Ecological Informatics 64 (2021) 101376

Contents lists available at ScienceDirect

Ecological Informatics

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

Machine learning prediction of mortality in the common desert shrub *Encelia farinosa*

Nicholas Q. Bitter *, James R. Ehleringer

University of Utah, School of Biological Sciences, 257 South 1400 East, Salt Lake City, UT 84112, United States of America

ARTICLE INFO

Keywords:
Random Forest
Logistic regression
Plant mortality
Plant survival
Arid ecosystems

ABSTRACT

Two populations of the common shrub *Encelia farinosa* in the northern and southern portions of the Mojave Desert have been surveyed each spring for nearly 40 years, providing an opportunity to assess highly variable shrub mortality in an arid ecosystem. Most of the newly established shrubs experienced mortality during the juvenile stage, with median survival time of about three years in both populations yet, a small number of shrubs lived for at least a dozen years or even decades. Applying machine learning techniques, we predicted shrub mortality at different life-history stages using random forest and logistic regression. First, we examined seedling survival to become yearlings (one-year old plants), finding that less than 3% of seedlings in both populations survived to become established yearling shrubs. Second, we predicted whether or not yearlings would die prior to reaching the mature adult stage (four years old). The models achieved an Area Under the Receiver Operating Characteristic (AUC) in the 0.80 range for the Oatman population (southern Mojave Desert) and 0.90 range for the Death Valley population (northern Mojave Desert). We found yearling characteristics of smaller shrub size, low leaf coverage, and location in specific microsites associated with experiencing mortality before reaching the mature stage. Third, using only the average juvenile plant characteristics over the first four years of life, we predicted whether or not new adult shrubs were likely to experience mortality within the next eight years. The performance in this application achieved AUC in the 0.72 range for both populations. We found adult *Encelia farinosa* shrubs that had juvenile characteristics of smaller size, flowered less frequently, and had smaller inter-plant distances for the Oatman population were associated with increased mortality within the next eight years. Overall, the size of the shrub was the most important feature for the mortality modeling applications. No significant difference in AUC was found for random forest and logistic regression.

1. Introduction

The ability to predict impending plant mortality has been a difficult modeling application in ecology (Hawkes, 2000; Keane et al., 2001; Monserud and Sterba, 1999; Wyckoff and Clark, 2002). The complexity of the underlying mechanisms and lack of test data have contributed to this challenge, leading mortality simulations prone to error (Hawkes, 2000). A majority of previous plant research has been centered around the prediction of forest/tree mortality. Shrubs and arid ecosystems have been relatively neglected (Götmark et al., 2016; Hawkes, 2000), despite the fact that shrubs are the dominant life form in many of Earth's ecosystems (Filazzola and Lortie, 2014). Often serving as keystone species, shrubs are critical to the health of ecosystems they reside in. A better understanding of both the primary drivers and prediction of shrub mortality in arid lands is crucial as desertification becomes increasingly prevalent (Fredrickson et al., 1998).

Logistic regression has been extensively used to model plant mortality (Cailleret et al., 2016; Cao, 2017; Collet and Le Moguedec, 2007; Monserud and Sterba, 1999; Woolley et al., 2012; Yang et al., 2003). Often the goal is to predict mortality for trees at the individual or population scale where dead or surviving trees are usually modeled as a classification task. Other machine learning models applied to plant mortality have effectively been used to model tree mortality (Bayat et al., 2019; da Rocha et al., 2018; Dobbertin and Biging, 1998; Fan et al., 2011; Guan and Gertner, 1991; Hasenauer et al., 2001; Hasenauer and Merkl, 1997; Rao et al., 2019; Reis et al., 2018; Shearman et al., 2019), often using Classification and Regression Trees (CART), Artificial Neural Networks (ANN), and random forest. An advantage of more complex machine learning models is they do not rely on prior assumptions and are well-equipped to handle non-linear relations and complex

* Corresponding author.
E-mail address: nicholas.bitter@utah.edu (N.Q. Bitter).

https://doi.org/10.1016/j.ecoinf.2021.101376
Received 28 February 2021; Received in revised form 17 July 2021; Accepted 17 July 2021
Available online 24 July 2021

interactions that are common in ecological data (Olden et al., 2008). Studies in tree mortality that have compared statistical models identified that machine learning models often outperform conventional logistic and linear models, and/or identify additional features related to plant mortality that are missed by logistic models (Bayat et al., 2019; Dobbertin and Biging, 1998; Guan and Gertner, 1991; Hasenauer and Merkl, 1997; King, 2003; Shearman et al., 2019). A large scale benchmark experiment comparing random forest and logistic regression found that random forest had higher performance in 69% of the datasets (Couronné et al., 2018). When compared to many other machine learning classifiers for species distribution modeling random forest was found to be the best performing in 29 out of 35 datasets (Lorena et al., 2011). However, there also disadvantageous to using more complex machine learning models because they are more difficult to interpret and computationally intensive.

Hawkes (2000) review of woody plant mortality algorithms identified several categories of predictors used in mortality models including age, size, competition, environmental abiotic factors, carbon balance, or a combination of these predictors. In the context of shrub mortality, studies have examined conditions including drought, competition, and soil texture (Bowers, 2005; Ehleringer and Sandquist, 2018; Hamerlynck and McAuliffe, 2008; Miriti et al., 1998, 2001, 2007; Paddock et al., 2013; Renne et al., 2019; Venturas et al., 2016; Winkler et al., 2019). Mortality among shrubs in arid ecosystems has largely been episodic (Bowers, 2005; Miriti et al., 2007; Watson et al., 1997). Shrub size and growth were identified as strong predictors of survival, with larger size favorable in more species (Paddock et al., 2013; Venturas et al., 2016). One possible reason for the size advantage in several species is larger shrubs may have better established root systems, providing a level of resilience in dry conditions. Additionally, juvenile and younger shrubs were associated with higher mortality rates, supporting shrub age as an indicator of expected mortality (Hegazy, 1992; Miriti et al., 2007). Close relations between the biological characteristics of juvenile plants and their success in establishment have also been identified (Cook, 1979). As expected, features associated with mortality include climate parameters, such as precipitation/drought (Bowers, 2005; Miriti et al., 2007; Paddock et al., 2013) where higher temperatures and lower precipitation were associated with higher mortality. Lower specific leaf areas were also identified as unfavorable for survival (Paddock et al., 2013). Spatial distribution differences and neighboring plants competing for water resources have been identified as factors influencing mortality rates and population structure (Miriti et al., 1998, 2001). Juvenile shrubs that had closer proximity to neighboring adult plants died more often than would be expected by chance (Miriti et al., 1998). Population growth has also favored isolated adult plants at a cost to juvenile plants. Additionally, historical conditions experienced by a shrub can affect its mortality in future events. Miriti et al. (2007) demonstrated that shrubs experiencing reduced growth for periods of up to fifteen years in the past can affect future mortality.

The aforementioned studies in shrub mortality and survival have focused largely on species in the American southwest including *Encelia farinosa, Amborsia deltoidea, Amborisa salsola, Amborsia dumosa,* and other species. The focus of this study is *Encelia farinosa,* a drought-deciduous shrub species native to the Mojave and Sonoran Deserts of southwestern North America (Ehleringer, 1984, 1988). In the Mojave Desert, this shrub is common on dry slope microhabitats at elevations up to 800 m, where it often forms mono-specific stands. In the Sonoran Desert, *E. farinosa* occurs on both flat and slope habitats. *Encelia* shrubs can be characterized into several life-history stages: seedling (0–1 years), juveniles (1–4 years), and adults (4+ years). This distinction from different life history stages lies in the difference in size, mortality rates, and flowering status which is uncommon among juveniles. Lastly, some adult *E. farinosa* shrubs can live in excess of thirty years (Ehleringer and Sandquist, 2018).

