

Species Habitat Suitability Models and Condition Assessment
for
Culturally Significant Plants
in
Bandelier National Monument, New Mexico



Natural Heritage New Mexico Report 427
February 2024



Species Habitat Suitability Models and Condition Assessment
for Culturally Significant Plants
in
Bandelier National Monument, New Mexico¹



Esteban Muldavin and Corrie Gonzalez
Natural Heritage New Mexico
Museum of Southwestern Biology
University of New Mexico, Albuquerque, New Mexico 87131
February, 2024



Summary

The goal of this project was to explore the distribution of a suite of culturally important plant species on Bandelier National Monument (BAND) and surrounding areas to help meet the park's obligations to support plant collection by designated tribes. Unfortunately, BAND has undergone significant wildfire and flooding events in the last few decades where wildfires have destroyed large areas of mature forest and have had numerous detrimental secondary impacts, especially flooding in riparian areas and exotic plant species expansion. In this context, the park was specifically interested in the occurrence, distribution, and abundance of a set of priority, culturally significant plant species to better understand their current status and potential sustainability into the future to inform the management.

To do this, we developed habitat suitability models (HSMs) based on known locations of the species, a set of environmental predictors, and a fire history of the park. For each species, we used five different computer machine-learning approaches and then combined them into an ensemble model that displays the highest probability locations for the species based on agreement among five underlying models. Where appropriate, we then intersected those distributions with fire-intensity maps and used other resource maps to better understand the status of the species to inform management decisions about resource collection. This comprehensive approach helps determine the most probable distribution of favorable environmental conditions for a species. Given the actual distribution of a species is likely unknown, accurate HSMs play a crucial role not only in guiding the harvest of plants for cultural purposes, but also in facilitating research, implementing conservation measures, and assisting land managers in identifying strategies to mitigate potential adverse impacts on species.

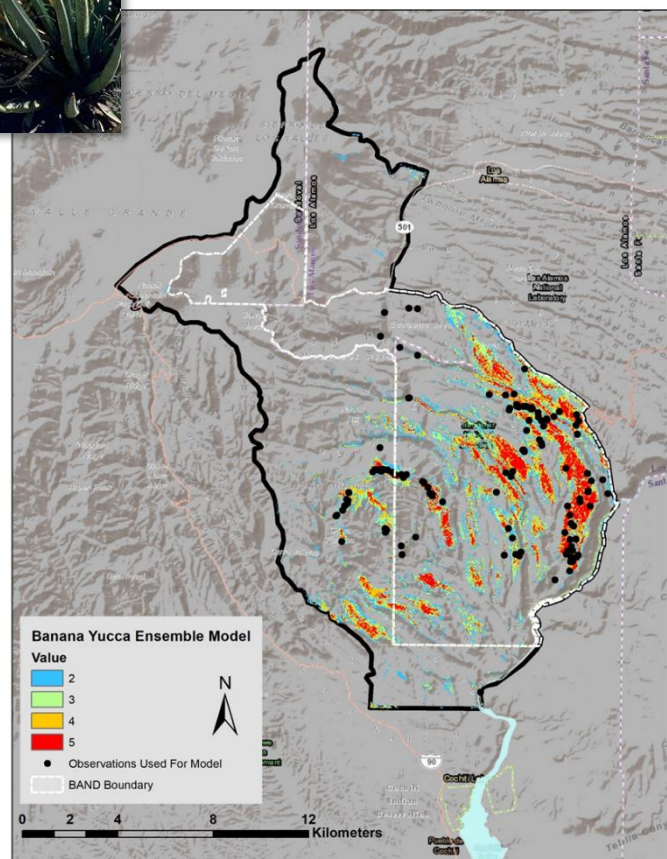
The priority species were: Douglas-fir, limber pine, banana yucca, chokecherry, common hop-tree, Gambel's oak, alderleaf mountain mahogany, Indian ricegrass, common yarrow, and amaranth. The models that we produced had good-to-excellent accuracy and reasonable patterns of distribution to aid management. For each species, we have provided summary fact sheets with a brief description of the species, mapped model outcomes, and interpretations.

¹ Natural Heritage New Mexico Report 427 submitted in partial full fulfillment of Grant No. P21AXC10652-00, Cooperative Research and Training Programs –Resources of the National Park System; Colorado Plateau Cooperative Ecosystems Studies Unit NPS-COPL-SESU-2019

These models represent the distribution of habitats based on data that spans about a 30-year period—a time when the park’s landscape has undergone major changes because of severe fire. We have attempted to account for this through understanding individual species responses to fire and explicit mapping of fire severity for those species that were most severely impacted (conifer trees). Beyond this, we suggest that the species be surveyed and monitored directly to evaluate what their actual responses to the fires have been. Lastly, in the face of climate change, the next step for modeling would be to take them into the future to guide their use and management for the long term.



Banana Yucca Habitat Suitability Model



Species Habitat Suitability Models and Condition Assessment for Culturally Significant Plants In Bandelier National Monument, New Mexico

Table of Contents

Summary.....	ii
Introduction.....	1
Study area.....	3
Methods	6
Modeling approach summary.....	6
Results and Discussion	9
Model output and accuracy.....	9
Model outcomes	10
Herbaceous plants.....	11
Indian rice-grass.....	11
Common yarrow	13
Amaranth	15
Shrubs	17
Banana Yucca	17
Black Chokecherry	19
Common hop-tree	21
Gambel’s oak	23
Alderleaf Mountain Mahogany.....	25
Conifer Trees.....	27
Douglas fir	27
Limber pine	30
Conclusions.....	33
References.....	34
Appendix 1 Modeling Details.....	35
Environmental Predictor Data	35
Pseudo-Absence (Background) Points	35
Modeling Algorithms.....	35
GLM-Based Variable Selection.....	36
Model Fitting and Ensemble Model Construction.....	37
Variable Importance.....	37

Accuracy Assessment	38
References	39

List of figures

Figure 1. The modeling domain for culturally significant species includes not only BAND but a buffer zone around the park on primarily US Forest Service lands and Valles Caldera National Preserve.....	2
Figure 2. In 2002-03, bark beetle (<i>Ips</i> sp.) and the associated fungal rot (<i>Ophiostoma</i> sp. and <i>Ceratocystis</i> sp.) killed thousands of piñon and ponderosa pines all across the mid elevations of the monument. In addition, fires, such as the 2000 Cerro Grande fire in the background, have had a significant impact on the vegetation of the park. (photo: Craig Allen).....	5
Figure 3. High and Moderate Severity Burned Areas within the modeling boundary. The inset map shows Las Conchas and Cerro Grande fire boundaries as well as high/moderate severity burned areas for the entirety of the fire boundaries.....	7
Figure 4. Indian Rice-grass (Sally and Andy Wasowski, Lady Bird Johnson Wildflower Center).	11
Figure 5. Indian rice grass potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 models.....	12
Figure 6. Common yarrow (Stephanie Brundage, Lady Bird Johnson Wildflower Center).	13
Figure 7. Common yarrow potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.	14
Figure 8. Illustration of two of potential amaranths on BAND	15
Figure 9. Amaranth species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.	16
Figure 10. Banana yucca (Norman G. Flaigg, Lady Bird Johnson Wildflower Center).	17
Figure 11. Banana Yucca species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.	18
Figure 12. Black chokecherry (Courtesy Sally and Andy Wasowski, Lady Bird Johnson Wildflower Center)	19
Figure 13. Black chokecherry species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.	20
Figure 14. Common hop-tree (USDA Plants)	21
Figure 15. Common hop-tree species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.	22
Figure 16. Gambel's Oak leaves and acorn (Courtesy James L. Reveal, Lady Bird Johnson Wildflower Center).....	23
Figure 17. Gambel's oak species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.	24
Figure 18. Alderleaf mountain mahogany (Courtesy Stephanie Brundage, Lady Bird Johnson Wildflower Center).....	25
Figure 19. Alderleaf mountain mahogany species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.....	26
Figure 20. Rocky Mountain Douglas-fir cones (Dave Powell, USDA Forest Service (retired), Bugwood.org).....	27

Figure 21. Douglas-fir species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.	28
Figure 22. Douglas-fir species potential suitable habitat ensemble model restricted to unburned or low-severity burn locations. Value reflects the co-occurrence of 2 to 5 model.	29
Figure 23. Limber pine needs and cones (James reveal, Lady Bird Johnson Wildflower Center). .	30
Figure 24. Limber pine species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.	31
Figure 25. Limber pine species potential suitable habitat ensemble model restricted to unburned or low-severity burn locations. Value reflects the co-occurrence of 2 to 5 model.....	32

List of Tables

Table 1. Modeled culturally significant plant species (CIPS) by lifeform, common name, and scientific name.	3
Table 2. List of 38 environmental predictor variables used in the modeling process.....	8
Table 3. Accuracy Measures: Percent Correctly Classified (PCC) and Area Under the Curve (AUC) for stacked Ensemble model for each species..	9
Table 4 Predicted suitable habitat area for each species by the number of co-occurring models	10

Introduction

Bandelier National Monument (BAND), in north-central New Mexico, is a premiere pre-conquest site of Native American structures and villages dating back 11,500 years, whose history is linked to current Puebloan tribes of the region (Figure 1).

In this respect, the park contains over 200 known culturally important plant species (CIPS) and since 2000 has a legislative requirement for plant collection by some associated tribes (Public Law 106-246), and now expanded authority for additional tribal collecting under the 2015 NPS final rule on plant gathering. Both authorities necessitate environmental analysis of potential effects of collecting on vegetation communities, among other issues. Unfortunately, BAND has undergone significant wildfire and flooding events in the last few decades because of past land-use practices and changing climatic conditions that may have affected these species. High fuel loads, combined with drought conditions, have led to a series of increasingly larger and higher severity fires in the Jemez Mountains. The most recent, the Las Conchas fire of 2011, was the largest wildfire in New Mexico history at the time and burned over 60% of the park (see Figure 1). Over the last 30 years, wildfires have destroyed large areas of mature forest and have had numerous detrimental secondary impacts, especially on riparian areas (e.g., exotic plant species expansion, post-fire floods) (Jacobs et al. 2015). In this context, new assessments on the status of culturally significant plant species were needed in response to these recent extreme disturbance events. Here, the park was specifically interested in the occurrence, distribution, and abundance of a set of priority, culturally significant plant species to better understand their current status and potential sustainability into the future to inform their management.

To help meet this need, we developed habitat suitability models (HSMs), also known as species distribution models, for 10 high-priority CIPS identified by BAND (Table 1). These species were chosen by the park based on their cultural significance (as identified in Stoffle et al. (2007) and other sources) and their relatively common status. HSMs are constructed using data from a set of environmental variables at known species locations (e.g., soils, elevation, precipitation, topography, etc.) to predict additional areas that may offer suitable environmental conditions and potentially be inhabited by the target species. We used five different algorithms and stacked them to create “ensemble” models showing the highest probability locations for a species based on concordance among the models.

Where appropriate, we then intersected those distributions with fire-intensity maps and used other resource maps to understand the status of the species to inform management decisions about resource collection. This comprehensive approach helps determine the most probable distribution of favorable environmental conditions for a species. Given that the actual distribution of a species is often unknown, accurate HSMs play a crucial role not only in guiding the harvest of plants for cultural purposes, but also in facilitating research, implementing conservation measures, and assisting land managers in identifying strategies to mitigate potential adverse impacts on species.

Acknowledgments.

John Leonard lent his modeling expertise to the project. Grace McCarthy and Amy Urbanovsky were responsible for gathering additional field data and aggregating the data for analysis.

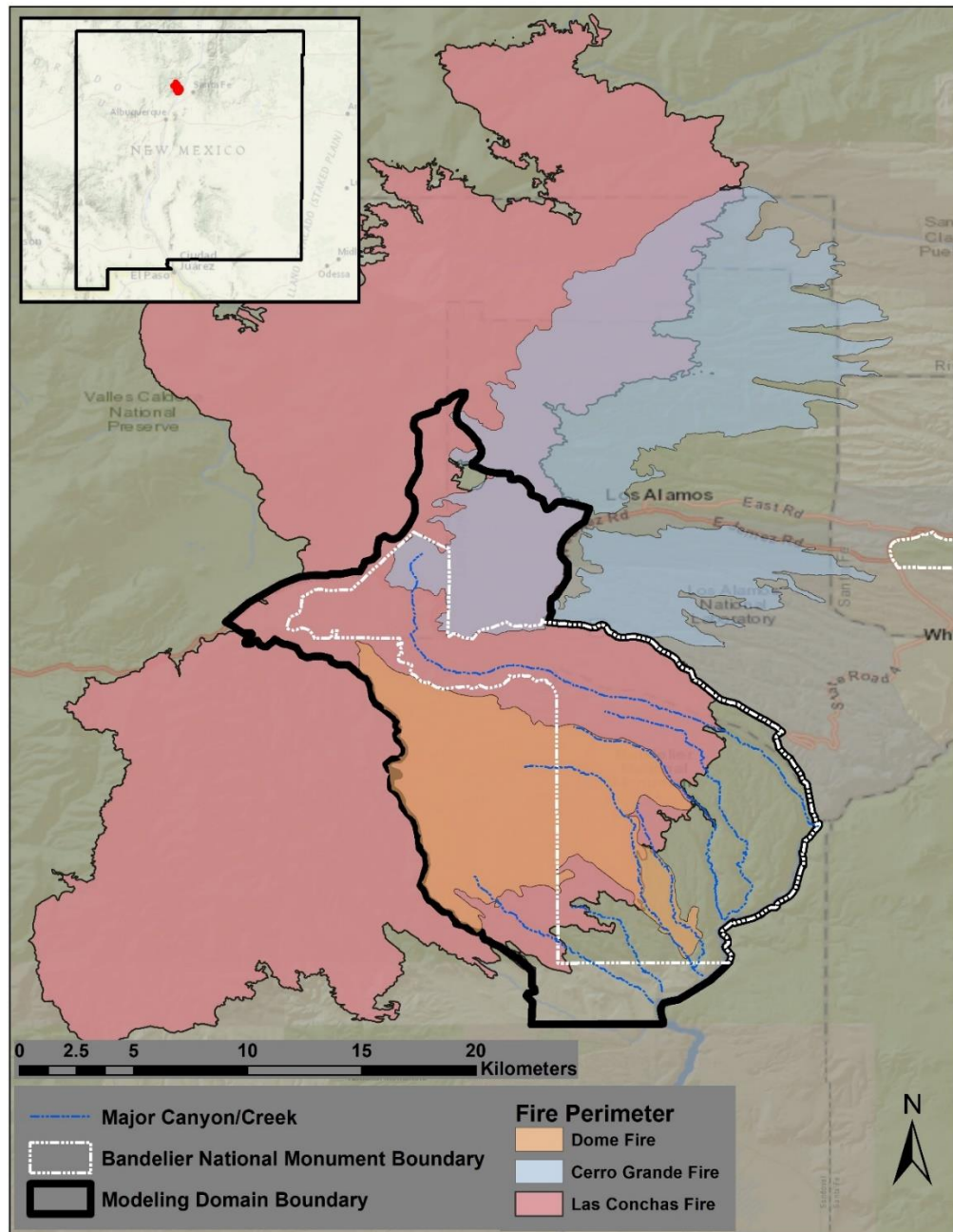


Figure 1. The modeling domain for culturally significant species includes not only BAND but a buffer zone around the park on primarily US Forest Service lands and Valles Caldera National Preserve.

