IDENTIFYING LANDTYPE PHASES FOR OREGON WHITE OAK RESTORATION IN

THE WILLAMETTE NATIONAL FOREST, OREGON

by

LINDSEY MABEL KURTZ

A THESIS

Presented to the Department of Landscape Architecture
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Master of Landscape Architecture

June 2022

THESIS APPROVAL PAGE

Student: Lindsey Mabel Kurtz

Title: Identifying Landtype Phases for Oregon White Oak Restoration in the Willamette National Forest, Oregon

This thesis has been accepted and approved in partial fulfillment of the requirements for the Master of Landscape Architecture degree in the Department of Landscape Architecture by:

Bart Johnson	Chairperson
Leslie Ryan	Member

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2022

THESIS ABSTRACT

Lindsey Mabel Kurtz

Master of Landscape Architecture

Department of Landscape Architecture

June 2022

Title: Identifying Landtype Phases for Oregon White Oak Restoration in the Willamette National Forest, Oregon

Ecological classification systems are used to understand and restore complex heterogeneous landscapes. We explored an ecological classification methodology to determine fine-grained land units by combining field and remote sensing data. Regression trees were used to create these land units, which we term landtype phases. Oregon white oak was chosen as a test case for the methodology because of its conservation importance, the paucity of knowledge about how to sustain it in heterogeneous landscapes, and its wide range of growing conditions. We identified two landtype phases, the moist margins of harsh meadows and cooler locations away from the meadows. The fieldwork-based variables used to identify and classify these landtype phases were translated into remote-sensing variables using LiDAR, which allowed landtype phase mapping. Our results demonstrate how an integration of field-based and LiDAR-based approaches can provide useful guidance for restoration while highlighting the need for improved translation among the two data types.

This thesis includes unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Lindsey Mabel Kurtz

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of Wisconsin-Eau Claire

DEGREES AWARDED:

Master of Landscape Architecture, 2022, University of Oregon
Bachelor of Science, Environmental Geography, 2018, University of Wisconsin-Eau Claire

AREAS OF SPECIAL INTEREST:

Geospatial Information Systems/Remote Sensing

Restoration Ecology

PROFESSIONAL EXPERIENCE:

Research Assistant, University of Oregon, June 2020 - Present

Graduate Teaching Employee, University of Oregon, January 2020 - Present

Museum Educator, University of Oregon, September 2019 - Present

GRANTS, AWARDS, AND HONORS:

A. Brian Mostue Memorial Scholarship in Landscape Architecture, University of Oregon, 2021

Barbara Fealy Student Scholarship in Landscape Architecture, University of Oregon, 2020

David S. Easly Grant, University of Oregon, 2021

Dennis Hickok Memorial Fund, University of Oregon, 2021

RM Tollefson Architecture Scholarship, University of Oregon, 2020

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I: Creating Landtype Phases

I intend to publish an adaptation of Chapter I of my thesis in an academic journal with co-authors Bart Johnson, Scott Bridgham, and Eyrie Horton. Bart Johnson and Eyrie Horton helped determine field, lab, and statistical procedures used in this work. Scott Bridgham provided data necessary for this project. I am responsible for all written text with Bart Johnson as the primary editor. I performed all analysis reported in the results and was a primary contributor to data collection and preparation.

1 | Introduction

One of the central concepts of landscape ecology is heterogeneity (With 2019). Landscape heterogeneity, defined here as the "variation in biotic and abiotic conditions across space and time" (Wiens 1997) is complex and has been found to be a result of strong interactions between physiography, soils, vegetation, (Barnes et al. 1982) and human management (Clewell & Aronson 2013). Barnes et al. (1982) demonstrated how a multi-factor site classification system could produce a mapped set of recurring ecosystem units for site management. Their system was based largely on using vegetation as an indicator of underlying site physical conditions and then using site physical conditions as the basis for mapping ecosystem units. In particular, they relied on how physiography often controls microenvironmental conditions and water movement and how landforms influence, and are influenced, by soil type. They also incorporated how soil characteristics such as moisture, nutrients, and pH control plant composition, size, and productivity as well as the reciprocal relationship that vegetation, through its composition, size, and productivity, integrates and reflects physiography and soil characteristics.

The study of these ecosystem interactions and relationships is the basis for the field of restoration ecology and the assisted recovery of degraded ecosystems - those that have lost biodiversity and have had disruptions in in their structure, composition and functionality through chronic human impacts (Society for Ecological Restoration 2022). However, applying knowledge of interrelationships in landscape heterogeneity is not easily done. Restoration ecologists often use classification systems to help make comprehensible the overwhelming complexity of a landscape's heterogeneity, including the interrelationships of environmental variables and vegetation distribution (Abella & Covington 2006). Classification systems accomplish this by dividing the landscape into smaller homogeneous units. For these systems to be of use to restoration ecologists, they must identify ecological units that will place the landscape back on a trajectory toward recovery within its full range of historical variability (Abella & Covington 2006; Moore et al. 1999; Morgan et al. 1994).

Recovering an ecosystem's range of historical variability is a better target than the re-creation of historic ecosystems because landscapes are not static; they have changed and will continue to change in the face of many factors, including climate change. In light of this, the common use of current vegetation as the basis for site classification systems (Abella et al. 2003; Falco & Waring 2020) is extremely problematic, especially if the reference ecosystem's vegetation has been degraded or destroyed. One way to use vegetation to understand the specific environmental conditions that supported an ecosystem's historic range of variability is through analyzing the landscape legacies of plants that were living under reference ecosystem or benchmark ecological conditions.

The legacies of past vegetation can be diverse and are not always easy to identify. Furthermore, the evidence of such legacies typically degrades over time. A frequent source of

13

such evidence in forested landscapes of North America that are targeted for restoration are trees that were alive prior to the time of Euro-American-settlement (Abella & Covington 2006), which is often considered to represent an inflection point of radical departure from historical vegetation and the natural and anthropogenic disturbances that sustained or altered it. Analysis of these older trees can tell us not only the range of historical tree densities and sizes but also their distribution across environmental gradients. One of the key decisions is to align the choice of legacies with statistical techniques suitable to classifying ecological site units and mapping them on a landscape.

A variety of statistical methods including principal coordinate analysis (PCA), cluster analysis, and classification and regression trees (CARTs) have been utilized in classification systems to create these ecological units (Abella et al. 2003; Mora & Iverson 2002; Falco & Waring 2020). CARTs in particular are adept at processing ecological data because they can handle non-linear relationships, missing values for both explanatory and response variables, and outliers (Moisen 2008). In addition, CART models can both describe current data and predict future data. Past studies have also shown the ability of CARTs to identify and predict species' habitat and use model values to translate classifications to entire study sites (De'ath & Fabricius 2000; Bourg et al. 2005).

The data used in classification systems is generally gathered from the field or from remote sensing technologies (Mora & Iverson 2002; Barnes et al. 1982; Falco & Waring 2020; Andrew & Ustin 2009). As geospatial technologies have improved, their use in classification systems has increased because of the ability to map across entire landscapes using tools such as light detection and ranging (LiDAR) (Andrew & Ustin 2009). However, field data still has advantages, particularly in providing fine-grained, validated data that may not be possible to

obtain via remote sensing (i.e., soil characteristics). It is therefore advantageous to harness the specificity and validation of fine-grained field data to the site-wide mapping capabilities of remote sensing. So, while classification systems have been created using CARTs with field data (De'ath & Fabricius 2000) or remote sensing data (Falco & Waring 2020; Bourg et al. 2005), there currently are few if any studies that have incorporated both field and remote sensing data in a singular classification system.