In this study, we examined seedling mortality of *E. farinosa* and modeled two different mortality applications at different life stages using random forest and logistic regression. In our first modeling application, we predicted whether or not yearlings (defined as 1-year-old shrubs) would die before reaching the mature stage (defined here as four years of age) based on plant physiological traits and spatial distribution at the first year of life. Juvenile plants, such as yearlings, are of particular ecological interest due to the critical role they serve in structuring adult populations (Cook, 1979). In our second modeling application, using only the average plant dimensions and spatial characteristics experienced over the juvenile stage, we predicted whether or not these new adults were likely to die within the next eight years or survive another eight years. Shrubs that achieve a long length of survival are of ecological interest because they contribute the greatest seed output for the birth of future generations. The primary objective of our study was to develop predictive models that can be used to inform researchers in the field of future population structure in the survey sites. The secondary objectives were to identify features of individual *E. farinosa* shrubs that were associated with longer-term survival or mortality at different life stages.

2. Methods

2.1. Data collection and processing

Annual field surveys of two populations of the drought-deciduous perennial shrub *E. farinosa* in the Mojave Desert have been conducted each March since the early 1980's (Ehleringer and Sandquist, 2018). The first population is in Death Valley National Park, approximately 21 km west of Shoshone, CA, and is hereafter referred to as the Death Valley population. The second population is located 10 km south of Oatman, AZ, and is hereafter referred to as the Oatman population. At both sites, *E. farinosa* shrubs form near monospecific stands, with *E. farinosa* comprising >95% of the perennial shrubs and have dominated these sites since population surveys began in 1981.

Within the annual surveys, all *E. farinosa* individuals are classified into one of three different life history categories: (a) seedlings – plants that have germinated within the current growing season, (b) yearlings – plants that have persisted for one year since germinating and remain as juveniles until attaining adulthood, and (c) adults – plants that have persisted for three years and are entering fourth year or older. In annual surveys, seedlings are counted, but not tagged (Ehleringer and Sandquist, 2018). Yearlings, having survived their first year, were easily identified because they have a woody base and were tagged with a unique identifier and their coordinates recorded on a Cartesian grid space (Ehleringer and Sandquist, 2018).

During each survey, several features were recorded on each tagged plant: (a) plant dimensions (width along the widest access, width along a perpendicular axis, and height); (b) the extent of canopy leaf cover development as 0–24%, 25–49%, 50–74%, and 75–100%); and flowering as yes or no. We also noted the presence and extent of any canopy dieback and the presence of dodder (*Cuscuta* sp.), a common parasitic plant, on individual shrubs. From these field data, we subsequently calculated shrub elliptical surface area (hereafter 'surface area'), shrub volume, the number of neighboring plants within 1.5 m, and the distance to nearest neighbor. When a shrub died, its death year was recorded and no further information was collected. We consider all shrubs that had known birth years throughout the 40-year study period so that the age of the shrub could be easily determined.

2.2. Models

We applied two models in the applications. The first is the widely used machine learning model Random forest, interfaced from R package Caret (Kuhn, 2020), with model tag 'rf', utilizing the 'RandomForest' package (Liaw and Wiener, 2002). Random forest algorithm proceeds as follows: ntree bootstrap samples from the original data are sampled. For each of the bootstrap samples, an unpruned classification or regression

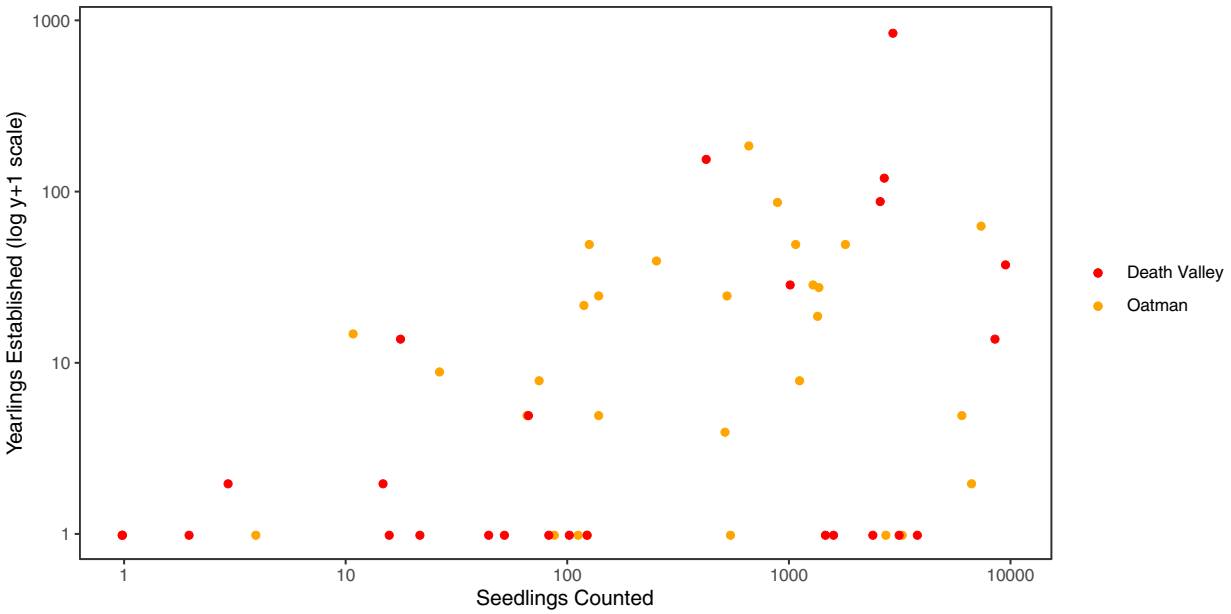

Fig. 1. Seedling establishment to yearlings. N + 1 scale displayed for log plotting.

tree is grown. Then, the predictors are randomly sampled (mtry) and the best split using the randomly sampled predictors is determined. Predictions are made by aggregating the predictions of *n* number of trees made (Liaw and Wiener, 2002). The next model is Logistic Regression, caret model tag 'glm', frequently used in plant mortality modeling. Logistic Regression is a parametric approach where the log odds of the outcome is modeled as a linear combination of predictor variables. Logistic regression makes underline assumptions about the data contrary to non-parametric random forest. The last model we refer to as Guess. In this model, we only predicted the most common outcome that was observed throughout the entire survey period. This can only predict one outcome and outputs no probabilities. To represent an overview of mortality of the shrub, we utilize a simple Kaplan-Meir curve using R package 'Survival', without comparison to any groups. All shrubs with known birth years between 1981 and 2020 were included in the survival curve. The survival time was represented as the number of years survived for newly established (age-1) shrubs that had known birth years. We model mortality for individual shrubs to occur within an interval, such as 3-year mortality for yearlings, using techniques commonly applied to plant mortality modeling and other human health related fields (Abbot, 1985; Hasenauer et al., 2001; Leung et al., 1997; Lund et al., 2019; Vock et al., 2016; Wallert et al., 2017; Yang et al., 2003).

