CHAPTER 3

A Comprehensive Distribution Map and Habitat Suitability Model for Great Basin Bristlecone Pine (*Pinus longaeva* D.K. Bailey)

David R. Burchfield[1], Otto W. De Groff[1], Stanley G. Kitchen[2], David A. Charlet[3], Douglas H. Page[4], Constance I. Millar[5], Douglas J. Merkler[6], Gregory W. Taylor[1], Héctor G. Ortiz[1], and Steven L. Petersen[1]

[1]Department of Plant and Wildlife Sciences, Brigham Young University, Provo, Utah, U.S.A.

[2]Shrub Sciences Laboratory, Rocky Mountain Research Station, USDA Forest Service, Provo, Utah, U.S.A.

[3]College of Southern Nevada, Henderson, Nevada, U.S.A.

[4]Society of American Foresters, Cedar City, Utah, U.S.A.

[5]Pacific Southwest Research Station, USDA Forest Service, Albany, California, U.S.A.

[6]Aerial M^2 LLC, Boulder City, Nevada, U.S.A.

ABSTRACT

Great Basin bristlecone pine (*Pinus longaeva* D.K. Bailey) is a long-lived subalpine conifer native to the western United States. Bristlecone pine is considered a keystone species of high-elevation Great Basin sky island forests, yet detailed distribution data and information about constraining environmental factors for the species are lacking. We compiled a distribution dataset consisting of 490 individual bristlecone pine forest stands covering nearly 100,000 ha in Utah, Nevada, and California to enable improved monitoring of the species and enhanced understanding of forest succession dynamics. These data were sourced using records from herbaria, an internet-based ecological database, government vegetation surveys and extensive field surveys. We field-verified approximately 60% of mapped bristlecone pine stands in cooperation with a network of collaborators. We used this occurrence data to train a Maximum

Entropy (MaxEnt) Habitat Suitability Model (HSM) covering the entire known distribution of bristlecone pine. Sixty environmental variables representing climatic, topographic, and edaphic factors were considered in the model and ranked by relative importance in constraining the MaxEnt model to the current known distribution of bristlecone pine. The model was regenerated using the 10 most influential variables that showed an insignificant correlation ($|r| < 0.7$) with all other variables to reduce possible effects of collinearity. A second model was produced using only climatic and topographic variables to fill coverage gaps in our soil variables, and the two resulting probability maps were merged. Both models performed well using the AUC metric of model fit--the models achieved values of 0.967 and 0.954, respectively. The variables that contributed the most predictive power to both models were mean dewpoint temperature and mean precipitation values for the months of January and February, respectively (> 80% predictive power), although other climatic, edaphic, and topographic factors were also shown to be useful for refining the distribution model. Our model's occurrence probability map will serve as a valuable resource for conservation of bristlecone pine populations and assessing risk in the face of multiple threats, such as fire disturbance, MPB, and WPBR. Additionally, we intend to project the model to characterize the potential response of the species distribution to anticipated climate change in the region.

INTRODUCTION

Great Basin bristlecone pine (*Pinus longaeva;* hereafter bristlecone pine) is a subalpine tree species native to the U.S. states of Utah, Nevada, and California. Bristlecone pine is a member of *Pinus* subsection *Balfourianae*. It shares this subsection with two other closely related North American pine species, the foxtail pine (*Pinus balfouriana*) and the Rocky Mountain bristlecone pine (*Pinus aristata*). Bristlecone pine is typically associated with limber pine (*P. flexilis*) in montane forested systems, referred to as the "Limber Pine-Bristlecone Pine zone" (Grayson 1993) and is considered a keystone species in montane forest communities in the Great Basin and western Colorado Plateau ecoregions (Eidson, Mock & Bentz 2018).

Bristlecone pine has shown tremendous scientific importance due to its longevity. Dendrochronologists have used ancient bristlecone pine wood to study climate patterns since the 1950s (Ferguson 1969). Individual trees in the White Mountains in eastern California have been dated through ring-counting and cross-dating techniques to be nearly 5000 years old (Currey 1965; Lanner & Conner 2001). Using ring-width analysis methods on wood from ancient trees (both living and dead), dendrochronologists have developed climatic chronologies to 8000+ years B.P. (Feng & Epstein 1994; Salzer, Pearson & Baisan 2019). Additionally, bristlecone pine wood has been used to calibrate radiocarbon dating methods (Ferguson & Graybill 1983).

Across much of its distribution, bristlecone pine frequently occurs in small, isolated patches, mainly in high-elevation, topographically rugged environments, and only rarely does this species form uniform forest stands that cover large areas (e.g. Mt. Charleston, Nevada, White Mountains, California) (Hiebert & Hamrick 1984; Bunn et al. 2018). However, the species distribution is not well-documented at finer spatial scales. The factors that control its distribution are also poorly understood and have not been studied extensively. We found two studies that attempted to model bristlecone pine distribution or establishment drivers. Gray (2017) utilized a raster-based random forest (RF) spatial modeling approach using bristlecone pine presence and absence data, and concluded that elevation, texture (derived from Landsat satellite imagery), and slope were important drivers of bristlecone pine distribution. Hankin & Bisbing (2021) examined the effects of snowpack conditions on the regeneration of bristlecone, limber, and whitebark pines after fire disturbance, and found that greater snowpack has a negative effect on bristlecone pine establishment. LaMarche & Stockton (1974) examined the effects of seasonal temperature and precipitation on bristlecone pine (*Pinus longaeva* and *P. aristata*) growth patterns (as opposed to establishment frequency) as measured by tree-ring widths, showing a wide range of growth responses to climatic factors that varied by site.

Multiple studies have also shown that soil characteristics are important for establishment and persistence of bristlecone pine—the species has an affinity for alkaline, calcareous soils, but its occurrence has occasionally been documented on other soil types as well (Wright & Mooney 1965; Currey 1965; Orlemann, Flinders & Allphin 2017). Additionally, ectomycorrhizal spore bank occurrence has been demonstrated to be important for establishment of bristlecone pine (Shemesh et al. 2019). Other edaphic and climatological variables that account for the species distribution at a global scale have not been explored in-depth, nor have distribution maps been published for the entire species range at a detailed scale. A comprehensive map of the species distribution is needed for improved management and conservation. Moreover, a detailed distribution map will lead to improved understanding of habitat requirements and site suitability for bristlecone pine establishment and survival. Furthermore, we propose a Habitat Suitability Modeling (HSM) approach to aid in characterization of environmental drivers of bristlecone pine distribution.

HSM, also known as Species Distribution Modeling (SDM) or Ecological Niche Modeling (ENM), is a commonly used tool in ecology and biogeography (Yackulic et al. 2013; Phillips et al. 2017). An HSM is a prediction of environmental suitability for a species within an area of interest (Phillips et al. 2017). The inputs to HSM algorithms are geo-referenced environmental variables (e.g. climate, topography, and soils) and species presence and/or absence data. Three of the most popular correlative HSM algorithms are BIOCLIM, random forest (RF), and Maximum Entropy (MaxEnt) (Evans et al. 2011; Booth et al. 2014; Phillips et al. 2017). MaxEnt has the advantage of using species presence data in tandem with environmental variables to quantify relative probability of occurrence—no absence data is required (Yackulic et al. 2013). This presence-only approach contrasts with other modeling methods that rely on both presence and absence data. Meaningful absence data is often challenging to collect, while there are many sources of presence data available (e.g. herbarium voucher records, iNaturalist).

MaxEnt provides as outputs both a map of relative occurrence rate (ROR) (Fithian & Hastie 2012; Merow, Smith & Silander 2013, also termed “suitability index” by Royle et al. 2021) and an analysis of variable contributions to the distribution model. Relative occurrence is expressed via a complementary log-log (cloglog) transform, which can be cautiously interpreted as occurrence probability provided certain statistical assumptions are met (Phillips et al. 2017). These assumptions include independence of presence samples (i.e. that they are a result of random or representative samples) and constant probability of detection across sample locations (Yackulic et al. 2013).

This study has two primary objectives:

1. To create a detailed spatial database of known stands of bristlecone pine using a variety of available data sources. These presence points will be mapped using GIS and used as an input for MaxEnt habitat suitability modeling.

2. To develop a model-based prediction for the distribution of bristlecone pine in relation to key environmental variables. We will use MaxEnt habitat suitability modeling to identify environmental factors that aid in characterization of the fundamental niche of the species (Roughgarden 1974). We predict that a combination of topographic, edaphic, and climatic variables will prove to be important drivers of the distribution and habitat niche of the species.

METHODS

Compilation of PILO occurrence data

In order to model the distribution of bristlecone pine, we compiled all available presence data for the species, building from historic mapping efforts, government vegetation surveys from the National Park Service and U.S. Forest Service, herbarium records, internet databases, and extensive field work. The overall distribution of Great Basin bristlecone pine had been mapped at least four times in the past, which provided historical context and a starting point for our mapping effort. Initially, bristlecone pine was not recognized as a separate species from the closely related Rocky Mountain bristlecone pine (*P. aristata*)—maps created before 1970 aggregated the distributions of the two species (Bailey 1970). E. N. Munns of the U.S. Forest Service published perhaps the first distribution map for *P. aristata* (including *P. longaeva*) in his 1938 atlas “The Distribution of Important Forest Trees of the United States” (Munns 1938). This map was merely an estimate of the species distribution, and the atlas did not list specific sources for the map data. Several locations are erroneously shown where bristlecone pine is now known to be absent (Figure 3.1).

In 1966, W. B. Critchfield and E. L. Little produced a map of the distribution of Rocky Mountain bristlecone pine (including Great Basin bristlecone pine) for their atlas "Geographic Distribution of the Pines of the World" (Critchfield & Little 1966). This map was copied in Little's well-known "Atlas of United States Trees" (Little & Viereck 1971). The authors listed only four unpublished sources of information for their map for bristlecone pine as a whole, including two from California and one from Utah. These sources are presumably U.S. Forest Service vegetation surveys conducted in those states. This map is both highly generalized and incomplete, displaying polygons for the locations of just 11 populations within the borders of Utah, Nevada, and California, with 17 additional point markers indicating "isolated occurrences" in those states (Figure 3.1).