Table 1. Modeled culturally significant plant species (CIPS) by lifeform, common name, scientific name, and the number observations available for modeling.

Species Modeled		Number of Obs
Trees		
Douglas-fir	<i>Pseudotsuga menziesii</i>	362
Limber pine	<i>Pinus flexilis</i>	45
Shrubs		
Banana yucca	<i>Yucca baccata</i>	120
Chokecherry	<i>Prunus virginiana</i>	89
Common hop-tree	<i>Ptelea trifoliata</i>	56
Gambel's oak	<i>Quercus gambelii</i>	389
Alderleaf mountain mahogany	<i>Cercocarpus montanus</i>	156
Grasses and forbs		
Indian ricegrass	<i>Achnatherum hymenoides</i>	62
Common yarrow	<i>Achillea millefolium</i>	117
Amaranth	<i>Amaranthus</i> spp.	139

Study area

The study area, or modeling domain, includes Bandelier National Monument and a buffer zone extending to the east and north, covering US Forest Service lands of the Santa Fe National Forest, Jemez District, and the Valles Caldera National Preserve (see Figure 1). This buffer zone, agreed upon by Natural Heritage New Mexico and BAND staff, helps to mitigate edge effects at the boundaries of Bandelier National Monument, thus improving prediction accuracy at the study area's edges. The western boundary follows the edge of Cochiti Canyon; to the northwest, the Sierra de los Valles; along the Los Alamos National Laboratory boundary to the northeast; and down to the Rio Grande in White Rock Canyon to Cochiti Pueblo. The total area of the study area is 25,700 ha (63,506 acres).

Elevations range from the Rio Grande in the bottom of White Rock Canyon at 1,600 m (5,300 ft) to the summit of Cerro Grande at 3,109 m (10,199 ft) in the Sierra de los Valles. The climate of is characterized by cool-to-cold, dry winters, and warm, wet (monsoonal) summers (Jacobs, Gatewood, and Allen 2002). In this essentially montane climate, the mean daily temperature extremes range between -10°C and 6°C (14°F and 21°F) in January and 13°C and 30°C (55°F and 86°F) in June, considered the coldest and warmest months of the year, respectively. The coldest recorded daily temperature at nearby Los Alamos was -27.2°C (-16°F) and the warmest 35°C (95°F). Jacobs, Gatewood, and Allen (2002) report that the mean annual precipitation is about 427 mm (16.8 in) as recorded at the fire lookout tower near park headquarters (unpublished weather records for 1925-2002, on file at Bandelier National Monument). Winter precipitation is delivered principally as snow by low-pressure systems that sweep from west to east across the Southwest, coalescing with moisture from the Pacific Ocean or the Gulf of Mexico. Winter precipitation is generally followed by a seasonal dry period during the months of May and June. The spring dry period is usually relieved by the onset of the Mexican monsoon; this weather

pattern typically delivers at least 40% of the annual precipitation July through September, and is associated mostly with short-duration, high-intensity thunderstorms. Each summer, as high pressure becomes entrenched off the coast of Baja California, low pressure in the Southwest feeds Pacific moisture across the region, fueling the development of afternoon thunderstorms. The magnitude, frequency, and tracking of individual large, intense thunderstorm cells during this period can account for large year-to-year variability in annual rainfall and between local areas.

With respect to the landscape and geology, Jacobs et al. (2002) reports:

“Bandelier National Monument is located on the southeast flank of the Jemez Mountains in north-central New Mexico (Jacobs et al. 2002). The volcanic Jemez Mountains were formed from a series of events dating from 13 million years B.P. to as recently as 50,000 years ago. The Valle Grande, the central feature of this landscape, was created by two major eruptions at 1.6 and 1.2 million years B.P.; cumulatively these deposited the 300-meter-thick Bandelier tuff in two distinct layers. The most recent eruption (El Cajete at 50 to 60,000 B.P.) covered the local landscape with many meters of pumice, much of which has been subsequently eroded and reworked, leaving pumice patches predominately on east-facing slopes and deep alluvial deposits on lower slopes (Goff et al. 1996). The eastern facing flank of the Jemez Mountains, called the Pajarito Plateau, has been incised into alternating landscape patterns of mesas separated by steep walled canyons.”

The vegetation communities of Bandelier are diverse. They range from cold desert shrublands and semiarid grasslands of the lowland canyons and foothills, through extensive one-seed juniper woodlands at mid-elevations (which before the 2003 die-off were codominated with two-needled pinyon) up to mixed conifer and subalpine forest and woodlands along with aspen at the highest elevations. See Muldavin et al. (2011) for detailed descriptions of the vegetation communities.

Historical fires and recent droughts have had a significant impact on vegetation patterns in the park. The first was the La Mesa Fire in the summer of 1977 that burned 6,354 ha (15,700 ac) on the plateaus of the central portion of the monument and the adjacent LANL lands. The Dome Fire occurred in the spring of 1996 and burned 6,677 ha (16,500 ac) over the western portion of the monument and adjacent USFS lands. The Cerro Grande Fire of 2000 was initiated as a prescribed fire in the upper Frijoles watershed within the monument, but control was lost and it burned some 16,909 ha (42,848 ac) altogether, mostly to the north on USFS, LANL, and private lands, with only about 494 ha (1,254 ac) within the park (based upon GIS coverages of the fire). The 2011 Las Conchas fire had a far greater impact burning 63,400 ha (160,000 ac) stretching from the tops of the watershed down into the juniper zone at mid elevation. The most recent fire was the 2022 Cerro Pelado fire which was restricted to about 717 ha (1,773 ac) within the upper reaches of the park. The fires had major impacts on vegetation patterns, leaving behind partially burned patches, areas of standing dead trees, and extensive open areas that are now in the process of revegetating.

Vegetation mortality due to recent drought also has had significant impacts on vegetation patterns at BAND (Figure 2). Water stress from dry years beginning in 1996 and culminating in severe drought from 2000 to 2004 killed many trees and also caused dieback of shrubs and

grasses at BAND. Tree mortality was caused by combinations of direct water stress, associated outbreaks of bark beetles (especially *Ips*, *Dendroctonus*, and *Scolystus* sp.), and fungal rots (*Ophiostoma* sp. and *Ceratocystis* sp.) carried by the beetles. This combination of drought and bark beetles killed trees of many species across elevational gradients in Bandelier (and throughout the Southwest), ranging from Douglas-fir and white fir up high down through ponderosa pine, piñon, and juniper on lower mesas (McDowell et al. 2010; Allen et al. 2010). Piñon mortality was particularly high in the Bandelier landscape, commonly killing over 95% of the trees in a given stand (cf. Breshears et al. 2005). While overall tree mortality rates were very high at many sites in Bandelier, young trees less than five cm (2 in) in basal diameter typically survived and are expected to support recovery of local forest and woodland canopies on most sites.



Figure 2. In 2002-03, bark beetle (*Ips* sp.) and the associated fungal rot (*Ophiostoma* sp. and *Ceratocystis* sp.) killed thousands of piñon and ponderosa pines all across the mid elevations of the monument. In addition, fires, such as the 2000 Cerro Grande fire in the background, have had a significant impact on the vegetation of the park. (photo: Craig Allen).

Methods

Modeling approach summary

For the modeling, we compiled species records sourced from the Natural Heritage New Mexico Ecology Plots database. These records predominantly originated from plots surveyed by NHHM during the vegetation mapping initiatives for Bandelier National Monument in 2000-2003 and subsequent post-fire monitoring plot surveys in 2015-2016. A small number of records were also gathered from SEINet and iNaturalist. Additionally, NHHM conducted species-specific surveys at Bandelier National Monument in September 2021.

Using these observations, we developed habitat suitability models (HSMs) for the 10 CIPS listed in Table 1 based on a suite of 38 high-resolution (30 m) environmental predictors (Table 2). As a first step, reduced the variable set for each species using generalized linear modeling to remove highly correlated predictors. We then used an ensemble approach whereby each species was modeled independently for potentially suitable habitat using five different statistical and machine learning algorithms: randomForest (RF), Generalized Linear Models (GLMs), Extreme Gradient Boosting (XGB), Multivariate Adaptive Regression Splines (MARS) and Bayesian Additive Regression Trees (BART). Each algorithm generates a probability surface of potential suitable habitat from 0 to 100. Lower probabilities were eliminated using the MaxKappa statistic as a threshold (commonly around 50% plus or minus 10% probability of occurrence).

In September and October 2023, we collected additional validation data strategically in areas bordering species ranges indicated by preliminary models. These data were employed to fine-tune the models, resulting in more inclusive models of suitable habitat.

For Douglas-fir and limber pine, we produced two models each—one that predicts overall potential habitat if there had been no fires, and one that takes into account the severity of recent fires that would limit the actual distribution of the trees in the near-term (the next 30 years or so depending on reproduction). To do this, we used Burned Area Emergency Response ([BAER](#)) fire severity maps for the Cerro Grande (2000) and Las Conchas (2011) fires to mask out areas in the potential habitat model where moderate- to high-severity fire had occurred (Figure 3). The expectation was that most if not all trees may have been killed in those locations (1997 Dome fire data was not available).

To simplify the interpretation of the models, for each 30 x 30 m raster cell in a model above the MaxKappa threshold was assigned a “1” as reflecting suitable habitat, and cells below or equal to the threshold were given a “0” as unsuitable for a species. The now binary (1 or 0) models were stacked to create an ensemble model where each cell was assigned a value of 1 to 5 where 5 denoted unanimous agreement among all five models regarding environmental suitability, whereas pixels with a value of 0 were unanimously considered unsuitable. Accordingly, the higher the count of concurring models, the higher likelihood of suitable habitat, i.e., values 4 and 5 to have the highest probability of a species being present. For BAND, we further eliminated cells from the ensemble model with scores of 1 for consideration as suitable habitat to deemphasize the bias of any given model.

To evaluate accuracy, 20% of the observations were randomly withheld to test each model. In addition, we also qualitatively assessed the efficacy of the ensemble models on how well they

encompassed the range of possible environmental conditions for each species based on spatial concordance with the model observations themselves and new independent locations gathered as part of an accuracy assessment survey conducted in summer 2023. See Appendix I for modeling details and variables.

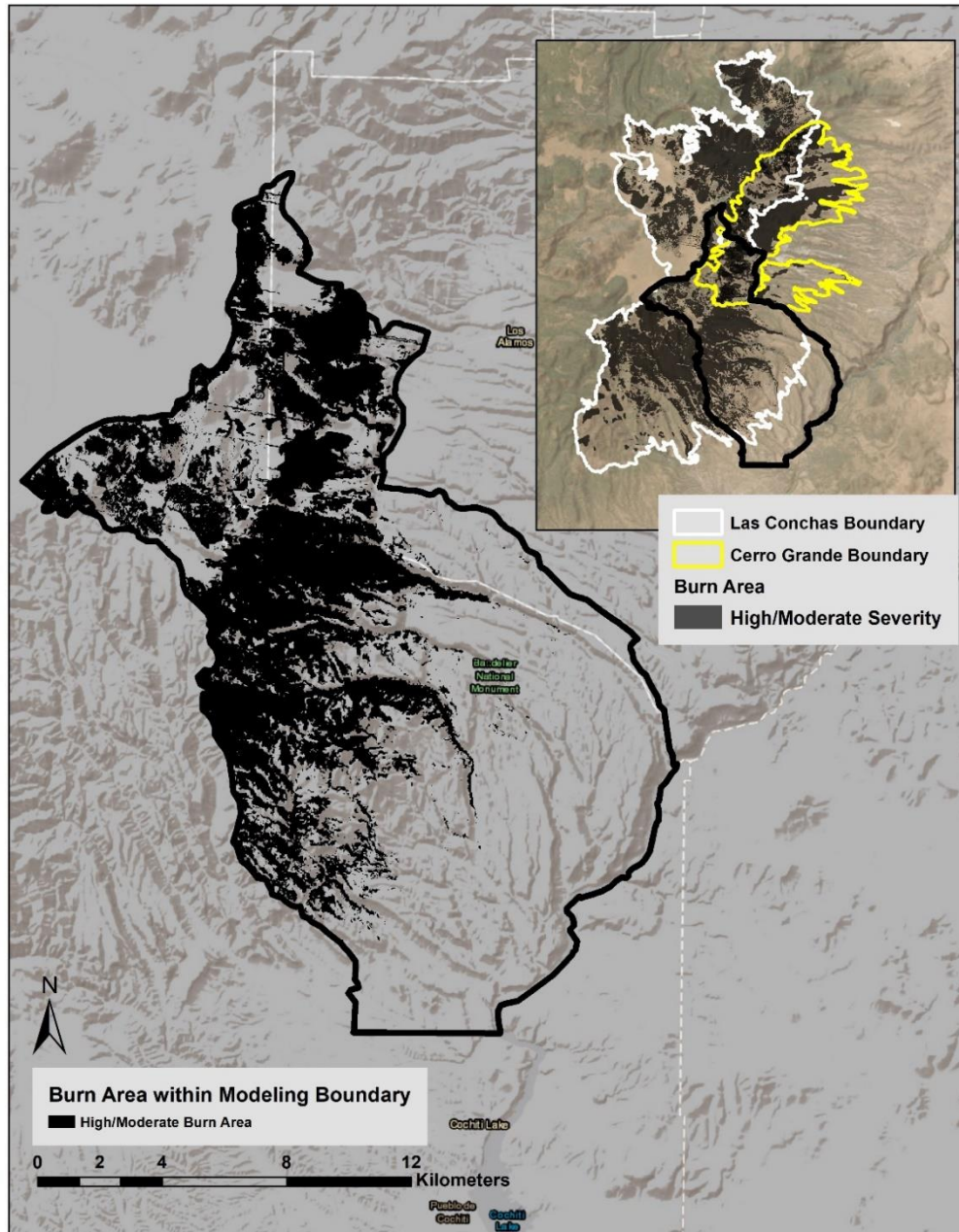


Figure 3. High and Moderate Severity Burned Areas within the modeling boundary. The inset map shows Las Conchas and Cerro Grande fire boundaries as well as high/moderate severity burned areas for the entirety of the fire boundaries.