2.3. Model testing

A train-test ratio of 70%–30% was performed where 70% of the data was used for training and 30% for testing. To illustrate feature importance for each of the applications, the Mean Decrease Accuracy from random forest was used for feature importance, where larger values indicate more predictive power (Breiman, 2001). The performance metrics in the test-set include Area Under the Receiver Operating Characteristic (AUC) and its confidence interval (Jin Huang and Ling, 2005), accuracy, sensitivity which we associated with the ability to predict shrubs that died, and specificity which we associated with the ability to predict shrubs that survived. AUC and its confidence interval were calculated with R package "pROC" (Robin et al., 2011). The confidence interval calculation was performed with the DeLong method using function 'ci.auc'. Since small differences in AUC can still be significant if the ROC curves are correlated (Robin et al., 2011) we use the function 'roc.test' in package 'pROC" to compare the classifier that appeared to have the highest AUC for each of the applications. R version 3.6.2 was used in this analysis (R Core Team, 2019). Testing for significance between survival groups was performed using Wilcoxon Rank Sum if continuous and Fisher's exact test for categorical.

2.3.1. Yearling-to-adulthood mortality

The objective of this application was to predict whether or not an individual yearling (age one) would die before reaching the mature adulthood stage (age four). Yearlings in the populations have episodic establishment linked to the presence of seedlings during El Niño events. In many years, few or no yearlings were present, while other years encompass the majority of yearlings. As there were no recent episodic births at the time of writing, the survival outcome of all yearlings reaching adulthood in both populations was known. In the Death Valley population, the majority of coordinate grid information was missing for yearlings born prior to 1997. Consequently, only the Death Valley yearlings born after 1997 were included for analysis. This does result in model performance and associations that are not representative of the entire population of Death Valley yearlings. For the Oatman population, 70% of the data was selected for training and the remainder 30% was selected for testing. For the Death Valley population, 70% of the data was selected for training and the remainder 30% was selected for testing. This resulted in 81 yearlings in the test for Death Valley and 235 yearlings in the test for Oatman. A 5-fold cross-validation was performed on the training data, tuning the number of variables to try at each split for random forest (mtry).Then the model with the highest performance on the 5-fold cross-validation procedure was applied on the test data. In this mortality application, classification threshold was 0.5. As a result, plants with probability of death greater than 50% were predicted to die before reaching adulthood.

2.3.2. Mortality within eight-years for adults

The objective of this application was to examine, for a shrub that did reach maturity, whether or not it would die within the next eight years or survive to an old age. Only the average plant size dimensions and spatial characteristics experienced over the first 4-years were used to generate a prediction. In the Death Valley population, 12.7% of the adult shrubs had missing values and were not included. In the Oatman population, 4.3% of the adult shrubs had missing values and were not included. Not including these observations is known to overestimate the risk of an event (e.g., mortality) occurring (Vock et al., 2016). Additionally, since it is much more common for a shrub to die before age

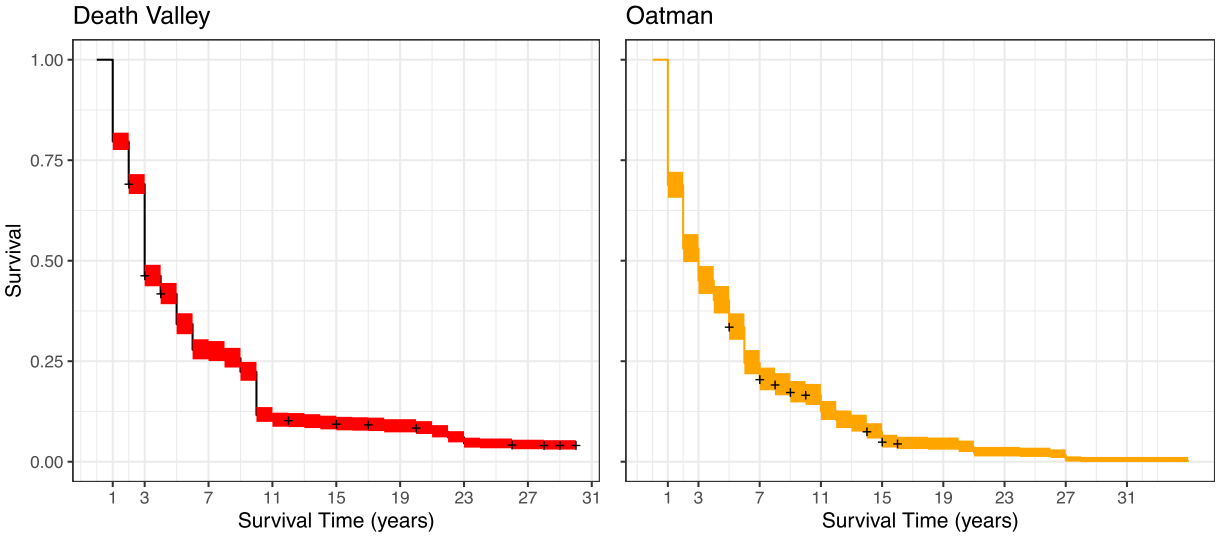

Fig. 2. Kaplan-Meier survival curves for *Encelia farinosa* after becoming established yearling shrubs (reached age-1). "+" indicates censured observations of shrubs that have not yet experienced mortality. Confidence intervals represented by shaded regions along the curve.

twelve then to survive to it, this application represented a class imbalance case. The data was split so that 70% was used for training and 30% was reserved for testing, resulting in 101 plants in the test set for Oatman and 132 plants for Death Valley. Then 5-fold cross-validation was performed on the training data, tuning the number of variables to try at each split for Random Forest (mtry). The classification threshold was set to 70% in this application where probabilities of experiencing mortality higher than 70% were predictions for mortality. Then model with the highest performance in the cross-validation procedure was tested on the test set.

3. Results and discussion

3.1. Seedling survival to become established yearlings

For the Death Valley population, 3.2% of the total seedlings counted over the entire survey period survived to become yearlings. In the Oatman population, 1.9% of the total seedlings counted survived to become yearlings. Fig. 1 shows the relationships between the number of seedlings counted in one year and the yearlings subsequently established in the following year. Extremely high variation was observed in the percentage of seedlings that did survive to the yearling stage at different times. In some cases, it was observed that over 40% of seedlings did survive to the yearling stage during some years when the number of seedlings counted was small. In other cases, over a thousand seedlings were observed yet, none of them survived to become established yearlings.

This very high mortality of seedlings is consistent with of other studies, with one-year seedling survival ranging from <1% to 10% or more in other species of shrubs (Hegazy, 1992; Maestre et al., 2001; Venturas et al., 2016). A variety of factors are likely to contribute to this level of mortality, such as seed predation (Louda, 1983), different soil characteristics across the grid (Maestre et al., 2001), and also the climate. Ehleringer and Sandquist (2018) found the germination of seedlings, leading to the presence of yearlings in the following year, was episodic and linked to El Niño events; survival from seedling to yearling stages increased with higher rainfall.

3.2. Mortality overview of established shrubs

Fig. 2 illustrates an overview of mortality, representing the survival time of *Encelia farinosa* shrubs after the seedling stage. It is clear that at the juvenile stage, observed during the first 3-years after establishment, high mortality is present noted by the steep decline in the estimated survival during these times. For example, a steep decline in survival is visible in the Oatman population where only 70% of established shrubs were estimated to survive one-year to reach age two.

Approximately 57% of the plants died during the juvenile stage,

Table 1
Statistical summary (mean,sd) of mortality before adulthood for all features used.

