediction accuracy. The overall accuracy of the model was most strongly impacted by the difficulty in prediction of the unburned areas (class 0), whose prediction accuracy was 35% using the Near Miss undersampling method. For this model run, the F1 score results followed the prediction accuracy ranking, with scores of 0.39, 0.53, 0.58, and 0.62 for burn classes 0 to 3, respectively. The feature importance values for the four input variables from Random Forest modeling with Near Miss undersampling were output as follows: 0.31, 0.26, 0.22, and 0.22 for NDVI in August 2020, NDVI in April 2020, slope, and aspect, respectively.

Random Forest modeling using the ENN undersampling method resulted in the highest accuracy for any of the models tested, with a 90% overall prediction accuracy. Although this method still retains a high amount of pixels values in class MBS 2, its data selection technique results in higher accuracy among all four classes, compared to even the subsetted random sampling method. Using the ENN undersampling method resulted in the moderate burn severity (MBS 2) with a 99% prediction accuracy, while the high burn severity class (HBS 3) had a 58% prediction accuracy, and the low burn severity (LBS 1) had a 21% prediction accuracy. The overall accuracy of the model was also impacted by the difficulty in prediction of the unburned areas (class 0), whose prediction accuracy was 34%. Similar to Near Miss undersampling results, the feature importance outputs from the ENN random forest run showed the significance of pre-fire NDVI data in relation to predicting fire burn severity classification. In this model result, the August and April NDVI variables had even higher feature importance, with a combined value of around 0.60, and both slope and aspect showing importance outputs of 0.20 each.

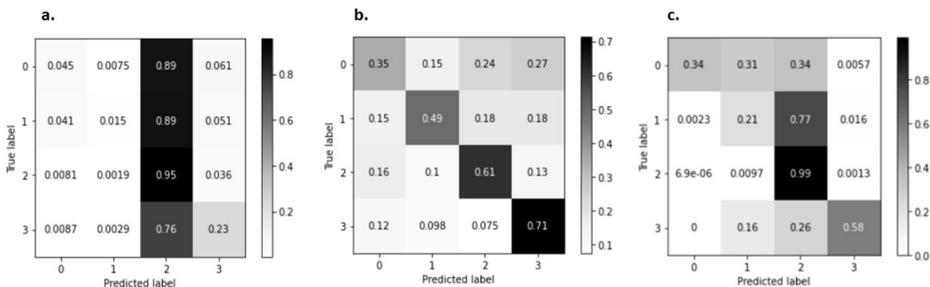


Figure 5. Classification accuracy (normalized confusion matrix) results for the Random Forest model on the (a) randomly sampled (one-third) subsets, and (b) Near Miss undersampling method for predicted burn severity classes in the SCU Complex fires. (c) ENN undersampling method for predicted burn severity classes in the SCU Complex fires. Near Miss and ENN show vastly improved performance for correct burn class predictions, seen along the main diagonal, compared to random sampling where class 2 is significantly overpredicted (column 3).

DISCUSSION

The principal finding from this study was that the overall burn severity patterns of the 2020 SCU Complex could be predicted from pre-fire vegetation green biomass, slope, and aspect variables with high accuracies of between 50% and 80% using Random Forest machine learning techniques. The August and April NDVI variables had the highest feature importance values, implying that the relative amount of dry biomass present at a given location was essential to predict how severely and completely the 2020 fires burned the vegetation cover and surface soils across this landscape. Since it was determined that pre-fire variables were predictive of fire severity, the results can be used to inform future fire mitigation activities. Specifically, the analysis of NDVI from Landsat in the months of April to June of any given year can be used to anticipate where the highest severity burning would occur in a central California woodland landscape where fire ignitions are frequent during the hottest days of the year. These machine learning methods have therefore advanced our understanding of the landscape attributes that influenced burn severity from a lightning fire complex in California mixed woodlands and grasslands.

The Near Miss undersampling technique selected for an equal number of pixels for each burn class was the most balanced and arguably most relevant machine learning result generated from the analysis of controls on the 2020 SCU Complex burn severity patterns. Because this undersampling technique was designed to select pixel datapoints from the majority class (MBS 2) with the smallest distance to the minority class(es), it generated a representation of the SCU Complex fire that would appear to be slightly less fragmented than the actual burned area landscape. This undersampling technique resulted in the best and most balanced combined prediction accuracy for burn classes 1–3, all with individual class accuracies between 49% and 71%.

On the other hand, while the ENN undersampling technique did not select for an equal number of pixels for each burn class, it predicted the MBS 2 class at a 99% accuracy level. Nonetheless, these strong results came at the expense of a much lower prediction accuracy for low burn severity (LBS 1) class at 21% accuracy. The ENN undersampling technique would have generated a representation of the SCU Complex fire that would appear to be smoother along edges and less ambiguous in terms of variations in burned area samples along the actual class boundaries. It would have sampled each burn class area from locations separated by a longer distance from any other burn severity class area to make the distinction between classes cleaner. While this was not the actual pattern of burn severity classes that resulted from the 2020 SCU Complex fires, the results demonstrated the change in accuracy that such a “smoothed along edges” burn pattern can have, compared to other more complex burn patterns.

