Ecological Modelling 414 (2019) 108855

Contents lists available at ScienceDirect

Ecological Modelling

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

Modelling post-fire tree mortality: Can random forest improve discrimination of imbalanced data?

Timothy M. Shearman[a,*], J. Morgan Varner[b], Sharon M. Hood[c], C. Alina Cansler[a,c], J. Kevin Hiers[b]

[a] School of Environmental and Forest Sciences, University of Washington, Seattle, WA, 98195, USA
[b] Tall Timbers Research Station, Tallahassee, FL, 32312, USA
[c] USDA Forest Service Rocky Mountain Research Station, Missoula, MT, 59808, USA

ARTICLE INFO

Keywords:
Fire effects
Logistic regression
Machine learning
Model evaluation
Model validation
Pinus palustris
Prescribed fire

ABSTRACT

Predicting post-fire tree mortality is a major area of research in fire-prone forests, woodlands, and savannas worldwide. Past research has relied overwhelmingly on logistic regression analysis (LR) that predicts post-fire tree status as a binary outcome (i.e. living or dead). One of the most problematic issues for LR (or any classification problem) occurs when there is a class imbalance in the training data. In these instances, predictions will be biased toward the majority class. Using a historical prescribed fire data set of longleaf pines (*Pinus palustris*) from northern Florida, USA, we compare results from standard LR and the machine-learning algorithm, random forest (RF). First, we demonstrate the class imbalance problem using simulated data. We then show how a balanced RF model can be used to alleviate the bias in the model and improve mortality prediction results. In the simulated example, LR model sensitivity and specificity was clearly biased based on the degree of imbalance between the classes. The balanced RF models had consistent sensitivity and specificity throughout the simulated data sets. Re-analyzing the original longleaf pine data set with a balanced RF model showed that although both LR and RF models had similar areas under the receiver operating curve (AUC), the RF model had better discrimination for predicting new observations of dead trees. Both LR and RF models identified duff consumption and percent crown scorch as important predictors of tree mortality, however the RF model also suggested pre-fire duff depth as an important predictor. Our analysis highlights LR limitations when data are imbalanced and supports using RF to develop post-fire tree mortality models. We suggest how RF can be incorporated into future tree mortality studies, as well as possible implementation in future decision-support tools.

1. Introduction

Tree mortality following fire is a major area of research in fire-prone forests, woodlands, and savannas worldwide (Michaletz and Johnson, 2007, O'Brien et al., 2018; Hood et al., 2018). Accurate prediction of individual tree mortality from measurable field variables such as tree size, injuries incurred, or fire behavior has tremendous utility for prescribed fire planning, risk assessment, as well as post-wildfire management decisions (Hood et al., 2007b; Varner et al., 2007; Reinhardt and Dickinson, 2010). Fire-induced mortality predictions are used prior to fire for prioritizing and designing prescriptions for fuels treatments (Agee and Skinner, 2005; Reinhardt and Dickinson, 2010). Prescribed fire managers need to predict potential tree mortality to determine what weather conditions and ignition patterns are needed to meet management objectives, and to assess fire impacts and determine post-fire plans if observed fire behavior exceeded intended bounds (Hiers et al., 2003; Ryan et al., 2013). After a wildfire, models are needed that identify dying trees for salvage and post-fire management decisions (Bevins, 1980).

Post-fire tree mortality can be a direct effect of fire causing immediate tree death (i.e. first-order effect), or an indirect effect related to interactions of fire-caused tree injury with climate or insects and disease causing delayed tree mortality years after the fire (Kane et al., 2017b; Hood et al., 2018). Accurate assessments of both immediate and delayed tree mortality are critical for predictions of broader second-order fire effects, such as vegetation changes, fuel dynamics, and erosion (Reinhardt and Dickinson, 2010; O'Brien et al., 2018). The complexities involved with the interaction of various processes such as fire, weather, land use, and seed availability, make modeling second-order fire effects difficult without understanding the first-order effects that

* Corresponding author.
E-mail address: tshearm@uw.edu (T.M. Shearman).

https://doi.org/10.1016/j.ecolmodel.2019.108855
Received 3 May 2019; Received in revised form 17 October 2019; Accepted 18 October 2019

contribute to them (Reinhardt et al., 2001). Improving post-fire tree mortality prediction therefore also improves our understanding of these broader outcomes.

Tree mortality is usually recorded as a binary outcome; either living or dead (with "dead" including top-kill of resprouting trees), and as a result, one obvious and often-used method for modeling tree mortality has been logistic regression (LR). In these models, the probability of a tree dying is conditional on a linear combination of a set of predictor variables through the logistic function:

$$P(Dead) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n)}}$$

where x_1 – x_n are the predictor variables and β_0 – β_n are coefficients estimated from observed data using maximum likelihood. Woolley et al. (2012) reviewed logistic regression mortality models generated for western North American conifers following fire and found the most commonly used predictors of post-fire tree mortality included measures of crown injury and bole injury, but studies have reported that including fire behavior and fuel consumption predictors significantly improved model performance. The most common measure of crown injury is the percentage of crown volume scorched, although crown length scorched and crown volume consumed (%) have also been included as predictors (Hood et al., 2010; Woolley et al., 2012; Hood et al., 2018). Measurements related to fire intensity include bole char height or crown scorch height–since these are related to flame length and its correlation with fireline intensity (Alexander, 1982). Indirect assessments of bole injury include bole char severity and tree defense measures such as bark thickness or surrogates for bark thickness such as tree diameter (Ryan and Reinhardt, 1988; Kobziar et al., 2006; Hood et al., 2008). Belowground injury to roots is typically measured via reductions in forest floor depth (Varner et al., 2007; Hood et al., 2010; O'Brien et al., 2010).