Features	Death Valley	Survived	p	Oatman	Survived	p
	Mortality before maturity			Mortality before maturity		
Yearling Coordinate X	11.29 (4.62)	9.07 (4.73)	$p < .01$	13.28 (7.64)	12.47 (7.55)	$p > .05$
Yearling Coordinate Y	11.35 (6.39)	11.36 (7.95)	$p > .05$	6.36 (3.87)	6.12 (3.84)	$p > .05$
Yearling Distance to nearest neighbor	0.52 (0.58)	0.58 (0.51)	$p < .05$	0.48 (0.44)	0.45 (0.38)	$p > .05$
Yearling Surface Area	0.01 (0.03)	0.06 (0.05)	$p < .01$	0.007 (0.01)	0.02 (0.03)	$p < .01$
Yearling Neighbors within 1.5 m	8.55 (6.96)	5.77 (4.41)	$p < .01$	8.60 (6.41)	9.28 (6.67)	$p > .05$
Yearling Flower	Not Flowering = 169, Flowering = 5	Not Flowering = 75, Flowering = 24	$p < .01$	Not Flowering = 438, Flowering = 2	Not Flowering = 329, Flowering = 16	$p < .01$
Yearling Leaf Cover	High = 106, low = 68	High = 98, low = 1	$p < .01$	High = 340, low = 100	High = 328, low = 17	$p < .01$

Categorical features represented as counts.

Table 2
Model performance for yearling mortality before adulthood prediction.

Population	Model	AUC	95% CI AUC	Specificity (%)	Sensitivity (%)	Accuracy (%)
Death Valley	RF	0.941	0.888–0.994	86.2	90.4	88.9
Death Valley	LR	0.918	0.858–0.979	69	92.3	84
Death Valley	Guess	0.500	–	0	100	64.2
Oatman	RF	0.809	0.753–0.865	67	79.5	74
Oatman	LR	0.78	0.72–0.84	46.6	88.6	70.2
Oatman	Guess	0.500	–	0	100	56.2

Specificity is the proportion of yearlings that survived to adulthood that were correctly predicted. Sensitivity is the proportion of yearlings that died before adulthood that were correctly predicted. Accuracy is the total number of correct predictions divided by the total number of predictions made. The model types were RF- Random Forest, LR- Logistic Regression.

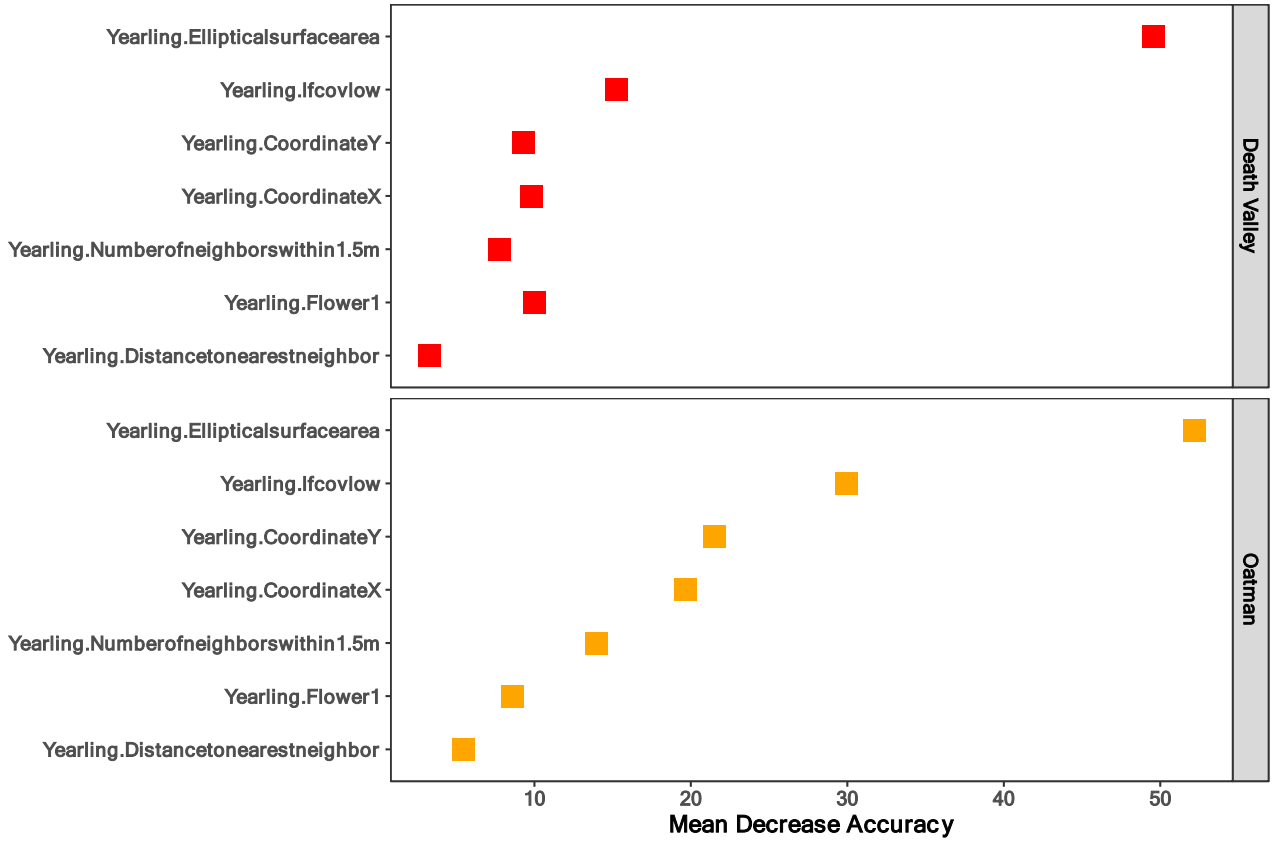

Fig. 3. Mean decrease accuracy values of different plant and microsite features for the mortality of yearlings before reaching adulthood.

failing to reach adulthood (age four) in both populations. The median survival time was about three years in both populations. However, for shrubs that did survive to older ages, the decline in proportion surviving decreased by a smaller amount compared to the juvenile life stage. The probability of surviving eleven years to age twelve was estimated to be 10% in Death Valley and 12% in Oatman. The shrubs that survived beyond eleven years likely possess biologically relevant characteristics of ecological interest, and undoubtedly provided a disproportionally higher proportion of seeds to future generations.

3.3. *Yearling mortality before adulthood prediction*

Table 1 shows a summary of the mortality groups for each population and all features included in the models.

In both populations, mortality before adulthood groups had smaller size, low leaf cover, and flowering. Flowering was not common among yearlings however, having high leaf cover was a common characteristic of the shrubs at this stage. High variation in plant measurements for the mortality groups was present in both populations for yearling surface area and yearling distance to nearest neighbor.

Table 2 illustrates the results of the models for yearling mortality before adulthood prediction.

When comparing the model-to-model performance in different populations, the AUC values of the models in the Death Valley population were significantly higher ($p < .05$) than in the Oatman population. No significant difference in AUC value was found within Death Valley or the Oatman population.

Fig. 3 shows the importance of features. Yearling surface area was by far the most important, followed by microsite (spatial coordinates) and competition features in the Oatman population. According to mean decrease accuracy importance, the X - Y coordinates of the Oatman population had greater importance than the X - Y coordinates in the Death Valley population. Additionally, leaf cover was more influential within the Oatman population than in the Death Valley population.

Fig. 4 illustrates the yearling field observations between survival and mortality before the mature stage.

