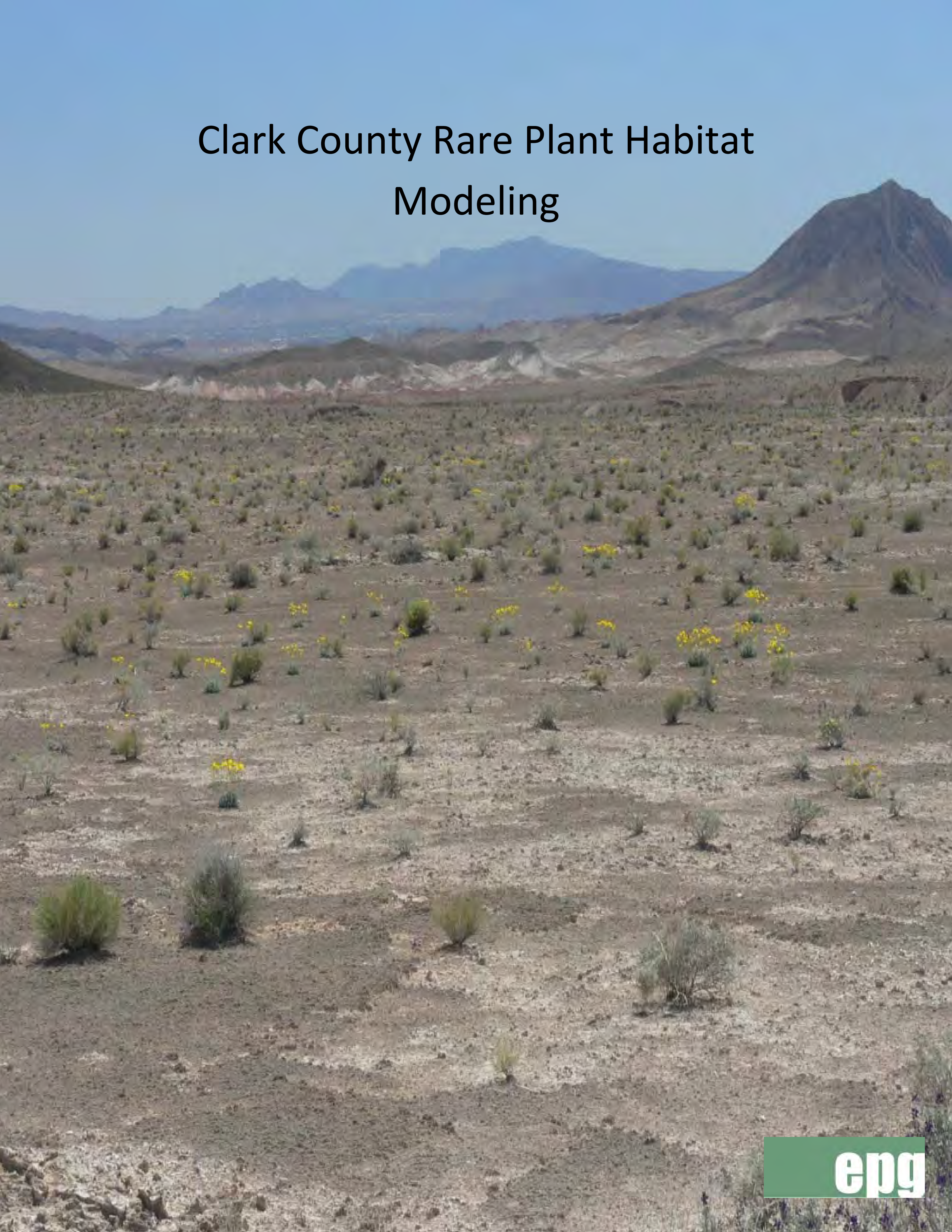
Clark County Rare Plant Habitat Modeling



Clark County Rare Plant Habitat Modeling

Prepared for

Bureau of Land Management
Las Vegas Field Office

Contract # GS10F0046N/L10PD04384

By

Mathew Hamilton
Environmental Planning Group, Inc.

