ARTICLE

Modeling spatial distributions of Amah Mutsun priority cultural plants to support Indigenous cultural revitalization

Annalise Taylor[1] [iD] | Alexii Sigona[1] | Maggi Kelly[1,2]

[1]Department of Environmental Sciences, Policy, and Management, University of California Berkeley, Berkeley, California, USA

[2]Informatics and GIS Statewide Program, University of California Division of Agriculture and Natural Resources, Davis, California, USA

Correspondence
Annalise Taylor
Email: annalisetaylor@berkeley.edu

Abstract

Along the Central Coast of California, USA, native plant biodiversity has depended on various forms of Indigenous stewardship such as burning, tilling, and gathering. Simultaneously, the Amah Mutsun Tribal Band (the Tribe) depends on these native ecosystems for cultural survivance. However, much of the knowledge related to the location and caretaking of cultural plants has become dormant in the community due to the immediate and ongoing effects of Euro-American colonization. We identified potential gathering areas by modeling the spatial distributions of 10 culturally important plants throughout the Tribe's stewardship area. We utilized community science datasets with an ensemble modeling approach that combines the results of five machine learning models to predict not only the distribution of each species, but also the relative certainty of those predictions spatially. Our results revealed that 265.2 km^2 (2.1%) of the Tribe's stewardship area is predicted habitat for seven or more of these cultural plants, and that the Tribe has potential access to approximately a third of these high-priority areas. Our findings will directly inform the Tribe's cultural revitalization and ecological stewardship programs. We show how geospatial models can support the revitalization of an Indigenous culture by renewing relationships with cultural plants.

KEYWORDS
cultural plants, cultural revitalization, ensemble modeling, Indigenous communities, machine learning models, species distribution modeling, target group background selection method

INTRODUCTION

Indigenous peoples throughout California, USA, have been shaping the region's ecosystems for millennia, stewarding landscapes through practices including burning, tilling, gathering, and planting (Anderson, 2013; Cuthrell et al., 2016). These reciprocal relationships between people and ecosystems are crucial for both the health and cultural sovereignty of Indigenous communities and the many ecosystems that depend on human stewardship (Baumflek et al., 2015; Kimmerer, 2011; Lake et al., 2017; Lopez, 2013). This "mutual caretaking between people and place" (Diver et al., 2019) has been restricted by both the immediate effects of Euro-American colonization (i.e., displacement and genocide) and its ongoing legacies (i.e., proprietization of land and systemic oppression of

Indigenous Peoples) (Sanchez et al., 2021). Despite these challenges, Indigenous communities in California are working to restore relationships with their homelands. In tandem, ecologists are increasingly recognizing the importance of active restoration and stewardship of native ecosystems, creating new opportunities for collaboration between tribes and land managers (Lightfoot et al., 2021).

For many people within the Amah Mutsun Tribal Band (hereafter AMTB or the Tribe), ecological restoration of their stewardship area along California's Central Coast (Figure 1) is a key goal of cultural revitalization efforts and is seen as a spiritual and moral obligation (Lopez, 2013). Past research with and by the Tribe has highlighted the importance of Amah Mutsun stewardship for maintaining healthy populations of native plants, particularly those dependent on disturbance (Anderson, 2013; Cuthrell, 2013). Amah Mutsun foodways, ceremonies, and medicines depend on relationships with the diverse plant and animal species found in native ecosystems (Cuthrell, 2013; Lopez, 2013). As a nonfederally recognized tribe, AMTB is not guaranteed property rights within their traditional homelands. Therefore, the Tribe is increasingly partnering with researchers, land-owning agencies, and conservation organizations to create opportunities for Amah Mutsun people to steward, gather, and restore their plant relatives as a means of healing both plants and people (Lightfoot et al., 2021). However, reinstating Amah Mutsun stewardship in many of these parks depends not only on formal access agreements, but also on the revitalization of dormant ethnobotanical knowledge
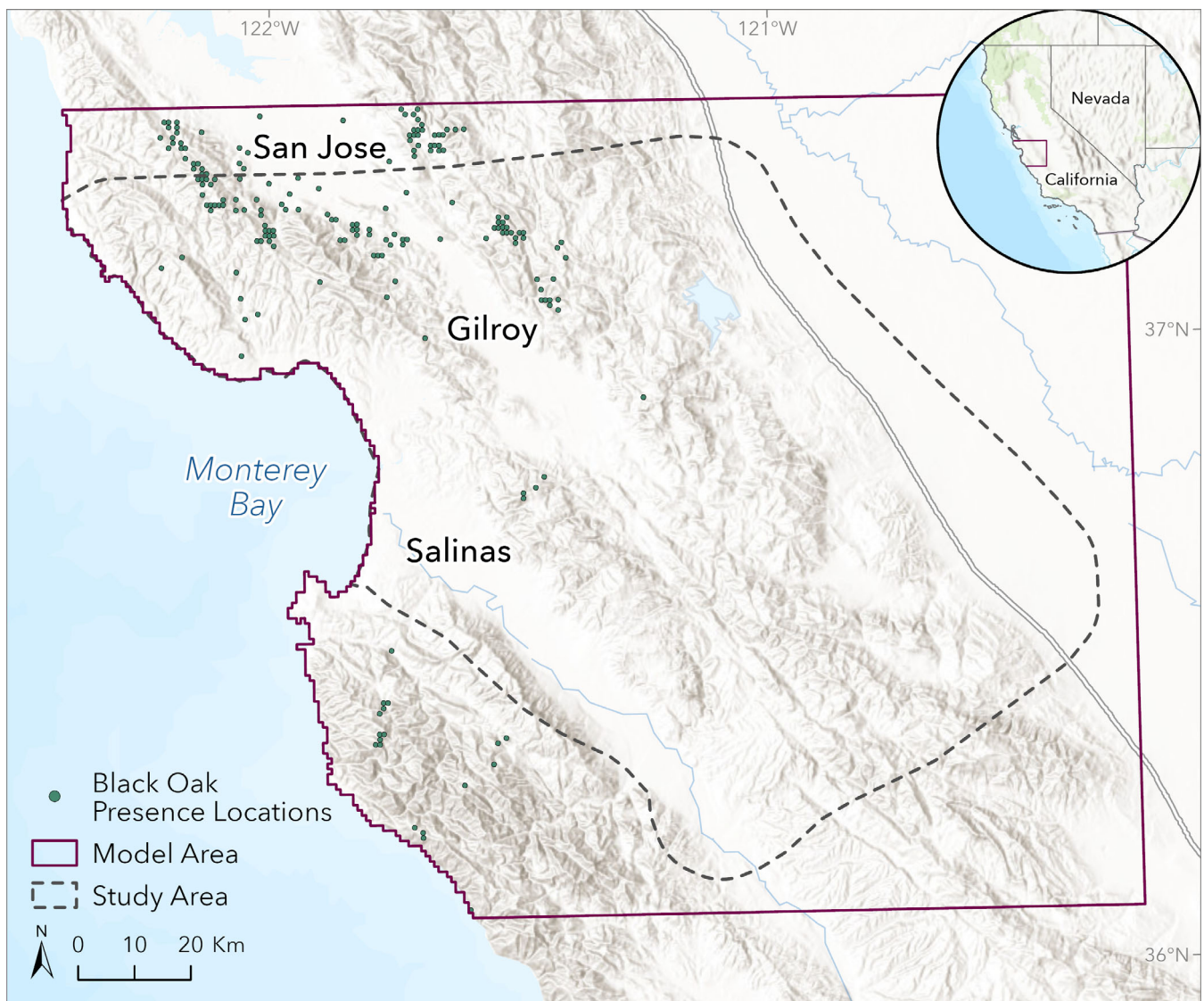

FIGURE 1 Model area, study area, and spatially filtered presence locations of California black oak (*Quercus kelloggii*) used with the model runs. The model area represents the area over which the models were run. Our final results were restricted to the study area, that is, the boundaries of the Amah Mutsun stewardship area. The inset map in the upper right shows the location of the model area within California, USA.

related to the uses, stewardship, and location of cultural plants within the community. Baumflek et al. (2015) assert that access to gathering areas enables both retention and intergenerational transmission of knowledge about plant use and ecology. Therefore, a critical next step is to build the Tribe's knowledge of the locations of cultural plants—defined here as plants that are used for food, medicine, ceremonies, basketry, and other materials—within their stewardship area. This region's particular history of colonization has meant that precise definitions of tribal political boundaries are often contentious and difficult to determine; our study therefore takes place within the Tribe's stewardship area, defined as the lands and waters that they are working to restore and steward.