In his 1970 taxonomic and geographic study of *Pinus* subsection *Balfourianae*, which identified Great Basin bristlecone pine as a separate species distinct from Rocky Mountain bristlecone pine and foxtail pine (*P. balfouriana*), D. K. Bailey created a map of *P. longaeva* using herbarium records and conducting site visits to verify the presence of the trees in several locations. He examined 600 specimens, aggregated them into localities, and geo-located them as point features on a paper map. Bailey also drew a rough polygon on his map representing an estimated absolute outer boundary of Great Basin bristlecone pine populations (Bailey 1970) (Figure 3.1).

Most recently, D. A. Charlet, in his book "Atlas of Nevada Conifers: A Phytogeographic Reference" (Charlet 1996), produced a map of known bristlecone pine stands across the state of Nevada. He mapped the locations of 72 herbarium collections in 17 mountain ranges in Nevada, which were plotted as point features on a map of Nevada with boundaries representing mountain range extents (Figure 3.1).

Each of these maps, while highly useful, showed limitations related to map scale or data type. The Munns and Little maps often drew broad, generalized polygons that covered entire mountain ranges, including some ranges where bristlecone pine does not occur. Bailey's map delineated an outer boundary of the distribution, but occurrences within the distribution were shown as point features instead of polygons representing population extents. Similarly, Charlet's map of the species distribution in Nevada, while highly precise, was based on herbarium occurrence records with only latitude and longitude values for each occurrence, precluding any sort of analysis where the extent of populations or individual stands might be important. Nevertheless, it is important to note that the quality of these maps has improved over time as the availability of vegetation distribution data has expanded, allowing researchers to map these data at finer spatial scales. These maps, along with a thorough literature review of bristlecone pine population locations, gave us a starting point for field visits that allowed us to produce species presence data in preparation for HSM.

We compiled stand-level distribution data for the entire known range of bristlecone pine, which encompasses portions of the states of Utah, Nevada, and California. A "stand," in a forestry context, is defined as "a contiguous community of trees uniform in composition, age and size class distribution, spatial arrangement, site quality, condition, or location to distinguish it from adjacent communities" (Nyland 2016). We opted to focus on the "composition" and "location" aspects of this definition in our mapping approach, identifying forested areas in which bristlecone pine is at least a minor but consistent component of the forest. The approach generally worked well due to the patchy, fragmented nature of most populations of the species. We did not identify a minimum stand size for inclusion in our map—stands were identified that were smaller than 1.0 hectare in area. The compilation of these stand data was a multi-year effort that involved gathering bristlecone pine distribution information from numerous sources, including vegetation surveys, herbarium records, iNaturalist observations (California Academy of Sciences, San Francisco, California), high-resolution aerial imagery (e.g. USDA NAIP imagery), and firsthand reports from scientists, foresters, and land managers. These known stand location polygons were digitized and assembled into an Esri geodatabase for display in an ArcGIS Online-based web mapping application (Esri, Redlands, California, USA). Apart from the stand polygons, attribute data were also assembled for each stand and recorded in the geodatabase, including a name, unique identifier, the mountain range or region, county, and state where the stand is found, the stand area, the source of the stand data, a web link to photographs of the stand (if available), and attribution to the photographer. Due to the difficulty of reaching many bristlecone pine stands, we also assigned an "access class" to each stand to quantify the ease of access, i.e. easy (a road passes through the stand), moderate (a road approaches to within 0.5 miles/0.8 km of the stand), and difficult (no road exists within 1.0 mile/1.6 km of the stand).

A total of 490 stand polygons were produced from the aforementioned mapping effort. Each of the stand polygons was tagged based on what data sources were used to create the stand. In cases where multiple sources were used, a secondary source tag was assigned (Figure 3.2). 296 (60.4%) of the stands were directly field verified by BYU Geospatial Habitat Analysis Laboratory (GHAL) staff or collaborators via site visits. An additional 64 stands (13.0%) were derived from geolocation information included with herbarium voucher records or iNaturalist observations and were verified to a high degree of confidence using aerial imagery. In sites directly field verified by GHAL staff, presence of bristlecone pine within the polygon area was confirmed, and the stand boundaries were validated to the extent this was possible by hiking around the stand and estimating the boundaries using binoculars, and in some cases, drones. Due to the rugged terrain in which many of the stands are found, it was often problematic to directly verify the boundaries, and in these cases they were estimated from aerial imagery by interpreting vegetation and soil patterns. 198 records from our database were either partially verified or derived from other sources deemed to be unreliable and received an "unverified" tag. These stands were excluded from the MaxEnt analysis. We acknowledge that there are likely to be errors of omission (type II errors) in the dataset due to under-estimation of bristlecone habitat areas. Errors of commission (type I errors) are also likely to be present, but in a more limited quantity due to the methodology we used.

Following completion of the stand mapping effort, we analyzed the current protected status of mapped stands using the USGS Protected Areas Database of the United States (USGS 2020). We used ArcGIS Pro to perform a geometric intersection of bristlecone pine stands with each of four GAP status classes to quantify the area of bristlecone pine habitat included in each class.

Maximum Entropy habitat suitability modeling

The MaxEnt HSM package required species presence data as an input, which necessitated the conversion of our stand polygons to a point-based table of samples consisting of latitude and longitude coordinates (WGS 84) for each sample. Prior to this conversion, we refined our presence dataset by eliminating observations that were not directly verified via site visits, herbarium vouchers with precise geolocation, or iNaturalist records. This greatly increased the reliability of the presence dataset for modeling purposes and reduced the number of largely redundant samples. Following the database cleanup process, we used the Generate Random Points tool in ArcGIS Pro 2.7.1 (Esri, Redlands, California, U.S.A.) to create a point feature class consisting of random points selected from within bristlecone pine stand polygon boundaries. We specified a minimum distance of 800 m between each randomly placed point to avoid excessive spatial clustering of presence samples. The 800 m spacing threshold was used because it was the same distance as the spatial resolution of our environmental predictor variables, and it allowed us to avoid the chance that multiple presence samples would fall onto the same cell of our predictor rasters. The smallest stand polygons received only one sample point, while the largest polygons were able to receive up to 1000 points using the 800 m spacing constraint. Latitude and longitude were calculated for each sample in the feature class, and the attribute table was exported as a comma-separated value (CSV) file for input into MaxEnt. This process yielded 371 samples representing locations where bristlecone pine was known to be present. Due to the comprehensive nature of the presence dataset and the random selection of sample points, we assumed that the input data was unbiased, which is an important assumption of MaxEnt modeling (Yackulic et al. 2013; Phillips et al. 2017).

The other primary input required in MaxEnt is a series of environmental predictor raster variables, which represent environmental phenomena that may constrain the distribution of the species of interest, including climate, soils, and topographic variables. Climate variables we selected as model inputs consisted of 30-year (1980–2010) annual and monthly normal values for total precipitation, mean dewpoint temperature, and minimum and maximum temperatures obtained from the PRISM Climate Group (PRISM 2019) at 800 m spatial resolution. These variables were selected in an exploratory sense because it was unknown which climatic factors would prove important for bristlecone pine occurrence. Additionally, Topographic Ruggedness Index (TRI) (Riley, DeGloria & Elliott 1999), slope, aspect, and total solar radiation (Fu & Rich 2002) values derived from PRISM elevation data were included as model inputs. These were calculated using tools in ArcGIS Pro and were derived from the PRISM elevation dataset. gSSURGO (Soil Survey Staff 2020) soil data for available water supply (AWS) (expressed as centimeters of water per centimeter of depth) (NRCS 2016), calcium carbonate ($CaCO_3$) concentration (percent of carbonates in the fraction of the soil less than 2.0 millimeters in size) (NRCS 2016), and pH were rasterized and resampled to match the 800 m spatial resolution of the PRISM datasets. AWS is a measure of the quantity of water that soil is capable of storing for use by plants, and is influenced by the soil's concentration of rock fragments, organic matter, and texture (Wieczorek 2014; NRCS 2016). This represented a condensed format for characterizing multiple variables that were deemed possibly important for establishment of bristlecone pine. Studies have indicated that bristlecone pine tends to grow on alkaline soils containing high concentrations of $CaCO_3$ (Wright & Mooney 1965; Bidartondo, Baar & Bruns 2001; Smithers & North 2020). $CaCO_3$ and pH were selected to test the importance of those influences on bristlecone pine distribution (Table 3.1).

Due to limitations of the gSSURGO dataset, there were coverage gaps among the soil variables. These gaps were primarily located in mountainous portions of California and central Utah in areas where detailed soil surveys have not been completed, and would produce coverage gaps in the resulting MaxEnt model probability maps. Because we were interested in the effects of these soil variables despite the coverage gaps, we decided to produce one model with the soil variables included, and a second model that only included the PRISM data and derived variables. Before input into MaxEnt, all predictor variable rasters were clipped to a modified version of Bailey's estimate of the extent of bristlecone pine (Bailey 1970) using ArcGIS Pro.

To aid in variable selection, an initial run of MaxEnt was performed using all variables to establish relative importance of the variables using the permutation importance table provided in the MaxEnt output. We used ENMTools (Warren et al. 2019) to produce a correlation matrix (Pearson's *R*) of the top variables to evaluate multicollinearity. Variables were discarded if they produced a significant correlation ($|r| \geq 0.7$) with a more important variable (higher on the permutation order table). We selected the 10 most important insignificantly correlated variables for inclusion in the final model to reduce possible negative effects of collinearity on model transferability (see Feng et al. 2019).