and

Sonja Kokos
Bureau of Reclamation
(formerly with Clark County Desert Conservation Program)

July 2011

Acknowledgements

The habitat modeling efforts and preparation of this report were funded by the Bureau of Land Management Nevada State Office through a 2010 challenge cost share award to the Las Vegas Field Office in partnership with the Clark County the Desert Conservation Program. We would like to thank the people who met with us to discuss the development and execution of this project, provided data necessary to complete this project, and provided comments on the draft report. This includes Fred Edwards of the Bureau of Land Management, Lee Bice of the Clark County Desert Conservation Program, and Larry Tinney and David Brickey of TerraSpectra Geomatics. A lot of the credit for this project goes to Sonja Kokos, who while with the Clark County Desert Conservation Program, developed the overall idea of the project and formulated the steps needed to complete the project. She was involved in all aspects of the initial modeling and field sampling efforts including overseeing and directing the work of multiple subcontractors. She also provided advice during the revision of the habitat models as part of this project.

Cover: Photo of Las Vegas bearpoppy habitat in the vicinity of the Pabco Mine and the Sunrise Instant Study Area, Clark County, NV. Photo taken by Mathew Hamilton in 2010.

Suggested Citation:

Hamilton, M.E. and S.R. Kokos. 2011. Clark County Rare Plant Habitat Modeling. Prepared for the Bureau of Land Management Las Vegas Field Office.

This page intentionally left blank.

Table of Contents

List of Tables

Appendices

1.0 EXECUTIVE SUMMARY

To understand the distribution of rare plants covered under the Clark County Multiple Species Habitat Conservation Plan, in 2009 the Desert Conservation Program and TerraSpectra Geomatics developed two coarse soil GIS models using ASTER Imagery, soil survey data, and geological data. Species specific habitat models for eight rare and endemic plant species were then created using the soil models and presence/absence data. The first group of models included models for three gypsum loving species: sticky ringstem (*Anulocaulis leiosolenus var. leiosolenus*); Las Vegas bearpoppy (*Arctomecon californica*); and Las Vegas buckwheat (*Eriogonum corymbosum var. nilesii*). The second group of models included models for five sand or potentially sand loving species: threecorner milkvetch (*Astragalus geyeri var. triquetrus*); Pahrump Valley buckwheat (*Eriogonom bifurcatum*); sticky buckwheat (*Eriogonum viscidulum);* Beaver Dam breadroot (*Pediomelum castoreum*); and white-margined beardtongue (*Penstemon albomarginatus*) (Kokos et al, 2009). The models were used to create a survey design to improve the knowledge of the distribution of the species within Clark County. Two additional species, white bearpoppy (*Arctomecon merriamii*) and two-tone beardtongue (*Penstemon bicolor* ssp. *bicolor*), were also targeted during the field surveys.

The original habitat models were purposely over-predictive so as to not bias the field surveys with prior knowledge about the species natural history. Once the surveys were completed, the goal was to refine the habitat models using the knowledge gathered during the field surveys. In addition, a new County-wide (Clark County) surficial geologic map has been published since the development of the original models (House et al, 2010). As part of the current project, the original gypsiferous and sandy soil models were updated by TerraSpectra to incorporate the new County-wide surficial geologic map along with other refinements (TerraSpectra, 2011). These new soil models were then used to create new habitat models for each of the species. These new species specific habitat models incorporated knowledge about the species elevation range along with the species occurrence patterns compared to the soil model detailed attributes. The individual models were then attributed with information about the species known occurrence and absence records. In addition to the soil based models, climate based models were created for each of the species using Maxent.

Overall, the project demonstrates the potential for and usefulness of creating soil based habitat models for many of the target species. A soil-based habitat model was not created for two-tone beardtongue because it is neither a gypsum nor sand loving species. Two-tone beardtongue is found on rocky and gravelly soils and no attempt was made to model this soil type because rocky soils are wide spread throughout Clark County and can occur in almost any geologic or soil map unit. The gypsum based model for white bearpoppy misses many of the known occurrences within Clark County. This species occurs not only on gypsum but also calcareous soils which were not modeled. For these two species, the climate based models may be the best option at this time.