Table 2. List of 38 environmental predictor variables used in the modeling process.

Layer	Description
clim_bio1_AnnMnTmp	Annual Mean Temperature
clim_bio10_MnTmpWQ	Mean Temperature of Warmest Quarter
clim_bio11_MnTmpCQ	Mean Temperature of Coldest Quarter
Clim_bio12_AnnPrec	Mean Annual Precipitation
clim_bio16_PrecWetQ	Precipitation of Wettest Quarter
clim_bio17_PrecDryQ	Precipitation of Driest Quarter
clim_ClimWatDeficit	Climatic Water Deficit
clim_GrowDegDays	Growing Degree Days
Landform_1_3x3	Warm Peak/Ridge 1 cell Square
Landform_10_10cell	Warm Lower Slope 10-cell Radius
Landform_11_10cell	Neutral Lower Slope 10-cell Radius
Landform_12_10cell	Cool Lower Slope 10-cell Radius
Landform_14_10cell	Valley 10-cell Radius
Landform_15_10cell	Narrow Valley 10-cell Radius
Landform_2_10cell	Neutral Peak/Ridge 10-cell Radius
Landform_3_10cell	Cool Peak/Ridge 10-cell Radius
Landform_5_10cell	Cliff 10-cell radius
Landform_6_10cell	Warm Upper Slope 10-cell Radius
Landform_7_10cell	Neutral Upper Slope 10-cell Radius
Landform_8_10cell	Cool Upper Slope 10-cell Radius
NED_2018_Esri_BeersAspect	Beer's Aspect*
NED_2018_Esri_Curvature	Slope Curvature
NED_2018_Esri_Elevation	Elevation
NED_2018_Esri_PlanCurvature	Plan Curvature**
NED_2018_Esri_Roughness_100r	Roughness 100-cell Circle
NED_2018_Esri_Roughness_10r	Roughness 10-cell Circle
NED_2018_Esri_Slope_Deg	Slope in Degrees
NED_2018_Esri_SolarRad_Equinox	Solar Radiation Equinox
NED_2018_Esri_SolarRad_SummerSol	Solar Radiation Summer Solstice
NED_2018_Esri_SolarRad_WinterSol	Solar Radiation Winter Solstice
NED_2018_Esri_TPI_100r_x1000	Topographic Position Index 100-cell Radius
NED_2018_Esri_TPI_10r_x1000	Topographic Position Index 10-cell Radius
NED_2018_Esri_TPI_3x3_x1000	Topographic Position Index 1-cell Square
NHD_2017_Esri_DistanceToStreamLine	Distance to Streams (Line)
NLCD_2011_ESRI_MeanConiferousForest	Mean Coniferous Forest Cover 100-cell Radius
NLCD_2011_ESRI_MeanUSFSCartoTreeCanopyPct	Mean Tree Canopy Cover 100-cell Radius
Polaris_Sep2018_Esri_Clay_x100	Soils Percent Clay
Polaris_Sep2018_Esri_N_x1000	Soils Pore Size Distribution (N)

* Beers Aspect: Beers et al. (1966) transformation of slope direction. Original scale is 0 (SW, most exposed) to 2 (NE, most sheltered), with values grading equivalently in both directions between the extremes.

** Plan Curvature: The curvature of a cell perpendicular to the direction of the maximum slope. Influences convergence and divergence of flow.

Results and Discussion

Model output and accuracy

We computed over 200 models to arrive at the optimal ensemble models for accuracy and meaningful species distributions. Model accuracies were evaluated in terms of Percent Correctly Classified (PCC) and Area Under the Curve (AUC) of each model was based on the 20% of the observations withheld from the modeling process for each species (Table 3). When AUC is 1.0 it means that model predicted suitable habitat 100% accurately with the withheld test observation samples. At 0.5, the model can no longer distinguish suitable from non-suitable habitat. Values between 0.7 and 0.8 are considered acceptable, 0.8-0.9 as excellent, and >0.9 as outstanding. Accordingly, all the models we developed are considered acceptable to excellent with respect to accuracy. Yet, how pseudo-negative sampling and thinning of known locations is done can affect the distributions. In choosing the final models for consideration, we sought to find a balance between overly constrained models, which yield little new information on distributions, and those that over-predicted habitat into areas well outside the known environmental range of a species.

Our most restricted models were for chokecherry and hop-tree, two species most commonly found in or near riparian areas, which is the most restricted habitat at BAND (Table 4). In contrast, Gambel's oak had the largest area distribution in the models, which was not surprising given that it is a common understory shrub in extensive ponderosa and mixed-conifer forests of the park, and often captures large areas after fires as a clonal persistent montane shrubland.

Table 3. Accuracy Measures: Percent Correctly Classified (PCC) and Area Under the Curve (AUC) for stacked Ensemble model for each species.

Species	PCC	AUC
Indian Ricegrass	0.70	0.70
Common Yarrow	0.87	0.85
Amaranthus species	0.83	0.78
Banana Yucca	0.78	0.82
Chokecherry	0.89	0.82
Common Hop-tree	0.91	0.82
Gambel's Oak	0.80	0.84
Alderleaf Mountain Mahogany	0.70	0.70
Douglas Fir - All Potential Habitat	0.82	0.80
Limber Pine - All Potential Habitat	0.89	0.89

Table 4. Predicted suitable habitat area for each species by the number of co-occurring models. Area with five models has the highest probability of suitable habitat. The total percentage of the total study area of 63,506 ac (% of area) for each model is provided. Ensemble model with values 2-5 are treated as potential suitable habitat and values 0-1 treated as non-habitat.

Species	Area (acres)					% of area
	2 Models	3 Models	4 Models	5 Models	Total	
Indian Ricegrass	3,039	2,768	1,582	1,952	9,341	14.7
Common Yarrow	3,435	3,212	2,051	1,878	10,576	16.6
Amaranthus species	4,025	2,941	1,087	766	8,819	13.9
Banana Yucca	3,236	2,692	2,399	2,523	10,850	17.1
Chokecherry	1,804	1,087	741	865	4,497	7.1
Common Hop-tree	2,176	1,035	699	784	4,694	7.4
Gambel's Oak	6,399	5,886	3,881	3,488	19,654	30.9
Mountain Mahogany	4,571	3,805	1,730	1,458	11,565	18.2
Douglas Fir	2,458	2,652	2,450	4,007	11,567	18.2
Douglas Fir - post fire	1,185	1,160	1,113	1,895	5,352	8.4
Limber Pine	3,242	2,438	2,485	4,582	12,748	20.0
Limber Pine - post fire	1,410	1,055	1,174	2,323	5,963	9.4

Model outcomes

Below we provide brief descriptions of each species followed by the ensemble model results for potential habitat suitability that include the mapped distribution and points used to create the model. Plant descriptions are derived from information provided by the [Lady Bird Johnson Wildflower Center](#) except for *Amaranthus* spp., which was courtesy of [Wikipedia](#). Fire effects, wildlife, and cultural uses are sourced from the [Native American Ethnobotany DB](#) and the [Fire Effects Information System \(FEIS\)](#).

Herbaceous plants

Indian rice-grass

Achnatherum hymenoides

Poaceae (Grass Family)

Synonym(s): *Eriocoma cuspidata*, *Eriocoma hymenoides*

Oryzopsis hymenoides, *Stipa hymenoides*

USDA Symbol: [ACHY](#)

Indian rice grass is a 1 to 2 ft. perennial cool-season bunchgrass, which can green up early in the spring depending on the precipitation. The sage-green, wiry foliage and ivory-colored seedheads give the grass an overall light, airy appearance (Figure 4). Foliage turns tan when dormant. It is very drought tolerant and typically grows in deep, sandy or rocky soils in the lower elevations of the park.

The seeds were ground for flour, mixed with water and cooked into a mush by the Montana Indians, Navajo, Hopi, Apache, Paiute, and others.

The leaves are highly palatable for many animals and the seeds are sought by birds and small mammals.

Model observation points were distributed in the lower elevation sector of the park and with most to eastern side (Figure 5). Based on all BAND observations and the regional distribution, we masked out predicted habitat above 7,750 feet (2,362 m). While the modeled distribution favored known locations, it also suggests that suitable habitat extends to the southwestern mesas, and canyon slopes and bottoms of the park as well further up in elevation into some of the recent fire zones. At these elevations, fire impacts are considered minimal to moderate depending on severity, and the species is known to reestablish after fire through seed dispersal and possibly by tillers (see [USFS Fire Effects species webpage](#)).



Figure 4. Indian Rice-grass (Sally and Andy Wasowski, Lady Bird Johnson Wildflower Center).

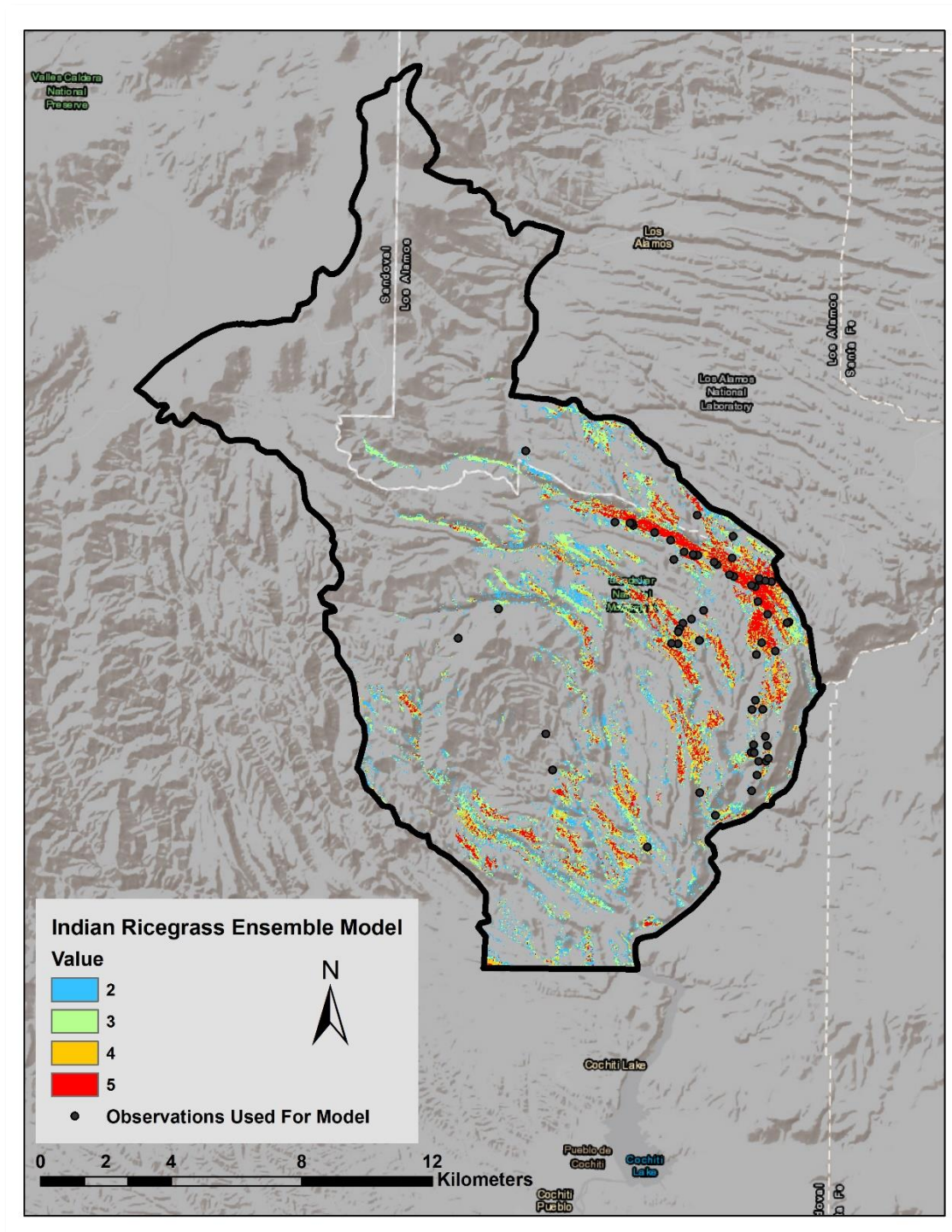


Figure 5. Indian rice grass potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 models.

Common yarrow*Achillea millefolium* var. *occidentalis*

Asteraceae (Aster Family)

USDA Symbol: [ACMI2](#)

Yarrow is a perennial herb that grows to 3 ft. and has no branches except near the top. The leaves are alternate, 3 to 5 in. long, with many leaflets on each side of the midrib (1- pinnately lobed); and these are further divided into smaller leaflets, giving them a delicate, fernlike, lacy appearance (Figure 6). Flower heads are arranged in large, compact clusters at the top of the stem, each cluster consisting of one or more flower heads. The flower head has 20 to 25 yellowish-white (rarely pink) ray flowers and similarly colored disc flowers. Flowering mid-July to mid-September. Fruit is a cypsela (pl. cypselae).



Figure 6. Common yarrow (Stephanie Brundage, Lady Bird Johnson Wildflower Center).

Yarrow has been used for medicinal purposes: to break a fever by increasing perspiration, to treat hemorrhaging and as a poultice for rashes. The plant has a distinctive fragrance and a tea has been used by Native Americans to relieve ear-, tooth-, and headaches; as an eyewash; to reduce swelling; and as a tonic or stimulant. Yarrow varies in taste and in potency depending on where it grows and at what stage of growth it is in. The best time to collect yarrow for tea is right before the flowers are produced, using only the new succulent leaves.