We therefore sought to develop a tractable methodology to determine mappable, fine-grained, land units for site restoration and management that integrated field data and remote sensing data (Figure 1). Regression trees, a variant of decision trees, were chosen as the statistical method to guide the creation of these land units, which we termed landtype phases to follow the ecological classification hierarchy created by the USFS (Boyce & Haney 1997).
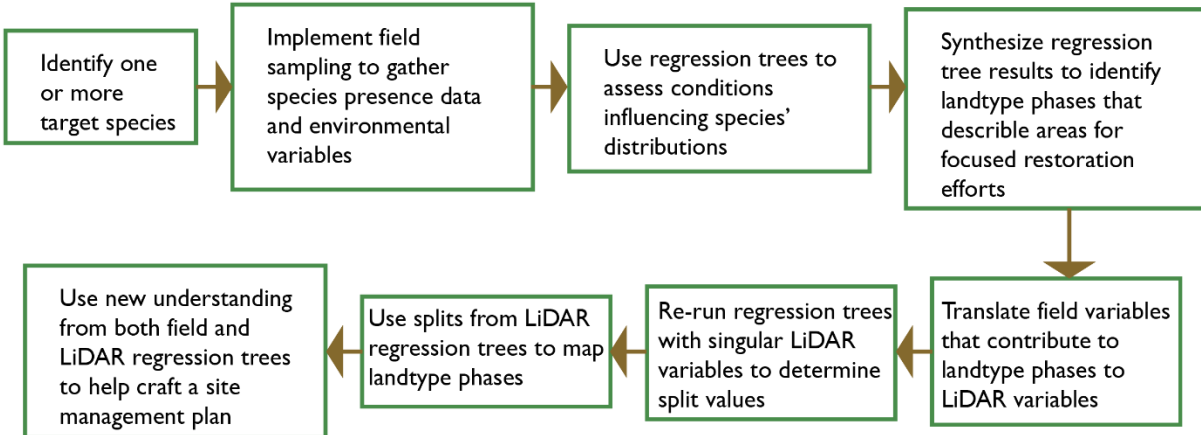

Figure 1. Flow chart of the proposed methodology from species identification to site management.

Oregon white oak (*Quercus garryana*) was chosen as a test case for the methodology because of its conservation importance, the paucity of knowledge about how to sustain it in the heterogeneous landscapes of the Western Oregon Cascades, and its wide range of growing

conditions. As the dominant tree species of formerly extensive savanna grasslands in the interior valleys west of the Cascade Mountains (Christy & Alverson 2011), its acorns were a key food resource of indigenous peoples and oaks provided many additional resources to wildlife (Vesely et al. 2004). However, 90 percent of these historic grasslands have been lost in the last 150 years due to agriculture, urbanization, and forest succession (Oregon Department of Fish and Wildlife 2016). Both oak savanna and oak woodland are listed as high priorities for conservation and restoration in the Oregon (Oregon Department of Fish and Wildlife 2016) and more broadly in the Pacific Northwest (U.S. Fish and Wildlife Service 2010). Oak savanna is of particular importance due to its high biodiversity of plants, mammals, birds and invertebrates (Vesely & Rosenberg 2010) and its importance to the cultures and livelihoods of the indigenous peoples of the region (Willamette Valley Oak and Prairie Cooperative 2020). It is therefore important to understand oak distribution on the landscape to restore these historic grasslands.

In localized areas of the western Oregon Cascades, successional infill by Douglas-fir has changed former oak-pine savanna grasslands to a closed conifer forest with the accompanying decline and death of oaks and pines (USDA Forest Service 2006). These areas are uncommon amid the otherwise conifer-dominant Cascade rainforest, where they provided key indigenous cultural services and wildlife resources, making them an important but particularly challenging conservation target. The historical composition and structure of open-canopied landscapes of the Cascades have been understood primarily through written historical accounts and research on existing tree stands, both of which suggest that these areas were still sparsely forested in the late 1800s due to a legacy of indigenous burning (USDA Forest Service 2006; Bailey & Kertis 2002). With the loss of these low-severity fire regimes, most areas of savanna quickly converted to closed canopy forest, leaving only partial and incomplete evidence of where oaks were growing.

Although Oregon white oak is thus a restoration target in portions of the Western Cascades, it is difficult to understand exactly where to restore it and how to maintain it because of its wide range of potential growing conditions from xeric to regularly flooded sites (Gucker 2007; Vesely et al. 2004) and a rich cultural and ecological history. Even if one knows the general outlines of an appropriate site, where within that site to place oaks is a priority. If Oregon white oak is to be restored as a keystone of savanna ecosystem restoration, it would be useful to determine where on the landscape it was able to establish and persist over long periods of time. Such knowledge may be critical to achieving ecological goals and is essential for efficient use of time and resources. Classification systems are a tool with the potential to help land managers in restoration identify ecological units where oaks are establishing and persisting on this heterogeneous landscape.

2 | Methods

Site Description

The area known today as Jim's Creek Restoration Area (Jim's Creek) has seen different ecosystems and human management over hundreds and likely thousands of years, with an apparent radical change in management and vegetation since the mid-nineteenth century (Figure 2).
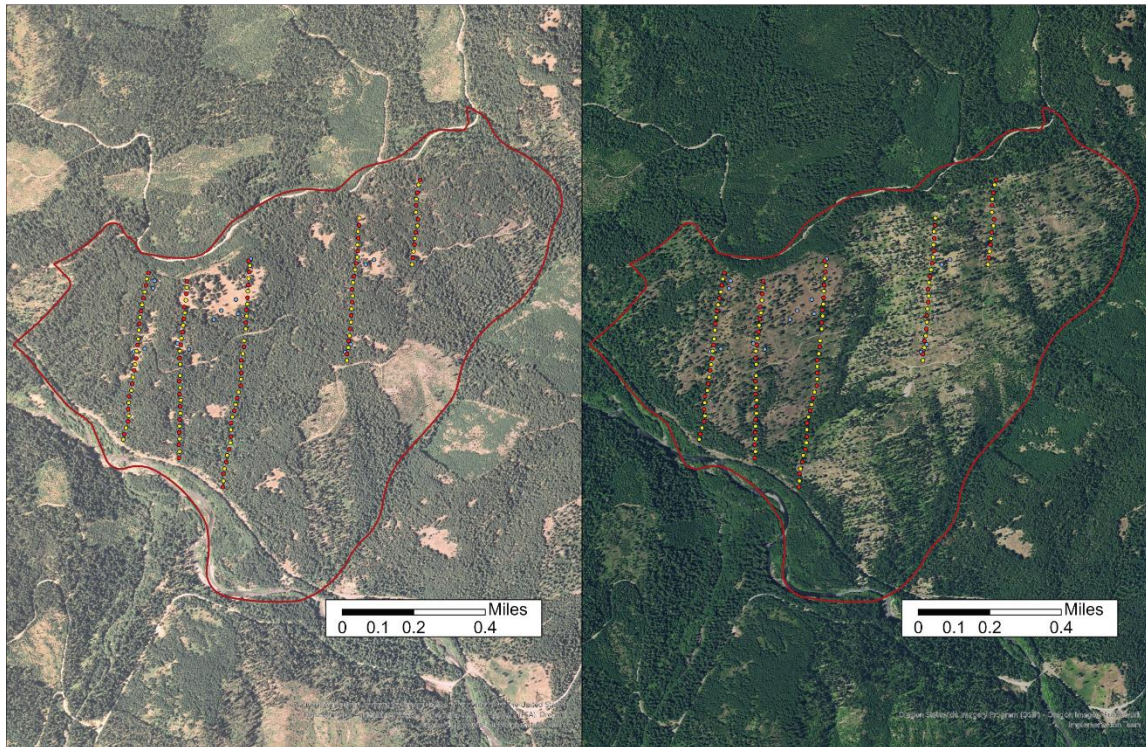
Figure 2. Map of Jim's Creek pre-restoration (left) and post-restoration (right) with 30-m, 60-m, and meadow and transition plots set in five 30-m belt transects covering all major environmental microhabitats.