This habitat modeling exercise highlighted several important issues for future modeling efforts. The quality of the models can be greatly influenced by the quality and spatial resolution of the inputs used, including the quality of the species occurrence data. Finer resolution maps are better than coarse resolution maps but even fine resolution maps are not helpful if the soil parameter of interest is not mapped. Issues with the quality of the plant species occurrence records can also impact the quality of the habitat maps. Data with errors or low positional accuracy can lead to habitat types inappropriately being included or dropped from the habitat model. Finally, this project highlights the need for absence data to be documented. Absence data does not definitively indicate that a species was never present at a site or that the site does not contain suitable habitat. Documented absence data, though, does provide management more information when trying to make decisions when an area has been modeled as potential suitable habitat.

2.0 INTRODUCTION

2.1 Background

2.1.1 History of Project

The project was originally proposed in 2004 by the Bureau of Land Management (BLM) as a rare plant inventory project as part of the 2005-2007 Implementation Plan and Budget for the Clark County Multiple Species Habitat Conservation Plan (MSHCP). The proposed project was accepted by the Clark County Desert Conservation Program (DCP) through the Round 6 funding cycle. When the funding became available in 2007, the BLM was unable to implement the project, so the project reverted back to the DCP for implementation.

Based on discussions with BLM and Fish and Wildlife Service (FWS), ten species were selected due to pressing conservation concerns. The species were sticky ringstem (*Anulocaulis leiosolenus var. leiosolenus*), Las Vegas bearpoppy (*Arctomecon californica*), white bearpoppy (*Arctomecon merriamii*), threecorner milkvetch (*Astragalus geyeri var. triquetrus*), Pahrump Valley buckwheat (*Eriogonum bifurcatum*), Las Vegas buckwheat (*Eriogonum corymbosum var. nilesii*), sticky buckwheat (*Eriogonum viscidulum*), Beaver Dam breadroot (*Pediomelum castoreum*), white-margined beardtongue (*Penstemon albomarginatus*), and yellow two-tone beardtongue (*Penstemon bicolor ssp. bicolor*).

For these ten species, the goal of the project was to address information gaps such as locating unknown populations or the extent of known populations in Clark County (County) (Kokos et al, 2009). This improved knowledge of the species distributions could be used to better identify potential areas of conservation by land managers.

2.2 Need for Habitat Models

In developing the sampling design for the DCP project, the surveys needed to be focused in areas with a potential to have the rare plants of interest instead of just randomly distributed throughout all areas of the County. In 2007, the Lower Elevation Rare Plant Conservation Management Strategy (LERPCMS) was completed by The Nature Conservancy (TNC, 2007). This document covered 7 out of the 10 species included in the DCP project. As part of this document, TNC produced polygons representing buffers around the known locations of each of the species. The DCP used this distribution information from the TNC document as a starting point for the inventory project. The DCP, though, determined that these distribution polygons were not suitable for use in developing a sampling design because they would have limited the surveys to areas already known to contain the species. This would not have resulted in increasing the distributional knowledge of the species within the County.

In order to maximize the potential for finding new populations of the rare plants while also looking at potential suitable habitat throughout the County, it was decided to develop predictive habitat models for the species that could be used to help limit where surveys would be performed. A requirement of project was that at least 20% of the surveys needed to be conducted in areas not known to contain the species.

2.3 Initial Modeling Efforts

2.3.1 SSURGO

Because the majority of the species of interest had distributions thought to be driven mostly by soil type, the Natural Resources Conservation Service (NRCS) Soil Survey Geographic Database (SSURGO) soil maps were investigated for their usefulness in developing quick predictive habitat models. Three separate SSURGO soil studies have been performed for different parts of the County by NRCS (NRCS 2007a, 2007b, 2009). No soil surveys had been performed for the northwest corner of the County by NRCS. This was not considered a concern because most of the species of interest were not known to occur in that corner of the County.