Species distribution models (SDMs) are frequently used to predict the actual or potential locations of a species (Guisan & Zimmermann, 2000). These models use a variety of statistical approaches to build a relationship between environmental or climatic variables and known presence locations of that species (Elith et al., 2006; Li & Wang, 2013). Previous work has used a single SDM to map the habitat of one or two ethnobotanical species (Baumflek et al., 2015; Gorman et al., 2008). By expanding both the number of models used and the number of species mapped, SDMs can create a more accurate and comprehensive picture of areas that are likely to contain multiple ethnobotanical species. Known locations of the target species are an integral input to an SDM; while a field survey of multiple species is not commonly feasible, large community science databases such as iNaturalist now enable modeling of multiple species across large areas using methods that account for their spatial biases (Di Cecco et al., 2021). Additionally, although any SDM has certain limitations and biases, an ensemble modeling approach minimizes the biases of any one model (Eisen et al., 2018).

While geospatial tools and data are useful, they have been used to exploit, extract, and reduce Indigenous ways of knowing (Baumflek et al., 2015; Brown & Kyttä, 2018; Reid & Sieber, 2020). Therefore, this research began with two years of discussions with the AMTB Tribal Council, the tribally held Amah Mutsun Land Trust, and associated researchers about our mutual research goals. Lead author Taylor is a non-native scientist trained in the ecological and geospatial sciences; coauthor Sigona is an interdisciplinary social scientist and an Amah Mutsun tribal member. Taylor and Sigona conducted 12 in-depth interviews with tribal elders and cultural practitioners regarding their relationships to land, culture, and the environment. Interviewees were identified in partnership with AMTB leadership and included tribal members with experience stewarding lands for cultural purposes. These semistructured interviews and our participation in tribal events helped us to build

relationships with a broader group of Amah Mutsun community members. This study was designed from the priorities expressed in those interviews—principally the restoration of ethnobotanical knowledge and reconnection with specific basketry and food plants—and represents one piece of our ongoing collaboration.

This study aims to support the Tribe's larger goal of restoring relationships between the Tribe and culturally important plants. Currently, the Tribe has potential access to more than 1000 km² of land within their stewardship area, with varying opportunities for gathering or stewardship of cultural plants. A complete field survey of these lands is not monetarily or physically feasible; therefore, we developed an ensemble distribution modeling method to identify and prioritize potential gathering areas with the ultimate goal of restoring ethnobotanical knowledge. Our three objectives were as follows:

1. Model the distribution of 10 cultural plants within the Amah Mutsun stewardship area;
2. identify areas where multiple cultural plants are likely to be found; and
3. evaluate which of these possible gathering locations are most accessible to the Tribe.

Our results will directly inform the Tribe's restoration and gathering programs and guide recommendations for agencies in their stewardship area. We recognize that the Amah Mutsun community relates to these plants as relatives. Due to the sensitive nature of these culturally important species, some of our spatial results are not public and are visible only to members of the Amah Mutsun community. With the permission of the Amah Mutsun Tribal Council, we have included our complete results for one of the 10 priority plants: California black oak (*Quercus kelloggii*). California black oaks are generally found in foothills or lower elevation mountains and their acorns are a preferred source for Mutsun acorn foods. We have also shared spatial results for the wavy-leafed soap plant (*Chlorogalum pomeridianum*), which can be prepared as food or used as a soap. For the eight other cultural species and the final results regarding potential gathering areas, we have included summarized results that do not indicate spatial locations.

METHODS

Model area and study area

All models (described below) were run within the maximum rectangular extent of the Amah Mutsun stewardship

area (the model area) and our final results were restricted to the actual boundaries of the stewardship area (the study area) (Figure 1). The study area includes regions of San Mateo, Santa Clara, Santa Cruz, San Benito, and Monterey counties within the Central Coast of California. This region is characterized by a Mediterranean climate with cool wet winters and warm dry summers and is subject to frequent periods of drought. Inland areas have greater temperature variations throughout the year (hotter summer temperatures and colder winter temperatures) than coastal areas.

Observation data

Species observation data included a combination of three confidential datasets and one public dataset. AMTB collected presence locations for 20 culturally significant species on preserves managed by the Midpeninsula Regional Open Space District located within San Mateo and Santa Clara counties from 2014 to 2019. We also incorporated plant observation data from the University of California Santa Cruz Younger Lagoon Reserve plant restoration team collected from 2014 to 2021 within Santa Cruz County, and from archeological surveys conducted in collaboration with the Tribe from 2014 to 2020 (Apodaca & Lightfoot, 2020; Younger Lagoon Reserve, 2021).

We combined these three confidential datasets with research-grade species observation data from the iNaturalist API using the pyinaturalist Python package for 15 of the Tribe's cultural priority species within the model area (iNaturalist Users, 2021). We used the coarsest resolution of our predictor variables (30 arc seconds, or approximately 740 × 920 m) to spatially filter the observation data using the Point to Raster and Raster to Point tools in ArcGIS Pro, which excluded duplicate points if they fell within the same pixel footprint (ESRI, 2021). We then selected the 10 species with at least 100 observations. These 10 plants are *Artemisia douglasiana*, *Calandrinia menziesii*, *C. pomeridianum*, *Clinopodium douglasii*, *Corylus cornuta* ssp. *californica*, *Q. kelloggii*, *Rubus parviflorus*, *Rubus ursinus*, *Sambucus nigra* ssp. *caerulea*, and *Vaccinium ovatum*. The ArcPy Python package was used to prepare each species' observation dataset for input into the models (ESRI, 2021).

Background points

We used presence-only observation data that were collected opportunistically in some cases and systematically in others. While this is a common approach in SDMs, uneven sampling effort and lack of recorded absence locations can lead to results that are spatially biased toward more accessible areas (Phillips et al., 2009). When the environmental ranges captured by observation data are biased, SDMs ultimately predict sampling effort rather than the habitat of a given species. To reduce this sampling bias and improve our models' predictive capacity, we used the target group background point selection method in which background points are generated from observation data for a broader set of species (Jarnevich et al., 2015; Phillips et al., 2009). We used the 145,000 most recent research-grade observation locations for all plant species within the model area from iNaturalist.org as our target group (iNaturalist Users, 2021). We then spatially filtered and restricted this to 8000 background points to maximize model speed and performance (Phillips & Dudík, 2008).

The target group sampling method was compared with another background point generation method that randomly places 8000 background points within a 95% kernel density estimate (KDE) of the presence location area. We compared the performance of these two background point generation methods using the average area under the receiver operating characteristic curve (AUC) value across all models and cross-validation runs for California black oak.

We also used the target group background points to assess the level of environmental bias in our sampling effort. To do this, we ran the 8000 background points through our SDMs as input presence locations; high AUC values (larger than 0.70) would indicate that the sampling effort had a high environmental bias (Phillips et al., 2009).

Predictor data

Environmental and climate datasets were standardized across the model area using ArcGIS Pro 2.9 (ESRI, 2021). WorldClim bioclimatic data are biologically meaningful climatic variables representing historical averages from 1970 to 2000 (Fick & Hijmans, 2017). These variables include annual metrics, seasonal metrics, and climatic extremes at 30 arc seconds spatial resolution (pixels are approximately 740 × 920 m within the study area). All 19 bioclimatic variables were used as potential model inputs (Table 1). Additionally, aspect, slope, and curvature (which indicates if a surface is concave, convex, or flat) were calculated from a $\frac{1}{3}$ arc second (approximately 10 m spatial resolution) digital elevation model provided by the U.S. Geological Survey (USGS, 2020). Elevation was also included as a predictor.

To capture the potential impact of past fires, we used CAL FIRE's fire perimeter data from 1911 to 2020, which

TABLE 1 Environmental and topographic datasets input as potential predictors and their sources.