We used the MaxEnt Java program (version 3.4.1) (Phillips et al. 2017) to generate our species distribution models for bristlecone pine. We configured the program to produce 10 replicates of the model, then average the resulting cloglog values into a composite map to reduce noise on the final output model probability map. Samples were set aside during each replicate run as test data used for *k*-fold cross-validation (Mosteller & Tukey 1968; Radosavljevic & Anderson 2014). Two separate models were produced using MaxEnt. The first included gSSURGO soil data as environmental variables, despite the large coverage gaps associated with those data ("with-soil model"). Additionally, a second model was generated that excluded the soil data ("no-soil model"). This allowed us to produce a MaxEnt probability map for areas that lacked gSSURGO coverage. It also allowed us to compare the relative strength of the two modeling approaches and determine whether certain climatic or topographic variables were important in either or both approaches.

The "Mosaic to New Raster" tool in ArcGIS Pro was used to merge the probability maps for the two models by giving the higher-quality model (with-soil) probability map mosaic priority over the lower-quality model (no-soil) probability map, but using the no-soil model to fill in the aforementioned gaps in the with-soil map. Additionally, scattered missing values (no-data cells) resulting from the input raster variables were filled using the "Elevation Void Fill" function in ArcGIS Pro. The resulting composite raster was resampled using a cubic convolution method in order to smooth the final output for improved visualization (Keys 1981).

RESULTS

Distribution map for Great Basin bristlecone pine

Using the stand polygon dataset, we produced a map of known bristlecone pine occurrences using a proportional point symbology to highlight differences in relative sizes of contiguous stands (Figure 3.3). We mapped 490 stand polygons encompassing 99,808 ha of forests in Utah, Nevada, and California that contained bristlecone pine as at least a minor component.

We segmented the total distribution into four regional groupings based upon apparent spatial clusters. These roughly followed physiographic boundaries in the case of the Utah High Plateaus (UHP) grouping (hydrographic Colorado Plateau), while the Central Great Basin (CGB), Southern Great Basin (SGB), and Western Great Basin (WGB) groupings were drawn based upon the large distances between bristlecone pine populations found in those groupings (Figure 3.3). Additionally, stands tended to be smaller ($\bar{x}$ = 167 ha) and more scattered among the CGB population when compared to the large contiguous stands prevalent in the SGB and WGB populations ($\bar{x}$ = 3578 and 833 ha, respectively). Mean stand size in the UHP stands was even smaller at 71 ha. To further test the robustness of these groupings in characterizing regional differences in environmental conditions within bristlecone pine populations, we randomly sampled the elevation values (n = 1000) within stand polygons in each grouping and performed a pairwise Wilcoxon rank sum test using R (R Core Team 2021; Wilcoxon 1946). We produced a box-and-whisker plot for visualization using ggplot2 (Wickham 2016) (Figure 3.4). Most pairs of groupings were significantly different, indicating that the regional groupings we selected displayed significant differences in environmental factors related to elevation. Other interactions between regional groupings and environmental variables would be explored after determining which variables proved most important in our MaxEnt models.

MaxEnt models

Variables used in model development

Following the initial run of MaxEnt to determine variable importance, we re-analyzed the dataset using only the top 10 insignificantly correlated variables (with-soil model) (Figure 3.5). To fill the coverage gaps found in the soil data, we created an additional model using the seven topographic and climatic variables that remained after removing the soil variables (AWS, $CaCO_3$, and pH) (no-soil model).

MaxEnt model results

Total model performance is characterized by the area under the receiver-operator curve (AUC) and represents the probability that a randomly chosen occurrence sample is ranked higher than a randomly selected background point (Merow, Smith & Silander 2013). An AUC value close to 1.0 represents a high probability and therefore high model performance (Syfert, Smith & Coomes 2013). Our MaxEnt model (mean of 10 replicates) with soil variables included achieved an average test AUC value of 0.967 with a standard deviation of 0.007. The 10 most influential variables in the model are displayed in Table 3.2. The MaxEnt model without soil variables included achieved an average test AUC value of 0.954, with a standard deviation of 0.008.

Variable importance and response

January mean dewpoint temperature and February mean precipitation were the most influential variables in both models, accounting for 90.3% and 82.3% of the total importance in the with-soil and no-soil models, respectively (Table 3.2).

MaxEnt output graphs showing variable response curves are shown in Figure 3.6 and Figure 3.7. The variables used in both models showed similar responses between models. After determining which variables MaxEnt showed to be most important for constraining bristlecone pine distribution, we randomly sampled values for the top two variables (January mean dewpoint temperature and February mean precipitation) in bristlecone pine stands using the same methodology previously discussed to determine whether the regional groupings showed significant differences. The pairwise Wilcoxon rank sum test again showed significance for both variables among most regional groupings, lending further support to the groupings we selected. We produced box-and-whisker plots to visualize the differences in these variables in each regional grouping (Figure 3.8).

Composite HSM probability map

The composite HSM probability map is shown in Figure 3.9. Modeled high-cloglog areas consistently aligned with observed topographic, climatic, and edaphic factors typical for bristlecone pine habitat.

Bristlecone pine protected status

A summary of GAP status information for mapped bristlecone pine stands is shown in Table 3.3. 81,243 ha (81.4%) of mapped bristlecone pine stands were listed as being currently managed for biodiversity (GAP 1 and GAP 2). However, a large proportion of protected stands are listed as being managed for multiple uses, including extractive and off-highway vehicle use (GAP 3). 1,785 ha are not found in protected areas (GAP 4).

DISCUSSION

Variable importance

January mean dewpoint temperature (tdewpoint01) and February mean precipitation (precip02) were the most influential variables in both models (Figure 3.10). This indicates that winter air dryness and precipitation are important factors in bristlecone persistence, likely resulting from a competitive advantage over other species that are less adapted to extreme cold and desiccation during winter months. Variable response curves show that occurrence probability increases logarithmically with February precipitation, which aligns with LaMarche & Stockton's (1974) observations of a positive correlation between mid-winter precipitation and bristlecone pine cambial growth in the White Mountains, California. The month of February is a key period for snowpack accumulation in high-elevation mountainous regions of the western U.S., and the interaction between bristlecone pine occurrence and may be related to long persistence of February precipitation in the form of snow, followed by a subsequent slow release of water and warming temperatures in the spring (see Harpold et al. 2015). The interaction between occurrence probability and January dewpoint temperature is more complex, represented by a narrow range of optimal dewpoint temperatures between -17 and -11°C, then a rapid decrease in occurrence probability (suitability) above -11°C. Bristlecone pine may outperform other species in these conditions that result in a higher probability of occurrence.

January mean dewpoint and February mean precipitation were highly correlated with other variables thought to be important in constraining the distribution of bristlecone pine. For example, January mean dewpoint showed a high negative correlation with elevation ($r = -0.92$) and could be considered a surrogate for elevation in the HSM. Additionally, the edaphic variables we used ($CaCO_3$ concentration, pH, and AWS) did not account for a large percentage of the total permutation importance, but they did appear to constrain mapped areas of high probability better than the climatic and topographic variables by themselves.

The two variables that contributed most to the MaxEnt model showed significant differences among the regional groupings we established for the overall range of bristlecone pine. This indicates differences in importance of environmental factors among genetic subpopulations of the species. Therefore, bristlecone pine's fundamental niche is likely to vary across space and among genetically different subpopulations. Additional analysis is needed to characterize these potential genetic differences among regional groupings.

MaxEnt cloglog map interpretation

The composite map of MaxEnt-derived cloglog values that combined values from the with-soil and no-soil models performed well based upon visual comparison with available bristlecone pine occurrence data. Map areas with high cloglog values generally showed locations where bristlecone pine presence has been documented (i.e. where the species' fundamental and realized niches aligned), with several exceptions. For example, the Diamond Mountains of Eureka County, Nevada appear on our map with generally high cloglog values along the crest of the range, but no bristlecone pine has yet been documented in this location (Charlet 1996). The Diamond Mountains may represent a region that provides bristlecone pine's fundamental niche, but unknown factors have prevented its establishment within this area. Conversely, an area of high cloglog values around the Chokecherry Benchmark in the southern Snake Range in eastern Nevada, where no bristlecone pine had been mapped, led us to a recent internet report of bristlecone pine occurrence there, confirming our model's prediction in that location (Gathright 2020). We anticipate that this model will aid in location of additional undocumented populations of bristlecone pine.

Model limitations

Our model has two limitations that should be mentioned. First, the spatial resolution of the PRISM datasets we used as environmental predictor variables (800 m) was appropriate for such a large area, but it could be too coarse to effectively characterize important fine-scale variation in variables such as total solar radiation, slope, and aspect. There could have been local variations in these phenomena that would explain bristlecone pine occurrence with greater discrimination, but those local variations could have been averaged out due to the relatively coarse scale we selected (see Levin 1992). Second, as previously discussed, there were large gaps in the gSSURGO soil data we used in the analysis. MaxEnt does not provide a means to ignore the gaps in one variable, so all variables get cropped to the extent of the variable that covers the smallest land area. This led to large holes in our bristlecone pine occurrence predictions in several areas that are known to contain bristlecone pine populations, resulting in the need to create a separate model without soil data. This second model was used to fill the gaps in the soil rasters, but its overall performance indicated lower reliability (Wright & Mooney 1965). Additionally, limitations in the accuracy of the gSSURGO data may be a source of error in model predictions.

We also acknowledge the likelihood that, across such a wide distribution area, certain subpopulations of bristlecone pine are likely to be more sensitive to specific environmental factors than other subpopulations. A blanket modeling treatment of the entire species distribution may preclude representation of important genetic groupings within the species that are constrained to different ecological niches. In order to study these subpopulations and their respective niches, it would be necessary to conduct extensive genetic testing on the species, identify genotypic groupings, and conduct modeling among each genetic group to determine the relative importance of environmental variables among the subpopulations.

Iterative modeling approach

These results represented the latest iteration of our modeling workflow. Our approach was to create MaxEnt distribution models, then search for areas of high probability where no bristlecone pine was known to occur, then field-verify those locations where possible (see Rhoden et al. 2017). This led us to document several populations that were previously unknown or undocumented in herbaria, e.g. O'Neal Peak, Nevada (Snake Range), Kern Mountains, Nevada, and Howell Peak, Utah (House Range). These newly verified locations were, in turn, used to improve our database of presence samples and to create new models. We plan to continue using this iterative modeling approach to find additional undocumented populations.