Having larger surface area was a significant characteristic of yearlings that survived to adulthood in both populations, suggesting that

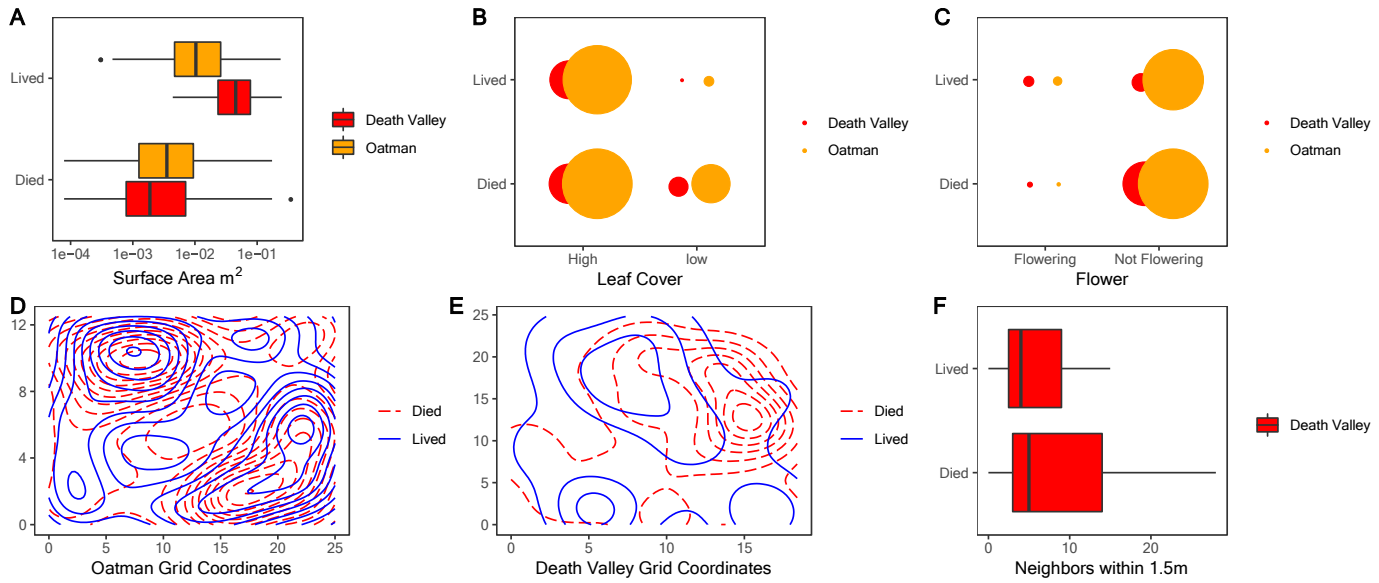

Fig. 4. Yearling field observations of important features influencing mortality before reaching adulthood. The 'Lived' group represents the yearlings that survived to adulthood and the 'Died' group represents the yearlings that died before reaching adulthood. Plate A: Yearling surface area presented on a log scale. Plate B: Yearling leaf cover represented as counts proportional to the size of the marker. Plate C: Yearling flowering status. Plates D and E: Oatman and Death Valley Yearling grid coordinates presented as density contours for visual effect grouped in mortality before adulthood (dashed red line), and survival to adulthood (solid blue). Plate F: The influence of neighboring shrubs on mortality of Death Valley yearlings. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Statistical summary (mean,sd) of 8-year mortality for shrubs that recently reach maturity.

Features	Death Valley	Survived	p	Oatman	Survived	P
	Mortality within eight-years			Mortality within eight-years		
CoordinateX	–	–		12.61 (7.22)	13.67 (8.44)	p > .05
CoordinateY	–	–		6.44 (3.91)	5.69 (3.39)	p > .05
Distancetonearestneighbor	–	–		0.41 (0.34)	0.53 (0.37)	p < .01
Ellipticalsurfacearea	0.06 (0.05)	0.08 (0.05)	p < .01	0.05 (0.068)	0.08 (0.09)	p < .01
Flower	High Flowering = 197, Low Flowering = 222	High Flowering = 93, Low Flowering = 22	p < .01	High Flowering = 66, Low Flowering = 183	High Flowering = 29, Low Flowering = 63	p > .05
Height	25.84 (7.82)	27.92 (6.80)	p < .01	21.76 (9.48)	26.36 (9.84)	p < .01
lfcov	High = 345, low = 74	High = 104, low = 11	p < .05	High = 170, low = 79	High = 57, low = 35	p > .05
Numberofneighborswithin1.5 m	–	–		9.53 (5.49)	7.36 (4.38)	p < .01

Categorical features represented as counts.

smaller yearlings were more susceptible to mortality. Lower leaf cover values were also characteristics of yearlings that died before achieving adulthood. One reason why the models may have achieved a higher performance in the Death Valley population was due to the larger difference between surface area in the shrubs that died before adulthood or survived, observed in Fig. 4 Panel A. It is also apparent there may be favorable and unfavorable microsite spots on the plot for establishment survival when visualizing the grid coordinates. While the upper left region of the Oatman grid (Fig. 4 Panel D) revealed no favorable or unfavorable outcomes for mortality before adulthood, the lower right region of the grid revealed two areas that may be associated with different mortality outcomes. The Death Valley grid coordinates (Fig. 4 panel E) show a greater yearling mortality in the upper right region of the grid, while yearlings that survived to adulthood were more commonly found in other regions of the grid. Reason as to why certain regions on the grid were associated with different mortality outcomes are unknown. However, from other studies it is known that rocks within the soil profile have a negative impact on plant competition for water (Hamerlynck and McAuliffe, 2008), which would be expected to influence mortality. Additionally Fig. 4 panel F shows more neighbors within 1.5 m associated with increased mortality before maturity in the Death Valley population, suggesting that competition with other shrubs may be influencing mortality. The mechanistic relationships as to why smaller yearlings with low leaf coverage at certain plot locations being predisposed to experiencing mortality before adulthood is unknown.

Despite the high performance of the models in this application the confidence intervals for the AUC were wide. This was likely due to the small test data size in the application. Yearling establishment was an uncommon event during the entire survey period. When it did occur, larger numbers of yearlings would be present at seemingly random times associated with El Niño events and higher rainfall (Ehleringer and Sandquist, 2018).

Table 4
New adult plants mortality within eight-years prediction based only on shrub dimensions during first four years of life.

Population	Model	AUC	95% CI AUC	Specificity (%)	Sensitivity (%)	Accuracy (%)
Death Valley	RF	0.641	0.539–0.743	32.4	83.2	72.3
Death Valley	LR	0.765	0.679–0.851	58.8	73.6	70.4
Death Valley	Guess	0.500	–	0	100	78.6
Oatman	RF	0.764	0.659–0.869	81.5	64.9	69.3
Oatman	LR	0.703	0.6–0.806	51.9	73	67.3
Oatman	Guess	0.500	–	0	100	73.3

Specificity is the proportion of shrubs that survived eight more years that were correctly predicted. Sensitivity is the proportion that died within eight years that were correctly predicted. Accuracy is the total number of correct predictions divided by the total number of predictions made. The model types were RF- Random Forest, LR- Logistic Regression.