Variable	Source
Annual mean temperature	WorldClim 2.0
Mean diurnal range	WorldClim 2.0
Isothermality	WorldClim 2.0
Temperature seasonality	WorldClim 2.0
Max temperature of warmest month	WorldClim 2.0
Min temperature of coldest month	WorldClim 2.0
Temperature annual range	WorldClim 2.0
Mean temperature of wettest quarter	WorldClim 2.0
Mean temperature of driest quarter	WorldClim 2.0
Mean temperature of warmest quarter	WorldClim 2.0
Mean temperature of coldest quarter	WorldClim 2.0
Annual precipitation	WorldClim 2.0
Precipitation of wettest month	WorldClim 2.0
Precipitation of driest month	WorldClim 2.0
Precipitation seasonality	WorldClim 2.0
Precipitation of wettest quarter	WorldClim 2.0
Precipitation of driest quarter	WorldClim 2.0
Precipitation of warmest quarter	WorldClim 2.0
Precipitation of coldest quarter	WorldClim 2.0
Elevation	USGS DEM
Curvature	USGS DEM
Aspect	USGS DEM
Slope	USGS DEM
Years since wildfire	CalFire, USFS
Years since prescribed fire	CalFire, USFS

Note: Original spatial resolution of WorldClim data was 30 arc seconds; all other datasets were 30 m (or 1.2 arc seconds) spatial resolution.
Abbreviations: DEM, digital elevation model; Max, maximum; Min, minimum; USFS, United States Forest Service; USGS, United States Geological Survey.

include both prescribed fires and wildfires larger than 0.04 km^2 (CAL FIRE and USFS, 2021). The polygon fire perimeters were converted to raster data with 30 m spatial resolution; if two or more fires overlapped in a given cell, the year of the more recent fire was used. Fire years were then converted to a raster representing years since the most recent fire, with areas with no recorded fire since 1911 conservatively assigned a value of 110 years. This process was conducted for prescribed fires and wildfires separately due to their differing impacts on vegetation.

The topographic and fire raster data were reprojected when necessary and resampled to match the spatial resolution, footprints, and projection of the bioclimatic variables. Modeling was conducted in the WGS84 geographic coordinate system (EPSG: 4326), a requirement of the modeling software. Table 1 lists all the predictor datasets and their sources.

Ensemble modeling

For each of the 10 species, location data and predictor data were input into five different SDM algorithms using the VisTrails Software for Assisted Habitat Modeling (SAHM, version 2.2.2) (Morisette et al., 2013). The five models used were boosted regression trees (BRT), random forest (RF), Maxent, multivariate adaptive regression splines (MARS), and a generalized linear model (GLM) with the default parameterization built into SAHM; the five selected models provided a mix of widely used regression and machine learning models (Elith et al., 2006). The five models were run for each of the 10 species using the following workflow (Figure 2). First, we removed collinear variables based on a combined correlation coefficient, calculated as the maximum value of the Pearson, Spearman, and Kendall correlation coefficients. Collinear variables were removed stepwise, starting with the variable with the greatest percentage of deviance explained (based on a univariate generalized additive model) and removing all of the variables that were highly correlated with it (correlation coefficient ≥0.75) until no highly correlated variable pairs remained. Finally, any variable with a percentage of deviance explained value of less than 1.0% was removed and each of the five models was run. This process was repeated independently for each species.

The resulting presence probability surfaces were converted to binary presence and absence classifications using the threshold at which the model's sensitivity equaled its specificity. We then evaluated the accuracy of each model using the mean AUC value, or area under the receiver operating characteristic curve, of 10-fold cross-validation runs. The AUC value is the probability that the model will rank a randomly chosen presence observation higher than a randomly chosen absence observation (Swets, 1988). We assessed how the mean AUC value of the cross-validation runs varied across plant functional types (trees, shrubs, annuals, and perennials). Each model was also evaluated via the percentage of correctly classified presences, the variable importance plot, variable response curves, and multivariate environmental similarity surface (MESS) maps, which indicate areas where a model is extrapolating into environmental conditions that were not represented in the training data (Elith et al., 2010).

For each species, the results of the five models were combined using the binary presence and absence classification maps. Any individual model for which the mean AUC
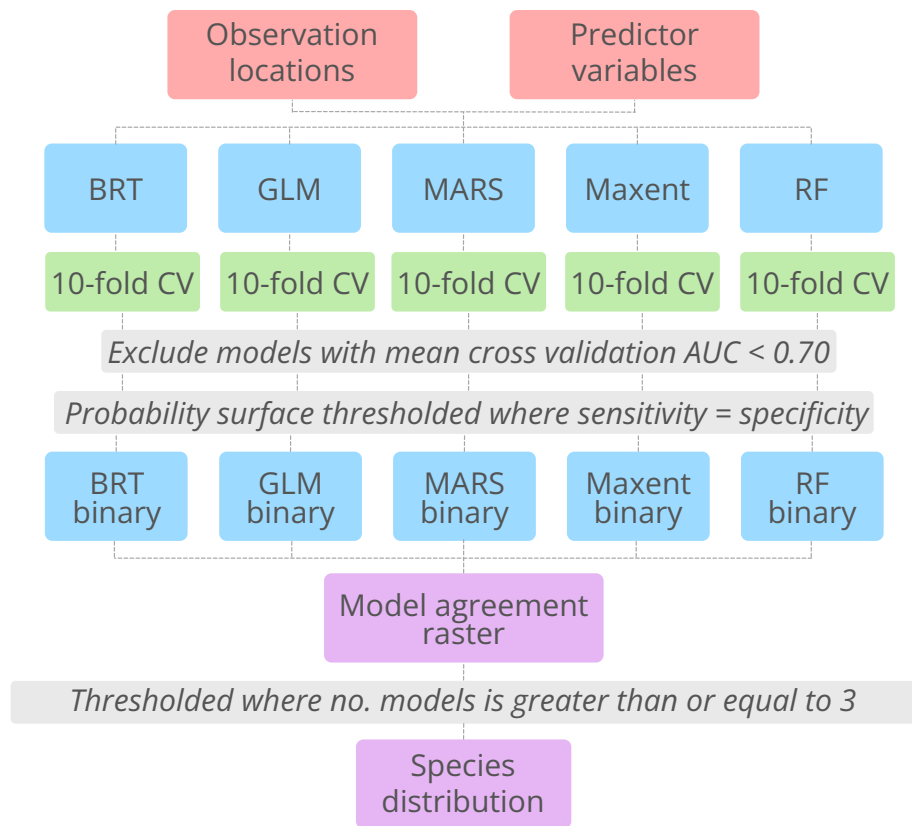
FIGURE 2 Diagram of the workflow conducted for each species. Red blocks represent input datasets, blue blocks represent model runs (cross-validation [CV] testing runs are shown in green), gray blocks represent threshold decisions made by the authors, and purple blocks represent model outputs. Models included boosted regression trees (BRT), a generalized linear model (GLM), multivariate adaptive regression spline (MARS), Maxent, and random forest (RF). AUC, area under the receiver operating characteristic curve.

value of the cross-validation runs was lower than 0.70 was excluded from the ensemble output. The remaining binary classifications were then added together to create an ensemble output indicating the total number of included models (0–5) predicting the species' presence within a given cell. To identify areas where multiple cultural species were likely to be present, we created binary species rasters from the ensemble output rasters. For each species, any pixels with at least three models (a majority) predicting that species' presence were set to 1 and all other pixels were set to 0. These binary species rasters were then summed to create a raster indicating the number of cultural species possibly present at that location (0–10). This predicted species-count map was used in subsequent analyses to identify potential gathering areas. Figure 2 summarizes this workflow.

Accessibility analysis

The Tribe has some formal access agreements with landowners, such as park agencies and conservation organizations, within their stewardship area. More common are informal discussions in which these land-owning entities are open to possible access agreements, but no formal agreement yet exists. Based on discussions with AMTB leadership and coauthor knowledge, we compiled areas where the Tribe has existing or potential gathering agreements from county parcel data and the California Protected Areas Database (CPAD, 2021). To calculate summary statistics and prioritize potential gathering areas, the final predicted species-count raster was clipped to the potential access areas.

RESULTS

Detailed results for California black oak, spatial results for wavy-leafed soap plant, and non-spatial results for the remaining eight species with sensitivity concerns are included here.