Predicting post-fire tree mortality is a classification problem; however, LR alone is not a classification algorithm. Each predicted observation is a probability of a positive outcome (in this case, a probability of mortality). To classify an observation, the user must select a threshold probability above which the tree is predicted to die. Typically, this threshold is set at 0.5 or 0.6 (Ryan and Reinhardt, 1988; Regelbrugge and Conard, 1993; Keyser et al., 2006; Thies et al., 2006; Hood et al., 2007b; Kane et al., 2017a), however Hood et al. (2007b) and Grayson et al. (2017) evaluated models using a range of thresholds. Adjusting the threshold probability affects the sensitivity (proportion of dead trees correctly predicted to die) and the specificity (proportion of live trees correctly predicted to live) of the model. Because of the dynamics between model threshold probability and sensitivity/specificity, LR model performance is often evaluated based on the Receiver Operating Characteristic (ROC) curve, which plots sensitivity against the false positive rate (1-specificity) over all possible threshold values (Saveland and Neuenschwander, 1990). The area under the ROC curve (AUC) provides an average estimate of the discrimination ability of the LR model, with AUC values ≥ 0.5 indicating more discrimination than random chance (Hosmer and Lemeshow, 2000). The higher the AUC, the better the model is at potentially discriminating between the two classes (live and dead trees).

Although the AUC is often used as a measure of the predictive performance of a model, it is not recommended to rely solely on this one statistic when comparing different models. The AUC only provides an estimate of the probability that a dead tree will have a higher predicted mortality probability than a live tree (Hosmer and Lemeshow, 2000). It is possible that a model has a high AUC but has poor goodness-of-fit (Lobo et al., 2008). Although the AUC removes the subjectivity of selecting a threshold by summarizing the accuracy over all possible thresholds, users may not be interested in all of these possibilities (Lobo et al., 2008; Ganio et al., 2015). It may be more important to consider partial areas under the curve by setting a fixed false positive rate (Ganio et al., 2015).

Validation is important for any predictive model, yet until recently was lacking for many of the published post-fire tree mortality models. Woolley et al. (2012) found that only ca. 11% of the 116 published models they reviewed were validated with data not used in the model fitting process. Recent evaluation efforts have shown mixed results in predictive accuracy (Hood et al., 2007b; Ganio and Progar, 2017; Kane et al., 2017a; Grayson et al., 2017; Hood and Lutes, 2017; Furniss et al., 2019). Mortality models are often site- and/or species-specific and therefore can perform poorly at predicting new data, which underscores the importance for validating models. For example, Ganio and Progar (2017) found that their logistic regression models for ponderosa pine (*Pinus ponderosa*) and Douglas-fir (*Pseudotsuga menziesii*) had low sensitivity, but high specificity. The disparity between sensitivity and specificity found in the validation of mortality models has led many to suggest that the threshold probability should be changed to reflect the management goals of the user (Hood et al., 2007a; Cansler et al., 2019). For example, if accurate prediction of dead trees is more important than accurate prediction of live trees, the user can adjust the threshold probability for the classification (Hood et al., 2010; Ganio and Progar, 2017; Grayson et al., 2017).

The pattern of low sensitivity and high specificity (or vice versa) is common in tree mortality models as well as in classification models in other disciplines where the response classes of the training data are heavily imbalanced (e.g., there are many more live trees than dead trees). In these so-called "rare events", models such as logistic regression underestimate the probability of the rare class (King and Zeng, 2001). For example, Alberg et al. (2004) demonstrated that when prevalence of a disease deviates away from 50%, sensitivity and specificity are conversely increased or decreased depending on the direction of the deviation. Post-fire tree mortality data sets typically suffer from class imbalance. For instance, in prescribed fires–where primary objectives typically include minimizing overstory mortality and the fire is generally a low-to-moderate intensity surface fire–logistic regression equations for post-fire tree mortality are likely to underestimate post-fire tree mortality.

In order to improve our ability to predict post-fire tree mortality, there is a major need to evaluate alternative analytical methods to the LR model (Barnard et al., 2019). One classification algorithm known to excel in predicting new observations is random forest (RF, Breiman, 2001a). Random forest is an extension of classification and regression trees (CART, see De'ath and Fabricius, 2000) that consists of a collection of CART style trees in which each tree casts a vote for which class (e.g., live or dead) is predicted (Breiman, 2001a,b). For each classification tree, a random set of observations are selected with replacement, as well as a random selection of predictor variables. Those observations not selected for fitting an individual classification tree are considered "out-of-bag" (OOB) samples and are used to assess error. The use of randomly selected predictors and OOB samples for validation reduces the bias in RF models and prevents overfitting the data (Breiman, 2001a). However, because RF models are a bootstrap sample of the training data, they are also sensitive to imbalanced classes. When classes are heavily imbalanced, the probability is high that some of the bootstrap samples will not contain any observations of the minority class (Chen et al., 2004). To solve this issue, Chen et al. (2004) proposed the balanced random forest algorithm, which draws an equal number of observations for each class (in our case, equal numbers of living and dead trees).

In this study, we investigate the utility of RF as an alternative to LR to model fire-induced tree mortality following prescribed fire. We test the impact of model selection on an existing data set of fire-induced mortality in long-unburned longleaf pine (*Pinus palustris* Mill.) forests in north Florida, USA (data from Varner et al., 2007). We demonstrate the class imbalance problem with post-fire mortality prediction using LR and RF on simulated data and compare variable importance and cross-validated predictions between the two analysis methods on the original data. We then discuss the advantages and disadvantages of RF analyses

Table 1
Mean (and S.D.) characteristics of longleaf pines (*Pinus palustris*) after prescribed fire in a long-unburned forest at Eglin Air Force Base in northern Florida.

Parameter	Live	Dead
n	400	43
DBH (cm)	30.9 (8.7)	34.9 (10.8)
Scorch (%)	31.5 (35.0)	69.7 (35.9)
Char Height (m)	2.4 (2.0)	4.3 (2.6)
Scorch Height (m)	9.0 (6.4)	14.0 (5.7)
Pre Duff Depth (cm)	6.4 (3.5)	8.7 (3.6)
Duff Consumption (%)	17.5 (21.9)	56.2 (31.7)

for post-fire tree mortality prediction and how to apply this model more broadly to enhance post-fire mortality prediction accuracy.