The DCP contracted with an independent soil scientist, Rick Van Remortel, to use SSURGO geographic information system (GIS) data to develop gypsum and sand coverages for the County based on both weighted averages and maximum values per soil map unit. In studying the results, several issues with using just SSURGO data were discovered. For gypsiferous soils, the soil surveys in the northeast portion of the County (NV608 – NRCS, 2009) tended to lump high gypsum soils in with badlands and did not report on their percent gypsum content. Therefore, the resulting gypsum map missed many of the map units that contained the gypsum plants in that area. The use of badland soils in addition to the percent gypsum map was investigated but this had the result of greatly over predicting gypsiferous soils in the area since the badland map unit type included many other types of soils such as sand dunes.

For the sandy soils, the known locations of species considered sand-loving plants occurred on a wide range of percent sand values. Thus use of this alone would have greatly over predicted the amount of possible habitat for these species.

2.3.2 Landsat Thematic Mapper Imagery

Because SSURGO by itself did not prove useful in developing initial habitat models, it was decided to investigate whether multispectral satellite imagery such as Landsat Thematic Mapper imagery (Landsat) could be used to identify gypsiferous and sandy soils. An initial visual inspection of Landsat images for the County with known plant locations overlain indicated that the imagery could prove useful in the development of the habitat models.

Based on the initial inspection of the Landsat imagery and the SSURGO “percent gypsum” map, DCP staff identified and visited several areas throughout the County. Specifically, an area in Gold Butte was identified that had no known species occurrences but SSURGO identified as having high gypsum content and the area visually matched other areas with known gypsum species occurrences on the Landsat imagery. Upon visiting the area, a population of Las Vegas buckwheat was encountered and documented. This reinforced the idea that multispectral satellite imagery could be used to help develop the initial predictive habitat models.

2.4 Creation of Sand and Gypsum Models

Because the DCP did not possess the software to analyze multispectral imagery, the DCP contracted TerraSpectra Geomatics (TerraSpectra) to further investigate the use of remote sensing imagery to identify areas of relatively higher gypsum and sand content.

An approach based on imagery from the Advanced Spaceborne Thermal Emission and Reflection (ASTER) satellite was selected for developing the gypsiferous and sandy soil maps. ASTER was specifically designed for geologic applications, with 14 spectral bands sensitive to specific regions of the electromagnetic spectrum in the visible and near infrared (VNIR), shortwave infrared (SWIR), and thermal infrared (TIR) . TerraSpectra acquired 14 individual ASTER scenes from 2004 (a drought year) and mosaiced them into a single countywide mosaic. Principal Components Analysis (PCA) was performed on the imagery using the VNIR and SWIR bands to create the gypsum classification; a PCA of the TIR bands was used to create a sandy soils classification.

Because sand is defined based on particle size, it cannot be mapped directly using remote sensing methods. Instead, TerraSpectra mapped quartz as a surrogate for sand since most of the sand in the County is silica and thus quartz based. Principal Component 2 (PC2) from the PCA of TIR imagery was found to be a good indicator of silica or quartz content. A binary threshold classification of PC2 for sandy versus non-sandy soils was performed using geologic maps of various scales and field visits to help control and then verify the classification.

For the gypsum classification, TerraSpectra used 20 gypsum training sites picked from known locations of gypsum plants, gypsum mining data from the United States Geological Survey (USGS) Mineral Resources Data System (MRDS) (USGS, 2005), site visits with DCP, and guidance from the DCP staff. These training sites were used to perform a supervised classification of the PCA results into gypsum and non-gypsum classes. Selected large scale (1:24k) geology maps were used to assist in the classification.

Because the ASTER imagery alone was not sufficient to uniquely identify sandy and gypsiferous soil models, the ASTER classifications were combined with geology maps and SSURGO soil maps (for the sand model) to create the final draft soil models. No single 1:100k geology map existed that covered the entire County so TerraSpectra had to combine six different geology maps with two different scales (1:100k and 1:250k) into a single composite map for the County. The map units on this composite map were then attributed as either gypsiferous, spring deposit, or non-gypsiferous, and high, medium, or low potential for sand-producing geologic units. This was done by reviewing the geologic unit descriptions from the individual map reports and by comparing the maps to selected 1:24k scale geology maps that contained information on gypsum bearing map units. For the gypsum model, spring deposits were used because they are sometimes gypsiferous and because of their known high correlation with the distribution of Las Vegas buckwheat.