Observation data

There were 401 research-grade iNaturalist observations of California black oak in the study area and no observations

from the three other data sources. One hundred eighty-three points were input into the models after excluding duplicates within each pixel footprint (Figure 1). For the remaining nine species, the number of total observations ranged from 231 to 1686 and the number of spatially filtered observations ranged from 109 to 686 (Appendix S1: Table S1).

Background points

The target group background point method improved the California black oak models' performance over the randomized KDE method as measured by the mean testing AUC value; we therefore used the target group background point method for our final analysis of all 10 species. The AUC values of the background point models (i.e., target group background points input as observation data) ranged from 0.515 to 0.694 (mean = 0.651), which indicated low environmental bias in the observation datasets used (an AUC value of 0.50 indicates no predictive capability) (Botella et al., 2020; Phillips et al., 2009). Despite low environmental bias, the target group background points were distributed much more densely in the western half of the model area (Appendix S1: Figure S1).

Ensemble modeling: California black oak

Each species was run with a different subset of the 25 potential predictor variables after stepwise exclusion of collinear variables. In the case of California black oak, the models were ultimately run with nine predictor variables: isothermality (mean diurnal range divided by the annual range in temperature), minimum temperature of

coldest month, precipitation of wettest month, years since prescribed burn, annual mean temperature, slope, precipitation seasonality, elevation, and aspect (Table 2). For four of the five models (GLM, MARS, Maxent, and RF), the three most important variables were isothermality, minimum temperature of the coldest month, and precipitation of the wettest month (Table 2). For the BRT model, the only two input variables selected were isothermality and precipitation of the wettest month (Table 2). Figures 3 and 4 display the California black oak binary presence and absence classification maps individually and combined, respectively.

All five models indicated high predictive capability in the training and cross-validation test runs, ranging from 0.853 (GLM) to 0.918 (BRT) in the training runs and 0.847 (GLM) to 0.869 (Maxent) in the testing runs (Table 3). The difference in AUC and the percent of presences correctly classified between the training and evaluation runs was small, indicating model consistency (Table 3, also see Appendix S1: Table S2).

Ensemble modeling: All species

The predictive capacity of this ensemble model approach varied across the 10 cultural species studied here (Table 4). Of the 50 individual models run, 47 models (94%) met our criteria for inclusion in the ensemble output. The three models that were excluded from our final results due to low predictive capacity used the GLM (2) and MARS (1) methods. Twenty-nine models (58%) had an AUC value greater than or equal to 0.80. Wavy-leafed soap plant had the highest mean AUC values in testing and training runs (Table 4). Figure 5

TABLE 2 Mean variable importance (%) across all 11 runs (training and 10-fold cross validation) of each model for California black oak (*Quercus kelloggii*).

	Variable importance (%)					
Variable	**BRT**	**GLM**	**MARS**	**Maxent**	**RF**	**Mean**
Isothermality (Bio 3)	21.17	21.48	24.12	25.54	12.09	20.88
Min temperature of coldest month (Bio 6)	...	5.20	13.55	7.90	3.99	7.66
Precipitation of wettest month (Bio 13)	12.29	5.08	5.19	10.80	4.82	7.63
Precipitation seasonality (Bio 15)	...	4.21	2.61	0.55	2.31	2.42
Annual mean temperature (Bio 1)	...	...	4.72	1.28	0.83	2.27
Slope	...	...	1.36	1.63	1.35	1.45
Elevation	...	...	0.63	0.69	1.61	0.98
Aspect	...	...	0.42	0.27	0.51	0.40
Years since prescribed fire	...	...	0.00	0.04	0.04	0.03

Note: The selected variables are listed in order of total mean importance across all five models.

Abbreviations: BRT, boosted regression trees; GLM, generalized linear model; MARS, multivariate adaptive regression spline; RF, random forest.
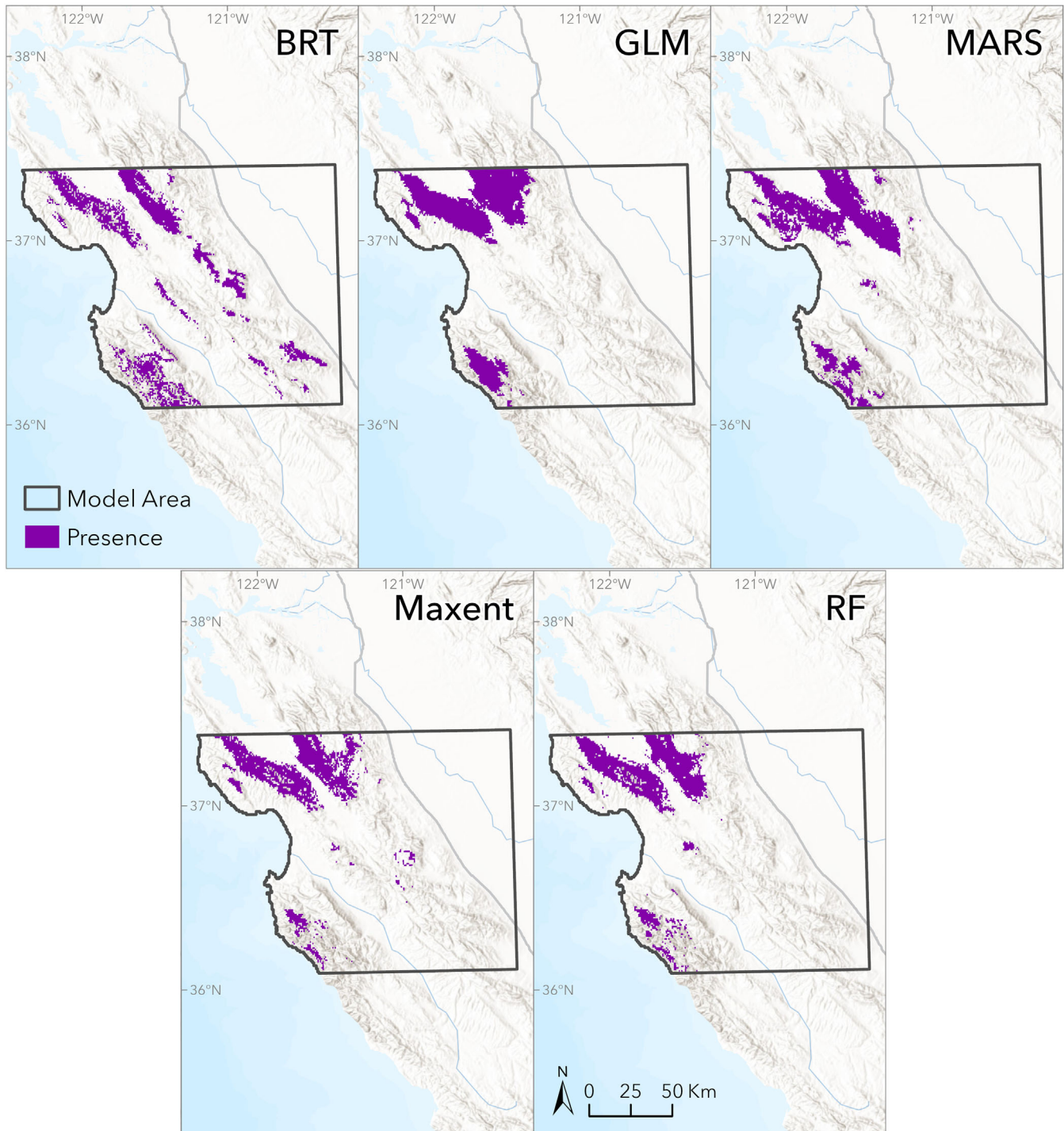
FIGURE 3   California black oak (*Quercus kelloggii*) presence classification results for each of the five species distribution models: boosted regression trees (BRT), generalized linear model (GLM), multivariate adaptive regression spline (MARS), Maxent, and random forest (RF).

shows the combined presence classifications for this species.