CONCLUSION

This study resulted in the most comprehensive distribution dataset for Great Basin bristlecone pine to date and used the resulting occurrence data to model the species distribution based upon 60 environmental variables. The resulting model showed that the species is sensitive to a combination of topographic, climatic, and edaphic variables that constrain the species to its current ecological niche. We intend to use this model to estimate the response of Great Basin bristlecone pine to warming climatic conditions. Bristlecone pine is a subalpine conifer with an already limited, fragmented distribution, and we would expect its range to contract under a warmer climate. Modifying the PRISM data to represent anticipated warming regimes and projecting our model to those conditions is a continuation of this work that we hope to pursue. We also plan to maintain the Great Basin bristlecone pine distribution database and continue to field-verify stands to iteratively improve the dataset. As data for newly documented stands become available, we will update the database. This will be an important dataset for forest scientists, researchers, and land managers as they monitor the overall health of the species in the face of multiple potential threats, including white pine blister rust (WPBR) and mountain pine beetle (MPB) infestations that have caused high mortality in other high-elevation pine species.

Mapped bristlecone pine stands are largely found within protected areas, including national parks, designated wilderness areas, and national forests. However, we estimate that 93% of the total mapped habitat area is contained within mixed-management areas where extractive activities could impact bristlecone pine populations. Further research should be performed to quantify the effects of these activities on overall health of the species and to determine whether additional protection is needed.

The bristlecone pine distribution database and associated web mapping application will continue to be hosted by the BYU Geospatial Habitat Analysis Laboratory (GHAL), and the database will be made freely available to interested parties. The permanent link for the bristlecone pine web mapping application is found at https://bristlecone.byu.edu.

ACKNOWLEDGEMENTS

We would like to thank the USDA Forest Service and the Charles Redd Center of Western Studies for their grants that funded field excursions to map bristlecone pine stands in Utah, Nevada, and California. We would also express gratitude for the multitude of volunteers and collaborators who shared their knowledge with us and supported us in cataloging, field-verifying, and photographing bristlecone pine stands.

LITERATURE CITED

Bailey, D. K. (1970). Phytogeography and taxonomy of *Pinus* subsection *Balfourianae*. *Annals of the Missouri Botanical Garden*, 210-249.

Bidartondo, M. I., Baar, J., & Bruns, T. D. (2001). Low ectomycorrhizal inoculum potential and diversity from soils in and near ancient forests of bristlecone pine (*Pinus longaeva*). *Canadian Journal of Botany*, 79(3), 293-299.

Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Diversity and Distributions*, *20*(1), 1-9.

Bunn, A. G., Salzer, M. W., Anchukaitis, K. J., Bruening, J. M., & Hughes, M. K. (2018). Spatiotemporal variability in the climate growth response of high elevation bristlecone pine in the White Mountains of California. *Geophysical Research Letters,* 45(24), 312-320.

Chambers, J. M. (2018). Graphical methods for data analysis. CRC Press.

Charlet, D. A. (1996). Atlas of Nevada conifers: a phytogeographic reference. University of Nevada Press.

Critchfield, W. B., & Little, E. L. (1966). *Geographic distribution of the pines of the world* (No. 991). US Department of Agriculture, Forest Service.

Currey, D. R. (1965). An ancient bristlecone pine stand in eastern Nevada. *Ecology*, *46*(4), 564-566.

Eidson, E. L., Mock, K. E., & Bentz, B. J. (2018). Low offspring survival in mountain pine beetle infesting the resistant Great Basin bristlecone pine supports the preference-performance hypothesis. *PloS one*, *13*(5), e0196732.

Evans, J. S., Murphy, M. A., Holden, Z. A., & Cushman, S. A. (2011). Modeling species distribution and change using random forest. In *Predictive species and habitat modeling in landscape ecology* (pp. 139-159). Springer, New York, NY.

Feng, X., & Epstein, S. (1994). Climatic implications of an 8000-year hydrogen isotope time series from bristlecone pine trees. *Science*, *265*(5175), 1079-1081.

Feng, X., Park, D. S., Liang, Y., Pandey, R., & Papeş, M. (2019). Collinearity in ecological niche modeling: confusions and challenges. *Ecology and Evolution*, 9(18), 10365-10376.

Ferguson, C. W. (1969). A 7104-year annual tree-ring chronology for bristlecone pine, *Pinus aristata*, from the White Mountains, California.

Ferguson, C. W., & Graybill, D. (1983). Dendrochronology of bristlecone pine: a progress report. *Radiocarbon*, *25*(2), 287-288.

Fithian, W., & Hastie, T. (2012). Statistical models for presence-only data: finite-sample equivalence and addressing observer bias. *Annals of Applied Statistics*.

Fu, P., & Rich, P. M. (2002). A geometric solar radiation model with applications in agriculture and forestry. *Computers and electronics in agriculture*, 37(1-3), 25-35.

Gathright, T. (2020). Ascent of Chokecherry Benchmark on 2020-07-17. Available online: https://www.peakbagger.com/climber/ascent.aspx?aid=1428639 (accessed on May 24, 2021).

Gray, C. A. (2017). Impact of Climate Variability on the Frequency and Severity of Ecological Disturbances in Great Basin Bristlecone Pine Sky Island Ecosystems.

Grayson, D. K. (1993). The Desert's Past: A Natural Prehistory of the Great Basin. Washington: Smithsonian Institution Press.

Hankin, L. E., & Bisbing, S. M. (2021). Let it snow? Spring snowpack and microsite characterize the regeneration niche of high-elevation pines. *Journal of Biogeography*.

Harpold, A. A., Molotch, N. P., Musselman, K. N., Bales, R. C., Kirchner, P. B., Litvak, M., & Brooks, P. D. (2015). Soil moisture response to snowmelt timing in mixed-conifer subalpine forests. *Hydrological Processes*, *29*(12), 2782-2798.

Hiebert, R. D., & Hamrick, J. L. (1984). An ecological study of bristlecone pine (*Pinus longaeva*) in Utah and eastern Nevada. *The Great Basin Naturalist*, 487-494.

Keys, R. (1981). Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6), 1153-1160.

LaMarche Jr, V. C., & Stockton, C. W. (1974). Chronologies from temperature-sensitive bristlecone pines at upper treeline in Western United States. *Tree-Ring Bulletin* 34:21-45.

Lanner, R. M., & Connor, K. F. (2001). Does bristlecone pine senesce? *Experimental Gerontology*, *36*(4-6), 675-685.

Levin, S. A. (1992). The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology*, 73(6), 1943-1967.

Little, E. L., & Viereck, L. A. (1971). *Atlas of United States Trees:(no. 1146). Conifers and important hardwoods, by EL Little, Jr* (Vol. 1, No. 1146). US Government Printing Office.

Little, E. L. (1979). Four varietal transfers of United States trees [*Persea borbonia pubescens, Populus fremontii mesetae, Quercus turbinella ajoensis, Pinus aristata longaeva*]. *Phytologia*.

Merow, C., Smith, M. J., & Silander Jr, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, 36(10), 1058-1069.

Mosteller, F., & Tukey, J. W. (1968). Data analysis, including statistics. *Handbook of social psychology*, *2*, 80-203.

Munns, E. N. (1938). *The distribution of important forest trees of the United States* (No. 287). US Government Printing Office.

Natural Resource Conservation Service (NRCS). (2016). Available Water Storage. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=nrcs142p2_053628. Accessed May 17, 2021.

Natural Resources Conservation Service (NRCS). (2016). Calcium Carbonate ($CaCO_3$). https://www.nrcs.usda.gov/wps/PA_NRCSConsumption/download?cid=nrcseprd1296354&ext=pdf. Accessed May 25, 2021.

Nyland, R. D. (2016). *Silviculture: concepts and applications*. Waveland Press.

Orlemann, A., Flinders, S. H., & Allphin, L. (2017). The discovery of great basin bristlecone pine, *Pinus longaeva*, in the Tushar Mountains of the Fishlake National Forest in central Utah, USA. *Western North American Naturalist*, *77*(1), 111-117.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, 40(7), 887-893.

PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created 29 Sept 2019.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, 41(4), 629-643.

Rhoden, C. M., Peterman, W. E., & Taylor, C. A. (2017). Maxent-directed field surveys identify new populations of narrowly endemic habitat specialists. *PeerJ*, *5*, e3632.

Riley, S. J., DeGloria, S. D., & Elliot, R. (1999). Index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences*, *5*(1-4), 23-27.

Roughgarden, J. (1974). The fundamental and realized niche of a solitary population. *The American Naturalist*, 108(960), 232-235.

Royle, J. A., Chandler, R. B., Yackulic, C., & Nichols, J. D. 2012. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods in Ecology and Evolugion* 3: 545-554.

Salzer, M. W., Pearson, C. L., & Baisan, C. H. (2019). Dating the Methuselah Walk bristlecone pine floating chronologies. *Tree-Ring Research*, *75*(1), 61-66.

Shemesh, H., Boaz, B. E., Millar, C. I., & Bruns, T. D. (2019). Symbiotic interactions above treeline of long-lived pines: Mycorrhizal advantage of limber pine (*Pinus flexilis*) over Great Basin bristlecone pine (*Pinus longaeva*) at the seedling stage. *Journal of Ecology* 108: 908-916.

Soil Survey Staff. (2020). The Gridded Soils Survey Geographic (gSSURGO) Database for Utah, Nevada, and California. United States Department of Agriculture, Natural Resources Conservation Service. Available online at https://gdg.sc.egov.usda.gov/. (202007 official release).

Smithers, B. V., & North, M. P. (2020). Mechanisms of species range shift: germination and early survival of Great Basin bristlecone pine and limber pine. *Plant and Soil*, 457(1), 167-183.