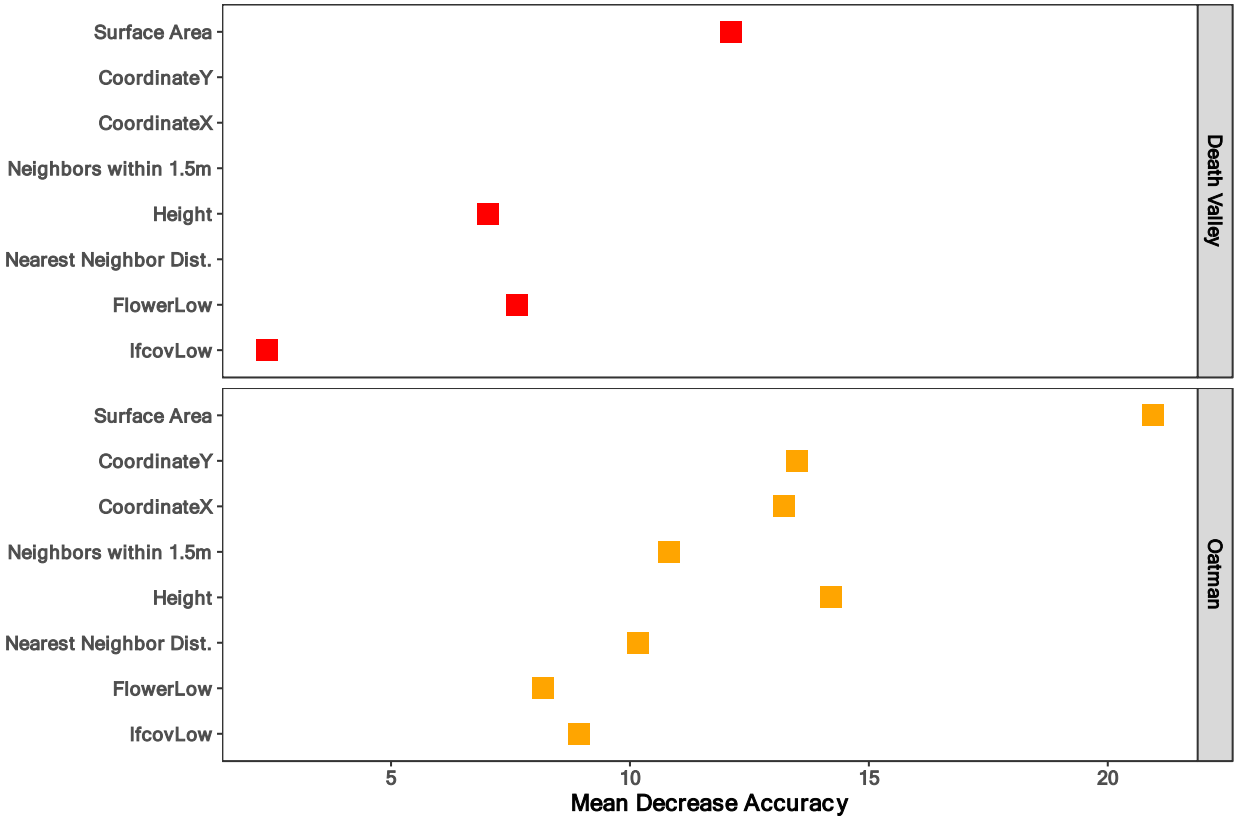

Fig. 5. Mean decrease accuracy values of different plant and microsite features for predicting mortality within eight-years.

3.4. Mortality within eight-years for adults

Table 3 shows the statistical summary of the mean juvenile shrub characteristics for the mortality groups in each population.

Table 4 shows the performance for predicting the adult shrubs that will likely experience mortality within the next eight years.

Table 4 illustrates the capability of the prediction. The AUC suggests a fair ability to predict longer term survival vs mortality within eight-years for new adult shrubs. There was also no significant difference in AUC between the populations. No significant different in AUC within the Death Valley population was found. Large uncertainty was found noted by the wide confidence intervals in AUC. This was indicative of the smaller test size for the application. This lack of test data is common in many ecological applications (Hawkes, 2000) and the uncertainty is reflected in the wide confidence interval.

Fig. 5 shows the morphological and microsite features of importance for mortality within eight years for new adults. Mean decrease accuracy values for the Death Valley population remain low compared to Oatman suggesting an overall lower predictive power from Random forest in the Death Valley population. It is also clear that surface area during the first four years of life was an important feature for predicting survival for another eight years in both populations. Flower was also an important feature. Leaf cover was of predictive importance in the Oatman population, but not in the Death Valley population. Neighbors within 1.5 m and XY coordinates were important features for prediction in the Oatman population. This suggests that not being able to utilize spatial information in the Death Valley population for this application may have resulted in a loss of predictive power.

Microsite (spatial coordinates) also played a role in the Oatman population. Fig. 5 shows field observations of important features for predicting eight-year mortality. Once again, we observed that plant surface area was a highly informative parameter for predicting survival. *E. farinosa* shrubs that achieved larger size over the first four years of life were better able to achieve greater longevity. Consequently, smaller young adult shrubs had shorter lifespans. This is consistent with shrub mortality literature that has identified smaller shrubs of several different species less able to survive harsh environmental conditions and more susceptible to mortality (Hamerlynck and McAuliffe, 2008; Venturas et al., 2016). Additionally, adult plants in Death Valley that flowered less often over the first four years of life, was associated with

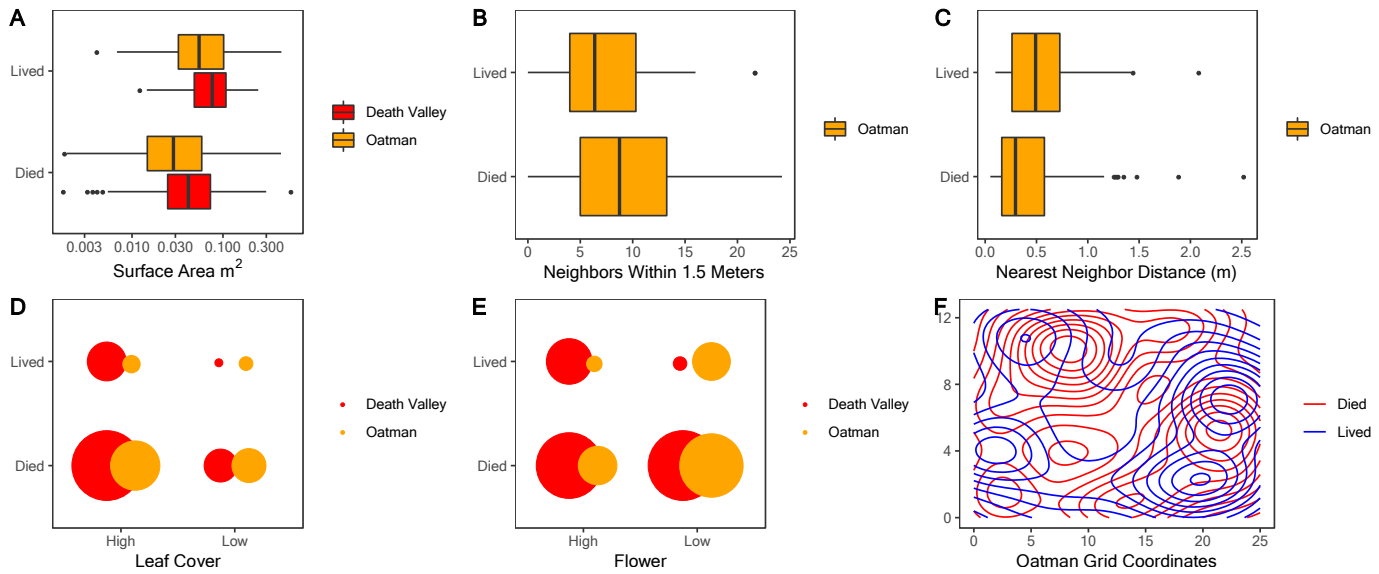

Fig. 6. Field observations averages over first four years of life for predicting mortality within eight years. The 'Died' group represents the new adult shrubs that did not have eight-year survival. The 'Lived' group represents the new adult shrubs that survived another eight years. Plate A: 4-year average Surface Area presented on a log scale. The influences of 4-year averages of neighbors within 1.5 m (Plate B), nearest neighbors (Plate C), leaf coverage represented as counts proportional to the size of the marker (Plate D), and flower (Plate E). Plate F shows density contours of shrub survival and mortality across the Oatman plot.

experiencing mortality within eight-years. This relationship may be due to larger plants flowering more often. The spatial features in the Oatman population also revealed clear relationships. Once again, the upper left and bottom right regions of the plot contain a majority of the shrubs in the Oatman population (Fig. 6 Panel F). However, plant microsites adjacent to these 'hot spots' for shrub growth may be more favorable for longer term survival. Additionally, young adult plants that grew with fewer nearby neighbors and with greater distance from nearest neighbors were also associated with a greater eight-year survival within the Oatman population (Fig. 5 Panel B and C). These characteristics suggest competitive interactions were apparent and that isolated adult plants were better able to achieve longer-term survival, likely because of less competition for water. A competitive interaction has been previously identified among adult *Encelia* shrubs also in the Oatman area (Ehleringer, 1984), where adult plants whose neighbors within two meters were removed, exhibited great growth rates, higher leaf areas, and great flowering than control plants. The impacts of competition in desert shrub populations have been more broadly discussed by (Miriti et al., 2001).