Of the five types of models run, RF most frequently resulted in the highest testing AUC value and GLM in the lowest testing AUC value (Table 4). Across all 10 cultural species, RF, Maxent, and BRT models had greater predictive capacity than the GLM and MARS models (Table 4). While we did not have a large enough sample size to statistically compare how AUC values varied across plant functional types, trees (0.858; $n = 1$) and shrubs (0.830; $n = 5$) had higher AUC values on average than perennial (0.817; $n = 3$) and annual plants (0.697; $n = 1$).
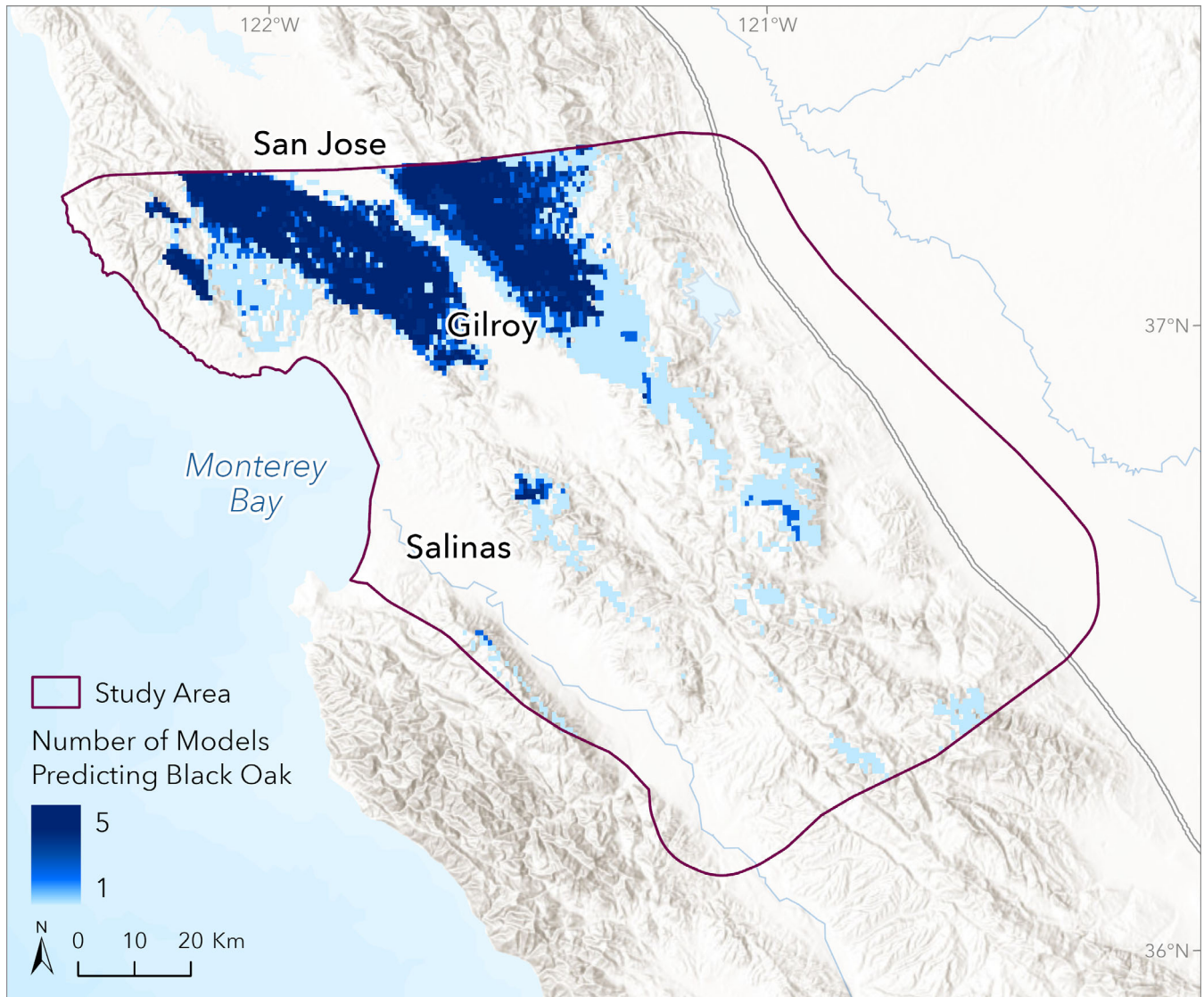
F I G U R E 4   Number of models (1–5) predicting California black oak (*Quercus kelloggii*) presence within the study area. All five models met the criteria for inclusion, that is, the mean area under the receiver operating characteristic curve value of the cross-validation runs was greater than 0.70.

T A B L E 3   Area under the receiver operating characteristic curve (AUC) values for the five California black oak (*Quercus kelloggii*) models for the training run and the average AUC value of the 10-fold cross-validation runs.

Type of model run	BRT	GLM	MARS	Maxent	RF	Mean
Training	0.918	0.853	0.865	0.868	0.894	0.880
Cross validation (mean)	0.860	0.847	0.854	0.869	0.858	0.858
Difference	−0.058	−0.007	−0.011	0.001	−0.036	−0.022

Note: The difference indicates the change in the AUC value in the cross-validation runs as compared with the training run, which can be used to evaluate model consistency.

Abbreviations: BRT, boosted regression trees; GLM, generalized linear model; MARS, multivariate adaptive regression spline; RF, random forest.

The MESS maps showed only minimal extrapolation (35 pixels) in the eastern edge of the model area across all 10 species, indicating that the available environment within the study area was well sampled.

Accessibility analysis

Areas with potential access by the Tribe made up 8.81% of the study area. The majority of these areas were

TABLE 4 Mean area under the receiver operating characteristic curve (AUC) values for the training and 10-fold cross-validation (CV) testing runs of the five models for each of the 10 cultural species.

Species	Mean AUC		Met ensemble criteria	Testing AUC by model	
	Training	CV testing		Highest	Lowest
Chlorogalum pomeridianum	0.939	0.905	5	RF	GLM
Corylus cornuta ssp. *californica*	0.919	0.894	5	RF	GLM
Vaccinium ovatum	0.903	0.885	5	Maxent	GLM
Quercus kelloggii	0.880	0.858	5	RF	GLM
Rubus parviflorus	0.875	0.854	5	Maxent	MARS
Clinopodium douglasii	0.851	0.821	5	RF	GLM
Rubus ursinus	0.798	0.776	5	Maxent	GLM
Sambucus nigra ssp. *caerulea*	0.779	0.743	5	BRT	GLM
Artemisia douglasiana	0.767	0.725	4	BRT	GLM
Calandrinia menziesii	0.744	0.697	3	RF	GLM

Note: The column labeled "Met ensemble criteria" indicates the number of models (of the five) for which the mean testing AUC was greater than 0.70 and was therefore included in that species' final model agreement raster. The table is sorted by the mean cross-validation testing AUC value.

Abbreviations: BRT, boosted regression trees; GLM, generalized linear model; MARS, multivariate adaptive regression spline; RF, random forest.

accessible on a possible case-by-case basis, meaning that access was not guaranteed and that a tribal member would need to request it and potentially coordinate a date and time for access. In total, 3.50% (39.87 km^2) required a request for access and gathering, 13.48% (153.75 km^2) allowed some form of access but prohibited gathering, and 83.03% (947.11 km^2) were indicated as potential but not guaranteed gathering access.

To prioritize within this large area, potential gathering places were defined as places where two or more cultural species were predicted to be present based on the predicted species-count map. Within the study area, 2501.5 km^2 or 19.33% of the area contained potential gathering places. Within the subset of lands where the Tribe may have some form of access, 609.35 km^2 or 53.63% contained potential gathering places. Places where seven or more cultural plants were predicted to be present totaled 265.2 km^2 (2.1%) of the study area; 75.85 km^2 (28.6%) or these high-priority areas fall within areas where the Tribe has some form of access. Tables 5 and 6 show the area (in square kilometers) breakdown by the predicted number of species present within these two areas (study area and potential access areas) respectively.

DISCUSSION

Using private and public observation data for 10 priority plant species and 25 predictor variables, we used an ensemble modeling approach to predict the potential distributions of each plant throughout the Amah Mutsun stewardship area. We then highlighted areas likely to contain multiple species of interest and analyzed how these predicted distributions overlapped with areas where the Tribe has various forms of access. Predicted distribution maps for each of the modeled species also highlight the best areas to gather specific plants. Our methodology was designed using best practices in the fields of geospatial modeling and Indigenous environmental studies in a number of key ways. In the realm of geospatial modeling, we employed an ensemble model approach and utilized target group presences as background points to reduce model bias. Based on best practices in the field of Indigenous environmental studies, we built our research partnership on the principles of free, prior, and informed consent from the Amah Mutsun Tribal Council. In addition, we conducted interviews with community members and built relationships with tribal leadership over several years prior to beginning this study, which allowed us to design culturally relevant research.