It is generally unpalatable to wildlife, but occasionally the flowers are consumed.

This is a common plant of mid- to high-elevation mixed conifer and ponderosa pine forests, montane grasslands and meadows, and montane shrublands from 5,900 to 10,000 ft. in the park. It occurs on mountain slopes extending down to the upper mesa tops and canyons following this elevation gradient (Figure 7). The model generally favors known locations, covering much of the upper elevations of the park along the Sierra de los Valles except northeast of the park across Los Alamos County. This may be due to differences in how soils were mapped between county surveys from which soils variables are derived. The probability of potential habitat declines with lower elevations, but is known to occur in relatively dry locations. The entire predicted habitat distribution lies within recent fire boundaries, and while severe fire can kill any plant, yarrow commonly increases following moderate fire sprouting from rhizomes seed germination (see [USFS Fire Effects species webpage](#)).

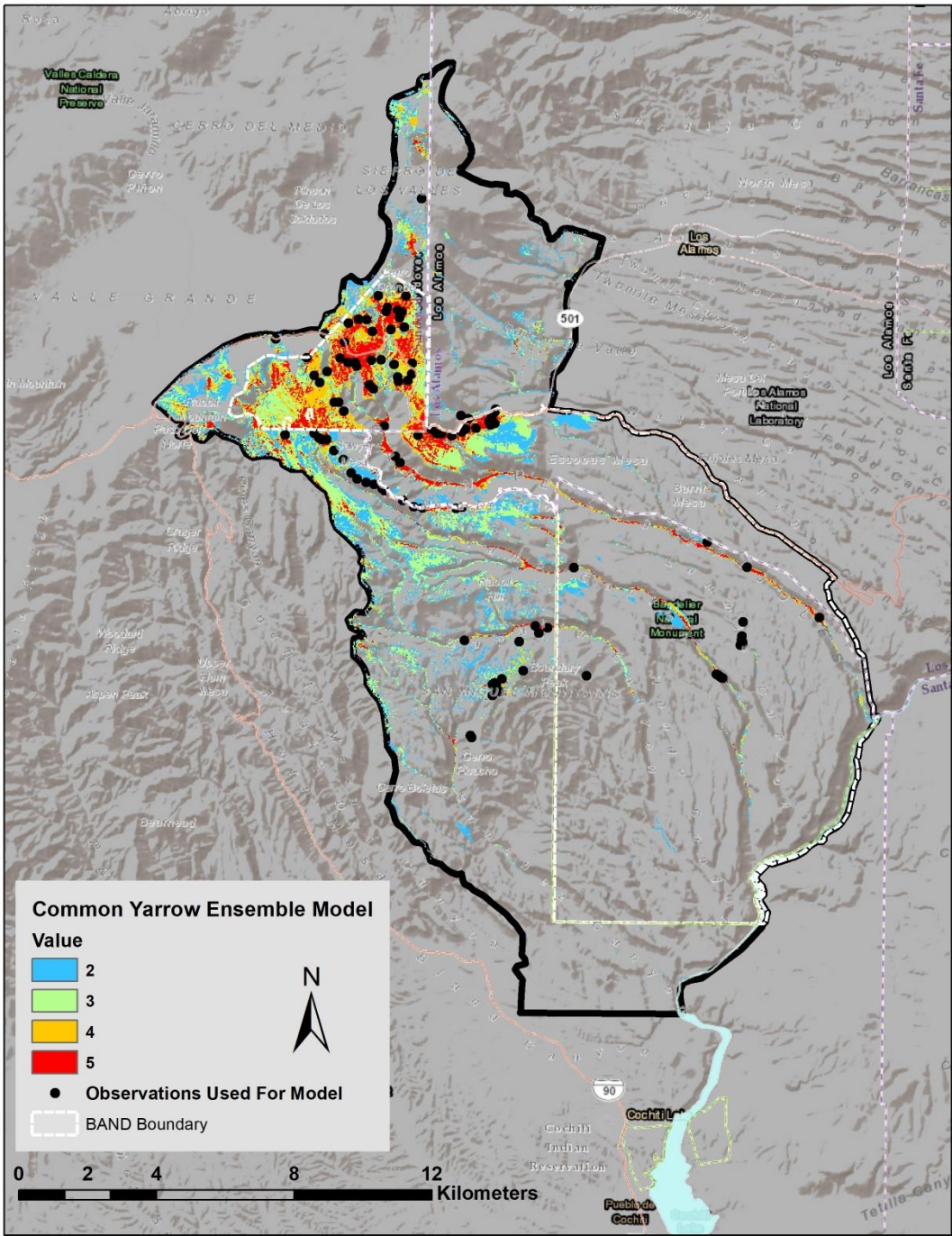


Figure 7. Common yarrow potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.

Amaranth

Amaranthus spp. Possible BAND species: *Amaranthus hybridus* (N), *Amaranthus blitoides* (I), *Amaranthus albus* (I), *Amaranthus wrightii* (N), and *Amaranthus retroflexus* (N).

Family: Chenopodiaceae (Goosefoot)

USDA Symbol: [AMARA](#)

Amaranthus is a cosmopolitan genus of annual or short-lived perennial plants collectively known as amaranths that on BAND can be either native or introduced (N or I above). Catkin-like cymes of densely packed flowers grow in summer or fall (Figure 8). Amaranth varies in flower, leaf, and stem color with a range of striking pigments from the spectrum of maroon to crimson and can grow three to eight feet tall with a cylindrical, succulent, fibrous stem that is hollow with grooves and bracteoles when mature.

Amaranth grain is collected and ground into flour, and the leaves eaten as greens.

The grain can also be used by wildlife, particularly birds.

This is a common, ruderal (weedy) plant of mid-elevations that is associated with ground disturbance and fire. Based on the plot data, on BAND it ranges from 6,100 to 8,280 ft in elevation with the average at 6,700 ft; on gentle (4%) to moderate (18%) slopes and predominantly warm aspects. The model expands the known plot-based distribution modestly, yet given its ruderal nature and wide distribution, suitable habitat may extend further, particularly after disturbance events such as fire (Figure 9). While fire will kill this annual plant, it can readily reappear from the seedbed the following season (see [USFS Fire Effects species webpage](#)).

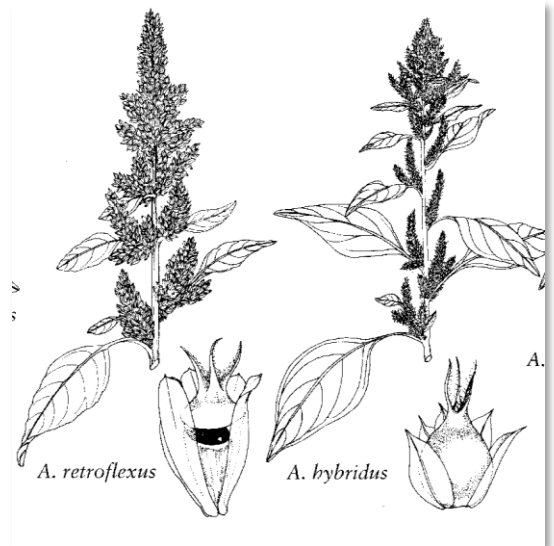


Figure 8. Illustration of two of potential amaranths on BAND

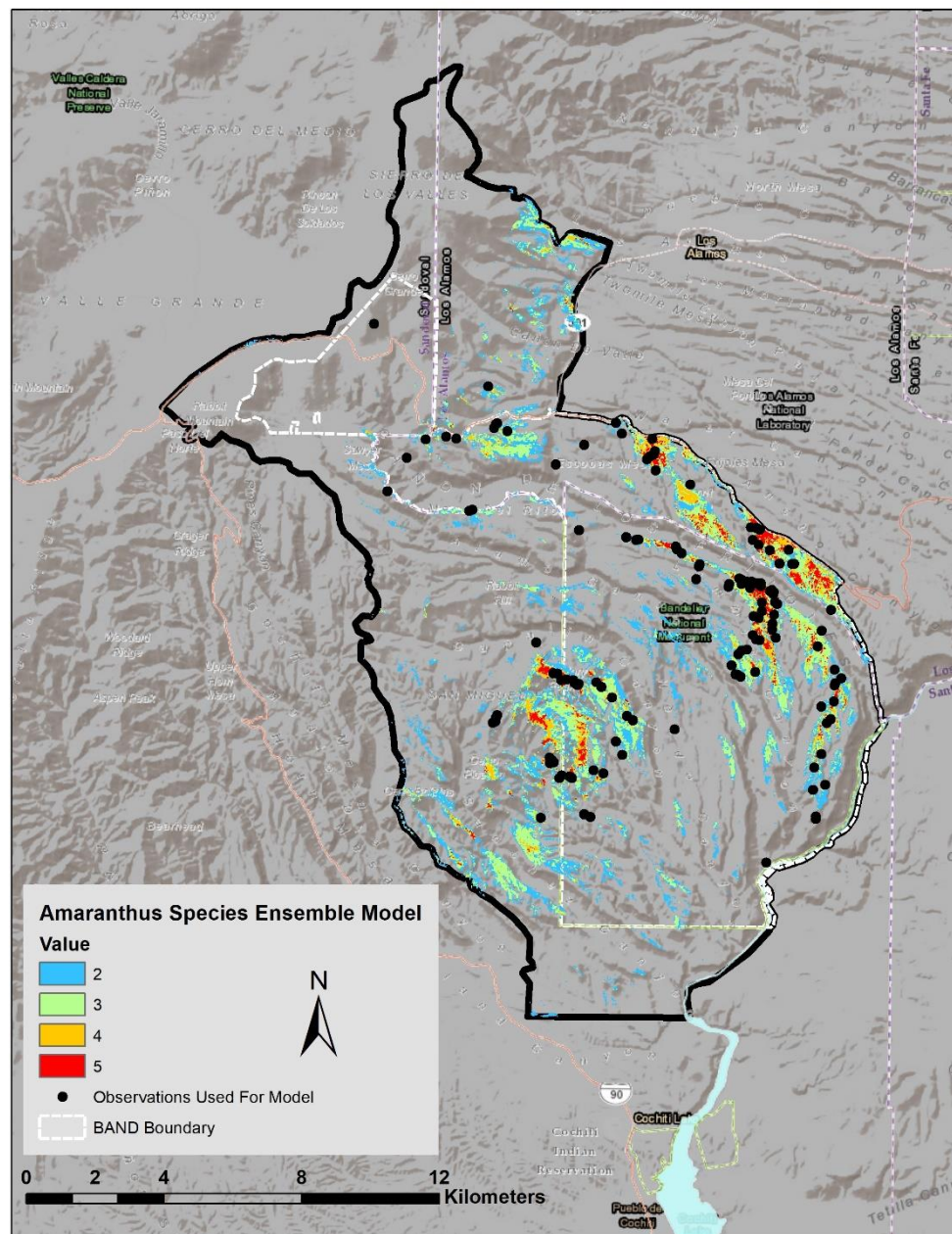


Figure 9. Amaranth species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.

Shrubs

Banana Yucca

Yucca baccata

Agavaceae (Agave Aster Family)

USDA Symbol: [YUBA](#)

Banana yucca usually occurs as a single, stemless shrub with succulent leathery leaf blades in a basal rosette. The plant sometimes grows in clumps with short, reclining stems (Figure 10). The narrow, spine-tipped leaves are up to 30 in. long and occur in an open cluster, which is often wider than the leaves are high. The flowering stem is up to 40 in. tall and bears large, pendant, fleshy, white flowers with a red-purple tinge. Fruits are fleshy and banana-shaped.



Figure 10. Banana yucca (Norman G. Flaigg, Lady Bird Johnson Wildflower Center).

It was used as a shampoo, as an antiemetic, for heartburn, as a cathartic, in childbirth, ceremonially, and as a drink. The fruits were roasted and eaten, as well as eaten raw, boiled, or baked, pulverized and dried in cakes; its use as a food is extensive throughout the region. It was also used for making fibers, rope, mats, sandals, and baskets; used also as an ingredient in dyes.

Banana yucca flower stalks are highly digestible, and are an important source of phosphorus and forage for deer and elk, and non-game birds. By consuming the fleshy outer portions of the fruit, rabbits and woodrats expose the seeds for dispersal. Deer often use banana yucca as a food source, searching out new leaves that sprout after fire and eating from the plant at a higher frequency than when it is not burned (see [USFS Fire Effects webpage](#)).

Banana yuccas occur at mid to low elevations 5,400 to 8,200 ft. on moderate to steep slopes and predominantly warm aspects. It is common associate in ponderosa and pinyon-juniper woodlands, Gambel's oak and wavyleaf oak/alderleaf mountain mahogany montane scrub, and occasionally into mesa top grasslands. The model expands beyond the known plot-based distribution primarily to mesa tops in the southwest portion of the study area, but the upper elevation limit may be too conservative (Figure 11). The majority of the distribution lies beyond the recent fire perimeters, but the species can resprout after fire from rhizomes and basal stem buds below the surface (see [USFS Fire Effects webpage](#)).