2. Materials & methods

2.1. Field data

We used the data set from Varner et al. (2007) as the basis for all of our data analyses (Table 1) to demonstrate the class imbalance problem with post-fire mortality prediction, compare results of LR and RF and discuss advantages and disadvantages of the two methods. The data are from long-unburned longleaf pine forests at Eglin Air Force Base in northwest Florida treated with prescribed fire. Stands consisted of a sparse remnant longleaf pine overstory of approximately 45–200 trees > 10 cm DBH ha^{-1} and midstories dominated by *Quercus laevis* Walt., *Q. geminata* Small, *Pinus clausa* (Chapman ex Engelm.) Vasey ex Sarg. var *immuginata*, and *Ilex vomitoria* Ait. (Varner et al., 2007). The climate of the area is subtropical, with warm humid summers and mild winters. Mean temperature of the study area is 19.7 °C and mean annual precipitation is 1580 mm (Overing et al., 1995). Prescribed fires were conducted in 2001 and 2002 from February to April. The original study was established as a randomized block experiment consisting of four blocks where day-of-burn duff moisture content varied among sites from dry (55% dry mass), moist (85% dry mass), and wet (115% dry mass). The unburned control treatments that were not burned in Varner et al. (2007) were not included in this study. Forest floor depth was measured at the base of 50 randomly selected longleaf pines > 15 cm diameter in each (~10 ha) burn unit by horizon including litter (Oi horizon) and duff (Oe and Oa organic soil horizons). At each tree, eight evenly spaced pins (20 cm tall) were buried flush with the litter surface prior to ignition and measured after the prescribed fire to estimate litter and duff consumption. Crown volume scorched (%) was estimated using the methods in Peterson (1985). Bole char heights were measured as the average of four measurements. Tree injury was assessed within six weeks of all burns and mortality was surveyed every six months for two years post-fire.

2.2. Simulated data analysis

To demonstrate the class imbalance problem, we simulated nine new data sets of 10,000 observations where each data set varied in the mortality rate (from 10% to 90% dead trees). For each live class and dead class in each new simulated data set, we also simulated two injuries—crown scorch (%) and duff consumption (%)–based a random sample of a normal distribution that had the same means and standard deviations of the original data (see Table 1 for means and standard deviations from Varner et al. (2007).

We fit LR models to each simulated data set using the *glm* function in the R package, "stats" (R Core Team 2018). We then fit a balanced RF model to the same data using the *randomForest* function in the package "randomForest", with 10,000 classification trees and the "sampsize" parameter set to randomly select 100 live and 100 dead trees for each classification tree (Liaw and Wiener, 2002). Crown volume scorched (%) and duff consumption (%) were the only predictor variables included in both LR and RF models. We repeated this process, creating new data sets and refitting the models 100 times. Model accuracy was determined by predicting the training data with the models and determining each models' sensitivity and specificity using the following formulas:

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

where TP and TN are the number of true positive and true negative predictions respectively, and FP and FN are the number of false positive and false negative predictions, respectively. For the LR models, a threshold of 0.5 was used to determine mortality.

2.3. Original data analysis

To explore variable importance between LR and RF generated models and to show how RF models can be useful in interpreting tree mortality, we reanalyzed the original data set of 443 longleaf pine trees using the *randomForest* function. In addition to crown scorch and duff consumption, we included DBH, char height, scorch height, pre-fire duff depth, and pre-fire litter depth as predictor variables. We fit LR models using crown scorch and duff consumption as predictors, since Varner et al. (2007) fit a full model and then followed standard model reductions procedures to develop a final model that only includes these two significant variables. We set the "sampsize" parameter to select 10 live and dead trees for each classification tree. Variable importance was determined using the *varImpPlot* function in the "randomForest" package. In RF models, variable importance is determined by calculating the mean decrease in accuracy when a variable is removed from the model by permuting the OOB data (Liaw and Wiener, 2002). We used the R package "pdp" to plot the partial dependence functions (Greenwell, 2017). Partial dependence functions plot the overall estimated change in the probability of mortality as a predictor varies over its distribution (Goldstein et al., 2015).

2.4. Cross validation

We used repeated 5-fold cross validation to compare RF and LR models on the original data. The data were randomly grouped into five equally sized subsets. Both LR and RF models were fit to four groups (the training data sets) and used to predict the data in the one group not used in model fitting (the test data set). The LR models used a threshold probability of 0.5 to determine mortality, while the RF models used the majority vote of classification trees. This process was repeated such that each group was used once as a test data set. For each model that was fit, we calculated the ROC curve and subsequent AUC using the *roc* and *auc* functions in the R package "pROC" (Robin et al., 2011). Probabilities for the RF models used to calculate the ROC curve were based on proportion of votes of the classification trees. The cross validation was repeated ten times with new random subsets assigned after each repetition. We used the *createFolds* function in the R package "caret" to subset the data for each repetition (Kuhn, 2008).

3. Results

3.1. Simulated data – Effects of class imbalance

Increasing the class imbalance had a striking effect on the LR post-fire longleaf pine mortality models (Fig. 1A). At the lowest mortality rate, sensitivity was low and specificity was high as the LR models predicted live trees with high accuracy and dead trees with low