We found that 2.1% of the Amah Mutsun stewardship area is potential habitat for seven or more of the cultural plant species included in our analysis; the Tribe has some form of access to approximately a third of these high-priority areas. This subset represents the highest priority for future investigation as potential gathering areas because access agreements or partnerships are already in place. For the remaining high-priority areas, the Tribe may use these maps to strategically reach out to other land-owning agencies or individuals as their gathering program expands.

Model accuracy

We found these models to be highly predictive of species locations. Of the 50 models run in this study, the majority
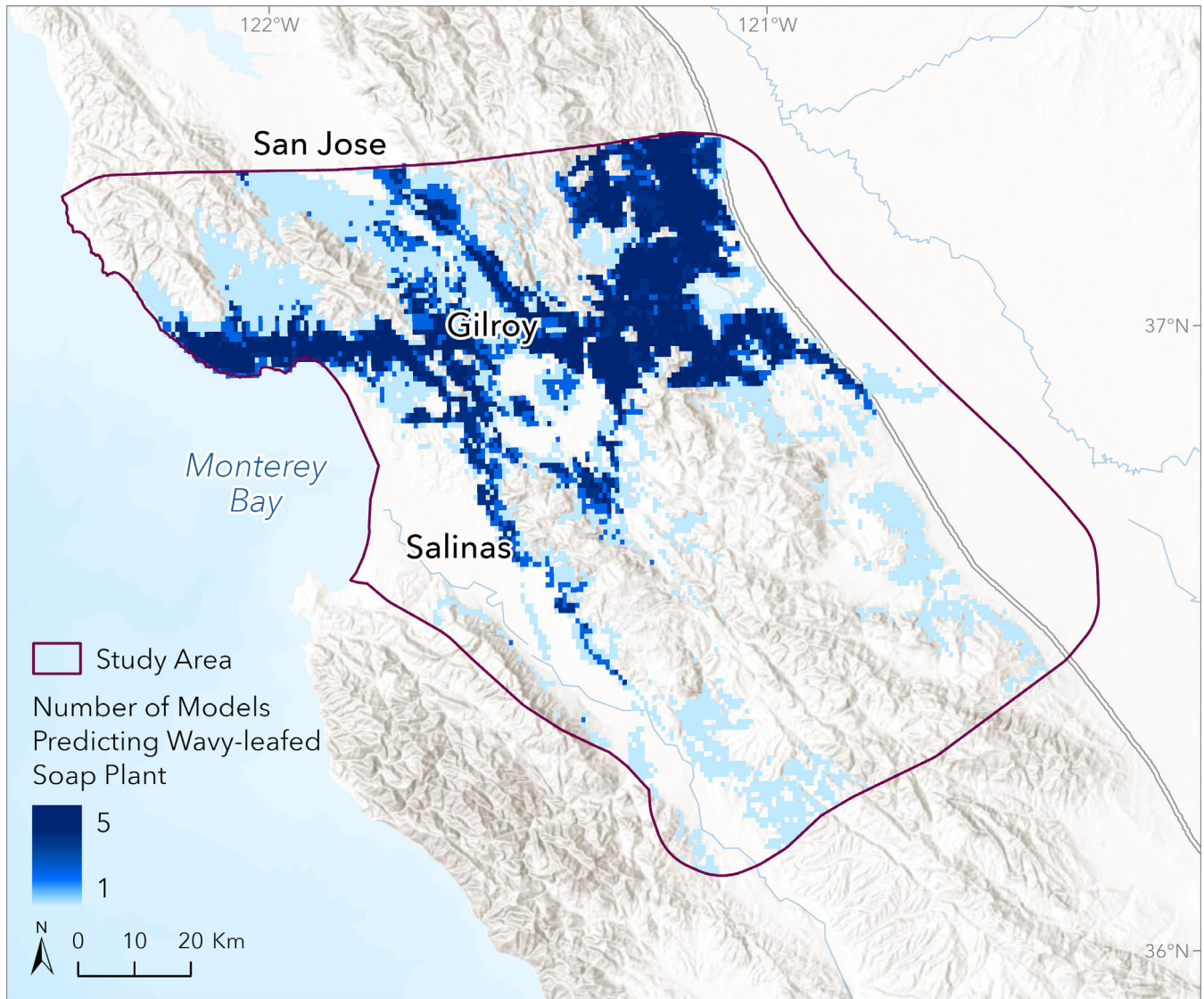
FIGURE 5 Number of models (1–5) predicting wavy-leafed soap plant (*Chlorogalum pomeridianum*) presence within the study area. All five models met the criteria for inclusion, that is, the mean area under the receiver operating characteristic curve value of the cross-validation runs was greater than 0.70.

(94%) had a mean AUC value greater than or equal to 0.70 in the cross-validation runs. There are a number of metrics and factors to consider when evaluating the predictive power of SDMs. First, in the case of presence-only species distribution modeling, the maximum achievable AUC value is less than 1, with widely distributed species having a lower maximum achievable AUC value (Wiley et al., 2003). It is therefore not advisable to compare AUC values across species without knowledge of the relative differences in their coverage within the study area (Wiley et al., 2003). AUC values are valuable for comparing different models and parameterizations with respect to each species individually. Across all 10 species, RF and then Maxent models most frequently had the highest predictivity while GLM most frequently had the lowest predictivity (Table 4). This may be because GLMs generally

do not capture complex ecological responses as well as the other methods used here (Elith et al., 2006). In the case of California black oak, the GLM and MARS models predicted a more widespread distribution relative to the other three modeling methods and had the two lowest percentages of correctly classified presences (Figure 3; Appendix S1: Table S2). Across all 10 species, we found that the remaining three models (RF, BRT, and Maxent) had the highest predictivity and may be more suitable for presence-only species distribution modeling (Table 4).

We also evaluated how well the relative importance of each predictor variable lined up with our expectations for each species. In the case of California black oak, isothermality and precipitation of the wettest month were retained by all five models and had the first and third highest mean variable importance (Table 2). Across the

TABLE 5 Area of potential gathering places within the study area, based on the number of species predicted by ensemble distribution modeling.

Predicted species count	Area (km^2)	Percentage
0	8254.4	63.77
1	2188.2	16.90
2	889.7	6.87
3	424.0	3.28
4	428.0	3.31
5	293.5	2.27
6	201.2	1.55
7	170.7	1.32
8	78.2	0.60
9	16.3	0.13
Total	12,944.1	100.00

TABLE 6 Area of potential gathering places, based on the number of species predicted by ensemble distribution modeling, within the areas where the Amah Mutsun Tribal Band may have some form of land access.

Predicted species count	Area (km^2)	Percentage
0	247.41	21.77
1	279.47	24.60
2	245.21	21.58
3	87.13	7.67
4	81.73	7.19
5	74.48	6.56
6	44.94	3.96
7	41.75	3.67
8	25.24	2.22
9	8.86	0.78
Total	1136.23	100.00

five models, California black oak habitat was more likely to be predicted in wetter areas and at higher elevations where temperatures are more variable throughout the year, which fits our expectations for the species.

Our methods were novel in that we included years since the most recent wildfire and years since the most recent prescribed fire as potential predictors. One of these factors was selected as an input into at least one of the five models for all but 1 of the 10 species (*R. ursinus*), and for 13 (prescribed fire) and 27 (wildfire) of the 50 total models. However, their mean variable importance tended to be low, as in the case of California black oak (Table 2). This does not necessarily indicate that fire or other forms of disturbance are not predictive of plant distribution, but rather show that it is difficult to incorporate a temporally

and spatially dynamic process into a static model. Policies of fire suppression and widespread urban development have meant that both wildfire and prescribed fire are rare within the study area, and this low occurrence makes it difficult to accurately evaluate the importance of fire in predicting plant habitat.

Our methods appeared to better predict the distribution of trees and shrubs as compared with annual and perennial plants, which may be due in part to variation in life-history strategies. For example, many annual plants depend on disturbance for survival and only 2 of the 25 potential predictors (years since prescribed fire and wildfire) reflected a form of disturbance. Our results are promising and justify future work to explore metrics of disturbance frequency in addition to disturbance presence, which may better predict species that are adapted to certain disturbance regimes.