Prior to Euro-American settlement, the 683-acre site, situated on what is now known as the upper Middle Fork of the Willamette River, functioned as a summer camp to local indigenous peoples and was likely subject to indigenous burning that maintained an open savanna and Oregon white oaks (USDA Forest Service 2006). The onset of Euro-American settlement saw the removal of indigenous peoples from the site and the open savanna transitioned into a closed canopy Douglas-fir forest in the absence of regular burning (USDA Forest Service 2006). Starting in 2005 UO researchers worked with the Middle Fork Ranger District of the United States Forest Service (USFS) and the Southern Willamette Forest Collaborative (SWFC) to restore Jim's Creek to its former open savanna ecosystem. Five random, stratified 30-m wide belt transects oriented north-south and covering 3.3 kilometers were placed to cover all major environmental microhabitats on the site. The transects are

composed of 128 plots with an additional 15 meadow and forest to meadow transition plots placed at various distances from the five transects. These plots serve as the basis for all data collected on-site. From 2008 to 2010, around ninety percent of the trees were removed excluding all Oregon white oaks, ponderosa pine (*Pinus ponderosa*), sugar pine (*Pinus lambertinii*), and Douglas-fir (*Pseudotsuga menziesii*) >75 cm diameter at breast height (DBH) (Johnson 2005). From 2010 to present the USFS has conducted prescribed burns on most of the site to maintain the historically open-canopied savanna structure.

Prior Completed Data Collection

Plot-level soil, site physiography, and vegetation data (including data on oak trees) was collected in prior studies. Maximum depth to obstruction (a surrogate for soil depth), pH, carbon, nitrogen, sand, silt and clay content were all collected at each 60-m plot (Murphy 2008). To calculate depth to obstruction, nine probes were drilled in each plot with 3.16" metal rods, with the maximum value being recorded. Site physiography variables were collected at each 60-m plot location in a 200-m^2 circular area. Of those variables, only a subset was included for analysis (Table 1). Vegetation structure and composition was also recorded at the ground, shrub, and canopy layer for each 60-m location in a 200-m^2 circular area.

Table 1. Table of topographic variables used in regression tree analysis. * Indicates a variable created post-data collection.

Variable Name	Units
Earthflow Presence	Center, Edge, No
Percent Slope	0-100
Slope Position	1-3 (Top, Middle, Lower)
Swale Presence	Yes, No
Microtopography Across Slope	Flat, Undulating, Convex, Concave
Microtopography Up and Down Slope	Flat, Undulating, Convex, Concave
Exposed Bedrock Percentage	0-100
Surface Rock Percentage	0-100
Heatload Index	0-1
Solar Orientation Index	0-20, low=hotter, SSW=0
Distance from Meadow Generalized*	Meadow, Infill Meadow, No Meadow, Side, Lower, Upper
Distance from Meadow Detailed*	Meadow, Infill Meadow, No Meadow, Transition Plot Side, Transition Plot Lower, Edge Side, Edgle Lower, Edge Upper

Field Data Collection

To determine a relative measure for soil moisture content across the 5 transects, poplar dowels were placed in plots for a minimum of 2 weeks, weighed, dried, and weighed once more and the resulting measurements generated a variable called dowel water content (DWC). At each 60-m plot, 5 ¼" diameter poplar dowels were placed in the ground; one in the center one meter to the east and four more in the cardinal directions offset one meter clockwise to avoid directly lying on the north-south transect line. The dowels were left in the soil for at least 2 weeks and at five times throughout the growing season in 2021 (5/09, 5/29, 7/02, 8/29, 10/03), the dowels were removed, placed in an airtight vial, and then replaced with a new set of dowels. In the lab

the vials and the dowels were weighted together and then the dowels were placed in paper bags and dried in a drying oven at 60 ℃ for at least 48 hours. The dried dowels and vials were weighed again and from these values the dowel water content (DWC) was calculated (Equation 1).

$$DWC = \frac{\text{wet weight} - \text{dry stick weight} - \text{dry tube weight}}{\text{dry stick weight} + \text{dry tube weight}}$$

Equation 1. Calculation of dowel water content.

To calibrate the DWC values to a measure of soil water tension, we used Watermark soil moisture sensors (Model 200S, Irrometer.com) to calculate the soil water tension of 14 samples of 7 representative soils from Jim's Creek (two samples per soil type). Each sample was packaged with a soil tensiometer set 2 inches above the bottom of the sample and 5 of the poplar dowels used in the field. When the tensiometer readings stabilized, the wet and dry dowel weights were recorded in the same manner as the field dowels. At each dowel measurement we also weighed and dried the soil samples to record the moisture of the samples as they dried down. This process was repeated five times and the results confirmed that DWC values tightly followed soil water tension values. The regression tree model uses rank order, so the ranked DWC values were an acceptable substitute for actual, usable soil moisture for plants and will thus be referred to as soil moisture values.

Data Analysis

1850 Oak Designation

The location and health of Oregon white oaks greater than 137 cm tall were collected in 900-m^2 sections of the five transects in 2005 (Johnson 2005). This data was supplemented with increment cores of live oaks across the spectrum of sizes at diameter at breast height (DBH). We used this oak data to create a dataset of older oaks circa 1850 (approximately 150 years old or more) which we call 1850 oaks. First, live oak ages taken from increment cores of a wide range of DBH, including the largest oaks on site, were plotted against DBH to create a linear regression of the age-DBH relationship. The oaks that were cored were easily identified as either an 1850 oak or not. For the un-cored oaks, the linear regression was used to estimate age. If the oak was alive, the estimated age based on the linear regression was considered the final age of the oak. If the oak was dead, additional years were added for its decay class. The decay classes start at 1 (just died) and end at 5 (very decayed). The decay classes were as follows: decay class 1: 5 years, decay class 2: 20 years, decay class 3: 35 years, decay class 4: 50 years, and decay class 5: 60 years. Based on the age-DBH relationship, we were able to determine cutoffs when oaks were always older than 150 years old and when oaks were always less than 150 years old. For the oaks between these two cutoffs, the proportion of oaks greater than and less than 150 years old were calculated for roughly 25-year increments. To attain these proportions, oaks were labeled either greater than or less than 150 years old based on field assessments of branching structure, number of trunks, and canopy shape indicating pre-Euro-American settlement growth patterns.

Five plot-level dependent variables were derived from the 1850 oak dataset. The first is the density of living and dead 1850 oaks in trees/ha. The other four are separated into live and dead oak density and basal area in m^2/ha. Basal area was calculated by summing the cross-

22

sectional area of each tree in the plot as measured by the DBH of each trunk (some trees has multiple trunks). All five variables were calculated using a plot expansion factor that accounted for the issue of distances measured along a slope overestimating the vertical area of a plot. The plot expansion factor thus adjusted densities and basal areas to account for the fact that all plots on slopes>0 were <900-m^2 in area. The final dataset consisted of all five 1850 oak variables with the DWC, soil, and site physiography data.

Regression Trees

Five regression trees using each of the five plot-level dependent 1850 oak variables were run using the RPART package (Recursive Partitioning and Regression Trees) in the R-Studio 4.0.3 software to determine the conditions that create landtype phases for oaks at Jim's Creek (RStudio Team 2020; Therneau et al. 2022). The tree is generated with an 'if' statement using one explanatory variable at a time to split the data starting with all observations of the dependent variable. Each split is created using the variable best able to split the dependent variable into two homogenous groups and explain the most variance of the dependent variable. Splitting continues until the tree is considered overgrown and is pruned back using v-fold cross-validation to a size that minimizes the cross-validation error. V-fold-cross-validation takes the data v times and splits it into 90 percent learning and 10 percent test data to determine the number of splits that still retains predictability (Moisen 2008). The ending tree size indicates that at the chosen size the data is not overfit and can predict new data. This is important for using the regression tree to predict oak conditions outside of Jim's Creek. The resulting nodes of the tree that were not split are named terminal nodes and represent varying conditions for either oak densities or oak biomass.

Identification of Landtype Phases

Landtype phases were determined from the five regression trees based on the split variables, the amount of variance each split explains, and over 15 years of field observations at Jim's Creek. For all regression trees, the most important variables that explained the most variance in the dataset were compared with each other to determine any similarities or trends. To aid in this step, the feature of the landscape that each variable represented was also determined. For example, if a regression tree determined soil clay content to be important, then we analyzed the plots with soil clay content and referenced previous studies on the site to identify any features that soil clay content represents. Once a trend or similarity was identified by multiple split variables, the areas on the landscape that those variables describe was checked with field observations. If the areas on the landscape were determined to be a distinct unit of the landscape, they were designated as a landtype phase.