It has been suggested that desert shrubs that survive to older ages are more likely to persist in more favorable microsites, have greater carbon reserves, and better developed root systems (Hamerlynck and McAuliffe, 2008). Miriti et al. (2001) also suggested that competitive interactions may influence survival during drought-like conditions, a common feature of arid land ecosystems. The data in Fig. 6 Plate B, C, and F indicate that pattern is apparent in the Oatman population.

4. Conclusion

In this study, we achieve a more predictive understanding of mortality at different stages in the life history of *Encelia farinosa*. We demonstrated that mortality at different life stages was predictable while illuminating which morphological, microsite, and inter-plant features contributed to mortality. In particular, the importance of shrub size for predicting mortality was apparent. Comparatively smaller *E. farinosa* shrubs were more susceptible to mortality at several different life stages.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study has been supported by NSF Grant DEB-1950025.

References

Abbot, R.D., 1985. Logistic regression in survival analysis. Am. J. Epidemiol. 121 (3), 465–471. https://doi.org/10.1093/oxfordjournals.aje.a114019.

Bayat, M., Ghorbanpour, M., Zare, R., Jaafari, A., Thai Pham, B., 2019. Application of artificial neural networks for predicting tree survival and mortality in the Hyrcanian forest of Iran. Comput. Electron. Agric. 164, 104929. https://doi.org/10.1016/j.compag.2019.104929.

Bowers, J.E., 2005. Effects of drought on shrub survival and longevity in the northern Sonoran Desert. J. Torrey Bot. Soc. 132 (3), 421–431 (JSTOR).

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Cailleret, M., Bigler, C., Bugmann, H., Camarero, J.J., Cufar, K., Davi, H., Mészáros, I., Minunno, F., Peltoniemi, M., Robert, E.M.R., Suarez, M.L., Tognetti, R., Martínez-Vilalta, J., 2016. Towards a common methodology for developing logistic tree mortality models based on ring-width data. Ecol. Appl. 26 (6), 1827–1841. https://doi.org/10.1890/15-1402.1.

Cao, Q.V., 2017. Evaluation of methods for modeling individual tree survival. For. Sci. 63 (4), 356–361. https://doi.org/10.5849/forsci.2016-001.

Collet, C., Le Moguedec, G., 2007. Individual seedling mortality as a function of size, growth and competition in naturally regenerated beech seedlings. Forestry 80 (4), 359–370. https://doi.org/10.1093/forestry/cpm016.

Cook, R.E., 1979. Patterns of juvenile mortality and recruitment in plants. In: Solbrig, O. T., Jain, S., Johnson, G.B., Raven, P.H. (Eds.), *Topics in Plant Population Biology* (pp. 207–231). https://doi.org/10.1007/978-1-349-04627-0_10. Macmillan Education UK.

Couronné, R., Probst, P., Boulesteix, A.-L., 2018. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics 19 (1), 270. https://doi.org/10.1186/s12859-018-2264-5.

da Rocha, S.J.S.S., Torres, C.M.M.E., Jacovine, L.A.G., Leite, H.G., Gelcer, E.M., Neves, K. M., Schettini, B.L.S., Villanova, P.H., da Silva, L.F., Reis, L.P., Zanuncio, J.C., 2018. Artificial neural networks: modeling tree survival and mortality in the Atlantic Forest biome in Brazil. Sci. Total Environ. 645, 655–661. https://doi.org/10.1016/j.scitotenv.2018.07.123.

Dobbertin, M., Biging, G.S., 1998. Using the non-parametric classifier CART to model forest tree mortality. For. Sci. 44 (4), 507–516. https://doi.org/10.1093/forestscience/44.4.507.

Ehleringer, J.R., 1984. Intraspecific competitive effects on water relations, growth and reproduction in Encelia farinosa. Oecologia 63 (2), 153–158 (JSTOR).