Our ensemble results lined up well with our expected distribution for each of the 10 species. Of the 10 species included in our analysis, the 5 most widely distributed species (according to the authors' ecological knowledge and CalFlora's estimated range maps within the study area) have the five lowest AUC values (CalFlora, 2021). This aligns with previous work that showed that the maximum achievable AUC value is lower for widely distributed species (Wiley et al., 2003). The MESS maps indicated very little extrapolation across all 50 models. However, our models systematically underpredicted plant distributions in the eastern and southern portions of the study area, which is likely driven by two interrelated factors. First, the sampling effort of our observation data was biased toward the western and northern quadrants of the study area where there are more public parks and trails (Appendix S1: Figure S1). Second, the inland areas in the south and east portions of the study area experience a different climate characterized by greater temperature extremes. Therefore, while it may be the case that some of these plants do not grow farther inland, it may also be that our input data do not accurately reflect how plants are distributed in this inland biome. A critical next step will be to collect field observations from these under-sampled areas to iteratively improve these models.

Limitations

Our methods work to mitigate the potential biases of SDMs in several ways. First, we used a target group background point selection method to account for potential environmental bias in the sampling effort of our observation data. Second, we used an ensemble modeling approach that uses majority agreement to assign any pixel as predicted habitat and excludes individual models

with lower predictive capacity. Third, we excluded species for which the number of observations was below 100. Fourth, we preprocessed our inputs to reduce pseudo-oversampling of observation data and ran each model using an uncorrelated and relevant subset of predictor datasets. We also chose to include two types of fire history (both wildfires and prescribed fire) as predictor variables in our model given the unique relationships between many Amah Mutsun cultural plants and fire.

Despite these mitigation efforts, it is important to acknowledge the assumptions and limitations that are inherent to SDMs. Specifically, SDMs assume that the presence locations for each species are a representative sample of its habitat, that the chosen predictor variables accurately capture the habitat constraints on each species, and that the spatial resolution of the models can capture each species' habitat (Jarnevich et al., 2015). In particular, while the bioclimatic variables are an extremely useful resource for distribution modeling, they reduced the spatial resolution of our analysis 25-fold. At this coarser resolution, our topographic predictor inputs did not reveal important microhabitats such as small ridges and valleys that may be important indicators of each plant's habitat.

In addition, we used presence-only models that leverage background points (as opposed to absence points) to model the environmental niche of each species. The iNaturalist observation data incorporated here are often collected opportunistically and we found that the sampling effort was biased toward popular or accessible areas (Appendix S1: Figure S1). We used the target group background point method to reduce the potential bias of our models resulting from this sampling bias and found that it increased the predictive capability of our models, which aligns with previous studies (Botella et al., 2020; Phillips et al., 2009). We found that the iNaturalist data were skewed not only toward accessible areas but also toward certain taxa, and that observations of native grasses were particularly sparse. Many native grass species are important cultural plants for the Tribe but were ultimately excluded due to an insufficient number of unique observation locations. Accurate identification of grass species is difficult and often requires specialized knowledge of grass anatomy, which may be less common among iNaturalist users. Lastly, these models identify areas where a cultural species is likely to be present, but do not indicate where each is likely to be most abundant. The implications of this are discussed in more detail in the next section.

While we acknowledge the limitations of our data and models, these predicted distributions are a valuable step toward rebuilding the Tribe's relationships with cultural plants. Our results will be used to direct valuable resources toward the highest priority areas and are not considered definitive species maps. Our ensemble modeling approach also allows us to map regions of more or less certainty, either by assessing how many models predicted presence in a given area, or by viewing the MESS maps to determine if any models were extrapolating into a given area.

Applications and future work

The primary next step will be to prioritize a subset of the areas identified as potential gathering areas for further investigation in the field. There are several existing factors that the Tribe may wish to use in this prioritization, including proximity to known Amah Mutsun sacred and cultural landscapes, ease of access (proximity to trails, parking lots, and tribal members' homes), and ADA accessibility. We are partnering with AMTB to create detailed maps showing these factors in relation to the potential gathering areas. Once refined, these potential gathering areas will become part of an interactive mapping tool (restricted to the Amah Mutsun community) that supports wider access to ethnobotanical resources, an expressed tribal priority.

In addition to directly supporting the Tribe's gathering program, this analysis can be used to protect culturally and ecologically sensitive areas. Specifically, these maps empower the Tribe to request changes in the management of these priority areas to exclude the use of pesticides and herbicides, or to conduct mowing and burning to the times of year best suited to sensitive species. In addition, there are multiple potential gathering areas in places where the Tribe does not yet have access or gathering rights; these maps can therefore be used to begin new partnerships, both with public agencies and private landowners.

Future work to analyze phenological patterns within the priority areas identified here could further refine the Tribe's stewardship programs. Stewardship and gathering of cultural plants require knowledge of both the location and phenology of each plant; the phenology determines not only when a plant may be ripe or ready to gather, but also the appropriate times for other stewardship activities such as cultural burning, mowing, or sowing of seeds. Given that the ideal timing of gathering or other stewardship activities may vary year to year and along environmental gradients, remote sensing methods that efficiently capture phenological signatures over large spatial and temporal scales can augment existing place-based knowledge held by tribal members.

The Tribe still faces many barriers in rebuilding relationships with cultural plants. In the absence of

Indigenous stewardship and reciprocal relationship, many of these plants will not produce materials of an abundance or quality high enough for cultural use or consumption. For example, California black oak acorns can be infested with acorn weevil in the absence of cultural burning (Anderson, 2013). Previous work with the Karuk and Yurok tribes also highlighted how California hazelnut (*C. cornuta* ssp. *californica*) required thinning or burning to produce basketry-quality shoots (Marks-Block et al., 2019). In areas where a cultural plant is present but not abundant enough for tribal members to gather it, additional stewardship or restoration may be necessary before gathering is possible. The minimum plant abundance necessary for gathering will vary by species and will likely be determined in partnership between the Tribe and relevant landowner. Second, while these maps serve as a guide for prioritizing new access partnerships, they do not do all the work of outreach, communication, and relationship building that is required to build and maintain those partnerships. Finally, there are challenges when balancing confidentiality and ease of use. Given the sensitive nature of the potential gathering area maps, it is critical that they be kept within the Amah Mutsun community. However, if they are kept so confidential that most community members cannot use them, they lose their purpose. An integral next step in our work will be to incorporate our results into an accessible mobile platform that can be kept internal to the Amah Mutsun community.

CONCLUSIONS

This work is a novel example of how geospatial modeling can be utilized by an Indigenous community to rebuild relationships with cultural plants across large areas of their homelands and directly contribute to land access and cultural revitalization. Our analysis paints a picture of the most accessible places for the restoration of relationships with native plants, which will help to direct limited time and resources to priority areas. The technical methods we developed represent cutting-edge modeling techniques and incorporate best practices in species distribution modeling. We compared five commonly used SDMs and found that RF and Maxent models performed best across 10 plant species in a presence-only modeling context. We also included past prescribed fire and wildfire as metrics of disturbance on the landscape and show their relevance for predicting plant habitat, which we hope will spark more work in this area. Furthermore, our methodology leverages publicly available data and an open-source program (SAHM), enabling us to quickly scale up this analysis to

include dozens of cultural plants at no cost. These methods are easily replicable and can be adopted by other Indigenous communities in their diverse efforts to reconnect with land. Finally, this work is an example of a partnership between spatial scientists and an Indigenous community in which the results of the study are directly applied toward cultural revitalization. We modeled a framework for integrating culturally sensitive information into geospatial research while respecting its confidentiality and the sovereignty of that Indigenous community.

CONFLICT OF INTEREST

Maggi Kelly declares no conflicts of interest. Annalise Taylor and Alexii Sigona have worked as independent contractors for the Amah Mutsun Land Trust (an organization separate from but related to the Amah Mustun Tribal Band) on projects unrelated to this research.

DATA AVAILABILITY STATEMENT

The maps showing the modeled locations of the eight sensitive species are available to qualified researchers through the Tribal Council of the Amah Mutsun Tribal Band by contacting info@amahmutsun.org. The raw data, derived data, and code supporting this research (Taylor et al., 2022), which are not sensitive, are available from Figshare: https://doi.org/10.6084/m9.figshare.20469156.v1.