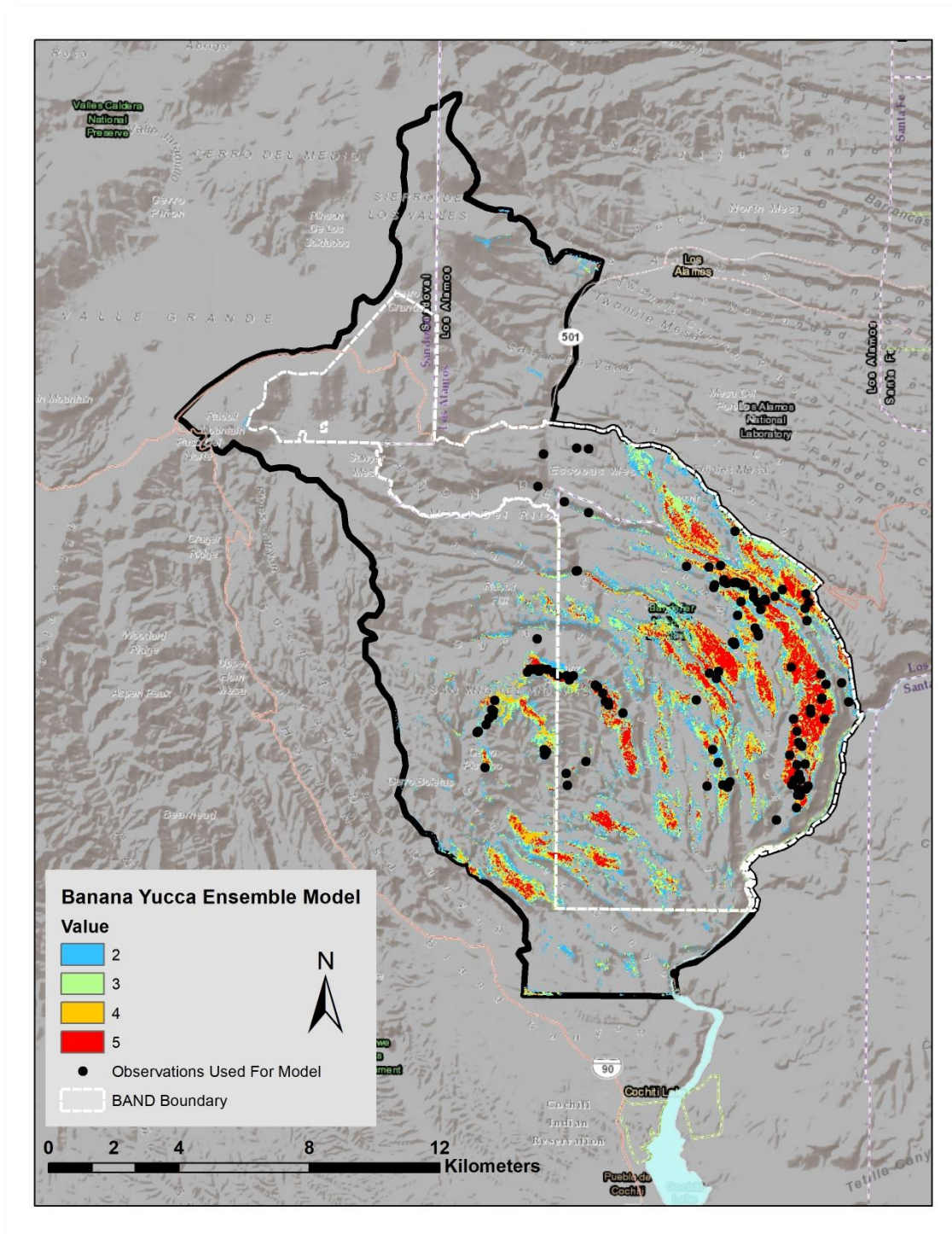


Figure 11. Banana Yucca species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.

Black Chokecherry*Prunus virginiana* var. *melanocarpa*

Rosaceae (Rose Family)

USDA Symbol: [PRVI](#)

Black chokecherry is a common species of the interior western mountain West. It is a large, deciduous shrub or small understory tree, that can grow to 20 to 30 ft. and often forms thickets. Dense clusters of white flowers are followed by red fruit ripening to dark purple from early to mid-summer (Figure 12).

As the common name suggests, chokecherries are astringent or puckery, especially when immature or raw, but are still often made into preserves and jelly. Chokecherry bark and roots have been used to make sedatives, blood-fortifying tonics, appetite stimulants, and medicinal teas.

Chokecherry is widely regarded as an important wildlife food plant and provides habitat. Fruits, leaves, and twigs are utilized. Large mammals including bears, moose, coyotes, bighorn sheep, pronghorn, elk, and deer use chokecherry as browse; it is also a food source for small mammals. The fruits are important food for many birds. Yet, new young growth, wilted leaves, or plant parts that are injured by frost or drought can be poisonous.

On BAND, chokecherry grows primarily in the riparian zone extending from mid to high elevations (5,900 to 9,000 ft.) and mostly in riparian areas but it can extend upslope into neighboring forests and montane shrublands. For the model, we clipped out the Rio Grande corridor through White Rock Canyon as unsuitable low-elevation habitat (Figure 13). In contrast, the model may be overly conservative in the upper elevation canyons and slopes. In addition, many of the canyons with habitat were burned in recent fires, but chokecherry is well adapted to disturbance by fire. Although susceptible to top-kill by fire, it resprouts from surviving root crowns and rhizomes and can proliferate (see [USFS Fire Effects species webpage](#)).



Figure 12. Black chokecherry (Courtesy Sally and Andy Wasowski, Lady Bird Johnson Wildflower Center)

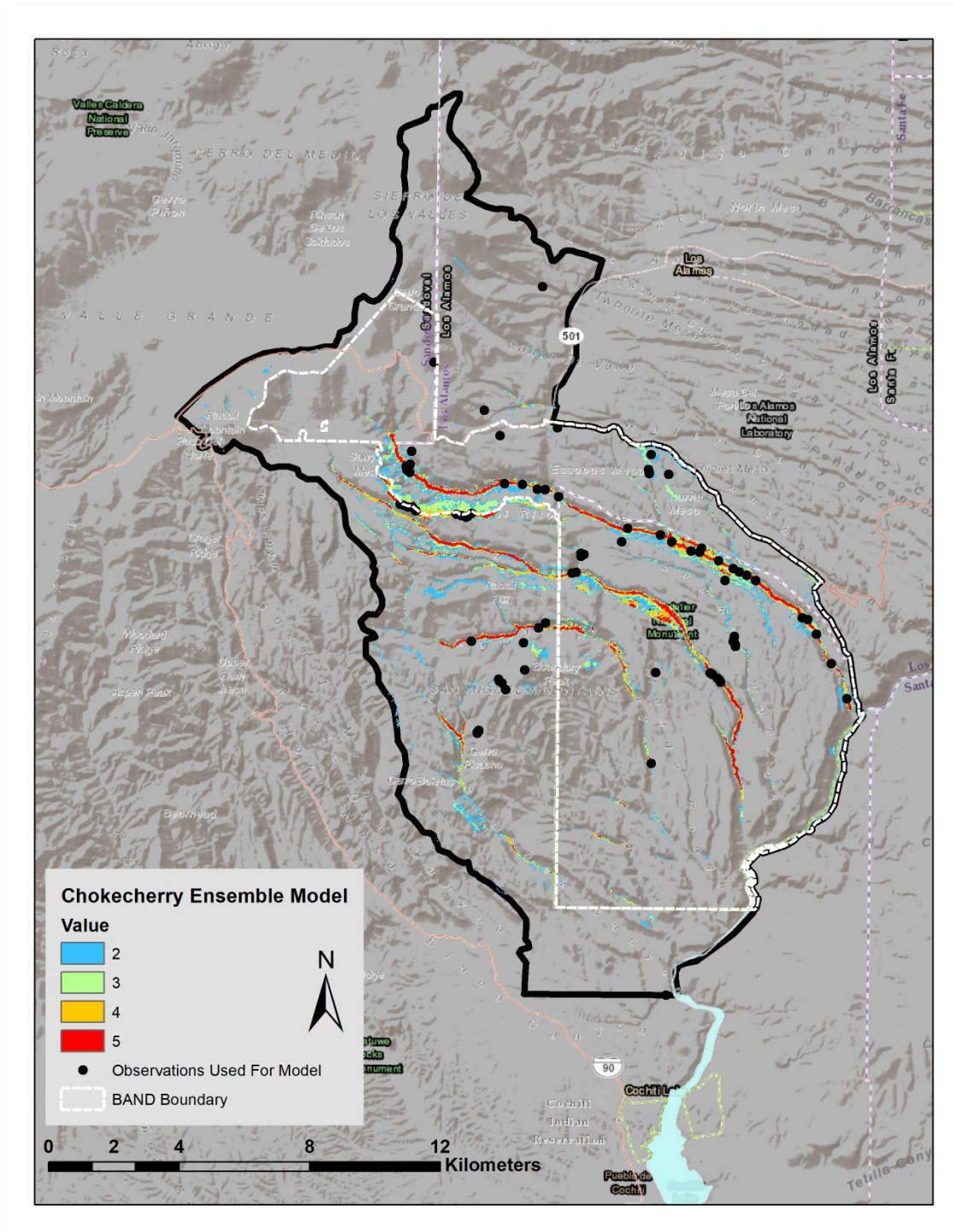


Figure 13. Black chokecherry species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.

Common hop-tree*Ptelea trifoliata*Rutaceae (Rue Family)USDA Symbol: [PTTR](#)

Common hop-tree is an aromatic shrub or small tree with a rounded crown. The trunk is slender and crooked, bearing interwoven, ascending branches. Bark, crushed foliage, and twigs have a slightly lemon-like, unpleasant (to some) musky odor. Trifoliate, deciduous leaves with leaflets on a petiole up to 2 in. long, the terminal leaflet up to 2.5 in. long, obovate, tapering more gradually to the base than to the tip, midrib of lateral leaflets off-center (Figure 14). Leaves are dark-green in summer, turning yellow in fall. Flowers small, greenish white, in clusters among the leaves, appearing in April. Fruit distinctive, wafer-like samara with broad wings, approximately 7/8 in. long by 3/4 in. wide.

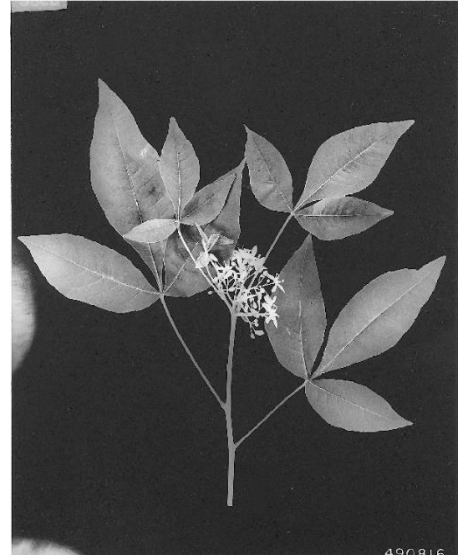


Figure 14. Common hop-tree (USDA Plants)

The common name refers to a reported use in earlier days of the bitter fruit as a substitute for hops in brewing beer. The bitter bark of the root, like other aromatic barks, has been used for home remedies.

Butterflies, bees, and other insects are known to make use of the plant and flowers, and the shrubs provide cover wildlife.

Common hop-tree is found in riparian zones and lower slopes of canyons on BAND from 5,500 to 8,500 ft. The model generally follows this environmental setting with potential habitat ascribed to most of the drainages with perennial or intermittent flows in the study area up to the mid elevations (Figure 15). While portions of the drainages were subject to recent fires, the impact of fire on this species has not been determined.

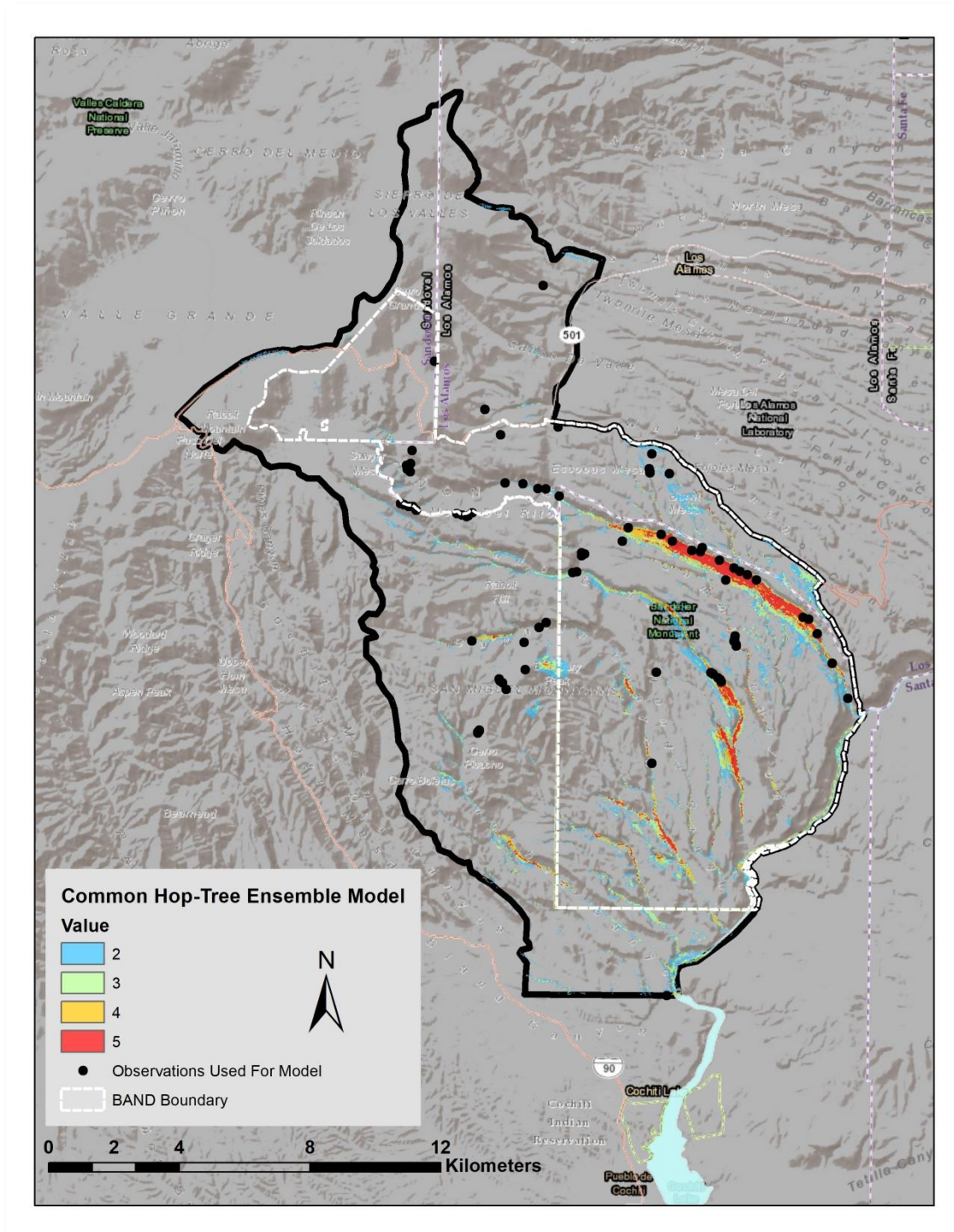


Figure 15. Common hop-tree species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.