Translation from field to LiDAR variable

LiDAR-derived variables with the highest correlation to key field variables that describe the two landtype phases were used to generate new split values to translate the field data results into a site-wide landtype phase map. We used the ArcGIS Geomorphometry & Gradient Metrics toolbox and ESRI ArcGIS Pro 2.8.0 to create LiDAR-derived variables which represented solar radiation, landscape complexity, or landform (Ironside et al. 2018; ESRI 2021; Evans et al. 2018). Each tool in the Geomorphometry and Gradient Metric toolbox created a raster whose values were summed in each 900-m^2 plot along the five transects for the LiDAR dataset. These variables were compared to the field variables from each landtype phase to find which had the highest correlation. The highest correlated LiDAR variables were isolated and run through a regression tree with their corresponding dependent variable in RPART to generate a new split to

use for mapping. The specific value given at each split was used in ArcGIS Pro to limit rasters from which the variable values were derived to the range given by the split value. This was done for each landtype phase (three variables for MMM, one variable for LCM) and the resulting rasters were combined and post-processed to create a map that represented the landtypes' conditions.

3 | Results

We asked a set of five questions about where oaks were most successful on this historical savanna and indigenous cultural site prior to Euro-American settlement (circa 1850), where those older savanna trees have persisted to present, and where they have died since the cessation of indigenous site use and loss of historical fire regimes.

1) In what kinds of locations did 1850 oaks grow?

2) Where have 1850 oaks survived to present?

3) Where did 1850 oaks suffer high mortality in the intervening decades?

4) Where are 1850 oaks currently growing best?

5) Where did 1850 oaks that died grow best?

In what kinds of locations did 1850 oaks grow?

A regression tree created using 1850 oak density (both alive and dead trees) produced seven terminal nodes with high explanatory power of oak distribution ($R^2_{adj} = 0.50$). The tree revealed higher densities of oaks on the side and lower edges of shallow-to-bedrock meadows with high shield season soil moisture (late spring and early fall) and lower oak densities as you move away from the side and lower edges of meadows with lower soil moisture. The most

important variables in explaining model variance were distance to shallow-to-bedrock meadows followed by shield season soil moisture as shown by the depth of the splits in the tree (Figure 3).

Figure 3. Regression tree of 1850 oak density pruned using cross-validation with table of node explanations. The nodes express the 1850 oak average density and the number of observations.

Where have 1850 oaks survived to present?

A regression tree created using density of living 1850 oaks produced four terminal nodes with high explanatory power of living 1850 oak distribution ($R^2_{adj} = 0.38$). In contrast to the previous tree, live 1850 oak locations appear to be determined by soil pH first followed by shield season soil moisture. Higher densities of oaks occur on sites with a soil pH between 6.4 and 6.7 and decrease with less soil moisture, increasing heat, and lower pH. The most important variable in explaining model variance was soil pH (Figure 4).
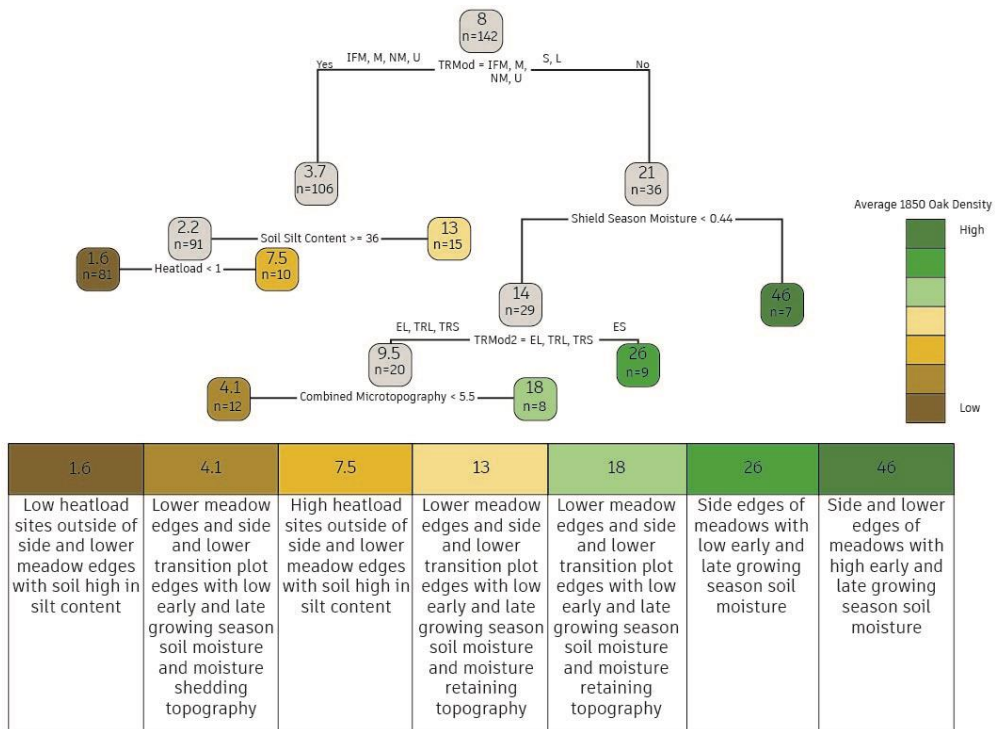
Figure 4. Regression tree of living 1850 oak density pruned using cross-validation with table of node descriptions. The nodes express the living 1850 oak average density and the number of observations.

Where did 1850 oaks suffer high mortality in the intervening decades?

A regression tree created using density of dead 1850 oaks produced three terminal nodes with high explanatory power of oak distribution (R^2_{adj} = 0.34). Heatload and soil moisture during the hot part of the growing season determined dead 1850 oak density. Higher densities of dead 1850 oaks occur on sites with a lower heatload and lower moisture during the hot part of the growing season with lower densities of dead 1850 oaks on warmer sites with more moisture. Heatload and moisture during the hot part of the growing season had equal influence in explaining model variance (Figure 5).
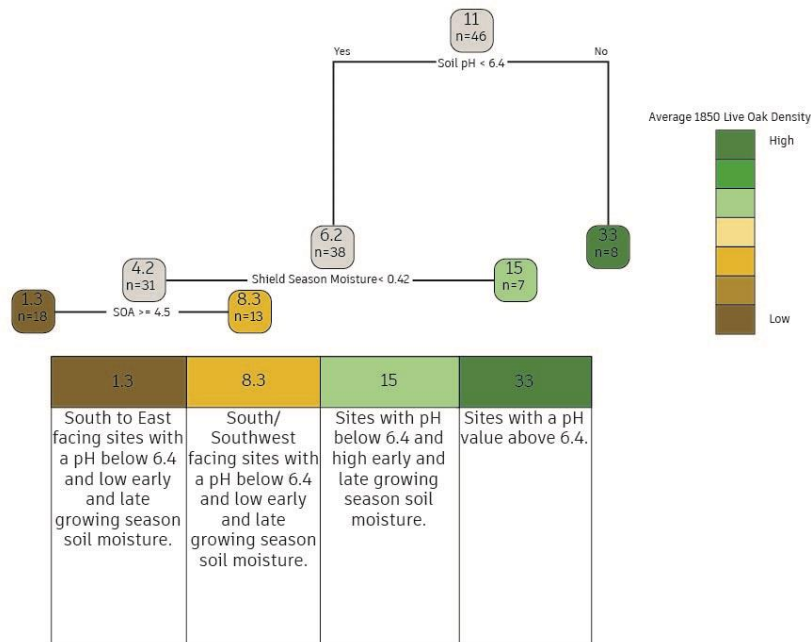
Figure 5. Regression tree of dead 1850 oak density pruned using cross-validation with a table of node descriptions. The nodes express the dead 1850 oak average density and the number of observations.

Where are 1850 oaks currently growing best?