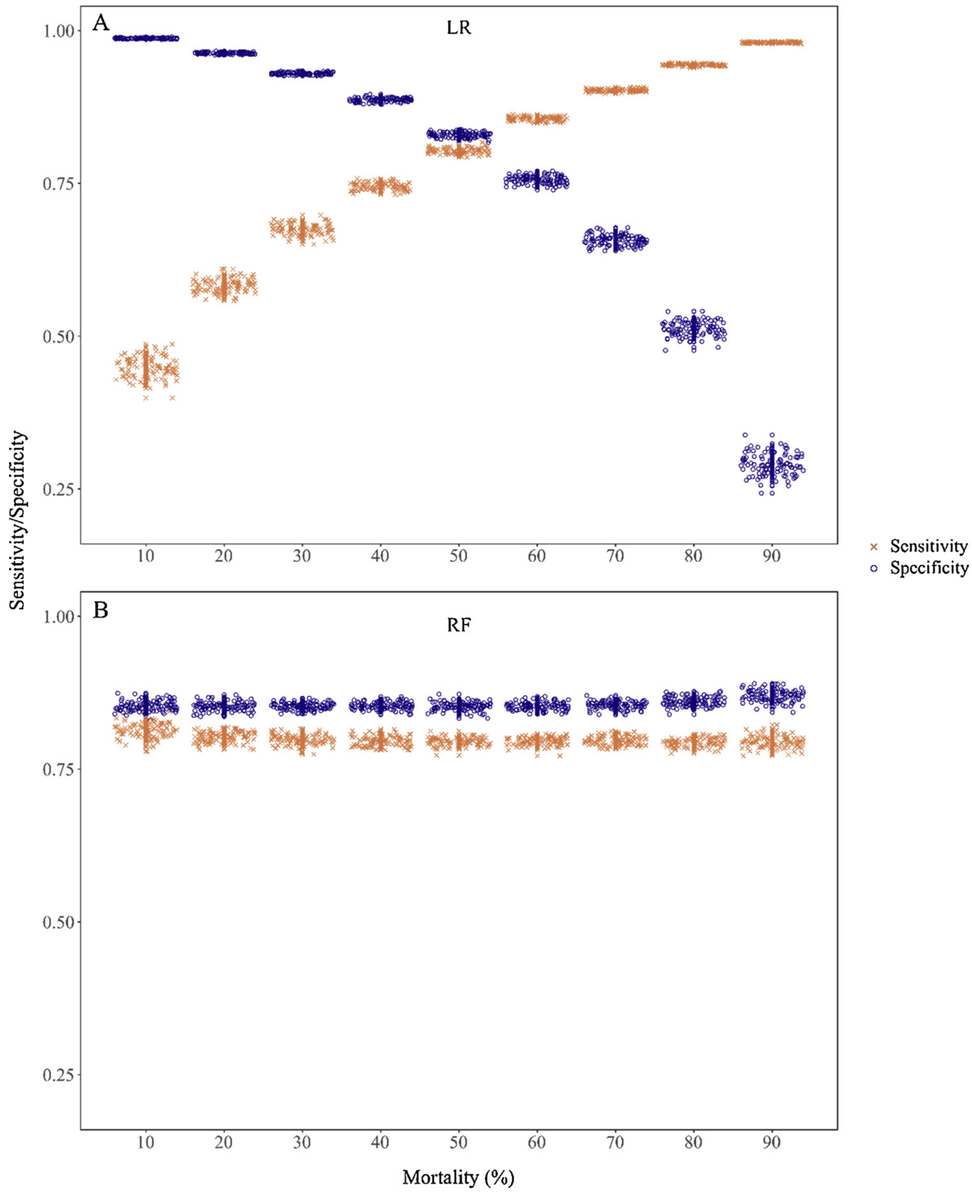

Fig. 1. Sensitivity and specificity of logistic regression (A) and balanced random forest (B) models on predicting post-fire tree mortality from duff consumption and crown volume scorch. Models were fit to simulated data that varied percent mortality but had equal means and standard deviations of predictor variables.

accuracy. Logistic regression models trained with balanced classes (50% mortality) had average sensitivity rates of 0.83 and average specificity rates of 0.80. At higher mortality rates, sensitivity was high and specificity was low. In contrast, the RF models had more consistent sensitivity and specificity rates than the LR models (Fig. 1B). Average accuracy for the RF models ranged from 0.85–0.87 for sensitivity and from 0.79–0.81 for specificity.

3.2. Original data– Variable importance to predict tree mortality

The RF model trained on the original longleaf pine data from Varner et al. (2007) provided additional insights into variable importance for predicting post-fire tree mortality. In agreement with the LR model, duff consumption was the most important predictor of post-fire tree mortality in long-unburned longleaf pine stands for the RF model (Fig. 2). Additionally, the RF model showed that crown volume scorched and pre-fire duff depth were also important positive predictors of fire-induced tree mortality. All other predictors (i.e. DBH, char height, scorch height, and pre-fire litter depth) in the model had little to no impact on model accuracy. The partial dependence plots showed that increasing duff consumption increased the average probability of tree mortality in the model from approximately 0.25 to 0.6 over the range of 0–100% duff consumption (Fig. 3A). Crown scorch also increased the probability of tree mortality from approximately 0.3 to 0.5 over the range of 0–100% scorch (Fig. 3B). Although pre-fire duff depth was nearly as important as fire-caused crown scorch, the partial dependence plot of duff depth on the probability of tree mortality was mostly flat with a slight increase in the probability of about 0.05 for trees with duff depths > 8 cm (Fig. 3C). The combined partial dependence of duff consumption and crown scorch shows that trees with high scorch and high duff consumption had a higher probability of mortality than trees with high scorch or high duff consumption alone (Fig. 4A). Likewise, the combination of a thick pre-fire duff layer and high duff consumption had high probability of mortality (Fig. 4B). Partial dependence plots for pre-fire duff and crown scorch showed minimal increases in mortality with increases of both predictor variables (Fig. 4C).

The 5-fold cross-validation of the LR and RF models fit to the original data from Varner et al. (2007) showed similar results to the

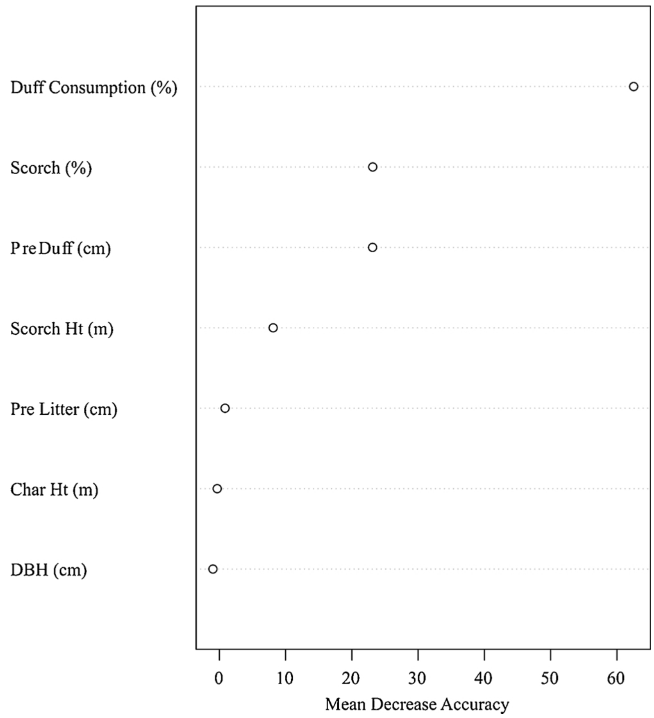

Fig. 2. Variable importance for a balanced random forest model predicting post-fire tree mortality in long-unburned longleaf pine forests. Average AUC using 5-fold cross validation was 0.85, with average sensitivity of 0.70 and average specificity of 0.80.