Gambel's oak*Quercus gambelii*Fagaceae (Beech Family)USDA Symbol: [QUGA](#)

Gambel's oak is a small, round-crowned tree or clump shrub, sometimes forming copses of trees as high as 30 ft. Its deeply-lobed, deciduous leaves are bright green above and paler below, turning brown or sometimes red in fall (Figure 16).

Acorns eaten whole and raw or parched, ground into flour or used to make mush. The wood is used in house construction, fence posts, and fuel.

Gambel's oak is an ecologically important species providing food and shelter for many wildlife species. Shrublands are browsed by big game and provide habitat for rabbits and rodents. Young pole stands provide sites for foliage-nesting birds. Mature growth forms provide maximum acorn yield for squirrels, birds, elk, and deer. Old stands containing large amounts of dead crown and hollow boles or limbs provide nesting sites for small mammals and birds.

Gambel's oak is common montane species of the southern Rocky Mountains and on BAND it is found on mountain slopes at 9,500 ft. extending down to 5,500 ft. in canyon riparian zones. In mixed-conifer or ponderosa pine forests is commonly a sub-canopy tree or shrub. But after severe fires, it resprouts as a clonal shrub and can form extensive stands of montane shrubland, that often re-burn (see [USFS Fire Effects species webpage](#)). The model is a moderate expansion on known habitat, particularly in the more remote areas of the study area (Figure 17).



Figure 16. Gambel's Oak leaves and acorn (Courtesy James L. Reveal, Lady Bird Johnson Wildflower Center).

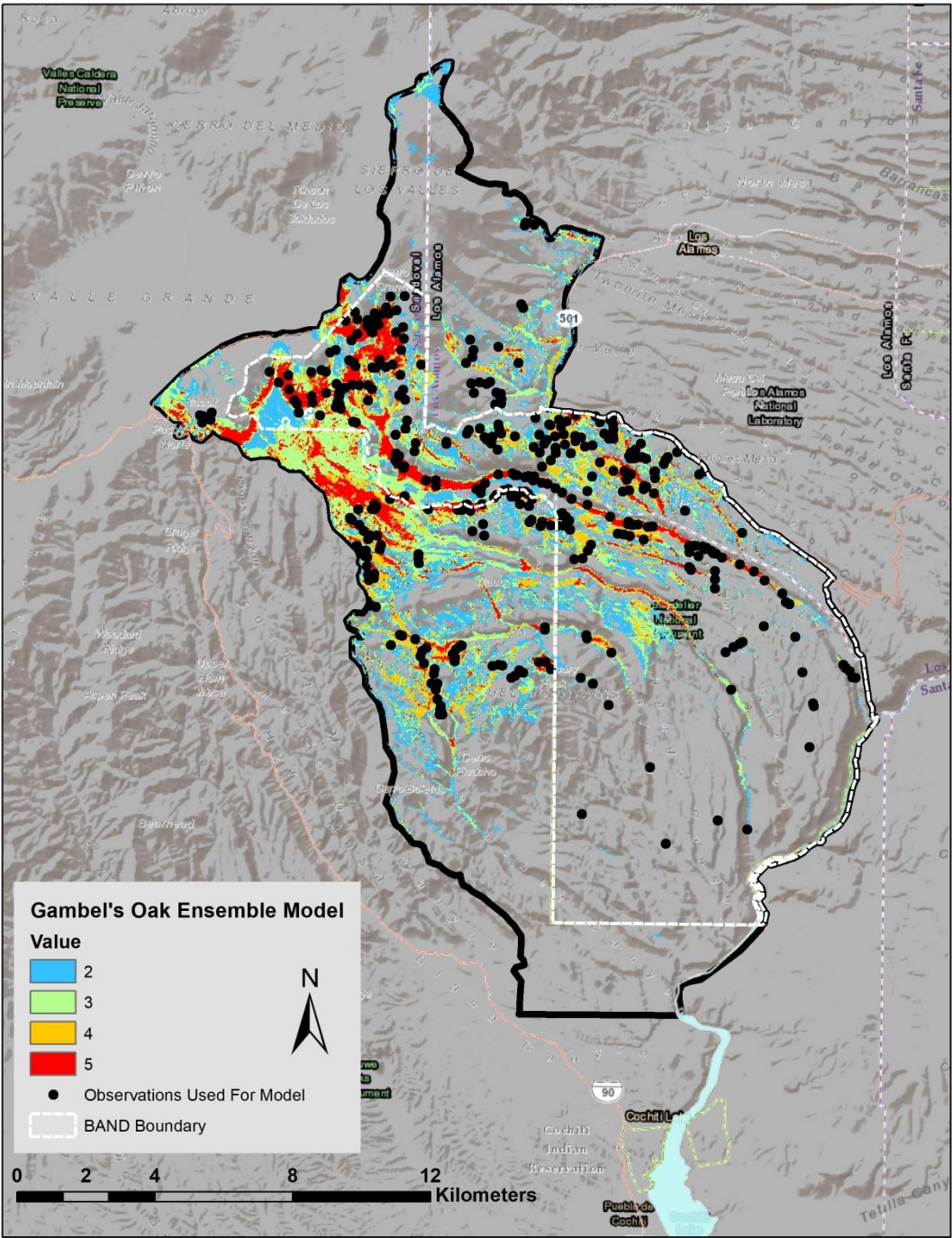


Figure 177. Gambel's oak species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.

Alderleaf Mountain Mahogany*Cercocarpus montanus*

Rosaceae (Rose Family)

USDA Symbol: [CEMO2](#)

Mountain Mahogany is an 8 to 20 ft. shrub or small tree with very attractive, almost-evergreen leaves - dark green on top and fuzzy silver underneath. Stems measure 0.2 to 2 in. in diameter and are sturdy and stiff (Figure 18). The main stem is typically less than 30 in. in diameter. Non-showy flowers are followed by feathery, silvery-white fruits, occurring from May to November. The common and scientific names both refer to the resemblance of the leaves to those of shrubby birches. The hairy fruits are carried long distances by the wind; animals harbor them in their fur, also aiding dispersal. After falling to the ground, the oddly shaped fruits twist into the soil.

The wood hard and durable and is used to make implements, and the roots can make a red dye.

Alderleaf mountain mahogany is considered highly valuable forage for wildlife, particularly deer and elk, and it provides cover and habitat to many game and song birds.

It is a common shrub in ponderosa pine forests and pinyon-juniper woodlands as well as forming extensive stands of montane shrubland (chaparral) vegetation along with wavyleaf oak (*Quercus pauciloba*) after fire. Its known elevation on BAND is between 5,450 and 8,000 ft. and is found on rocky hillsides, cliffs, mesas and drier slopes. Given this elevation range, we masked out predicted habitat above 8,000 ft. Accordingly, the model reflects the lower elevation, drier conditions where the species is found (Figure 19). The majority of distribution lies with recent fire boundaries, but while alderleaf mountain-mahogany is often top-killed by fire, it quickly recolonizes burned sites through root crown or rhizome sprouts (see [USFS Fire Effects species webpage](#)).



Figure 18. Alderleaf mountain mahogany (Courtesy Stephanie Brundage, Lady Bird Johnson Wildflower Center).

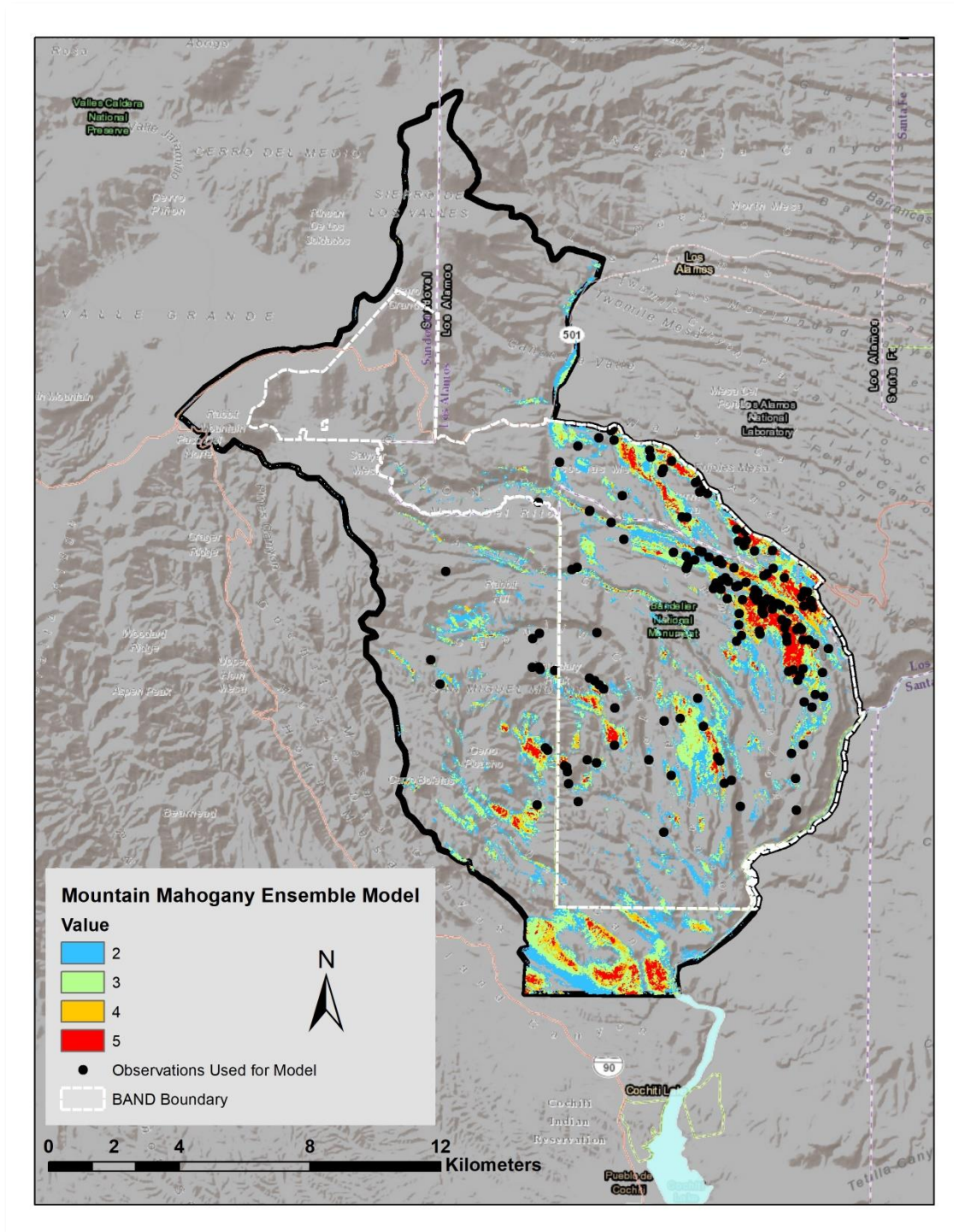


Figure 19. Alderleaf mountain mahogany species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.

Conifer Trees

Douglas fir

Pseudotsuga menziesii var. *glauca*

Pinaceae (Pine Family)

USDA Symbol: [PSME](#)

This evergreen tree can reach 150 ft. or more in New Mexico. The Rocky Mountain Douglas-fir (var. *glauca*), of the Rocky Mountain region, is a medium-sized to large tree with flattened, blue-green needles and cones 4-7 cm long with bracts bent upward (Figure 20).

One of the world's most important timber species, its branches are important in Native American ceremonies, and the leaves and sap are used for medicinal purposes.

Douglas-fire forest provide cover and foliage is consumed by deer and elk. Numerous songbirds and small mammals extract seeds from Douglas-fir cones or forage for seeds on the ground, and birds use tree cavities for nesting (see [USFS Fire Effects species webpage](#)).

On BAND, Douglas-fir ranges from 5,900 to 9,990 ft in elevation as is the dominant or co-dominant species in mixed-conifer forests along with white fir (*Abies concolor*), ponderosa pine (*Pinus ponderosa*), and limber pine (*Pinus flexilis*). Habitats range from dry-mesic upper slopes on warm aspects to moist-mesic lower slopes and canyon bottoms along streams. The model potential habitat model reflects this distribution (Figure 21), but when the moderate to high-severity of recent fire is considered, coverage is reduced by 54% (Figure 22). Following large, stand-replacing fires, Rocky Mountain Douglas-fir seedling establishment can be slow. Seedlings are restricted to the burn edge or near surviving trees within the main burn (see [USFS Fire Effects species webpage](#)).



Figure 20. Rocky Mountain Douglas-fir cones (Dave Powell, USDA Forest Service (retired), Bugwood.org).