Syfert, M. M., Smith, M. J., & Coomes, D. A. (2013). The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PloS one*, 8(2), e55158.

U.S. Geological Survey (USGS) Gap Analysis Project (GAP). (2020). Protected Areas Database of the United States (PAD-US) 2.1: U.S. Geological Survey data release, https://doi.org/10.5066/P92QM3NT.

Warren, D. L., Beaumont, L. J., Dinnage, R., & Baumgartner, J. (2019). New methods for measuring ENM breadth and overlap in environmental space. *Ecography*, *42*(3), 444-446.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wieczorek, M. (2014). Area-and depth-weighted averages of selected SSURGO variables for the conterminous United States and District of Columbia (No. 866). U.S. Geological Survey.

Wilcoxon, F. (1946). Individual comparisons of grouped data by ranking methods. *Journal of Economic Entomology*, *39*(2), 269-270.

Wright, R. D., & Mooney, H. A. (1965). Substrate-oriented distribution of bristlecone pine in the White Mountains of California. *American Midland Naturalist*, 257-284.

Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Campbell Grant, E. H., & Veran, S. (2013). Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution*, *4*(3), 236-243.

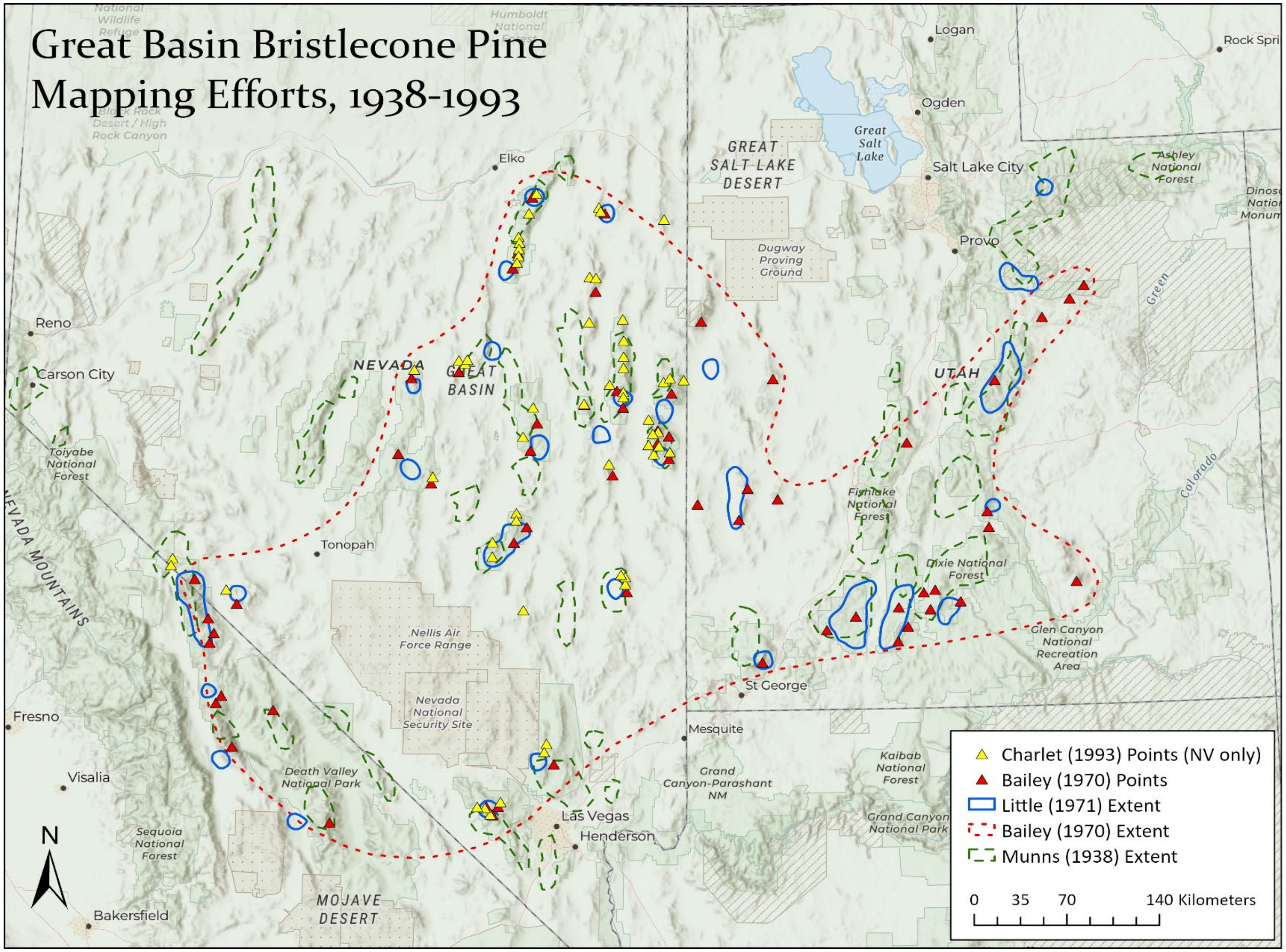

Figure 3.1: Historic map data for Great Basin bristlecone pine (1938-1993).

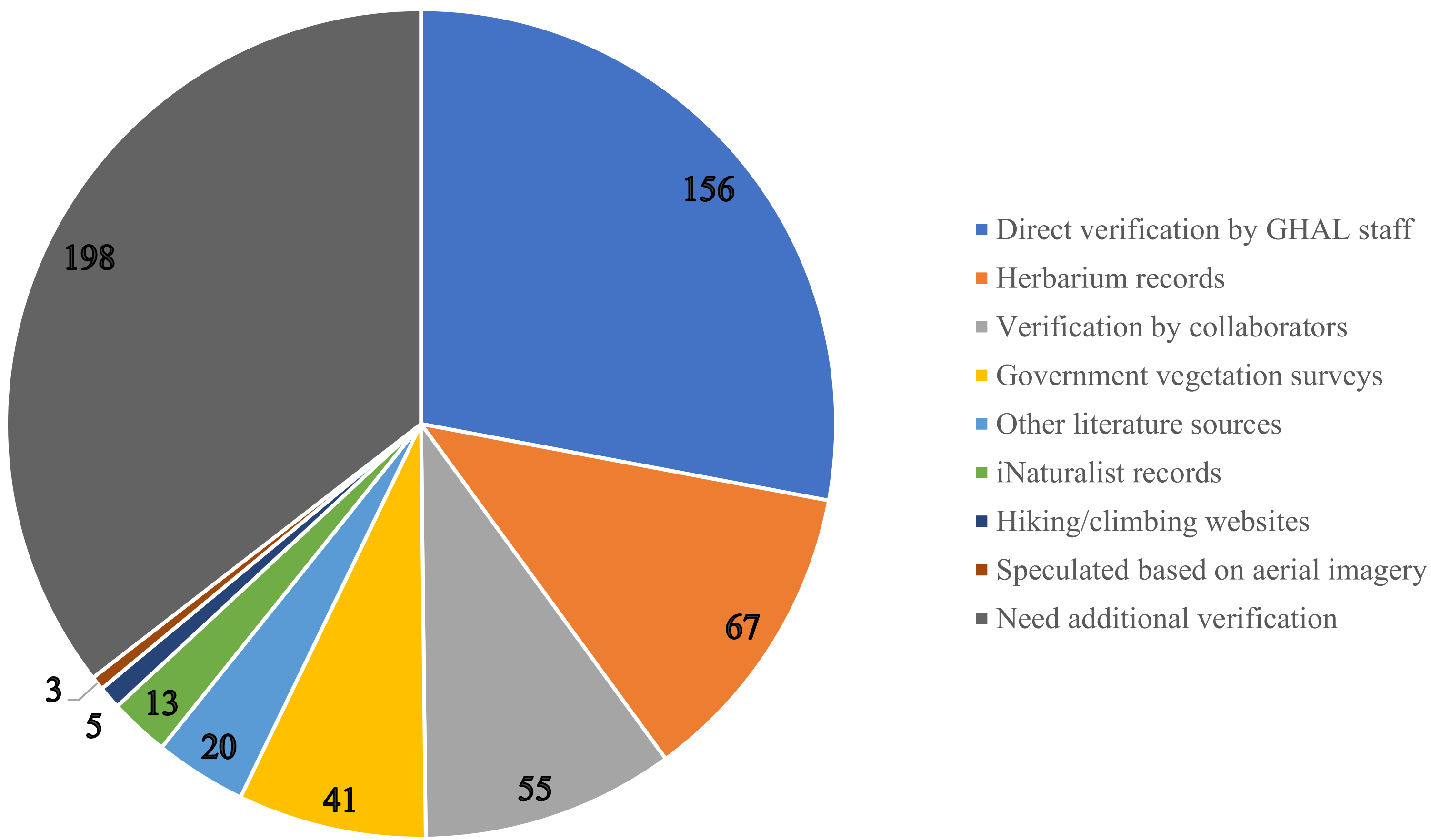

Figure 3.2: Pie chart of source tags assigned to mapped bristlecone pine stand polygons. 558 tags were assigned to 490 polygons, meaning that ~14% of stand polygons were derived from more than one source. 198 polygons (~40%) were tagged as "need additional verification" and were excluded from the HSM analysis.