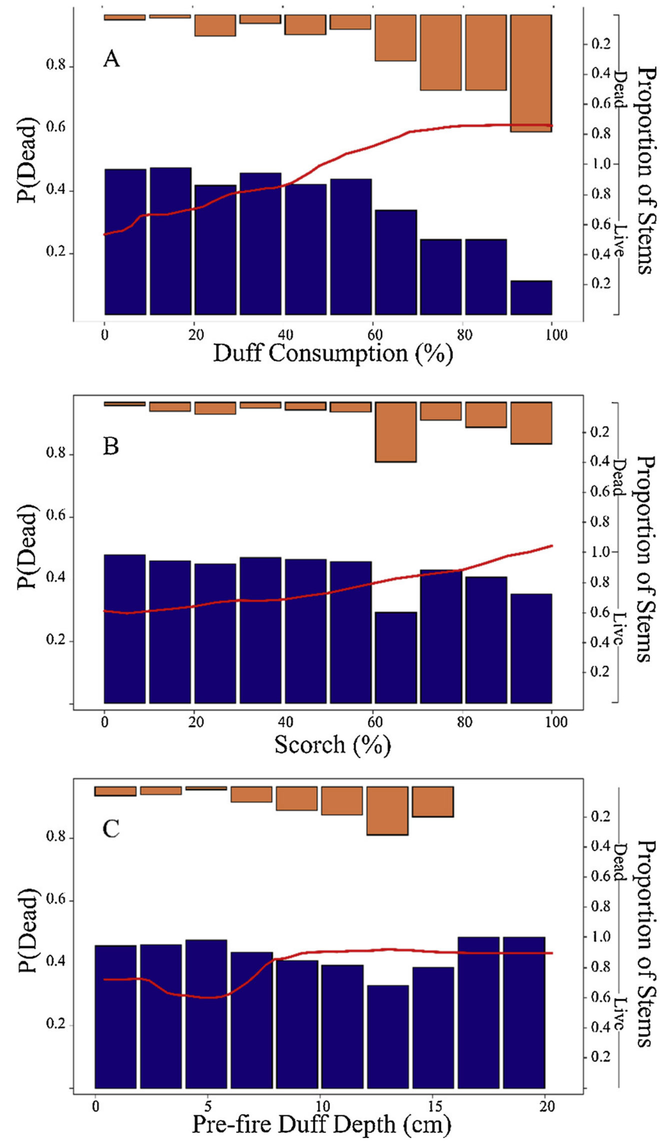

Fig. 3. Partial dependence plots showing the probability of post-fire tree mortality (red line, left scale) for different levels of A) percent of duff consumed, B) percent of crown volume scorched, C) pre-fire duff depth. Bars represent proportions (second y-axis) of live (bottom bars) and dead (top bars) longleaf pine trees in respective size classes. Partial plots are estimated based on a balanced random forest model, holding all other predictor variables constant (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

simulation data at the lowest mortality level (10%). The LR models had an average sensitivity of 0.32 and average specificity of 0.99 (Fig. 5). The RF models were more consistent, with average sensitivity of 0.70 and average specificity of 0.80 (Fig. 5). The AUC values from the ROC curves were similar for both model types, with an average AUC of 0.87 for LR and 0.85 for RF (Fig. 5).

4. Discussion

Traditional post-fire tree mortality LR models trained with data containing heavily imbalanced response classes can severely underpredict the minority class (in most prescribed fire cases, dead trees). These results are consistent with the findings in other disciplines that use LR to model binary outcomes, such as disease diagnostic tests in medical fields (Alberg et al., 2004). To our knowledge, no studies in the post-fire tree mortality literature have specifically dealt with the problem of imbalance classes on the training data of LR models. All of the current fire modeling tools available to practitioners rely on LR for the prediction of tree mortality. It is likely that the class imbalance problem is a major contributor to the poor predictive performance observed in tree mortality models validated in the literature (Grayson et al., 2017, Ganio and Progar, 2017; Hood and Lutes, 2017, Kane et al., 2017a). In our study, we show that balanced RF models can be a potential solution to the class imbalance problem in post-fire tree mortality models.

King and Zeng (2001) proposed that the class imbalance problem can be mitigated during the data collection stage of a study. They suggest using an “endogenous stratified” sampling strategy as a more efficient sampling design for models with rare events. For example, if post-fire tree mortality is low, collect all (or a subset of) dead tree observations and then a random equal sample of live tree observations. This sampling strategy not only solves the class imbalance issue but can also save on resources for data collection (King and Zeng, 2001). Hood et al. (2010) used a slightly different sampling approach that resulted in approximately equal number of live and dead trees to develop LR mortality models for California conifers following wildfire. They sampled trees across a matrix of low-to-high crown and bole injury ranges for different tree sizes and species, targeting approximately 30 sample trees within each combination of species, size class, crown injury class, and bole injury class. Using 10-fold cross validation, Hood et al. (2010) evaluated the accuracy of their models and found false positive and negative rates varied but all were higher than 0.6 and most were higher than 0.7. Although Hood et al. (2010) validated their models with data not used in the model fitting process using cross validation (as we did with ours), they note that the data are still from the same fires and ideally the models should be tested on new data from fires on separate sites. Grayson et al. (2017), however, included the models from Hood et al. (2010) in their validation of models in the literature on a new data set and found that they were not among the best fitting models. This might suggest that one or more additional predictors are missing to extend their applicability to other geographic areas or for different forest conditions (e.g., predictors incorporating