Ehleringer, J.R., 1988. Comparative ecophysiology of Encelia farinosa and Encelia frutescens. Oecologia 76 (4), 553–561 (JSTOR).
Ehleringer, J.R., Sandquist, D.R., 2018. A tale of ENSO, PDO, and increasing aridity impacts on drought-deciduous shrubs in the Death Valley region. Oecologia 187 (4), 879–895. https://doi.org/10.1007/s00442-018-4200-9.
Fan, Z., Fan, X., Spetich, M.A., Shifley, S.R., Moser, W.K., Jensen, R.G., Kabrick, J.M., 2011. Developing a stand Hazard index for oak decline in upland oak forests of the Ozark highlands, Missouri. North. J. Appl. For. 28 (1), 19–26. https://doi.org/10.1093/njaf/28.1.19.
Filazzola, A., Lortie, C.J., 2014. A systematic review and conceptual framework for the mechanistic pathways of nurse plants. Glob. Ecol. Biogeogr. 23 (12), 1335–1345. https://doi.org/10.1111/geb.12202.
Fredrickson, E., Havstad Kris, M., Estell, R., Hyder, P., 1998. Perspectives on desertification: South-Western United States. J. Arid Environ. 39 (2), 191–207. https://doi.org/10.1006/jare.1998.0390.
Götmark, F., Götmark, E., Jensen, A.M., 2016. Why be a shrub? A basic model and hypotheses for the adaptive values of a common growth form. Front. Plant Sci. 7 https://doi.org/10.3389/fpls.2016.01095.
Guan, B.T., Gertner, G., 1991. Using a parallel distributed processing system to model individual tree mortality. For. Sci. 37 (3), 871–885. https://doi.org/10.1093/forestscience/37.3.871.
Hamerlynck, E.P., McAuliffe, J.R., 2008. Soil-dependent canopy die-back and plant mortality in two Mojave Desert shrubs. J. Arid Environ. 72 (10), 1793–1802. https://doi.org/10.1016/j.jaridenv.2008.05.002.
Hasenauer, H., Merkl, D., 1997. Forest tree mortality simulation in uneven-aged stands using connectionist networks. In: Proceedings of the International Conference on Engineering Applications of Neural Networks, pp. 341–348.
Hasenauer, H., Merkl, D., Weingartner, M., 2001. Estimating tree mortality of Norway spruce stands with neural networks. Adv. Environ. Res. 5 (4), 405–414. https://doi.org/10.1016/S1093-0191(01)00092-2.
Hawkes, C., 2000. Woody plant mortality algorithms: description, problems and progress. Ecol. Model. 126 (2), 225–248. https://doi.org/10.1016/S0304-3800(00)00267-2.
Hegazy, A.K., 1992. Age-specific survival, mortality and reproduction, and prospects for conservation of Limonium delicatulum. J. Appl. Ecol. 29 (3), 549–557. JSTOR. https://doi.org/10.2307/2404462.
Huang, Jin, Ling, C.X., 2005. Using AUC and accuracy in evaluating learning algorithms. IEEE Trans. Knowl. Data Eng. 17 (3), 299–310. https://doi.org/10.1109/TKDE.2005.50.
Keane, R.E., Austin, M., Field, C., Huth, A., Lexer, M.J., Peters, D., Solomon, A., Wyckoff, P., 2001. Tree mortality in gap models: Application to climate change. Clim. Chang. 51 (3–4), 509–540.
King, S.L., 2003. Using ROC curves to compare neural networks and logistic regression for modeling individual noncatastrophic tree mortality. In: Van Sambeek, J.W., Dawson Jeffery, O., Ponder Jr., Felix, Loewenstein Edward, F., Fralish James, S. (Eds.), Proceedings of the 13th Central Hardwood Forest Conference; Gen. Tech. Rep. NC-234, 234. U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN, pp. 349–358. https://www.fs.usda.gov/treesearch/pubs/15774.
Kuhn, M., 2020. caret: Classification and Regression Training. https://CRAN.R-project.org/package=caret.
Leung, K.-M., Elashoff, R.M., Afifi, A.A., 1997. Censoring issues in survival analysis. Annu. Rev. Public Health 18 (1), 83–104. https://doi.org/10.1146/annurev.publhealth.18.1.83.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2 (3), 18–22.
Lorena, A.C., Jacintho, L.F.O., Siqueira, M.F., Giovanni, R.D., Lohmann, L.G., de Carvalho, Yamamoto, M., 2011. Comparing machine learning classifiers in potential distribution modelling. Expert Syst. Applicat. 38 (5), 5268–5275. https://doi.org/10.1016/j.eswa.2010.10.031.
Louda, S.M., 1983. Seed predation and seedling mortality in the recruitment of a shrub, Haplopappus Venetus (Asteraceae), along a climatic gradient. Ecology 64 (3), 511–521. JSTOR. https://doi.org/10.2307/1939971.
Lund, J.L., Kuo, T.-M., Brookhart, M.A., Meyer, A.-M., Dalton, A.F., Kistler, C.E., Wheeler, S.B., Lewis, C.L., 2019. Development and validation of a five-year mortality prediction model using regularized regression and Medicare data. Pharmacoepidemiol. Drug Saf. 28 (5), 584–592. https://doi.org/10.1002/pds.4769.
Maestre, F.T., Bautista, S., Cortina, J., Bellot, J., 2001. Potential for using facilitation by grasses to establish shrubs on a semiarid degraded steppe. Ecol. Appl. 11 (6), 1641–1655. https://doi.org/10.1890/1051-0761(2001)011[1641:PFUFBG]2.0.CO;2.
Miriti, M.N., Howe, H.F., Wright, S.J., 1998. Spatial patterns of mortality in a Colorado desert plant community. Plant Ecol. 136 (1), 41–51. https://doi.org/10.1023/A:1009711311970.
Miriti, M.N., Wright, S.J., Howe, H.F., 2001. The effects of neighbors on the demography of a dominant desert shrub (ambrosia Dumosa). Ecol. Monogr. 71 (4), 491–509. https://doi.org/10.1890/0012-9615(2001)071[0491:TEONOT]2.0.CO;2.
Miriti, M.N., Rodríguez-Buriticá, S., Wright, S.J., Howe, H.F., 2007. Episodic death across species of desert shrubs. Ecology 88 (1), 32–36. https://doi.org/10.1890/0012-9658(2007)88[32:EDASOD]2.0.CO;2.
Monserud, R.A., Sterba, H., 1999. Modeling individual tree mortality for Austrian forest species. For. Ecol. Manag. 113 (2), 109–123. https://doi.org/10.1016/S0378-1127(98)00419-8.
Olden, J.D., Lawler, J.J., Poff, N.L., 2008. Machine learning methods without tears: a primer for ecologists. Q. Rev. Biol. 83 (2), 171–193. https://doi.org/10.1086/587826.
Paddock, W., Davis, S., Pratt, R., Jacobsen, A., Tobin, M., López-Portillo, J., Ewers, F., 2013. Factors determining mortality of adult chaparral shrubs in an extreme drought year in California. Aliso 31 (1), 49–57. https://doi.org/10.5642/aliso.20133101.08.
R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/.
Rao, K., Anderegg, W.R.L., Sala, A., Martínez-Vilalta, J., Konings, A.G., 2019. Satellite-based vegetation optical depth as an indicator of drought-driven tree mortality. Remote Sens. Environ. 227, 125–136. https://doi.org/10.1016/j.rse.2019.03.026.
Reis, L.P., de Souza, A.L., dos Reis, P.C.M., Mazzei, L., Soares, C.P.B., Miquelino Eleto Torres, C.M., da Silva, L.F., Ruschel, A.R., Rêgo, L.J.S., Leite, H.G., 2018. Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest. Ecol. Eng. 112, 140–147. https://doi.org/10.1016/j.ecoleng.2017.12.014.
Renne, R.R., Schlaepfer, D.R., Palmquist, K.A., Bradford, J.B., Burke, I.C., Lauenroth, W.K., 2019. Soil and stand structure explain shrub mortality patterns following global change–type drought and extreme precipitation. Ecology 100 (12), e02889. https://doi.org/10.1002/ecy.2889.
Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., Müller, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12 (1), 77. https://doi.org/10.1186/1471-2105-12-77.
Shearman, T.M., Varner, J.M., Hood, S.M., Cansler, C.A., Hiers, J.K., 2019. Modelling post-fire tree mortality: can random forest improve discrimination of imbalanced data? Ecol. Model. 414, 108855. https://doi.org/10.1016/j.ecolmodel.2019.108855.
Venturas, M.D., MacKinnon, E.D., Dario, H.L., Jacobsen, A.L., Pratt, R.B., Davis, S.D., 2016. Chaparral shrub hydraulic traits, size, and life history types relate to species mortality during California's historic drought of 2014. PLoS One 11 (7), e0159145. https://doi.org/10.1371/journal.pone.0159145.
Vock, D.M., Wolfson, J., Bandyopadhyay, S., Adomavicius, G., Johnson, P.E., Vazquez-Benitez, G., O'Connor, P.J., 2016. Adapting machine learning techniques to censored time-to-event health record data: a general-purpose approach using inverse probability of censoring weighting. J. Biomed. Inform. 61, 119–131. https://doi.org/10.1016/j.jbi.2016.03.009.
Wallert, J., Tomasoni, M., Madison, G., Held, C., 2017. Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data. BMC Med. Inform. Decis. Making 17 (1), 99. https://doi.org/10.1186/s12911-017-0500-y.
Watson, I.W., Westoby, M., Holm, A. McR, 1997. Continuous and episodic components of demographic change in arid zone shrubs: models of two Eremophila species from Western Australia compared with published data on other species. J. Ecol. 85 (6), 833–846. JSTOR. https://doi.org/10.2307/2960605.
Winkler, D.E., Belnap, J., Hoover, D., Reed, S.C., Duniway, M.C., 2019. Shrub persistence and increased grass mortality in response to drought in dryland systems. Glob. Chang. Biol. 25 (9), 3121–3135. https://doi.org/10.1111/gcb.14667.
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), 1–35. https://doi.org/10.1071/WF09039.
Wyckoff, P.H., Clark, J.S., 2002. The relationship between growth and mortality for seven co-occurring tree species in the southern Appalachian Mountains. J. Ecol. 90 (4), 604–615.
Yang, Y., Titus, S.J., Huang, S., 2003. Modeling individual tree mortality for white spruce in Alberta. Ecol. Model. 163 (3), 209–222. https://doi.org/10.1016/S0304-3800(03)00008-5.