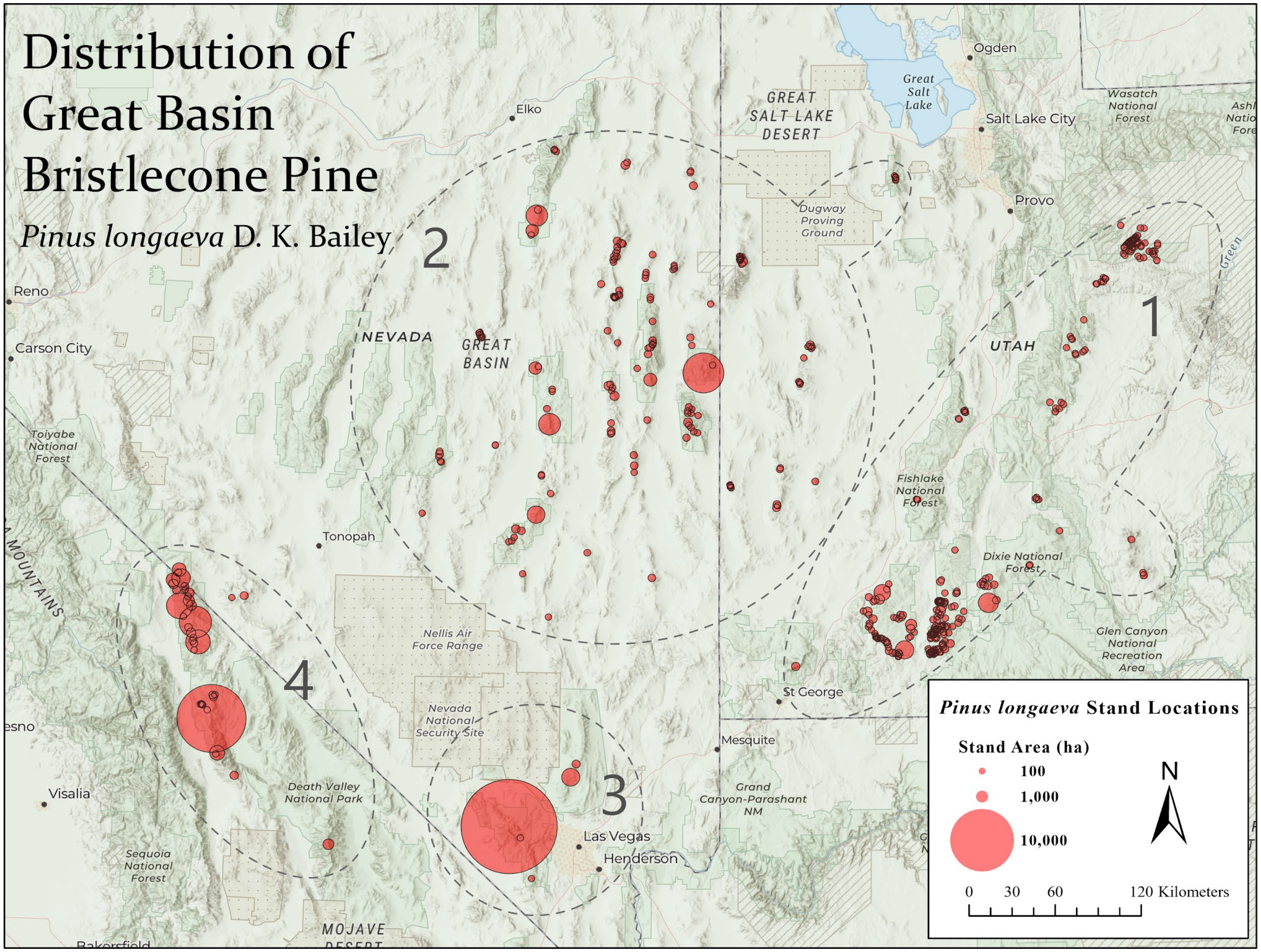

Figure 3.3: Proportional point map showing the distribution of stands of Great Basin bristlecone pine (*Pinus longaeva* D. K. Bailey). Dashed lines enclose regional groupings of bristlecone pine populations: Utah High Plateaus (1), Central Great Basin (2), Southern Great Basin (3), Western Great Basin (4). Basemap data courtesy of Esri.

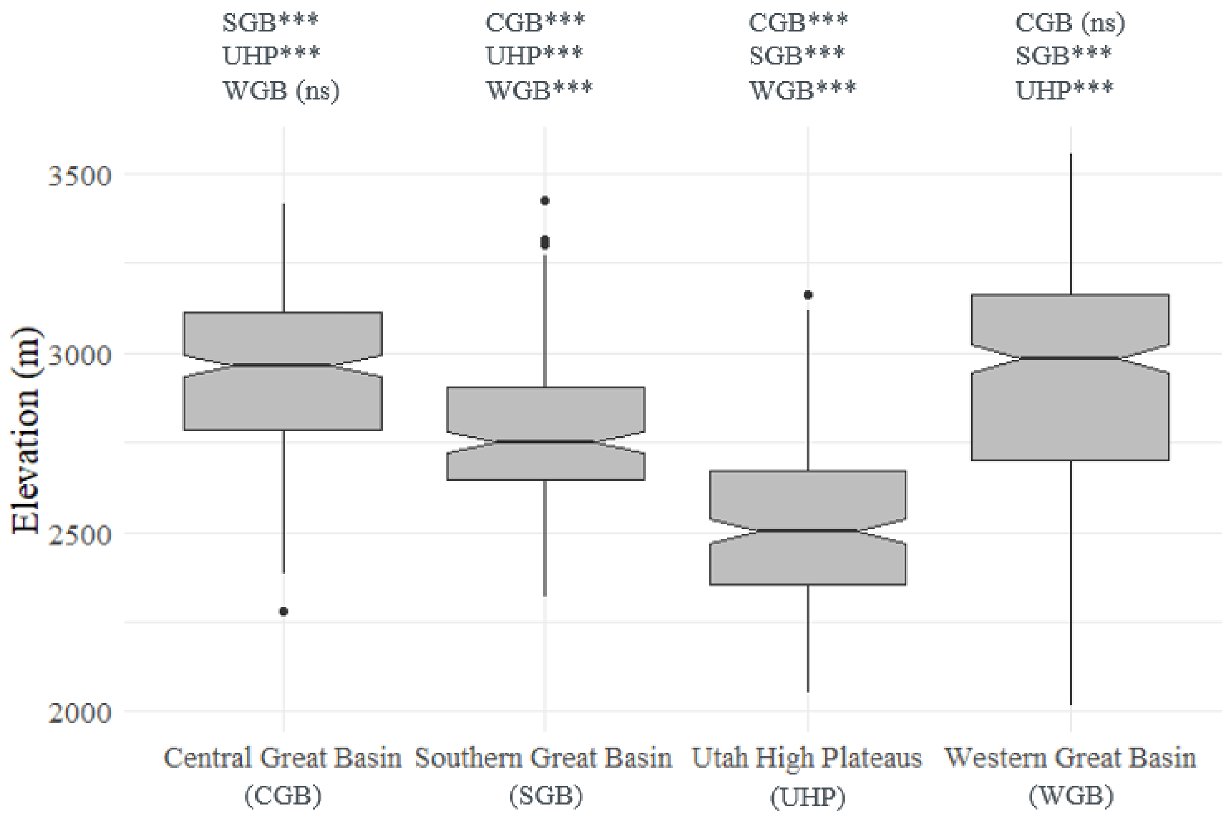

Figure 3.4: Notched box-and-whisker plot showing elevation distributions for randomly selected points within bristlecone pine stands in each regional grouping. Notches represent the confidence interval around the median (Chambers 2018). Asterisks denote significance level when comparing each group with other groups using the pairwise Wilcoxon rank sum test. *** denotes significance at the 0.01 level; "ns" means not significant.

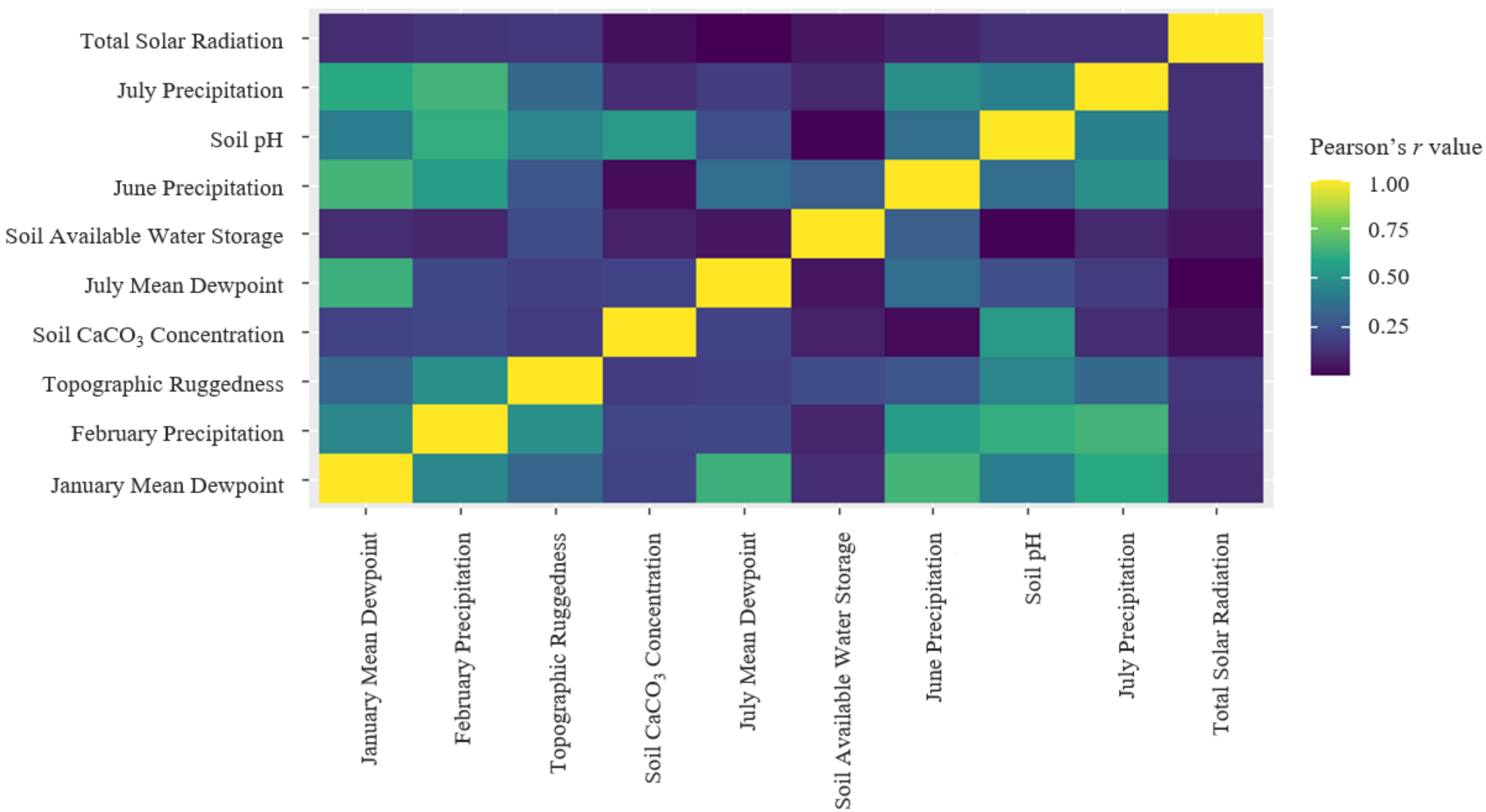

Figure 3.5: Correlation heat map of the 10 selected variables. Cells in the heat map are colorized based on the Pearson's r value between variables.