A regression tree created using basal area of living 1850 oaks produced four terminal nodes with high explanatory power of oak distribution (R^2_{adj} = 0.56). Soil clay content, distance from shallow-to-bedrock meadows and slope position determined living 1850 oak basal area. Higher living 1850 oak basal area occurs on sites with a high soil clay content and living 1850 oak basal area decreased as clay content decreased and one moves away from side and lower edges of meadows. Soil clay content had by far the most explanatory power in the model's variance (Figure 6).
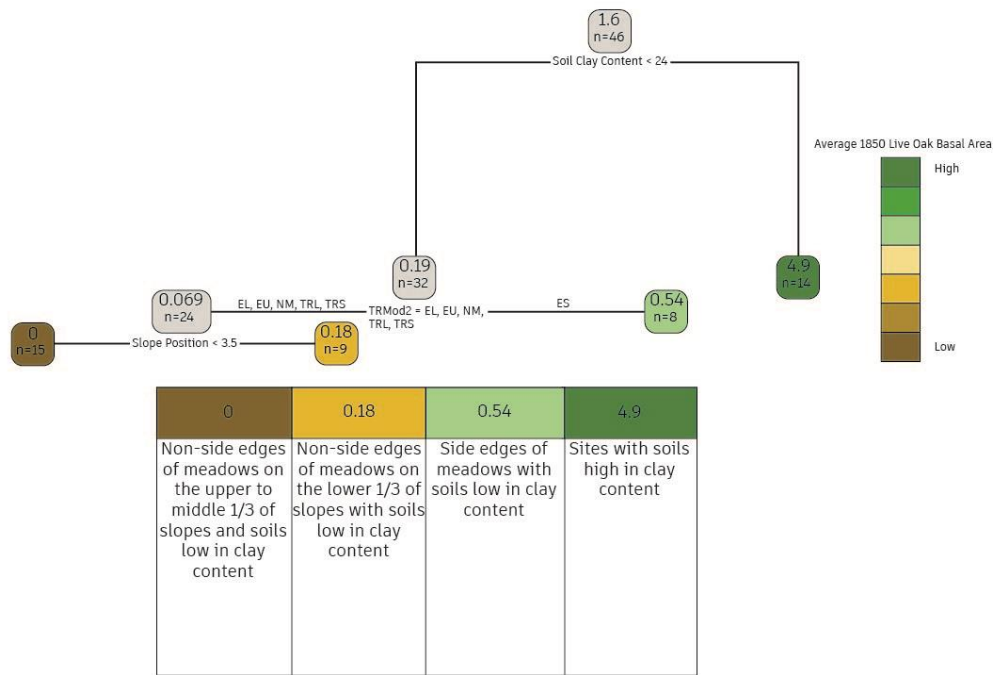
Figure 6. Regression tree of living 1850 oak basal area using cross-validation with a table of node descriptions. The nodes express the living 1850 oak average basal area and the number of observations.

Where did 1850 oaks that died grow best?

A regression tree created using basal area of dead 1850 oaks produced four terminal nodes with moderate explanatory power of oak distribution (R^2_{adj} = 0.24). Soil depth, heatload, and soil pH determined dead 1850 oak basal area with the highest basal area occurring on shallow soils and basal area decreasing as soil depth increases and heatload increases. Soil depth explained the most model variance (Figure 7).
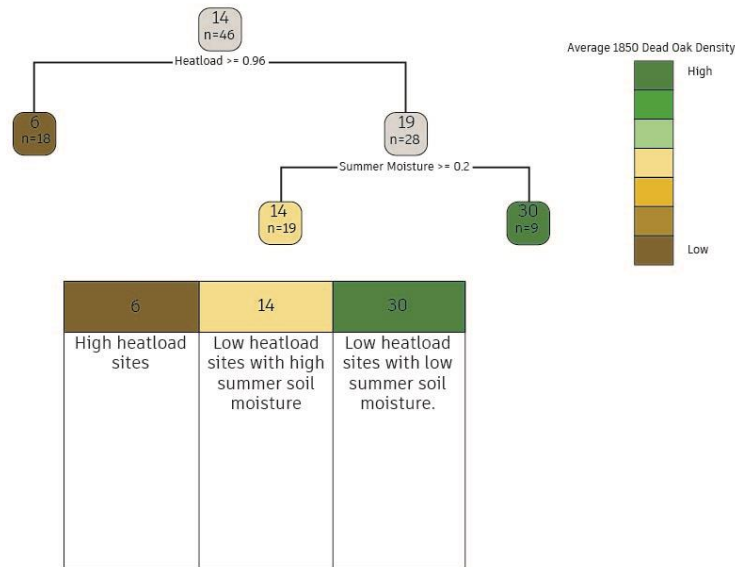
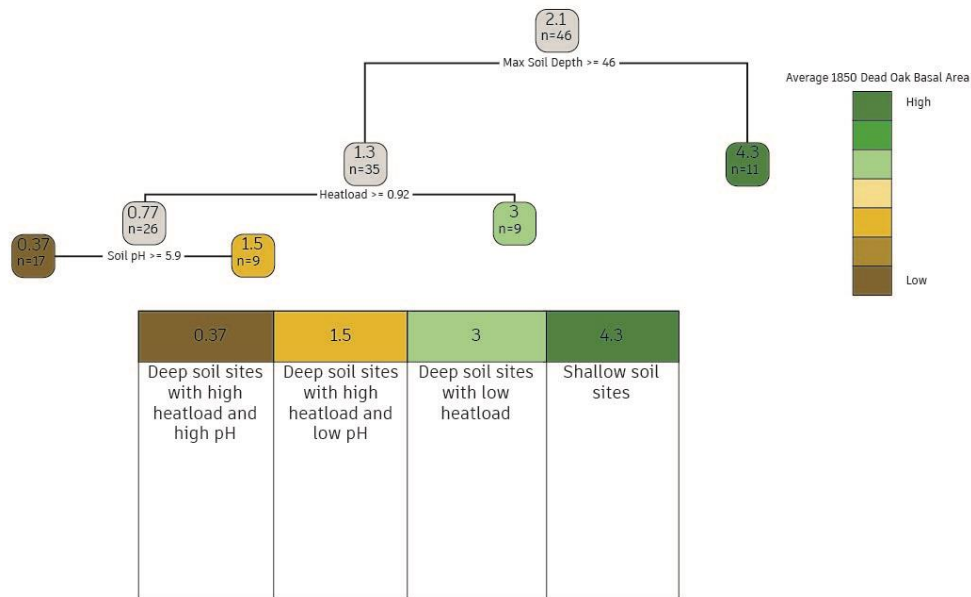**Figure 7.** Regression tree of dead 1850 oak basal area pruned using cross-validation and a table of node descriptions. The nodes express the dead 1850 oak average basal area and the number of observations.

Landtype Phases

Analysis of the five regression trees generated using the 1850 oak dataset revealed two landtype phases present on Jim's Creek; Moist Meadow Margins (MMM) and the Cool Landscape Matrix (CLM) (Table 2).

Table 2. Two landtype phases identified at Jim's Creek with a short description and key variables creating each landtype phase.

Landtype Phase	Description	Key Variables
Moist Meadow Margins (MMM)	Side and lower edges of shallow-to-bedrock meadows with high shield season moisture	TRMod, TRMod2, Shield Season Moisture, Soil Clay Content, Soil pH, Max Soil Depth
Cool Landscape Matrix (CLM)	Cooler sites away from the meadow edges	Heatload

1850 oak density, living 1850 oak density, living 1850 oak basal area regression trees, and one split in the dead 1850 oak basal area regression tree support the MMM landtype phase. In the 1850 oak density regression tree, the split that explained the most variance was 'distance from meadow generalized' followed by shield season soil moisture and 'distance from meadow detailed' (Figure 3). In the living 1850 oak density regression tree, pH explained almost all variance (Figure 4). In the living 1850 oak basal area regression tree, soil clay content explained almost all variance (Figure 6). It is also worth noting that in the dead 1850 oak basal area regression tree the split explaining the most variance was soil depth, which also supports the MMM landtype phase. From these splits, we determined that the areas with historically high densities of surviving oaks are the side and lower edges of shallow-to-bedrock meadows with high shield season moisture. A previous study at Jim's Creek revealed high soil pH, high soil clay content, and shallow soil depth were all associated with the meadows at Jim's Creek

(Murphy 2008). The distance from meadow variables identified the side and lower edges of the meadows as the areas with the highest densities of living oaks. This landtype phase is called Moist Meadow Margins (MMM).