Examining more closely the influence of the unburned class (0) in the equally-distributed (Near Miss undersampling) model run illustrated the overall difficulty of predicting areas of that did not burn during the 2020 SCU Complex fires. If patches of unburned pixels that were scattered throughout the entire SCU fire-affected area were ignored, and the model strictly focused on predicting burn severity classes 1 to 3, this adjustment would increase the Random Forest model’s overall performance by up to 35%, from approximately 54% to nearly 80% prediction accuracy. Nonetheless, nearly 15,000 ha (or 9% of the entire SCU Complex coverage area) were unburned within the 2020 fire perimeter. Many of the larger patches of unburned area shown in Fig. 1 were located along creek bottom lands that were evidently spared from the rapid spread of the fire. These (relatively) lower elevation and

presumably slightly wetter creek-side locations proved to be among the most difficult for the machine learning model to determine as either burned or unburned. In addition, the location of unburned areas could largely be a consequence of the random strike points of lightning that occurred on 16 August 2020, completely unrelated to vegetation cover, slope, or aspect.

To begin to put the results from this study of the burn patterns from the 2020 SCU Complex fires into a broader regional perspective, it is worth noting that Estes et al. (2017) reported that shrub vegetation was more likely to burn at higher severity than mixed hardwood/conifer or hardwood vegetation in northern California wildfires. Likewise, we found that the pre-fire cover density of evergreen vegetation was the most important variable to predict burn severity classes within the SCU Complex. Estes et al. (2017) also reported that upper- and mid-slopes tended to burn at higher fire severity than lower-slopes in the Klamath Mountains of northern California. East- and southeast-facing aspects tended to burn at higher severity than other aspects in this region.

Compared to analysis results of several other large wildfires in central California over the past decade, the 160,498 ha SCU Complex fires had a substantially higher MBS fraction (of 83%) than did the 104,131 ha Rim Fire, which burned through the in the Stanislaus National Forest of the central Sierra Nevada in 2013. The MBS fraction was 22% within the Rim Fire burned perimeter, while its HBS fraction was estimated at 34% (Potter 2014c), which was much higher than the 12% cover of HBS area estimated for the SCU Complex fires. It was also reported by Potter (2014b) that most of the HBS areas in the Rim Fire were located in areas where high levels of pre-fire fuels were quantified by 2013 Landsat NDVI imagery.

The Diablo Range landscape lends itself to more moderate burn severity impacts than other more heavily wooded forests of the Sierra Nevada. The trees in these oak savannah and mixed woodland-grasslands are more sparsely distributed in than were dense stands of conifers that burned in the Rim Fire. Moreover, Fry (2008) reported that oak mortality was low following prescribed burning in the northern Diablo Range of Santa Clara County.

In an analysis of the 20 largest wildfires that burned near the California central coast since 1984, Potter (2017) reported that the fraction of HBS area to total area burned ranged from a minimum of 0 to a maximum of 73%, with an average of 21%. Again, this typical HBS fraction from this collection of recent Pacific coast wildfires was much higher than the 12% cover of HBS area estimated for the SCU Complex fires. The acreage of HBS patches was found to increase exponentially and significantly ($p < 0.01$) with total area burned in each of these 20 coastal fires, but since the 2020 SCU Complex fire area was larger than any of these coastal fires before it, the SCU Complex does not fit the pattern cited by Potter (2017) that wildfires in central California experience their most rapid rate of increase in acreage of HBS area when the total fire size exceeds 48,500 ha (120,000 acres). It is plausible that the SCU Complex burned mainly at MBS from start to finish of the 2020 fire period and did not expand in the fraction of HBS coverage as it progressed.

Potter (2016) recounted that the Soberanes Fire that burned in 2016 in Monterey County on the California central coast resulted in a HBS fraction of 22% of the total area impacted, whereas final moderate burn severity (MBS) area comprised about 10% of the total area burned of approximately 53,470 ha (132,130 acres). Therefore, the Soberanes Fire was typical of most wildfires on California central coast in terms of MBS and HBS fractions, and contrasts again with the SCU Complex fire that had a much lower fraction of HBS coverage.

Conclusions

Although the SCU Complex fires burned mostly at a moderate burn severity level during August and September of 2020, which was out-of-the-ordinary for a large wildfire in central California woodlands, the amount of dry biomass present, as detected from Landsat satellite data, was the most important input variable used to predict how severely these fires burned the vegetation cover and surface soils across the steep watersheds of the southern Diablo Range. The input variables used in this study to predict burn severity levels and locations are readily available for any fire-prone region around the globe. Our study results suggest that Random Forest machine learning can be applied with confidence to predict and map potential medium and high burn severity areas accurately in advance of future fires for partially wooded landscapes in central California. While burn severity patterns can be measured post-fire, the factors that contributed to variations in burn severity levels cannot be assessed post-fire if those factors have been severely altered by the fire event, as is the case of vegetation cover. Knowing where the greatest risk for high burn severity is present on the landscape in terms of vegetation biomass can be a valuable piece of information for local resource managers.