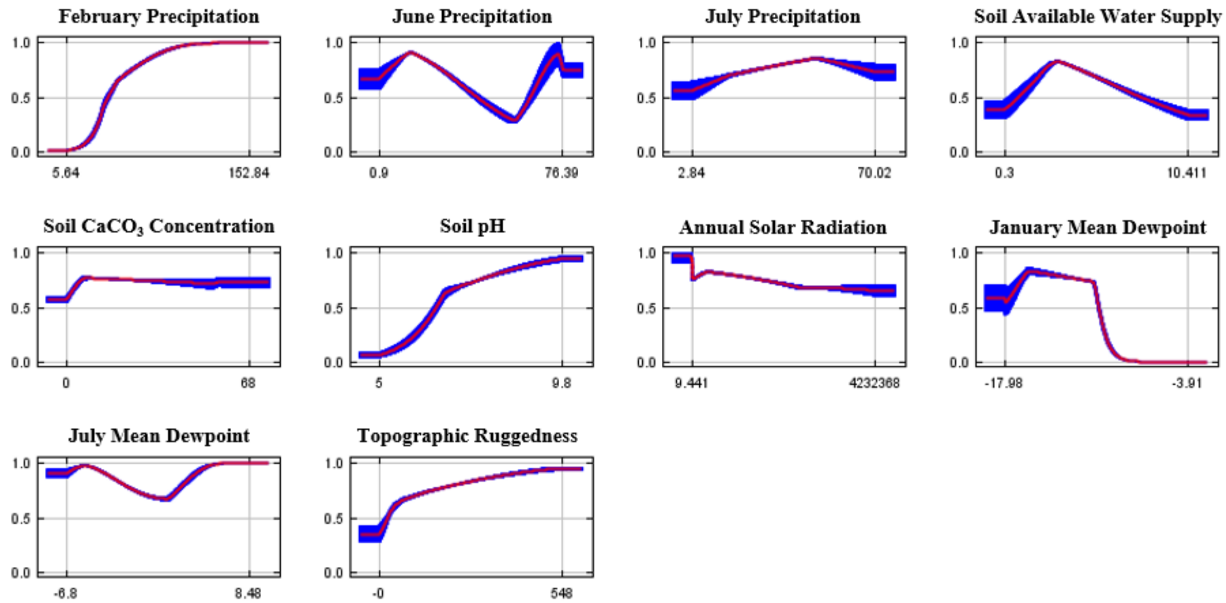

Figure 3.6: Variable response curves for with-soil model.

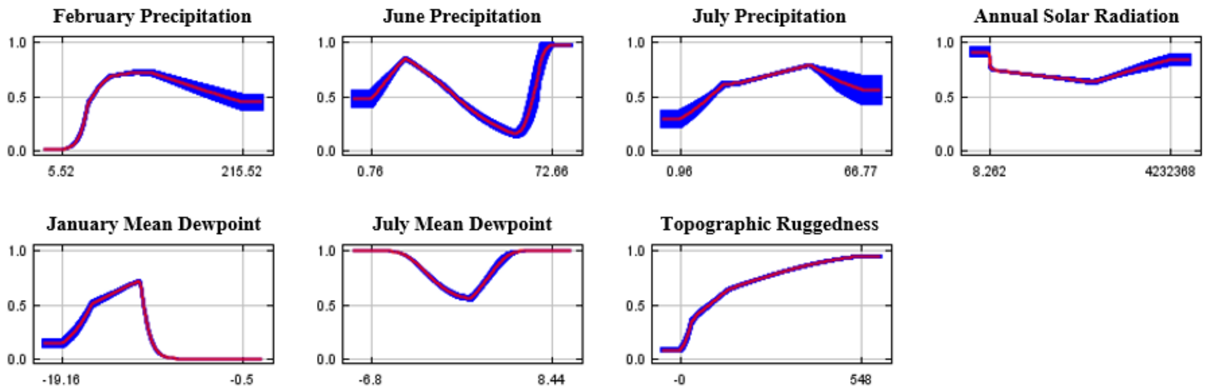

Figure 3.7: Variable response curves for no-soil model.

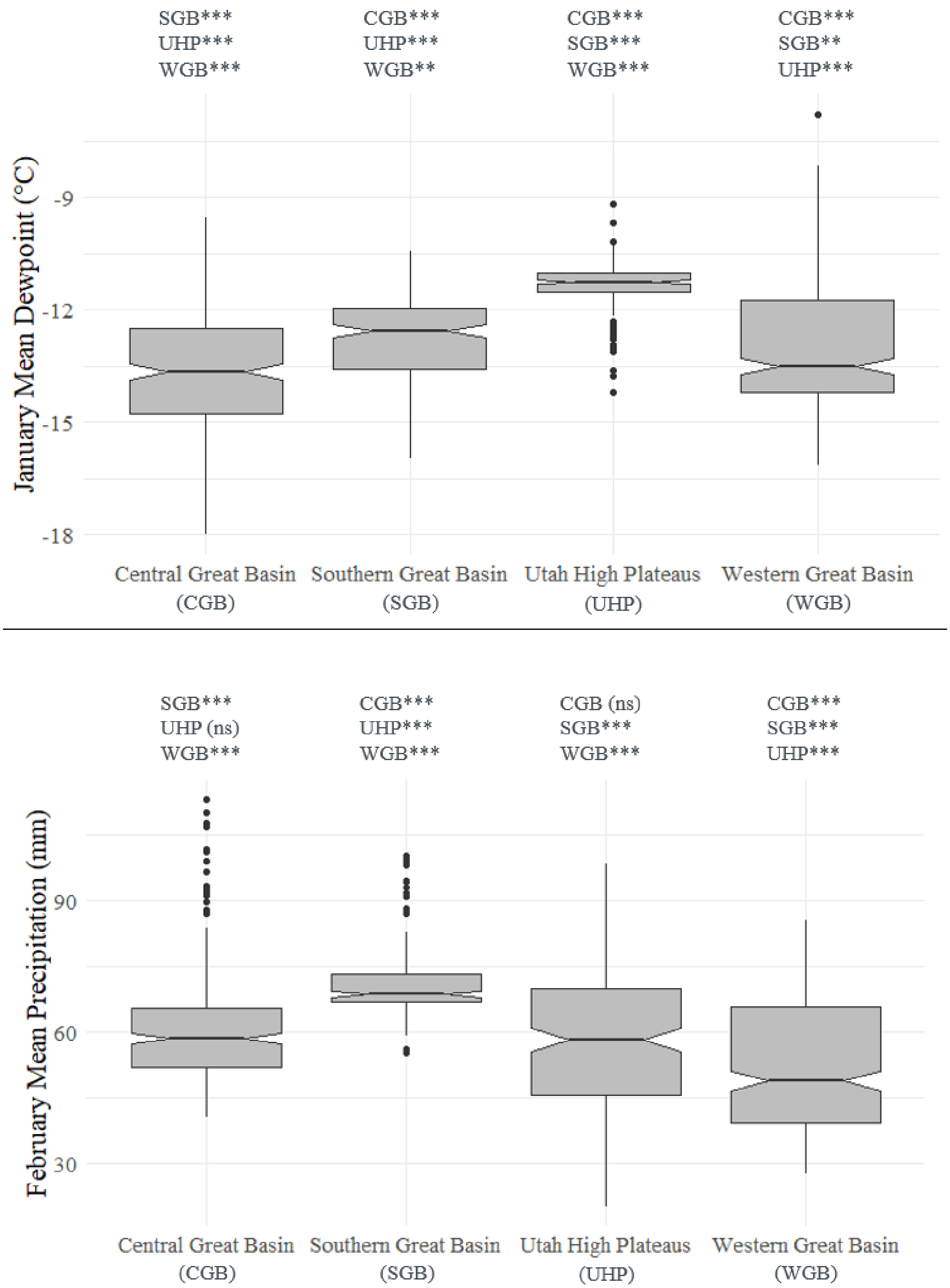

Figure 3.8: Notched box-and-whisker plots of January mean dewpoint temperature (top) and February mean precipitation (bottom). Asterisks denote significance level when comparing each group with other groups using the pairwise Wilcoxon rank sum test. *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; "ns" means not significant.