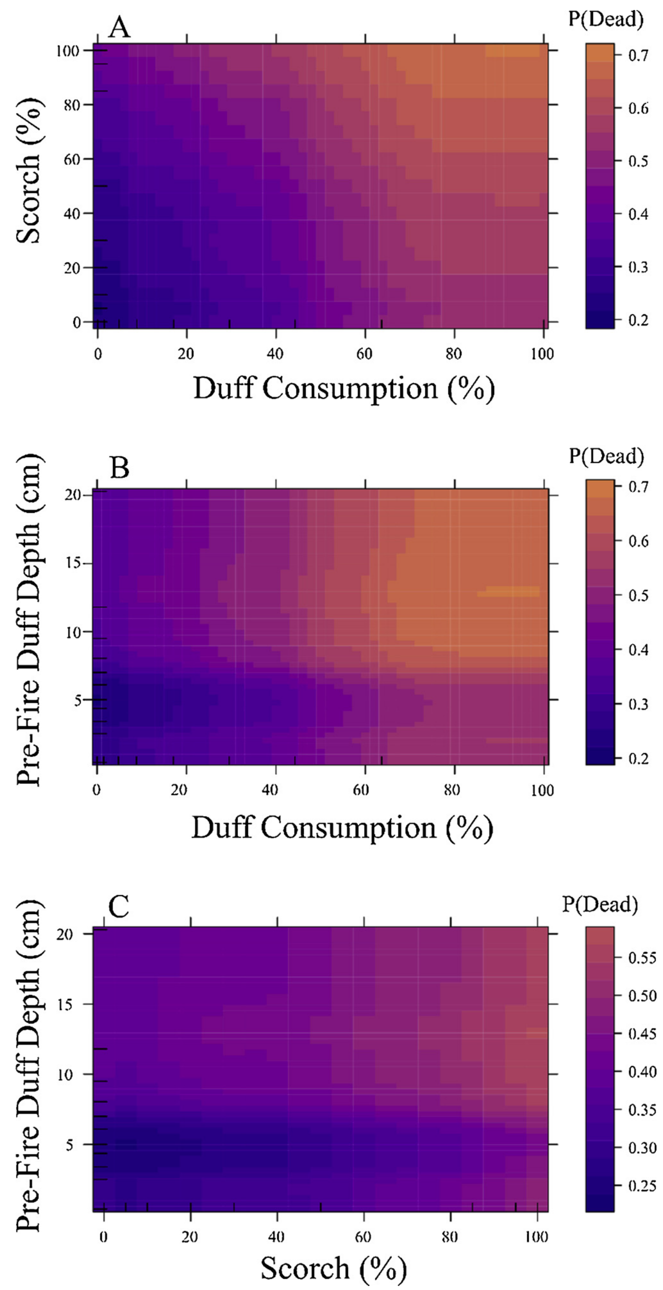

Fig. 4. Partial dependence plots showing the effect of two predictors on the probability of post-fire tree mortality: A) Percent duff consumption and percent crown volume scorched, B) percent duff consumption and pre-fire duff depth, and C) percent crown volume scorch and pre-fire duff depth. Probabilities are based on a balanced random forest model of longleaf pine tree mortality at Eglin Air Force base in Florida, USA, holding all other predictors constant.

spatial dispersion or competition). Random forest models may be a better option in predicting new data due to the ability to incorporate a large number of predictor variables. Studies and model development that require pre-fire measures of the trees or local fuels that may be modified by the fire (as with pre-fire duff depth in the longleaf pine data set) would complicate the endogenous stratification approach, but extensive pre-fire measures of these conditions or modeled *a priori* relationships could be used to ameliorate this issue.

In their original study using LR, Varner et al. (2007) found that the best predictors of individual mortality of longleaf pine were the percent of duff consumed in the fire and the percent crown volume scorched, with scorch only significant in the dry duff moisture treatment. Post-fire tree mortality was significantly higher in the dry duff moisture treatment than any other treatment (Varner et al., 2007). Duff consumption was six times greater in the dry duff moisture treatment than in the wet treatment and three times greater than in the moist treatment (Varner et al., 2007). As with many prescribed fire studies, overall tree mortality was low, with mean mortality rates of 0.5%, 3.0%, and 20.5% for the wet, moist, and dry treatments respectively.

The original longleaf pine mortality data had a class imbalance of 400 live trees to 43 dead trees. As a result, the cross-validated LR model, using a threshold of 0.5, performed poorly at predicting dead trees (low sensitivity) and predicted live trees with high accuracy (high specificity). Several studies suggest that a different threshold value should be used to accommodate for the differences in sensitivity and specificity based on management objectives (Hood et al., 2010; Ganio and Progar, 2017; Grayson et al., 2017; Cansler et al., 2019). The fact that the AUC values of both LR and RF models were nearly identical indicates that the potential discriminatory ability of both models is similar and that adjusting the threshold value of the LR models could yield similar accuracy to the RF model as adjusting the threshold does not affect the AUC. This solution, as our study illustrates, sidesteps an important point: the mortality probabilities estimated by LR are likely to be biased. We suggest using balanced RF as a better solution as there is no need to specify a single threshold criterion. By default, RF models use the majority vote of each classification tree in the "forest" to determine the predicted outcome. The proportion of votes in a RF model can also be used in the same manner as a threshold probability in LR models, should management applications require a different decision criterion than the majority.

Random forest models have been shown to provide exceptional predictive performance (Breiman, 2001a,b; Prasad et al., 2006; Rodriguez-Galiano et al., 2012). For example, Rodriguez-Galiano et al. (2012) found that RF correctly classified remotely sensed land cover data with an accuracy of 92%. Random forest models can also effectively predict fire impacts across large landscapes, including fire severity and post-fire forest structure (Kane et al., 2015a,b). Random forest models also have several appealing statistical qualities in that they do not make any distributional assumptions about the data, they can handle a high number of predictor variables, and they are not sensitive to missing values (Cutler et al., 2007). Additionally, RF models can evaluate more than two classes as well as continuous response variables. Although multinomial LR can be used to model more than two response classes, and linear regression can be used to model continuous response variables, it is still often necessary to reduce the dimensionality of the data in LR and other traditional statistical methods especially if there are high correlations among predictor variables. Breiman (2001b) argues that by reducing the dimensionality of the data, one also reduces potentially valuable information for prediction. By including the full suite of predictors in our RF model, we identified a relationship with pre-fire duff depth that was not evident in the LR model of Varner et al. (2007). Past and future studies on post-fire tree mortality may benefit by using RF and including additional potential predictors, as well as considering model performance for continuous responses such as growth or estimates of vigor instead of coarse binary responses of live or dead.