The 1850 oak density, dead 1850 oak density, and dead 1850 oak basal area regression trees support the (M) landtype phase. In the dead 1850 oak density regression tree, heatload explained the most variance followed by summer soil moisture. In the dead 1850 oak basal area regression tree, soil depth explained the most variance followed by heatload and pH. Field observations at Jim's Creek showed dead Oregon white oaks growing away from the meadows in closed-canopy Douglas-fir forest at much lower densities. From the splits and field observations we determined that historically Oregon white oaks also grew in cooler sites (low heatload) away from the meadow edges, but at lower densities. From these split variables, comparisons to the MMM landtype phase, and field observations, we identified the second landtype phase CLM.

Landtype phase map using LiDAR-derived variables

A combination of correlations and regression trees produced a map showing two major landtype phases identified through synthesizing answers to the previous five questions. The first landtype phase, Moist Meadow Margins (MMM) and the second landtype phase, the Cool Landscape Matrix (CLM). Both landtype phase layers match up to their representative locations and highlight other areas that potentially have similar conditions (Figure 8).
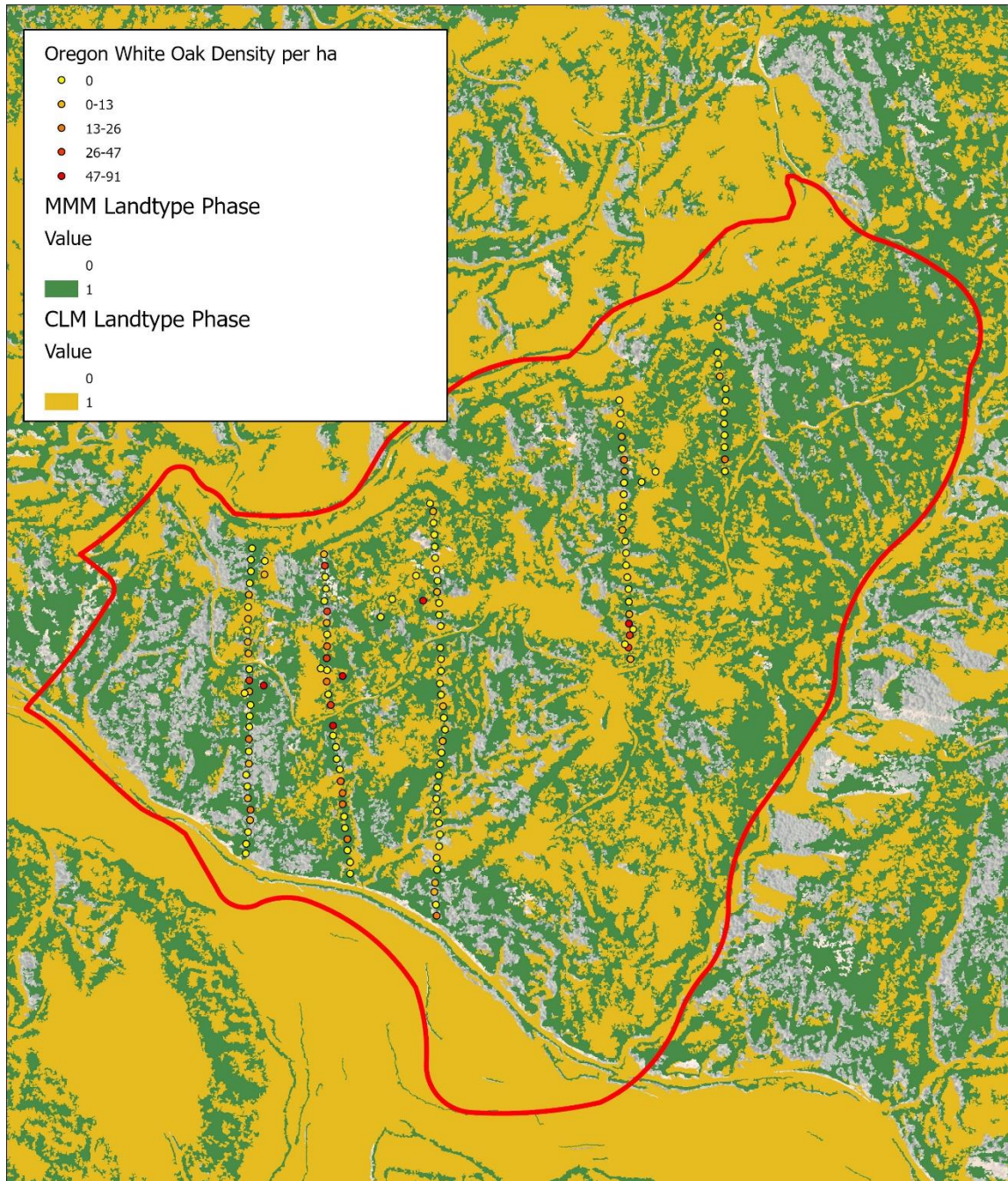
Figure 8. Landtype phase map showing the MMM landtype phase in green and the CLM landtype phase in yellow with 1850 oak densities by plot.

4 | Discussion

We sought to develop a methodology to determine mappable, fine-grained, landtype

phases from field and LiDAR data to assist ecological restoration efforts in heterogeneous

landscapes with rich cultural and ecological histories. Regression trees, a variant of decision trees, were chosen as the statistical method to identify the underlying biophysical conditions that differentiate where Oregon white oaks are most likely to be successful in the landscape and then use those conditions to classify landtype phases. The challenge was to determine where on the landscape oaks have been able to establish and persist over long periods of time so both time and funds used in the restoration are maximized.

As we show below, assessing both living and dead 1850 oaks provided insights that would not have been possible from only living oaks. At the same time, any assessment of the density or basal area of trees long dead must bear the caveat that the trail of evidence erodes over time. Specifically, the signs of a dead tree vanish due to decay or consumption by fire, and this has direct impacts on our assessments of dead trees. First, the density of dead oaks that were alive in 1850 are limited to larger trees whose skeletons are still visible as standing snags or as logs on the ground. Oaks, however, decay extremely slowly and because of their distinctive grain (think of an oak floor) are recognizable at even advanced stages of decay. Furthermore, we saw little evidence of any substantial fire on the landscape since the mid-1800s based on visual assessments of fire scars or recently charred bark, and particularly by the abrupt and near continuous recruitment of trees since the mid-1800s requiring ongoing recruitment of conifer seedlings and saplings which are unlikely to survive a fire (Day 2005). Thus, we expect that our counts of dead 1850 oaks are underestimates of their actual numbers but that the error is likely to be relatively evenly distributed across the site.

It is also important to differentiate our use of the variables of density and basal area as metrics of oak success and as functions of establishment, survival, and growth. Density, as well as providing a measure of abundance for assessments of each category of 1850 oaks, can also

serve as a comparator metric of relative abundances between the two categories since all trees were assessed as greater than 150 years old. In comparison, we use basal area as an additional metric of relative success for each category because it is known to be a good surrogate for biomass but do not compare categories directly since living trees have continued to grow. Because regression tree algorithms are based on the ranks of each observation rather than their actual values, we posit that our use of basal area for dead trees, which died at different times, still provides a useful means to incorporate the notion that a small number of large trees may be considered as a higher or at least different level of success than many small trees. We bear these distinctions and caveats in mind in the interpretation of our data.

We used reconstructions of the distribution of 1850 oaks at Jim's Creek to assess the fine-scale landscape conditions that had supported oaks over long periods of time, and thus were likely candidates for restoration and management. We asked five major questions to identify where oaks that had been growing prior to 1850 were most abundant and grew largest, and where they have survived to present day versus where they have died following the cessation of indigenous use of the site and subsequent successional infill by Douglas-fir. Breaking up our initial dataset of older oaks to answer these questions in the regression trees allowed us to parse out the nuances of oak distribution on Jim's Creek that may have otherwise been lost in the larger dataset.

Overall, regression trees successfully identified oak distributions for each category of interest and explained moderate to high amounts of the variation in each category. However, we note that the dead density regression tree did not hold up to cross-validation due to the small sample size and weak pattern of the data. In this situation, the tree was manually generated using the pre-pruning process of setting the maximum depth of the tree (how many levels of splits).