The final gypsiferous soil model was then produced by combining the results of the ASTER gypsum classification with the compiled and attributed gypsiferous geologic map. The combination of the ASTER gypsum model and the attributed geologic map resulted in 6 possible combinations. These were: ASTER classed gypsum on gypsiferous geology; not ASTER classed gypsum on gypsiferous geology; ASTER classed gypsum on spring deposits; not ASTER classed gypsum on spring deposits; ASTER classed gypsum on non-gypsiferous geology; and not ASTER classed gypsum on non-gypsiferous geology.

Since many of the geologic units have the potential to be sandy, refinement of the sand model even further was desired. The SSURGO maps were recoded into whether a map unit had an area weighted average percent sand within the top 1 foot of soil above or below 75%. In addition, it was decided to recode geology maps into not just sand/non-sand but also their basic geologic type (e.g., younger alluvium, tertiary clastic bedrock, etc.) This allowed the model to have multiple potential sand unit types to choose from when developing the individual species predictive habitat models. The final sand model was a combination of the ASTER quartz classification, the geology maps coded into general geologic formation type, and the SSURGO percent sand map. This resulted in 28 possible combinations.

2.5 Creation of Species Predictive Habitat Models

Once the gypsiferous and sandy soil models were developed by TerraSpectra, the DCP used the models to create individual predictive habitat models for each of the species of interest. This was done by using GIS software to tabulate the number of known occurrences for each species by polygon type from either the sand or gypsum soil models. The polygon types were then classed as high, medium, low, or no probability of occurrence for each species (Tables 2-1 and 2-2). This was done by analyzing the number of occurrences per polygon type. The analysis did take into account the fact that the occurrence data is heavily biased by survey effort with some areas having hundreds of points due to intensive sampling efforts and other areas only having single records. In addition, many of the species had occurrence records on a wide range of soil types (or for the gypsum model, all soil types), including those modeled to not be gypsiferous or sandy soils. Thus, professional judgment was used in ranking suitability.

Table 2-1. Gypsum Species Predictive Habitat Models

Species	ASTER Classed Gypsiferous Geologic Unit	ASTER Classes Non-Gypsiferous Unit	ASTER Classed Spring Deposit	Not ASTER Classed Gypsiferous Unit	Not ASTER Classed Spring Deposit
Las Vegas bearpoppy	High	Medium	Low	High	Medium
Las Vegas buckwheat	Medium	Medium	High	Medium	High
Sticky ringstem	High	Medium	Low	High	Medium

Table 2-2. Sand Species Predictive Habitat Models.

Species	Not Quartz Classfied, Eolian	Not Quartz Classified, Mixed Eolian and Alluvium	Not Quartz Classified, Younger Alluvium, SSURGO sand >75%	Quartz Classified, Drought Exposed Lake Bed	Quartz Classified, Eolian	Quartz Classified, Mixed Eolian and Alluvium	Quartz Classified, Non-Clastic Bedrock, SSURGO sand >75%	Quartz Classified, Older Alluvium	Quartz Classified, Quartz Sand Veneer Over Calcrete	Quartz Classified, Tertiary Clastic Bedrock	Quartz Classified, Younger Alluvium
Threecorner milkvetch[1]	-	-	-	-	High	-	Medium	Medium	Medium	Low	High
Sticky buckwheat	-	-	-	High	Low	-	-	High	-	High	Medium
Pahrump Valley buckwheat	-	High	-	-	-	-	-	-	-	-	-
White-margined penstemon	High	High	High	-	-	High	-	-	-	-	Medium

1. *Model used for Beaver Dam Breadroot also.*

Because some species had similar predictive habitat models, only one model was created for each group of species. Models were created for Las Vegas buckwheat, Las Vegas bearpoppy (including sticky ringstem), white-margined penstemon, Pahrump Valley buckwheat, threecorner milkvetch (including Beaver Dam breadroot), and sticky buckwheat. For some species a low probability model was not created.