One of the criticisms of RF models is that they are a "black box" with the model using complex and largely unknown mechanisms to produce predictions from the predictor variables (Breiman, 2001b). Partial dependence functions can help with some of the interpretation, but RF models are generally poor at interpretability and are not a tool to be used for hypothesis testing (Breiman, 2001b, Cutler et al., 2007). Random forest models do not generate p-values nor estimate confidence intervals, which may be a drawback for some users. However, Cutler et al. (2007) suggest that the results of RF models can be used as inputs to more traditional multivariate statistical techniques. Regardless, because predicting post-fire tree mortality is a classification problem, accurate prediction is more important for many applications than

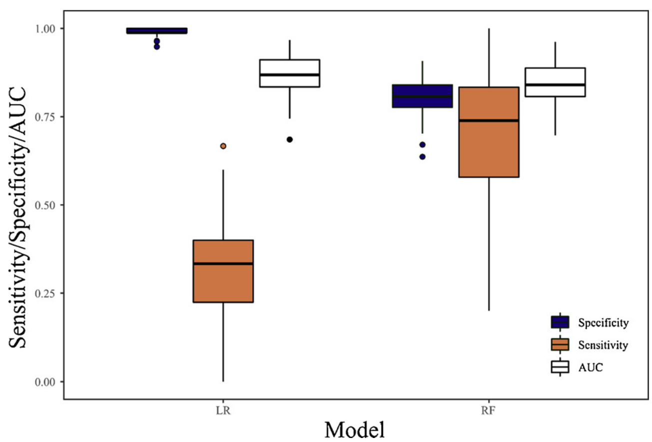

Fig. 5. Sensitivity, specificity, and area under the receiver operating characteristic curve (AUC) for logistic regression (LR) and random forest (RF) models. Models were evaluated using 5-fold cross validation with ten repetitions. See text for full model details.

interpretability.

Decision-support tools designed to assist land managers with predicting fire-caused tree mortality, such as the software program First Order Fire Effects Model (FOFEM; Lutes, 2018) may improve by incorporating balanced RF models in future versions. There are several potential benefits for land managers including the ability to enter new data and have the model update predictions to improve local estimates. For example, RF could potentially allow users to add new predictors to the model without the need to check that all the assumptions of LR are met and without the need to explicitly state interactive effects. Future decision-support tools could therefore "learn" to provide more accurate predictions specific to different environments. At the same time, the data that are provided by the users can be sent to the researchers that develop the software for evaluation. An application such as this could greatly increase our ability to rapidly incorporate new data into predictive modeling systems and increase our understanding of the underlying mechanisms of fire-induced tree mortality (Hood et al., 2018).

Author contributions

TMS and JMV conceived the ideas. JMV and JKH collected the original data. TMS analyzed the data. TMS, JMV, SMH, and CAC contributed to interpretation of the data. TMS and JMV led the writing of the manuscript. All authors contributed to critically revising drafts of the manuscript and gave final approval for publication.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

Funding for this project was provided by the US Joint Fire Science Program (projects 01-1-03-11 and 16-1-04-8) and the National Fire Plan. This work was stimulated by thoughtful questions from R. Mitchell and F.E. Putz. The authors declare no conflicts of interest.

References

Agee, J.K., Skinner, C.N., 2005. Basic principles of forest fuel reduction treatments. For. Ecol. Manage. 211, 83–96.

Alberg, A.J., Park, J.W., Hager, B.W., Brock, M.V., Diener-West, M., 2004. The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. J. Gen. Intern. Med. 19, 460–465.

Alexander, M.E., 1982. Calculating and interpreting forest fire intensities. Can. J. Bot. 60, 349–357.

Barnard, D.M., Germino, M.J., Pilliod, D.S., Arkle, R.S., Applestein, C., Davidson, B.E., Fisk, M.R., 2019. Can't see the random forest for the decision trees: selecting predictive models for restoration ecology. Restor. Ecol. https://doi.org/10.1111/rec.12938.

Bevins, C.D., 1980. Estimating Survival and Salvage Potential of Fire-scarred Douglas-Fir. U.S. Dept. of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, Utah Research Note INT-RN-287.

Breiman, L., 2001a. Random forests. Mach. Learn. 45, 5–32.

Breiman, L., 2001b. Statistical modeling: the two cultures. Stat. Sci. 16, 199–215.

Cansler, C.A., Hood, S.M., Varner, J.M., van Mantgem, P., 2019. Evaluating and optimizing the use of logistic regression for tree mortality models in the first order fire effects model (FOFEM). in Press In: Proceedings of the Fire Continuum Conference. 5/24/2018, Missoula, MT. USDA Forest Service General Technical Report.

Chen, C., Liaw, A., Breiman, L., 2004. Using Random Forest to Learn Imbalanced Data. Dept. Statistics, Univ. California, Berkeley, CA, Tech. Rep, pp. 666.

Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792.

De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81, 3178–3192.

Furniss, T.J., Larson, A.J., Kane, V.R., Lutz, J.A., 2019. Multi-scale assessment of post-fire tree mortality models. Int. J. Wildland Fire 28, 46–61.

Ganio, L.M., Woolley, T., Shaw, D.C., Fitzgerald, S.A., 2015. The discriminatory ability of postfire tree mortality logistic regression models. For. Sci. 61, 344–352.

Ganio, L.M., Progar, R.A., 2017. Mortality predictions of fire-injured large Douglas-fir and ponderosa pine in Oregon and Washington, USA. For. Ecol. Manage. 390, 47–67.

Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J. Comput. Graph. Stat. 24, 44–65.

Grayson, L.M., Progar, R.A., Hood, S.M., 2017. Predicting post-fire tree mortality for 14 conifers in the Pacific Northwest, USA: model evaluation, development, and thresholds. For. Ecol. Manage. 399, 213–226.

Greenwell, B.R., 2017. Pdp: and R package for constructing partial dependence plots. R J. 9, 421–436. https://journal.r-project.org/archive/2017/RJ-2017-016/index.html.

Hiers, J.K., Laine, S.C., Bachant, J.J., Furman, J.H., Greene Jr., W.W., Compton, V., 2003. Simple spatial modeling tool for prioritizing prescribed burning activities at the landscape scale. Conserv. Biol. 17, 1571–1578.