35

The depth was chosen based on variables chosen for splits, results from regression trees, and field observations. It should also be noted that due to small sample sizes all the trees had low predictability and at times required multiple iterations of the model to get a significant tree.

Despite these caveats, the trees aligned well with over 15 years of field observations on the site, giving us confidence that the results are useful for Oregon white oak restoration at Jim's Creek and similar habitats. Using the five regression trees to address the five questions posed about pre-1850 oak distribution, we synthesized the results to identify two landtype phases. We recommend these mapped areas as a tool for conducting reliable and effective oak restoration and management.

Landtype Phases from Regression Trees

Our analysis of landtype phases started with five questions, each of which was based on a regression tree using either the full dataset or a subset of the oak data into live and dead density and live and dead basal area (biomass). The overall trend from the regression trees appeared to be a distinction between cool former open savanna areas now dominated by successional infill by Douglas-fir, and the sides and lower edges of edaphically controlled shallow-to-bedrock meadows with relatively high soil moisture during the late spring to early fall growing season. This time period covers a dry-down from the wet winter into the depth of the Mediterranean summer drought, which is a critical determinant of regional vegetation distribution. The dead density and dead basal area regression trees reflect the successional infill conditions, while the live density and biomass trees reflect the meadow edges. Dead oaks found on the site were in closed-canopy Douglas-fir forests, which established on the more productive and less stressful areas of the site. The live oaks were found on the edges of the meadows where conditions were more stressful but because of this Douglas-fir were not able to establish and shade out the oaks.

36

We propose these two types of areas as two landtype phases for Jim's Creek. We call the side and lower edges of meadows with high soil moisture Moist Meadow Margins (MMM) and the cool former open savanna areas Cool Landscape Matrix (CLM) (Table 2). The MMM landtype phase was found to have concentrated densities of oaks and the CLM landtype phase to have diffused oaks throughout their respective areas.

Moist Meadow Margin (MMM) Landtype Phase

The MMM landtype was supported by several of the regression tree splits: soil clay content, distance to meadow generalized, shield season moisture, maximum depth to obstruction, and soil pH. All the previously mentioned variables except shield season soil moisture are indicators of meadow edge topography (Murphy 2008). The side and lower edges of the meadows at Jim's Creek are characterized by shallower soil, increases in soil clay content, and high pH which all move in the opposite direction as one moves from the treeless meadows to the areas of successional infill (Murphy 2008). We propose that pH is also an indicator of meadows because the bedrock at Jim's Creek is basalt, which has been proven to increase nearby soil pH (Shamshuddin et al. 2015). Furthermore, portions of these meadow edges appear to be enriched by surface and subsurface runoff, somewhat offsetting the droughty conditions of the shallow soils. These key splits were mainly in the 1850 oak density and the live 1850 oak density and basal area regression trees. This landtype is discrete, spatially limited, and consists of mainly the live trees because the meadow edges are more stressful for trees compared to more productive areas farther from the edges and thus, the stress-tolerant oaks persist by avoiding successional infill by Douglas-fir.

Management needs for the MMM landtype are relatively low. The higher stress conditions that inhibit successional infill means there is less need for frequent thinning, and

longer prescribed fire return intervals may be sufficient to hold back successional infill due to reduced rates of tree recruitment. Simultaneously, the reduced herbaceous fuels of these xeric sites are more likely to promote low-severity prescribed burns (Peterson & Reich 2001), making it easier and safer to implement prescribed fire with less risk to oaks and other desired target trees. While prescribed fire should be implemented across the entire landscape, fire may be applied with longer return intervals in this landtype phase, or it can be included in more frequent fires with less risk.

Cool Landscape Matrix (CLM) Landtype Phase

While the MMM landtype phase represents zones on the side and lower edges of shallow-to-bedrock meadows that appear to have been the most reliable for long-term oak establishment and survival, some oaks were able to occupy less-stressful areas of the historical savanna in the mid-1800s at low densities until they were overtopped by dense Douglas-fir. We note in particular the high densities of dead 1850 oaks on sites with low evapotranspirational demand (low heatload) that would generally be less stressful during the summer drought. While the highest densities of dead 1850 oaks were in the subset of these areas with low summer soil moisture, they were still moderately high on areas with high summer soil moisture. When we further consider where the dead oaks achieved highest biomass, we see that shallow soils followed by deep soils with low heatload were the most prominent areas. This evidence led us to identify a second landtytpe phase, the cool landscape matrix (CLM).

Management needs for the CLM are more involved than those for the MMM landtype phase. As noted earlier, the stress tolerance of oaks confers advantages in stressful sites, but on more productive sites leads to disadvantages when faster-growing trees are present. The CLM landtype phase consists of more productive sites away from the MMM landtype phase. These

areas required extensive thinning of faster growing conifers prior to restoration at Jim's Creek. Thinning opens the canopy and reduces the competition for resources so oaks can once again establish. However, the open savanna on these productive sites must be sustained through prescribed burning to prevent the establishment of new Douglas-fir seedlings. At the same time, because more productive sites produce more fuels, managers may need to burn at higher fuel moistures, burn more frequently to keep fuel loads low, remove fuels prior to fire around oaks and particularly oak seedlings and saplings, or a do a combination of these tasks so the fires do not kill the oaks.

While such management of the CLM landtype phase may be more complex and costly than for the MMM landtype phase, we argue that it is an important part of restoring the full range of historical variability for oaks. Future analysis may confirm our expectations that oaks in these less stressful sites may grow faster and reach maturity sooner than oaks grown in the more stressful MMM sites, although the added moisture collection of the MMM sites could compensate for their higher heatload and shallow soils. Even if this is not the case, it may be that the deeper soils of sites away from the meadow edges provide a buffer during extended periods of extreme drought when even the MMM sites may be under extreme stress.

The designation of these two landtype phases supports fine-grained site management strategies by linking evidence from past oak distribution to key landscape characteristics that may have enabled oaks to persist across a range of site conditions over long periods of time. Furthermore, it allows us to designate where on the landscape different management strategies may need to be enacted in response to the underlying environmental conditions.

Regression Trees as a Method of Ecological Classification

This study revealed several issues that may arise when using regression trees with both field and LiDAR data for ecological classification. For one, regression trees are data hungry. Our study only had 142 plots, which was reduced even further in the live and dead oak datasets. A regression tree needs a large dataset to pull patterns from the data and to create trees that can be used to predict future data (i.e., they hold up to v-fold cross-validation). Regression trees are also hard to compare against one another because each tree uses different variables to split at different locations. For example, in our tree with 1850 oak densities, distance to meadow was a significant factor and used at the first split. When the data was split into live and dead densities of trees, different variables were chosen as the first split because they explained more of the variance in those datasets. This does not mean that distance to meadow is not still an important factor, but we cannot tell at first glance by looking at the regression tree. One would need to look at the plots that ended up at the terminal nodes to determine if there was still a difference in the nodes due to distance to meadow. For this reason, similarities and relationships between regression trees and their data are also hard to determine.

There are also some positive takeaways from the use of regression trees in ecological classification. Ecological data is complex. For species' presence data, there are often many zeros, and of the observations that do have species' presence, the data often does not follow parametric assumptions or have a linear relationship. Regression trees handle all these things and gives exact values for each variable when splitting the data, which allows for discrete units to be determined and mapped on the landscape.

Future work with regression trees as a method for creating classifications should utilize a large dataset if possible. More data has the potential to show patterns on the landscape,

particularly if the species of concern has a diffuse presence on the landscape. The translation from field-data to LiDAR-data could also be improved by focusing on variables that easily translate between the two datasets, such as heatload or slope position.

LiDAR as a Method to Map Landtype Phases

The LiDAR-based maps show potential locations of the two identified landtype phases across the entire Jim's Creek site. However, there are several challenges that arose in the process of translating field variables to LiDAR variables.

One of these challenges is to determine which variables to use in mapping the landtype phases and how to recreate the field variable regression tree. We extrapolated approximately 30 LiDAR variables from a LiDAR DEM to use in the regression trees. To determine which variables from this dataset to use, correlations were run between the LiDAR dataset and the variables in the field-data regression trees. Due to poor correlations and difficulty in interpreting the LiDAR splits and variable meanings, each LiDAR correlate was run individually on the appropriate oak dataset and the number from the first split was used in the mapping of the landtype phases. This creates a disconnect from the original field data regression trees.