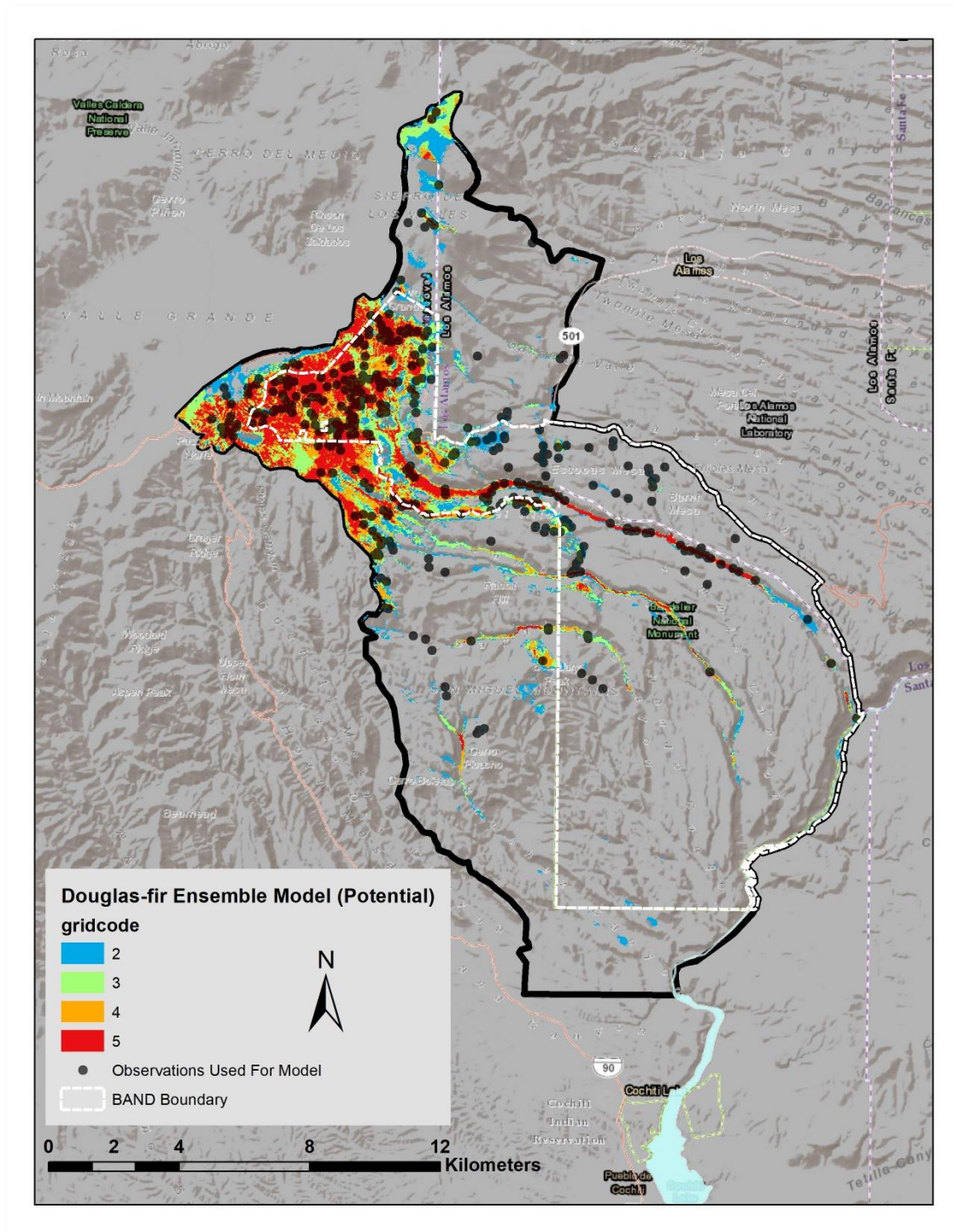


Figure 21. Douglas-fir species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.

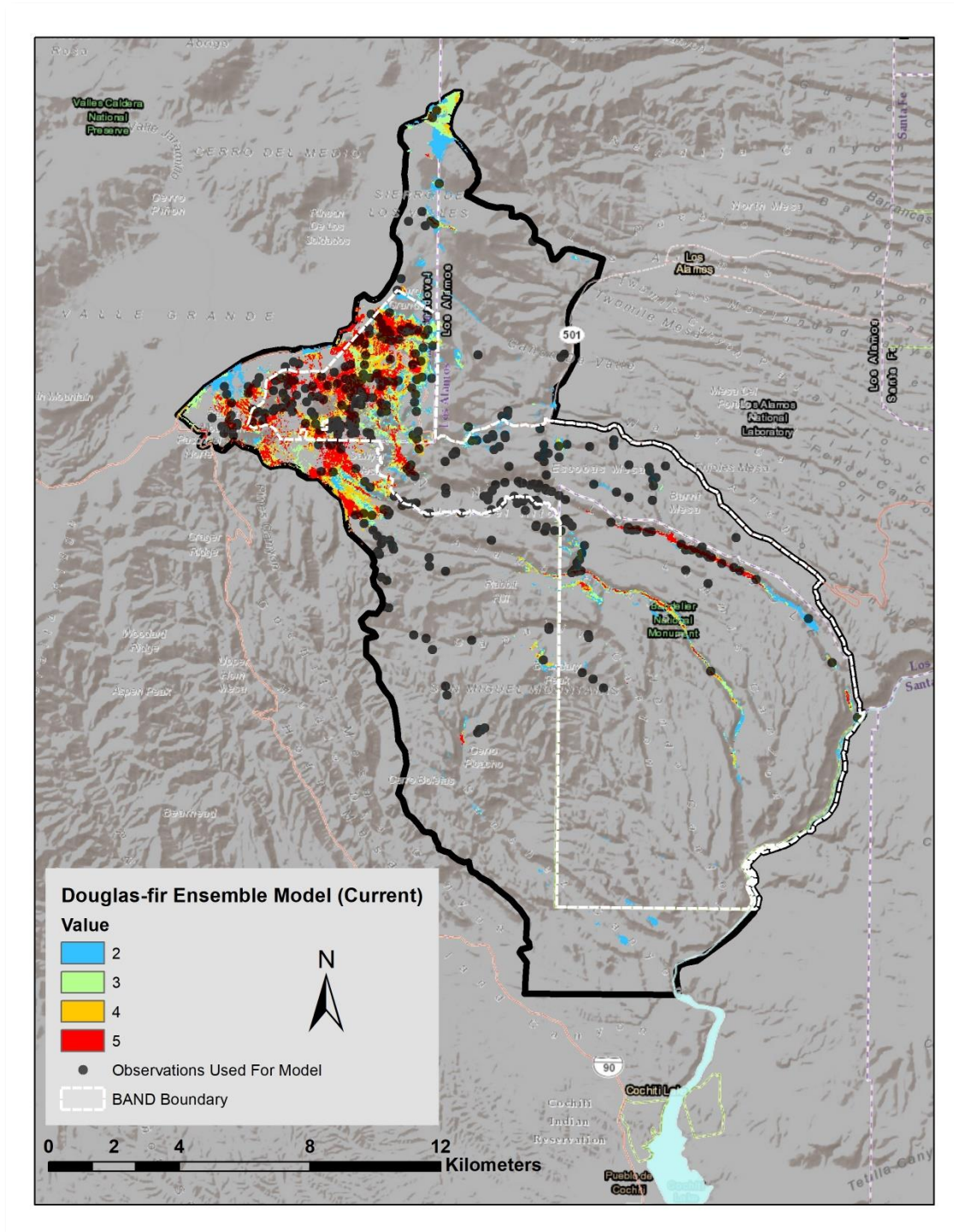


Figure 22. Douglas-fir species potential suitable habitat ensemble model restricted to unburned or low-severity burn locations. Value reflects the co-occurrence of 2 to 5 model.

Limber pine

Pinus flexilis

Pinaceae (Pine Family)

USDA Symbol: [PIFL2](#)

This is a 30 - 60 ft. (sometimes taller) pine with a short trunk and dark blue-green needles in clusters of five. Irregular branching and tapering limbs create a broad, flat-topped tree at maturity (Figure 23). It is pyramidal in youth. Cones are long. The names refer to the very tough and flexible twigs, which can sometimes be twisted into a knot.

Its seeds are roasted and hulled or sometimes the seeds are ground, shell and all and eaten.

Birds and mammals consume the large seeds and cache the seeds for later use.

On BAND, limber pine ranges from mid to high elevation (7,350 to 9,800 ft) and is a common associate in mixed-conifer and ponderosa pine forests, but seldom the dominant (except on the dry margins or in montane grasslands). It is most common on dry-mesic upper slopes and ridges, but it can occur in more mesic conditions. The potential-habitat model generally reflects this distribution (Figure 24), but when the moderate to high-severity of recent fire is considered, coverage is reduced by 53% (Figure 25). Limber pine is often killed by fire because of its relatively thin bark. Limber pine regeneration on burns is largely from germinants of Clark's nutcrackers seed caches (see [USFS Fire Effects species webpage](#)).



Figure 23. Limber pine needs and cones (James reveal, Lady Bird Johnson Wildflower Center).

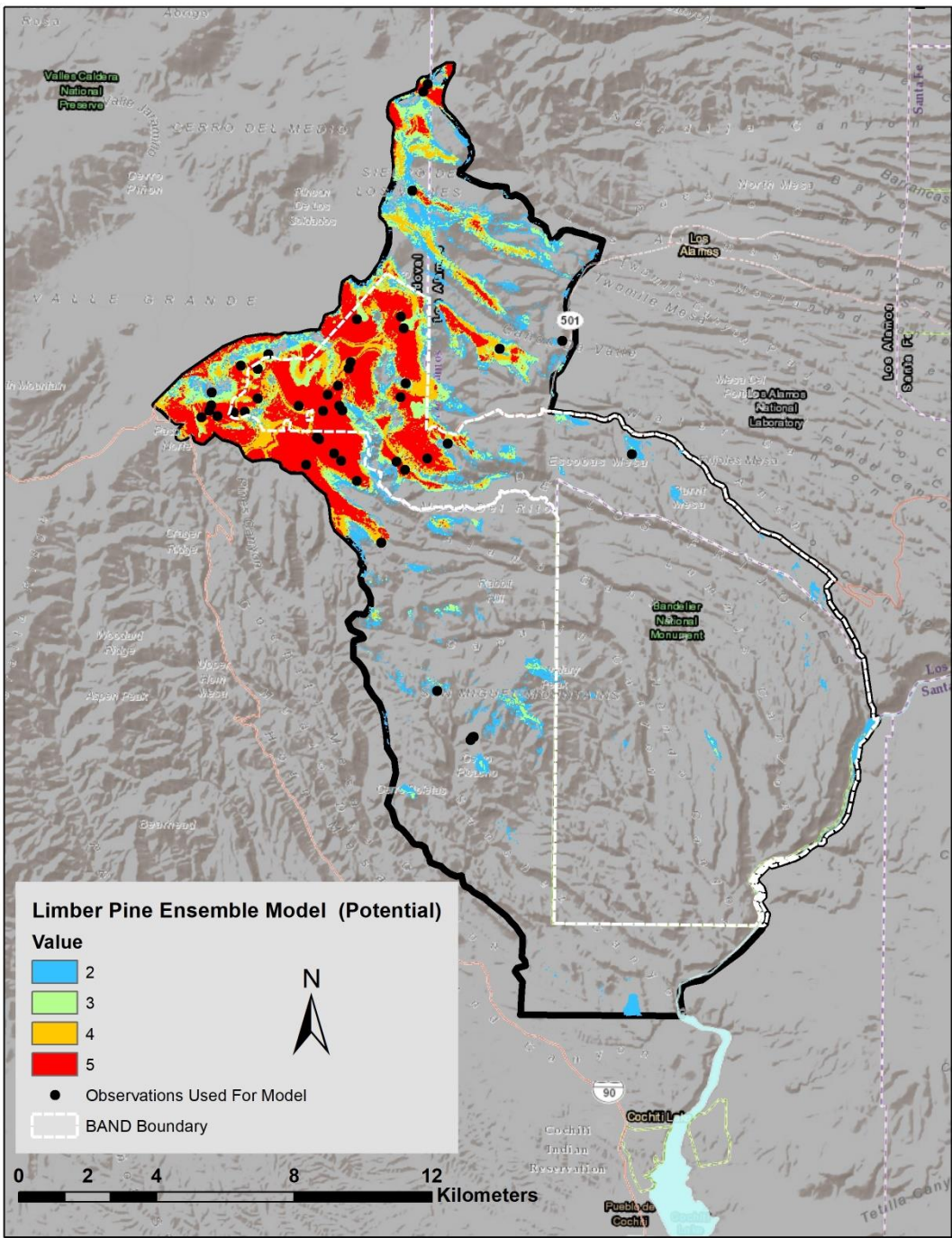


Figure 24. Limber pine species potential suitable habitat ensemble model. Value reflects the co-occurrence of 2 to 5 model.

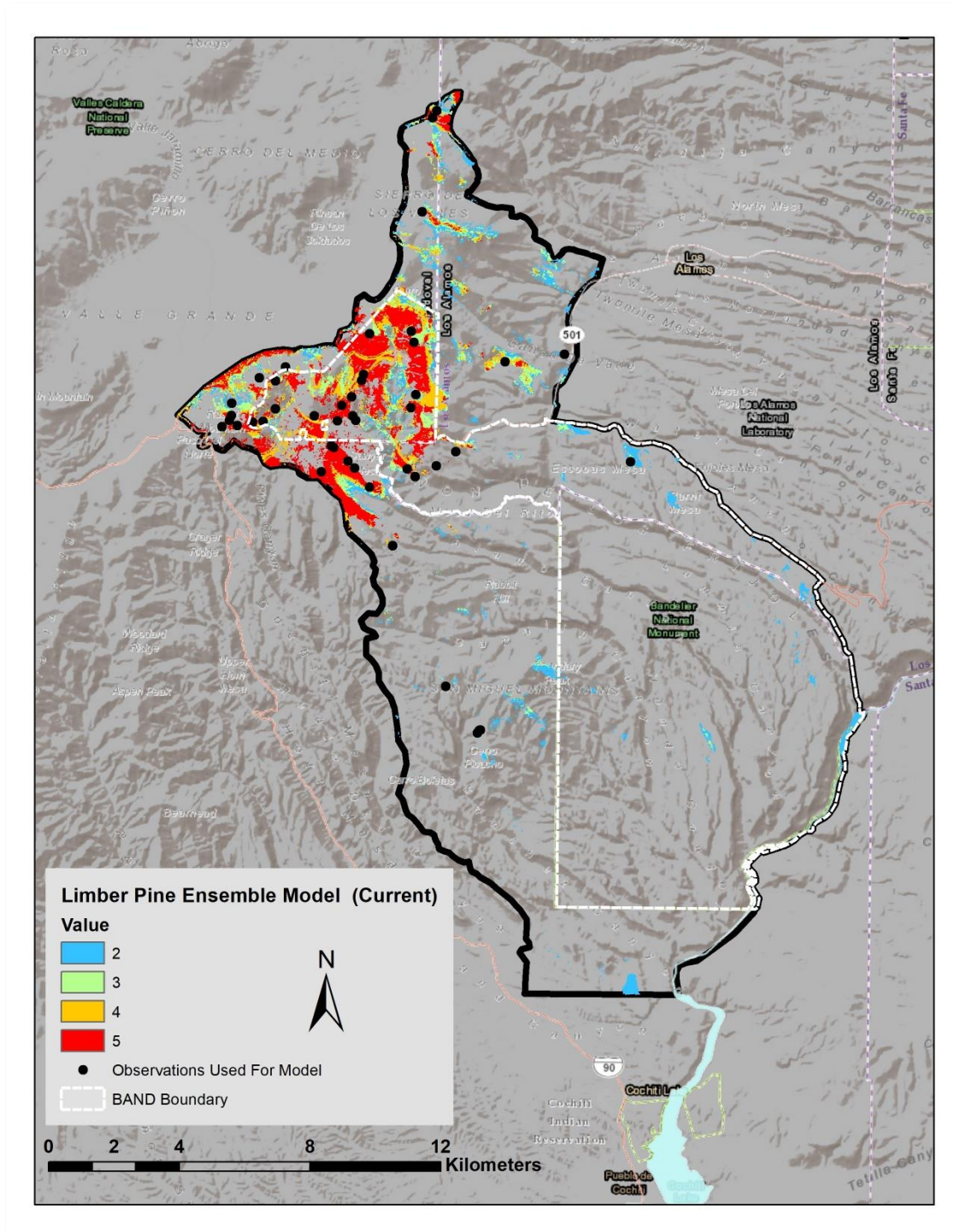


Figure 25. Limber pine species potential suitable habitat ensemble model restricted to unburned or low-severity burn locations. Value reflects the co-occurrence of 2 to 5 model

Conclusions

The intent of the models presented here is to provide first-order guidance for the sustainability of plants of cultural interest with respect to collection and their overall management within the park and surrounding areas. Given that models are only as good as their input variables, the models we present are relatively conservative to account for that variability in the input data, but species habitats may extend somewhat beyond the bounds presented here. Accordingly, the species distributions may need to be field verified for site-specific activities depending on a given project goals.