LITERATURE CITED

- Boucher, J., A. Beaudoin, C. Hébert, L. Guindon, and É. Bauce. 2016. Assessing the potential of the differenced normalized burn ratio (dNBR) for estimating burn severity in eastern Canadian boreal forests. *International Journal of Wildland Fire* 26:32–45.
- Breiman, L. 2001. Random forests. *Machine Learning* 45:5–32.
- Estes, B. L., E. E. Knapp, C. N. Skinner, J. D. Miller, and H. K. Preisler. 2017. Factors influencing fire severity under moderate burning conditions in the Klamath Mountains, northern California, USA. *Ecosphere* 8(5):e01794.
- Forestry Canada Fire Danger Group (FCFDG). 1992. Development of the Canadian forest fire behavior prediction system. Forestry Canada, Ottawa, Ontario, Canada.
- French, N. H. F., E. S. Kasischke, R. J. Hall, K. A. Murphy, D. L. Verbyla, E. E. Hoy, and J. L. Allen. 2008. Using Landsat data to assess fire and burn severity in the North American boreal forest region: an overview and summary of results. *International Journal of Wildland Fire* 17:443–462.
- Fry, D. L. 2008. Prescribed fire effects on deciduous oak woodland stand structure, Northern Diablo Range, California. *Rangeland Ecology and Management* 61:294–301.
- Krawchuk, M. A., S. L. Haire, J. D. Coop, M.-A. Parisien, E. Whitman, G. W. Chong, and C. Miller. 2016. Topographic and fire weather controls of fire refugia in forested ecosystems of northwestern North America. *Ecosphere* 7:e01632.
- Liaw, A., and M. Weiner. 2018. Breiman and Cutler's Random Forests for Classification and Regression. R Package Version 4.6–7. Available from: <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
- California Department of Forestry and Fire Protection (CALFIRE). 2020. SCU Lightning Complex, Cal Fire Incidents. Available from: <https://www.fire.ca.gov/incidents/2020/8/18/scu-lightning-complex/>
- Key, C. H., and N. C. Benson. 2006. Landscape assessment: sampling and analysis meth-

- ods. USDA Forest Service General Technical Report RMRS-GTR-164-CD. LA1-LA51. USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO, USA.
- Leevy, J., T. Khoshgofaar, R. Bauder, and N. Seliya. 2018. A survey on addressing high-class imbalance in big data. *Journal of Big Data* 5:42.
- Lydersen, J. M., B. M. Collins, M. L. Brooks, J. R. Matchett, K. L. Shive, N. A. Povak, V. R. Kane, and D. F. Smith. 2017. Evidence of fuels management and fire weather influencing fire severity in an extreme fire event. *Ecological Applications* 27:2013–2030.
- Miller, J., and A. Thode. 2007. Quantifying burn severity in a heterogeneous landscape with a relative version of the delta normalized burn ratio (dNBR). *Remote Sensing of the Environment* 109:66–80.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.
- Potter, C. 2014a. Monitoring the production of central California coastal rangelands using satellite remote sensing. *Journal of Coastal Conservation* 18:213–220.
- Potter, C. 2014b. Microclimate influences on vegetation water availability and net primary production in coastal ecosystems of Central California. *Landscape Ecology* 29(4):677–687.
- Potter, C. 2014c. Geographic analysis of burn severity for the 2013 California Rim Fire. *Natural Resources* 5:1–10.
- Potter, C. 2016. Landscape patterns of burn severity in the Soberanes Fire of 2016. *Journal of Geography & Natural Disasters* S6:005.
- Potter, C. 2017. Fire-climate history and landscape patterns of high burn severity areas on the California southern and central coast. *Journal of Coastal Conservation* 21:393–404.
- Radočaj, D., M. Jurišić, and M. Gašparović. 2021. A wildfire growth prediction and evaluation approach using Landsat and MODIS data. *Journal of Environmental Management* 304:114351.
- Stahle, D. W., R. D. Griffin, D. M. Meko, M. D. Therrell, J. R. Edmondson, M. K. Cleaveland, L. N. Stahle, D. J. Burnette, J. T. Abatzoglou, K. T. Redmond, and M. D. Dettinger. 2013. The ancient blue oak woodlands of California: longevity and hydroclimatic history. *Earth Interactions* 17(12):1–23.
- Verbesselt, J., R. Hyndman, G. Newnham, and D. Culvenor. 2010. Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment* 114:106–115.
- Watershed Emergency Response Team (WERT). 2020. SCU Lightning Complex, California. CA-SCU-005740, Department of Forestry and Fire Protection, Sacramento, CA, USA.
- White, K. L. 1966. Structure and composition of foothill woodland in central coastal California. *Ecology* 47:229–237.
- Whitman, E., M. A. Parisien, D. K. Thompson, R. J. Hall, R. S. Skakun, M. D. Flannigan. 2018. Variability and drivers of burn severity in the northwestern Canadian boreal

forest. *Ecosphere* 9(2):e02128.

Submitted 11 May 2021

Accepted 23 June 2021

Associate Editor were C. Walters and T. Ryan