Predictive habitat models for white bearpoppy and yellow two-tone beardtongue could not be created using either the sandy or gypsiferous soil models. Yellow two-tone beardtongue is found in disturbed rocky or gravelly habitats, such as washes and along roads. White bearpoppy does occur on gypsum soils but also occurs on calcareous soils that were not modeled (NNHP, 2001). It also can occur in small outcrops of gypsiferous soil that are smaller than the spatial resolution of the imagery or minimum mapping units of the compiled geologic map and soil survey maps. Therefore, most of the known occurrences of white bearpoppy were not in areas predicted to contain gypsiferous soils based on the gypsum model created by TerraSpectra.

2.6 Field Survey Sample Design

Once the predictive habitat models for each species were created, the next step was to develop a stratified sampling design in order to survey areas throughout the County as potential suitable habitat but for which there were no known existing GIS occurrence data.

2.6.1 Division of County into Geographic Units

The first step in developing the sampling design was to split the County into 18 smaller geographic units to ensure sampling points were more evenly distributed throughout the County and to ensure some points occurred in each geographic unit. These geographic areas were not equal in shape and size but instead were based on the desired level of effort for each area. Areas for which access would be an issue due to land ownership or terrain were delineated as separate units. These included tribal lands, the urbanized Las Vegas valley, the Nellis Air Force Range, a portion of the USFWS wildlife refuge surrounded by Nellis Air Force Range property, and the Spring Mountains. This left 13 geographic units to be sampled.

2.6.2 Creation of Random Sampling Locations

Once the County was divided into geographic units, the DCP tasked TNC to use a Generalized Random Tessellation Stratified (GRTS) design to create the sample points for the field surveys. DCP directed TNC to stratify the sample points in each geographic unit with 70% of the points within a unit being in high probability habitat, 20% in medium, and 10% in low. Because the geographic units each contained different total area amounts for the different high/medium/low predictive habitat types for each species, it was decided to further stratify the number of sample points in each geographic unit by the amount of habitat within the unit. This was done by quantifying the acreage of predicted habitat within each geographic unit for each species. The geographic units were then grouped into either large or small units to be surveyed. Thus for the geographic units with a large amount of habitat, TNC, using GRTS, created 70 sample points in high, 20 in medium, and 10 in low probability habitat predicted by the species habitat models. For units with a small amount of habitat, TNC created 35 high, 10 medium, and 5 low. For species without a low probability model, only high and medium probability points were created. Due to a lack of a habitat type in some units for some species, not all species had sample points in all of the high/medium/low categories. To prevent sampling from known locations, a 4 hectare buffer was created around existing known locations and these buffered areas were removed from the models before creating the sampling locations with GRTS.

Once the set of potential survey locations were created by TNC, the DCP reviewed each point for access issues. In all areas except for the Sheep Range, points at high elevation or rough terrain were removed, mainly due to potential access issues. For the Sheep Range, elevation was not used as a filter because of the potential for white bearpoppy at higher elevations and the general overall higher elevations of the area. From the remaining set of survey points, the DCP selected the first 7 high, 2 medium, and 1 low predicted habitat points for each species group from each geographic unit to create the official set of points for the field surveys. Since not all species groups had points for all three probability levels in each geographic unit, some species groups did not have 10 survey points per geographic unit. For the Las Vegas buckwheat model, high probability habitat was originally spring deposits but because this habitat type was uncommon or small in most of the geographic units, placing a large number of sample points in this habitat type would have resulted in repeat sampling of the same area. Thus for the field surveys, the Las Vegas buckwheat model survey points were flipped so that 7 points were in the medium and 2 were in the high probability habitat type.

Although GRTS is supposed to ensure that the survey points are well distributed spatially, it did create overlapping points in habitat types that were rare or where few polygons were available to distribute the survey points. The final sample point pool consisted of a total of 547 points. The Gold Butte geographic unit received a double allocation of points for the Las Vegas buckwheat and Las Vegas bearpoppy models as part of additional funding to more intensively survey that geographic unit.