These models represent the distribution of habitats based on data that spans about a 30-year period—a time when the park's landscape has undergone major changes because of severe fire. We have attempted to account for this through understanding individual species responses to fire and explicit mapping of fire severity for those species that were most severely impacted (conifer trees). Beyond this, we suggest that the species be surveyed and monitored directly to evaluate what their actual responses to the fires have been. Lastly, in the face of climate change, the next step for modeling would be to take them into the future to guide their use and management for the long term.

References

- Allen, C. D., A. K. Macalady, H. Chenchouni, D. Bachelet, N. McDowell, M. Vennetier, T. Kitzberger, A. Rigling, D. D. Breshears, E. H. Hogg, P. Gonzalez, R. Fensham, Z. Zhang, J. H. Lim, J. Castro, N. Demidova, G. Allard, S. W. Running, A. Semerci, and N. Cobb. 2010. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management* 259:660–684.
- Breshears D.D., N.S. Cobb, P.M. Rich PM, K.P Price, C.D. Allen CD, R.G. Balice, W.H. Romme WH, J.H. Kastens, M.L. Floyd, J. Belnap, J.J. Anderson, O.B. Myers, C.W. Meyer CW. 2005. Regional vegetation die-off in response to global-change-type drought. *Proceedings of the National Academy of Sciences USA* 102: 15144–15148.
- Goff, F., B.S. Kues, M.A. Rogers, M.A., L.D. McFadden, and J.N. Gardner, eds. 1996. The Jemez Mountain Region. New Mexico Geological Society 47th Annual Field Conference Guidebook. New Mexico Geological Society, Inc.
- Jacobs, B. F, R.G. Gatewood, and C.D. Allen. 2002. Watershed restoration in degraded piñon-juniper woodlands: a paired watershed study 1996-98(9). Final report to: USGS-BRD Research / NPS-Natural Resource Preservation Program.
- Jacobs, B.F. 2015. Restoration of degraded transitional (piñon–juniper) woodland sites improves ecohydrologic condition and primes understory resilience to subsequent disturbance. *Ecohydrology* (wileyonlinelibrary.com) DOI: 10.1002/eco.1591
- McDowell, N., C. D. Allen, and L. Marshall. 2010. Growth, carbon isotope discrimination, and climate-induced mortality across a *Pinus ponderosa* elevation transect. *Global Change Biology* 16:399–415.
- Muldavin, E., A. Kennedy, C. Jackson, P. Neville, T. Neville, K. Schulz, and M. Reid. 2011. Vegetation classification and map: Bandelier National Monument. Natural Resource Technical Report NPS/SCPN/NRTR— 2011/438. National Park Service, Fort Collins, Colorado.

Appendix 1 Modeling Details

Environmental Predictor Data

For all 10 HSMs, the environmental predictor data consisted of 38 high-resolution (30 m) layers created by [Natureserve's Map of Biodiversity Importance \(MoBI\)](#) (see Table 2). All 38 MoBI predictor layers were clipped to the predefined modeling domain boundary and combined into a raster stack in the R package "raster" (Hijmans and van Etten 2016) for efficient processing.

Pseudo-Absence (Background) Points

Spatial modeling of species requires both presence and absence data. For most species, we lack actual absence data and must generate "pseudo-absence" points across more or less randomly across the modeling domain. For each species, we employed two techniques to generate pseudo-absence points "presence thinning" and "thickening" outlined by Vollerling et al. (2019). The "presence thinning" method uses random background point placement and reduces the clustering of actual presence points by thinning records by a certain distance (in this case, 90 to 200 m), this minimizes overfitting risk and sampling bias while preventing absence points from being placed too close to presences, ensuring environmental dissimilarity. We used the "randomPoints" function from the R package "dismo" for this purpose.

For the "thickening" method, background points representing pseudo-absence are strategically added by mimicking presence point density and selecting ecologically similar locations, which addresses sampling bias and enhancing balance. To implement this, used the "sample_background" function in the R package "flexsdm" with the method set to "thickening." For both methods, presence points were buffered by 90 m, and pseudo-absence points were sampled beyond buffered zones but within the modeling domain. This ensured that points would not be placed in the same 30 x 30-m environmental cell space of the predictor variables.

Both methods were executed at least twice, incorporating varying numbers of pseudo-absences (two, four, and sometimes three times the number of presence points for each species), resulting in at least four different ensemble models. Subsequently, we evaluated all resulting ensemble models to determine which one best represented each species based on their individual ecologies and abiotic requirements.

Modeling Algorithms

For each species, we modeled the distribution of potentially suitable habitat using five different statistical and machine learning algorithms: randomForest (RF), Generalized Linear Models (GLMs), Extreme Gradient Boosting (XGB), Multivariate Adaptive Regression Splines (MARS) and Bayesian Additive Regression Trees (BART).

The RF algorithm, developed by Breiman in 2001, is a tree-based method for classification and regression. In this approach, individual trees are trained on subsets of input data, which include both dependent variable and predictor variables. The classification is determined based on the

features each tree was trained on. The complete ensemble of trees in the model collectively contributes to the final classification through a "voting" mechanism. The predicted class for the Random Forests model is the one that receives the most votes across all trees. We utilized the "randomForest" R package developed by Liaw and Weiner in 2002, set the number of classification trees grown (ntree) to 500 and the number of variables available for splitting at each node (mtry) to the square root of the total number of predictor variables.

We utilized the R function "glm" for Generalized Linear Models (GLM), to fit a logistic regression models to each species. The dependent variable was the training point's identity (presence or pseudo-absence), and selected variables from the glm-based selection (outlined above) were used as predictors, without incorporating interaction effects.

XGB, is similar to RF in the sense that it is a tree-based classification technique and trains a forest of classification trees using a subset of the training data and of the predictors. However, unlike RF, XGB attempts to enhance tree information by learning from mistakes in earlier tree fitting attempts. The XGB model was fit using the "xgbTree" approach in the "caret" R package (Kuhn 2008). Model parameters used: n rounds (number of trees or iterations) to 100 (default); max_depth (maximum tree depth) to 10, 15, 20, and 25; colsample_bytree (percentage of features used for building each tree) between 0.5 and 0.9; eta (step size shrinkage) to 0.1; gamma (regularization tuning) to 0 (default), and minimum_child_weight (minimum leaf weight) to 1 (default). For more information on setting model parameters see [XGB parameters documentation](#).

MARS constructs an effective, piecewise linear model, offering a versatile alternative to polynomial regression for capturing non-linear relationships (Friedman 1991). This method identifies cutpoints or knots where the linear relationship between dependent and independent variables changes. Our implementation of this algorithm utilized the "earth" method in the "caret" R package (Kuhn 2008). We set the degree to 1 (default, indicating the maximum degree of interactions for building an additive model) and nprune, representing the maximum number of terms in a pruned model, to 40. In other words, nprune serves as an upper limit on the model size.

BART is a Bayesian nonparametric method that is suitable for several different outcomes, in this case, continuous. It is a complex machine learning approach, described as a black-box model, in which the relationships between inputs and outputs are not explicitly clear. BART works by employing an arbitrary random function, f , for outcome fitting based on covariates, and its regularization tree-branching penalty prevents overfitting, ensuring a generally good fit (Sparapani et al., 2021). We implemented this method using the "bart" function to from the R package "embarcadero" (Carlson 2020).

GLM-Based Variable Selection

To identify a set of essential, minimally correlated variables for each species, we employed a variable selection technique based on generalized linear models (GLM) that involved fitting a preliminary, basic logistic regression model for each species. In this analysis, each environmental predictor was treated as the independent variable and presence (1 = presence, 0 = pseudo-

absence) was the dependent variable. We assessed model fit using Akaike Information Criterion (AIC) values, where lower AIC values indicate greater variable importance. After ranking models and their constituent variables based on AIC, we proceeded to systematically evaluate pairwise Spearman's correlation coefficients between each variable – starting with the variable possessing the lowest AIC value and subsequently comparing it to all other variables. If the correlation coefficient between two variables was >0.7 , we discarded the variable with the higher AIC value. Subsequently, we continued the correlation coefficient evaluation and variable removal process with the variable possessing the second-best AIC value. This process continued iteratively until we had evaluated all variables in the list. This approach allowed us to preserve the variables that were deemed most informative based on their AIC values, while also enabling us to create a list of variables with limited correlation to one another.

Model Fitting and Ensemble Model Construction

Once we selected the top environmental variables associated with each species, we fit all five modeling algorithms to the occurrence and environmental data for each algorithm. This method involved training each model on a dataset consisting of 80% of presence and absence (i.e., “pseudo-absence”) records, with the remaining 20% reserved for model accuracy assessment. Because the initial output of the selected algorithms is a continuous raster surface ranging from 0 to 1, a threshold must be used to convert these continuous surfaces into binary maps that display suitable vs. unsuitable environmental conditions. We used the MaxKappa value as the optimal threshold for all models, with the goal of optimizing both sensitivity and specificity. Raster cells above the MaxKappa threshold were classified as suitable (1) and cells below or equal to the threshold were classified as unsuitable (0). The ensemble was crafted by stacking binary models generated by all five algorithms. Pixels assigned a value of 5 denoted a unanimous agreement among all models regarding environmental suitability, whereas pixels with a value of 0 were unanimously considered unsuitable. We designated habitat suitability for all models of all species within the range of values 2-5, with values 4 and 5 carrying the highest likelihood of occurrence for each specific species.

Variable Importance

We assessed variable importance for each modeling algorithm in the species-specific final models. For GLM, MARS, XGB, and RF, we used the functions “vi” and “vip” in the R package “vip” to calculate the top contributing variables for each model. For the BART algorithm we used the function “varimp” in the R package “embarcadero” to calculate and create a variable importance plot. Each algorithm's variable importance is represented by a distinct index specific to that particular modeling algorithm. The variable importance plots for each algorithm, corresponding to each species, are available upon request.

Accuracy Assessment

To gauge the collective accuracy of the ensemble model for each species, we computed the average of continuous outputs across all five model algorithms. We then applied a threshold using the MaxKappa value, resulting in a binary ensemble model where a value of 1 denotes suitable habitat and 0 designates non-suitable habitat. We used the R package “PresenceAbsence” (Freeman and Moisen 2008) to calculate the following measures of accuracy for the thresholded ensemble model: Percent Correctly Classified (PCC), Area Under the Receiver Operating Characteristic (ROC) Curve (AUC). PCC, assesses overall classification accuracy, measuring the percentage of cases correctly classified out of the total cases, reflecting the model's ability to accurately identify both positive and negative instances. Generally, a higher PCC value signifies better accuracy, with values nearing 100% indicating high precision, while values close to 50% suggest performance similar to random chance. The ROC curve displays model sensitivity (true positive rate) against 1-specificity (false positive rate). The AUC is used as an overall measure of model performance that is not sensitive to model thresholds (Manel et al. 2001). Strong models have high AUC values, with values > 0.8 indicating high accuracy (Swets 1988).

References

- Beers, T. W, P. E. Dress, and L. C. Wensel. 1966. Aspect transformation in site productivity research. *Journal of Forestry* 64(10):691-692.
- Breiman, L. 2001. Random Forests. *Machine Learning* 45:5-32.
- Carlson, C. J. (2020). embarcadero: Species distribution modelling with Bayesian additive regression trees in r. *Methods in Ecology and Evolution*, 11(7), 850–858.
<https://doi.org/10.1111/2041-210X.13389>
- Freeman, E. A., and G. Moisen. 2008. PresenceAbsence: An R Package for Presence-Absence Model Analysis. *Journal of Statistical Software* 23(11):1-31.
- Friedman, J. H. 1991. Multivariate Adaptive Regression Splines. *The Annals of Statistics* 19(1):1-141. DOI: 10.1214/aos/1176347963
- Hijmans, R. J., and J. van Etten. 2016. Raster: Geographic analysis and modeling with raster data. R package version 2.5-8 <https://CRAN.R-project.org/package=raster>.
- Kuhn, M. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* 28(5):1-26. doi:<http://dx.doi.org/10.18637/jss.v028.i05>
- Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. *R News* 2:18-22.
- Manel, S., H. C. Williams, and S. J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38:921-931.
- Sparapani, R., Spanbauer, C., & McCulloch, R. (2021). Nonparametric Machine Learning and Efficient Computation with Bayesian Additive Regression Trees: The BART R Package. *Journal of Statistical Software*, 97(1), 1–66. <https://doi.org/10.18637/jss.v097.i01>
- Stoffle, R.W, and N. O'Meara, R. Toupal, M. Buttram, J. Dumbauld (2007). A Study of Natural Resource Use among Culturally Affiliated Pueblo Communities. Report on file Bandelier National Monument, Los Alamos, NM.
- Swets, J.A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240:1285-1293.
- Vollering, J., R. Halvorsen, I. Auestad, and K. Rydgren. 2019. Bunching up the background betters bias in Habitat Suitability Models. *Ecography* 42(10):1717-1727.
<https://doi.org/10.1111/ecog.04503>