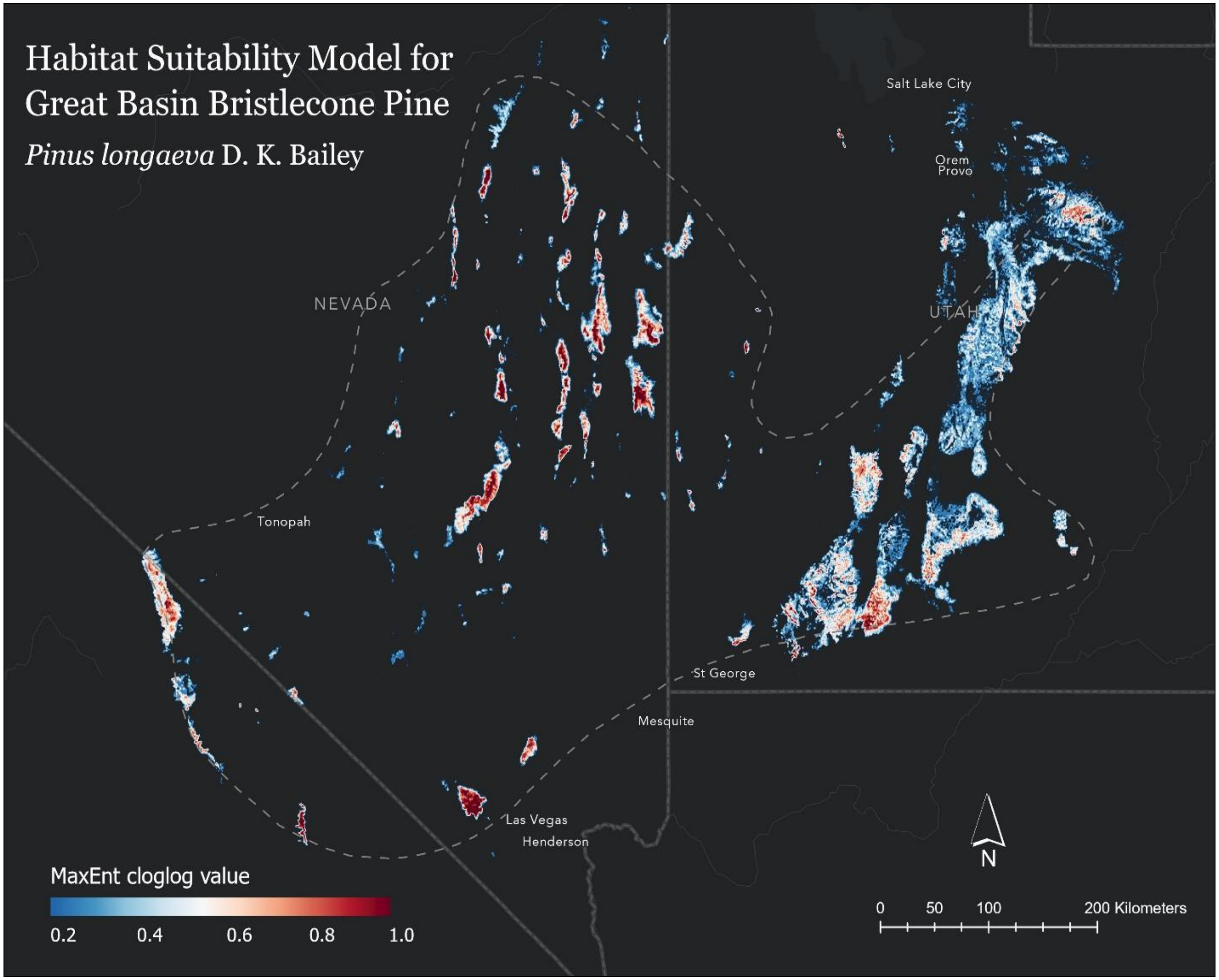

Figure 3.9: MaxEnt-derived composite HSM probability map for bristlecone pine. Bailey's estimated outer extent of bristlecone pine distribution is shown as a dashed gray line (Bailey 1970).

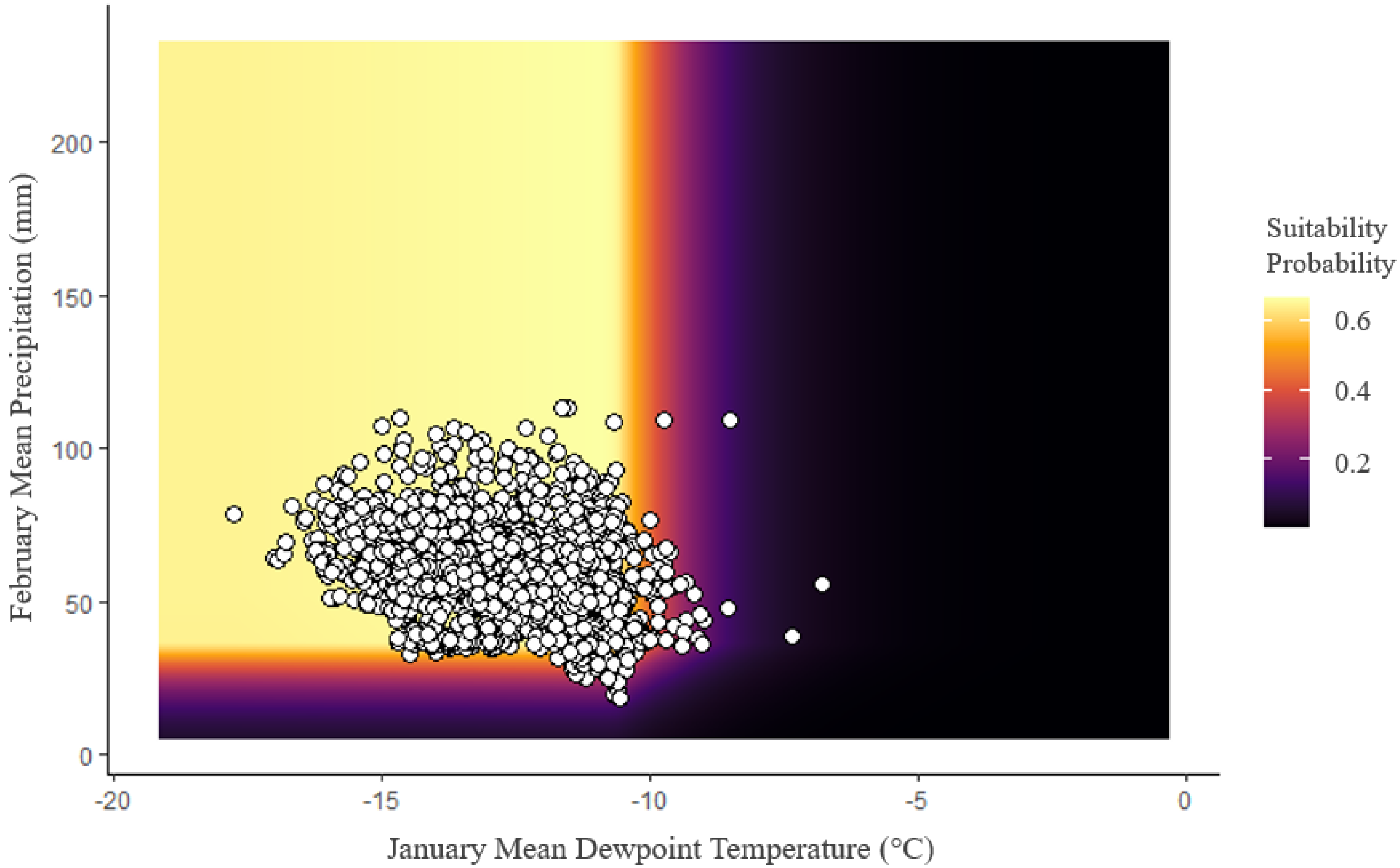

Figure 3.10: Plot of predicted suitability for January mean dewpoint temperature versus February mean precipitation. White points represent the locations of MaxEnt bristlecone pine presence training data in environment space.

TABLES

Table 3.1: MaxEnt environmental predictor variables.

Environmental Predictor Variables	**Number of variables**	**Variable code**
PRISM elevation (m)	1	elevation
Slope (derived from PRISM elevation) (°)	1	slope
Aspect (derived from PRISM elevation) (°)	1	aspect
Topographic Ruggedness Index (derived from PRISM elevation)	1	tri
Annual total solar radiation (derived from PRISM elevation) (Wh/m^2)	1	solar_radiation
PRISM 30 year normal monthly and annual precipitation (mm)	13	precip_XX
PRISM 30 year normal monthly and annual dewpoint temperature (°C)	13	dewpoint_XX
PRISM 30 year normal monthly and annual maximum temperature (°C)	13	tmax_XX
PRISM 30 year normal monthly and annual minimum temperature (°C)	13	tmin_XX
USDA/NRCS gSSURGO available water storage, 0-50 cm (cm/cm)	1	soil_aws
USDA/NRCS gSSURGO calcium carbonate ($CaCO_3$), 0-50 cm (cm/cm)	1	soil_caco3
USDA/NRCS gSSURGO pH (1-to-1 water), 0-50 cm	1	soil_pH
Total:	60	

Table 3.2: Variable permutation importance for MaxEnt model with soil data vs. model with no soil data.

Variable	Importance (with-soils model)	Importance (no-soils model)
January Mean Dewpoint	76.5%	68.1%
February Precipitation	13.8%	14.2%
June Precipitation	2.0%	3.1%
July Mean Dewpoint	1.8%	6.4%
Topographic Ruggedness	1.6%	6.9%
Soil Available Water Storage	1.6%	N/A
Soil pH	0.9%	N/A
Soil $CaCO_3$ Concentration	0.9%	N/A
Total Solar Radiation	0.7%	0.4%
July Precipitation	0.3%	1.0%

Table 3.3: Protected status information for mapped stands of bristlecone pine. Spatial overlap between GAP 2 and GAP 3 classes results in the total exceeding 100%. Data courtesy of the USGS Protected Areas Database of the United States.

GAP Status Code	Hectares	Percent of Total	GAP Code Description
1	54,242.9	54.3%	Managed for biodiversity - disturbance events proceed or are mimicked
2	27,000.2	27.1%	Managed for biodiversity - disturbance events suppressed
3	92,949.5	93.1%	Managed for multiple uses - subject to extractive (e.g. mining or logging) or OHV use
4	1,785.1	1.8%	No known mandate for biodiversity protection
Total Mapped Habitat Area:	99,810.3		