ORCID

Annalise Taylor https://orcid.org/0000-0002-0336-5983

REFERENCES

Anderson, K. M. 2013. *Tending the Wild: Native American Knowledge and the Management of California's Natural Resources*. Berkeley, CA: University of California Press.

Apodaca, A., and K. Lightfoot. 2020. "Cultural Plant Locations—Survey Data." University of California. Unpublished Dataset. https://anthropology.berkeley.edu/kent-lightfoot.

Baumflek, M., S. DeGloria, and K.-A. Kassam. 2015. "Habitat Modeling for Health Sovereignty: Increasing Indigenous Access to Medicinal Plants in Northern Maine, USA." *Applied Geography* 56: 83–94.

Botella, C., A. Joly, P. Monestiez, P. Bonnet, and F. Munoz. 2020. "Bias in Presence-Only Niche Models Related to Sampling Effort and Species Niches: Lessons for Background Point Selection." *PLoS One* 15: e0232078.

Brown, G., and M. Kyttä. 2018. "Key Issues and Priorities in Participatory Mapping: Toward Integration or Increased Specialization?" *Applied Geography* 95: 1–8.

CAL FIRE and USFS. 2021. "California Fire Perimeters." Dataset. https://frap.fire.ca.gov/mapping/gis-data/.

CalFlora. 2021. "Maps of CalFlora Potential Habitat by Climate Model." https://www.calflora.org/entry/help/dgrid-help.html#design.

California Protected Areas Database (CPAD). 2021. "CPAD Version 2021b." Dataset. www.calands.org.

Cuthrell, R. Q. 2013. "An Eco-Archaeological Study of Late Holocene Indigenous Foodways and Landscape Management Practices at Quiroste Valley Cultural Preserve, San Mateo County, California." Doctoral diss., University of California.

Cuthrell, R. Q., C. Striplen, M. Hylkema, and K. G. Lightfoot. 2016. "A Land of Fire: Anthropogenic Burning on the Central Coast of California." In *Contemporary Issues in California Archaeology*, edited by T. L. Jones and J. E. Perry, 153–72. London: Routledge.

Di Cecco, G. J., V. Barve, M. W. Belitz, B. J. Stucky, R. P. Guralnick, and A. H. Hurlbert. 2021. "Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science." *Bioscience* 71: 1179–88.

Diver, S., M. Vaughan, M. Baker-Médard, and H. Lukacs. 2019. "Recognizing 'Reciprocal Relations' to Restore Community Access to Land and Water." *International Journal of the Commons* 13: 400–29.

Eisen, R. J., S. Feirer, K. A. Padgett, M. B. Hahn, A. J. Monaghan, V. L. Kramer, R. S. Lane, and M. Kelly. 2018. "Modeling Climate Suitability of the Western Blacklegged Tick in California." *Journal of Medical Entomology* 55: 1133–42.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, et al. 2006. "Novel Methods Improve Prediction of Species' Distributions from Occurrence Data." *Ecography* 29: 129–51.

Elith, J., M. Kearney, and S. Phillips. 2010. "The Art of Modelling Range-Shifting Species." *Methods in Ecology and Evolution* 1(4): 330–42.

Environmental Systems Research Institute (ESRI). 2021. "ArcGIS Pro. V. 2.9." Environmental Systems Research Institute. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview.

Fick, S. E., and R. J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37: 4302–15.

Gorman, J., D. Pearson, and P. Whitehead. 2008. "Assisting Australian Indigenous Resource Management and Sustainable Utilization of Species through the Use of GIS and Environmental Modeling Techniques." *Journal of Environmental Management* 86: 104–13.

Guisan, A., and N. E. Zimmermann. 2000. "Predictive Habitat Distribution Models in Ecology." *Ecological Modelling* 135: 147–86.

iNaturalist Users. 2021. "iNaturalist Research-Grade Observations." Dataset. https://doi.org/10.15468/ab3s5x.

Jarnevich, C. S., T. J. Stohlgren, S. Kumar, J. T. Morisette, and T. R. Holcombe. 2015. "Caveats for Correlative Species Distribution Modeling." *Ecological Informatics* 29: 6–15.

Kimmerer, R. 2011. "Restoration and Reciprocity: The Contributions of Traditional Ecological Knowledge." In *Human Dimensions of Ecological Restoration: Integrating Science, Nature, and Culture*, edited by D. Egan, E. E. Hjerpe, and J. Abrams, 257–76. Washington, DC: Island Press/Center for Resource Economics.

Lake, F. K., V. Wright, P. Morgan, M. McFadzen, D. McWethy, and C. Stevens-Rumann. 2017. "Returning Fire to the Land: Celebrating Traditional Knowledge and Fire." *Journal of Forestry* 115: 343–53.

Li, X., and Y. Wang. 2013. "Applying Various Algorithms for Species Distribution Modelling." *Integrative Zoology* 8: 124–35.

Lightfoot, K., R. Cuthrell, M. Hylkema, V. Lopez, D. Gifford-Gonzalez, J. R. G. Michael, S. Gabriel, et al. 2021. "The Ecoarchaeological Investigation of Indigenous Stewardship Practices on the Santa Cruz Coast." *Journal of California and Great Basin Anthropology* 41: 187–205.

Lopez, V. 2013. "The Amah Mutsun Tribal Band: Reflections on Collaborative Archaeology." *California Archaeology* 5: 221–3.

Marks-Block, T., F. K. Lake, and L. M. Curran. 2019. "Effects of Understory Fire Management Treatments on California Hazelnut, an Ecocultural Resource of the Karuk and Yurok Indians in the Pacific Northwest." *Forest Ecology and Management* 450: 117517.

Morisette, J. T., C. S. Jarnevich, T. R. Holcombe, C. B. Talbert, D. Ignizio, M. K. Talbert, C. Silva, D. Koop, A. Swanson, and N. E. Young. 2013. "VisTrails SAHM: Visualization and Workflow Management for Species Habitat Modeling." *Ecography* 36: 129–35.

Phillips, S. J., and M. Dudík. 2008. "Modeling of Species Distributions with Maxent: New Extensions and a Comprehensive Evaluation." *Ecography* 31: 161–75.

Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. "Sample Selection Bias and Presence-Only Distribution Models: Implications for Background and Pseudo-Absence Data." *Ecological Applications* 19: 181–97.

Reid, G., and R. E. Sieber. 2020. "Learning from Critiques of GIS for Assessing the Geoweb and Indigenous Knowledges." *GeoJournal* 87: 875–93.

Sanchez, G. M., M. A. Grone, A. J. Apodaca, R. S. Byram, V. Lopez, and R. A. Jewett. 2021. "Sensing the Past: Perspectives on Collaborative Archaeology and Ground Penetrating Radar Techniques from Coastal California." *Remote Sensing* 13: 285.

Swets, J. A. 1988. "Measuring the Accuracy of Diagnostic Systems." *Science* 240: 1285–93.

Taylor, A., A. Sigona, and M. Kelly. 2022. "Data and Supporting Code: Modeling Spatial Distributions of Amah Mutsun Priority Cultural Plants to Support Indigenous Cultural Revitalization." Dataset. Figshare. https://doi.org/10.6084/m9.figshare.20469156.v1.

United States Geological Survey (USGS). 2020. "3D Elevation Program 10-Meter Resolution Digital Elevation Model." Dataset. https://data.usgs.gov/datacatalog/data/USGS:3a81321b-c153-416f-98b7-cc8e5f0e17c3.

Wiley, E. O., K. M. McNyset, A. T. Peterson, C. R. Robins, and A. M. Stewart. 2003. "Niche Modeling Perspective on Geographic Range Predictions in the Marine Environment Using a Machine-Learning Algorithm." *Oceanography* 16: 120–7.

Younger Lagoon Reserve. 2021. "Native Seed Collection Data." Unpublished Dataset. https://youngerlagoonreserve.ucsc.edu/.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Taylor, Annalise, Alexii Sigona, and Maggi Kelly. 2023. "Modeling Spatial Distributions of Amah Mutsun Priority Cultural Plants to Support Indigenous Cultural Revitalization." *Ecosphere* 14(1): e4374. https://doi.org/10.1002/ecs2.4374