Hood, S.M., Bentz, B., Gibson, K., Ryan, K.C., DeNitto, G., 2007a. Assessing Post-fire Douglas-fir Mortality and Douglas-fir Beetle Attacks in the Northern Rocky Mountains. RMRS-GTR-199. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

Hood, S.M., McHugh, C.W., Ryan, K.C., Reinhardt, E., Smith, S.L., 2007b. Evaluation of a post-fire tree mortality model for western USA conifers. Int. J. Wildland Fire 16, 679–689.

Hood, S.M., Cluck, D.R., Smith, S.L., Ryan, K.C., 2008. Using bark char codes to predict post-fire cambium mortality. Fire Ecol. 4, 57–73.

Hood, S.M., Smith, S.L., Cluck, D.R., 2010. Predicting mortality for five California conifers following wildfire. For. Ecol. Manage. 260, 750–762.

Hood, S.M., Lutes, D., 2017. Predicting post-fire tree mortality for 12 western US conifers using the First Order Fire Effects Model (FOFEM). Fire Ecol. 13, 66–84.

Hood, S.M., Varner, J.M., van Mantgem, P.J., Cansler, C.A., 2018. Fire and tree death: understanding and improving modeling of fire-induced tree mortality. Environ. Res. Lett. 13, 113004.

Hosmer, D.W., Lemeshow, S., 2000. Applied Logistic Regression. John Wiley and Sons, New York.

Kane, J.M., van Mantgem, P.J., Lalemand, L.B., Keifer, M., 2017a. Higher sensitivity and lower specificity in post-fire mortality model validation of 11 western US tree species. Int. J. Wildland Fire 26, 444–454.

Kane, J.M., Varner, J.M., Metz, M.R., van Mantgem, P.J., 2017b. Characterizing interactions between fire and other disturbances and their impacts on tree mortality in western U.S. Forests. For. Ecol. Manage. 405, 188–199.

Kane, V.R., Lutz, J.A., Cansler, C.A., Povak, N.A., Churchill, D., Smith, D.F., Kane, J.T., North, M.P., 2015a. Water balance and topography predict fire and forest structure patterns. For. Ecol. Manage. 338, 1–13.

Kane, V.R., Cansler, C.A., Povak, N.A., Kane, J.T., McGaughey, R.J., Lutz, J.A., Churchill, D.J., North, M.P., 2015b. Mixed severity fire effects within the Rim fire: relative importance of local climate, fire weather, topography, and forest structure. For. Ecol. Manage. 358, 62–79.

Keyser, T.L., Smith, F.W., Lentile, L.B., Shepperd, W.D., 2006. Modeling postfire mortality of ponderosa pine following a mixed-severity wildfire in the Black Hills: the role of tree morphology and direct fire effects. For. Sci. 52, 530–539.

King, G., Zeng, L., 2001. Logistic regression in rare events data. Political Anal. 9, 137–163. http://gking.harvard.edu/files/abs/0s-abs.shtml.

Kobziar, L., Moghaddas, J., Stephens, S.L., 2006. Tree mortality patterns following prescribed fires in a mixed conifer forest. Can. J. For. Res. 36, 3222–3238.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28, 1–26.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2, 18–22.

Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17, 145–151.

Lutes, D., 2018. FOFEM 6.5 User Guide USDA Forest Service. Rocky Mountain Research Station, Fort Collins.

Michaletz, S.T., Johnson, E.A., 2007. How forest fires kill trees: a review of the fundamental biophysical processes. Scand. J. For. Res. 22, 500–515.

O'Brien, J.J., Hiers, J.K., Mitchell, R.J., Varner, J.M., Mordecai, K., 2010. Acute physiological stress and mortality following fire in a long-unburned longleaf pine

ecosystem. Fire Ecol. 6, 1–12.
O'Brien, J.J., Hiers, J.K., Varner, J.M., Hoffman, C.M., Dickinson, M.B., Michaletz, S.T., Loudermilk, E.L., Butler, B.W., 2018. Advances in mechanistic approaches to quantifying biophysical fire effects. Curr. For. Rep. 4, 161–177.
Overing, J.D., Weeks, H.H., Wilson, J.P., Sullivan, J., Ford, R.D., 1995. Soil Survey of Okaloosa County, Florida. USDA Natural Resource Conservation Service, Washington. D.C.
Peterson, D.L., 1985. Crown scorch volume and scorch height: estimates of postfire tree condition. Can. J. For. Res. 15, 596–598.
Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181–199.
R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Regelbrugge, J.C., Conard, S.G., 1993. Modeling tree mortality following wildfire in *Pinus ponderosa* forests in the central Sierra-Nevada of California. Int. J. Wildland Fire 3, 139–148.
Reinhardt, E.D., Keane, R.E., Brown, J.K., 2001. Modeling fire effects. Int. J. Wildland Fire 10, 373–380.
Reinhardt, E.D., Dickinson, M.B., 2010. First-order fire effects models for land management: overview and issues. Fire Ecol. 6, 131–142.
Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J., Muller, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinf. 12, 77.
Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. Isprs J. Photogrammetry Remote Sens. 67, 93–104.
Ryan, K.C., Reinhardt, E.D., 1988. Predicting postfire mortality of seven western conifers. Can. J. For. Res. 18, 1291–1297.
Ryan, K.C., Knapp, E.E., Varner, J.M., 2013. Prescribed fire in North American forests and woodlands: history, current practice, and challenges. Front. Ecol. Environ. 11 (s1), e15–e24.
Saveland, J.M., Neuenschwander, L.F., 1990. A signal detection framework to evaluate models of tree mortality following fire damage. For. Sci. 36, 66–76.
Thies, W.G., Westlind, D.J., Loewen, M., Brenner, G., 2006. Prediction of delayed mortality of fire-damaged ponderosa pine following prescribed fires in eastern Oregon, USA. Int. J. Wildland Fire 15, 19–29.
Varner, J.M., Hiers, J.K., Ottmar, R.D., Gordon, D.R., Putz, F.E., Wade, D.D., 2007. Overstory tree mortality resulting from reintroducing fire to long-unburned longleaf pine forests: the importance of duff moisture. Can. J. For. Res. 37, 1349–1358.
Woolley, T., Shaw, D.C., Ganio, L.M., Fitzgerald, S., 2012. A review of logistic regression models used to predict post-fire tree mortality of western North American conifers. Int. J. Wildland Fire 21, 1–35.