An important next step is to improve the variable correlations so that the field tree can be recreated with the LiDAR variables. This would improve the split numbers and create a more accurate representation of the statistically significant results in the field data regression trees which in turn would make the maps more useful to land managers. Additionally, more LiDAR derivatives can be explored and correlated to the field variables. Finally, oak locations should be mapped with GPS and compared to the landtype phases to check the accuracy of the landtype phase maps.

Although challenges in using both field and LiDAR variables exist, there are advantages as well. The field variable regression trees describe the conditions of the landtype phases based on the data collected in the 142 plots. When these results were translated to LiDAR variables and used to create a map of the whole site, the results are not just of moist meadow margins, for example, but rather of all locations on the site that match the conditions that were associated with meadow margins and subsequently high 1850 oak density. This can be used to identify more restoration sites beyond just the moist meadow margins. A caveat to this is that the correlations are not perfect, and even if they are improved, just because the map identifies locations as being suitable for oaks does not mean oaks will establish or persist there. If sites are identified for restoration and management with the LiDAR maps, those sites should be evaluated in the field with the information provided by the field-data regression trees to determine if they are indeed good candidate sites for oak restoration. Therefore, it is recommended that the LiDAR and field regression tree results be used together to create the best site management recommendations. Overall, the translation from plot-level, field variables to site-wide, LiDAR variables and subsequent mapping of landtype phases shows promise for creating fine-grained, mappable land units for ecological restoration and land management.

5 | Conclusion

At its core, classification is a simplification. The goal of the classification we present in this paper is to create a framework to understand a complex, heterogeneous landscape and try to make sense of it so that concrete restoration actions can be enacted on the landscape. It shows promise in creating generalizable yet detailed information about species' suitable habitat locations and mapping them. We believe that for the best management value the field data regression trees and LiDAR-based maps should be used in conjunction with one another. The

LiDAR-based maps provide site-wide information that is good for initial planning of

management. The detailed information present in the field data regression trees should be used to

ground truth the proposed management based on the LIDAR-data maps. The use of regression

trees combined with LiDAR mapping allows for both detailed and general species' habitat

information that can inform and improve ecological restoration projects.

6 | Literature Cited

Abella SR, Covington WW (2006) Forest ecosystems of an Arizona Pinus ponderosa landscape: multifactor classification and implications for ecological restoration. Journal of Biogeography 33:1368–1383

Abella SR, Shelburne VB, MacDonald NW (2003) Multifactor classification of forest landscape ecosystems of Jocassee Gorges, southern Appalachian Mountains, South Carolina. Canadian Journal of Forest Research 33:1933–1946

Andrew ME, Ustin SL (2009) Habitat suitability modelling of an invasive plant with advanced remote sensing data. Diversity and Distributions 15:627–640

Bailey T, Kertis J (2002) Jim's Creek Savanna: The Potential for Restoration of an Oregon White Oak and Ponderosa Pine Savanna. United States Forest Service, Middle Fork Ranger District

Barnes BV, Pregitzer KS, Spies TA, Spooner VH (1982) Ecological Forest Site Classification. Journal of Forestry 80:493–498

Bourg N, McShea W, Gill D (2005) Putting a cart before the search: successful habitat prediction for a rare forest herb. Ecology 86:2793–2804

Boyce MS, Haney AW (1997) Ecosystem Management: Applications for Sustainable Forest and Wildlife Resources. Yale University Press

Christy JA, Alverson ER (2011) Historical Vegetation of the Willamette Valley, Oregon, circa 1850. Northwest Science 85:93–107

Clewell AF, Aronson J (2013) Ecological Restoration, Second Edition: Principles, Values, and Structure of an Emerging Profession. Island Press

Day J (2005) Historical Savanna Structure and Succession at Jim's Creek, Willamette National Forest, Oregon. Master of Science, University of Oregon, Willamette National Forest

De'ath G, Fabricius KE (2000) Classification and Regression Trees: A Powerful yet Simple Technique for Ecological Data Analysis. Ecology 81:3178–3192

ESRI (2021) ArcGIS Pro. ESRI

Evans J, Oakleaf J, Cushman S (2018) An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling.

Falco G, Waring KM (2020) Community Classification of Piñon-Juniper Vegetation in the Four Corners Region, USA. Forest Science 66:687–699

Gucker C (2007) Quercus garryana. USFS Fire Effects Information System

Ironside K, Mattson D, Arundel T, Theimer T, Holton B, Peters M, Edwards T, Hansen J (2018) Geomorphometry in Landscape Ecology: Issues of Scale, Physiography, and Application. Environment and Ecology Research 6:397

Johnson B (2005) Final Report - Jim's Creek Study. University of Oregon

Moisen GG (2008) Classification and regression trees. In: Jørgensen, Sven Erik; Fath, Brian D. (Editor-in-Chief). Encyclopedia of Ecology, volume 1. Oxford, U.K.: Elsevier. p. 582-588. 582–588

Moore MM, Covington WW, Fule PZ (1999) Reference Conditions and Ecological Restoration:

    A Southwestern Ponderosa Pine Perspective. Ecological Applications 9:1266–1277

Mora F, Iverson L (2002) A Spatially Constrained Ecological Classification: Rationale,

    Methodology and Implementation. Plant Ecology 158:153–169

Morgan P, Aplet G, Haufler J, Humphries H, Moore M, Wilson W (1994) Historical Range of

    Variability. Journal of Sustainable Forestry 2:87–111

Murphy M (2008) EDAPHIC CONTROLS OVER SUCCESSION IN FORMER OAK

    SAVANNA, WILLAMETTE VALLEY, OREGON. Master of Science, University of

    Oregon

Oregon Department of Fish and Wildlife (2016) Oregon Conservation Strategy. Salem, Oregon

Peterson DW, Reich PB (2001) Prescribed Fire in Oak Savanna: Fire Frequency Effects on Stand

    Structure and Dynamics. Ecological Applications 11:914–927

RStudio Team (2020) RStudio: Integrated Development for R. RStudio, PBC, Boston, MA

Shamshuddin J, Panhwar Q, Ismail R, Ishak C, Naher U, Hakeem K (2015) Sustaining Cocoa

    Production on Oxisols in Malaysia. In: Crop Production and Global Environmental

    Issues. Khalid, H, editor. Springer International.

Society for Ecological Restoration (2022) What is Ecological Restoration? Society for

    Ecological Restoration: Restoration Resource Center

Therneau T, Atkinson B, Ripley B (2022) Recursive Partitioning and Regression Trees.

U.S. Fish and Wildlife Service (2010) Recovery Plan for the Prairie Species of Western Oregon
and Southwestern Washington. Fish and Wildlife Service, Portland, Oregon

USDA Forest Service (2006) Jim's Creek Savanna Restoration Project Environmental
Assessment. USDA Forest Service, Middle Fork Ranger District of the Willamette
National Forest, Westfir, Oregon

Vesely David, Vesely D, OKeefe R, Tucker G, Conservancy AB, Conservancy (U.S.) N,
University OS, Oregon, States U, States U, States U (2004) A landowner's guide for
restoring and managing Oregon white oak habitats. Bureau of Land Management, Salem
District, [Salem, Or.]

Vesely DG, Rosenberg DK (2010) Wildlife Conservation in the Willamette Valley's Remnant
Prairies and Oak Habitats: A Research Synthesis. ORegon Wildlife Institute, Corvallis,
Oregon

Wiens JA (1997) The Emerging Role of Patchiness in Conservation Biology. In: The Ecological
Basis of Conservation: Heterogeneity, Ecosystems, and Biodiversity. Pickett, STA,
Ostfeld, RS, Shachak, M, & Likens, GE, editors. Springer US, Boston, MA pp. 93–107.

Willamette Valley Oak and Prairie Cooperative (2020) Willamette Valley Oak and Prairie
Cooperative Strategic Action Plan. Willamette Valley

With KA (2019) An Introduction to Landscape Ecology: Foundations and Core Concepts. In:
Essentials of Landscape Ecology. Oxford University Press, Oxford.