2.7 Field Surveys

Once the sample points were determined, DCP contracted and directed ICF Jones & Stokes (ICF) to conduct the field surveys. ICF was required to not only survey each location for all 10 target species but also looked for and documented any of the target species that were encountered while en route to the survey locations. These observations were documented as incidental observations. At each survey location, the DCP required ICF to record the physical attributes of the plot including basic soil type of the plot (e.g., sand, gypsum, rocky, etc.) and the predominant slope and aspect. ICF was directed to document the survey location through photo points. ICF also documented the presence of other non-target gypsum or sand species that could be used as indicators of potential suitable habitat for the target species. The list of indicator species was provided by the DCP. Detailed information of the data recorded at each plot and for each target species occurrence can be found in the Data Management Plan, Work Plan, and Final Report submitted by ICF to the DCP as part of that contract (ICF 2009a, 2009b, 2010).

2.7.1 Field Survey Results

Table 2-3 documents the field survey results for the surveys conducted in 2009 and 2010 by ICF. All ten of the target species were encountered.

Table 2-3. Field Survey Results.

Species	Plot Observations (Plots[1])	Incidental Observations	Total
Sticky ringstem	2	3	5
Las Vegas bearpoppy	20 (17)	18	38
White bearpoppy	9 (8)	4	13
Las Vegas buckwheat	1	2	3
Threecorner milkvetch	6	2	8
Pahrump Valley buckwheat	3	0	3
Sticky buckwheat	5	1	6
Beaver Dam breadroot	16 (14)	9	25
White-margined penstemon	1	0	1
Yellow Two-tone beardtongue	2	2	4
Total Number of Observations	65 (56)	41	106

1. The number in parentheses represents the number of different plots with at least one observation. One plot could have more than one observation per species or more than one species.

Table 2-4 shows the occurrences broken down by their high, medium, and low model levels and provides a different way of looking at how well the models performed. The table includes both the plot and incidental species observations. The survey points were provided to ICF without the descriptive information that would have shown what species was targeted for a given survey plot. Within each survey plot every target species was surveyed for regardless of habitat type. The results in the “non” category of Table 4 describe occurrences documented in habitat not included in the species specific predictive habitat model. When analyzing the results in the table, it should be kept in mind that 70% of the survey locations were in the high categories and that more occurrences in this category could be simply due to survey effort. The results, though, do support the use of the high, medium, and low levels to help maximize the probability of finding the target species.

Table 2-4. Survey Results by Model.

Species	High	Medium	Low	Non
Sticky ringstem	4 + 1 edge[1]	0	0	0
Las Vegas bearpoppy	16 + 3 edge	6 + 6 edge	5	2
White bearpoppy	8 + 2 edge	2	0	1
Las Vegas buckwheat	2	1	0	0
Threecorner milkvetch	3 + 1 edge	3	0	1[2]
Pahrump Valley buckwheat	1	0	0	2[3]
Sticky buckwheat	5	0	0	1[3]
Beaver Dam breadroot	12 + 1 edge	8 + 2 edge	1	1[4]
White-margined penstemon	1	0	0	0
Total Number of Observations	52 + 8 edge	20 + 8 edge	6	8

1. Edge indicates that an occurrence was documented just outside of a polygon of the corresponding habitat type and due to GPS accuracy or model resolution, likely occurred in that habitat type.
2. Occurrence was documented in a high habitat type for sticky buckwheat.
3. Occurrences were documented in gypsum habitat.
4. Occurrence was documented just outside of a high habitat polygon for white-margined penstemon.

This page intentionally left blank.

3.0 CURRENT PROJECT

3.1 Purpose

As stated above, the original draft predictive habitat models were intentionally left broad to prevent biasing the models toward what was already known about the species distribution and natural history. The goal of the current project is to inform resource managers as to the location of suitable and potentially suitable habitat for the species within the County by creating new spatial habitat models.

3.2 Review of Draft Predictive Habitat Models

The draft predictive habitat models from the DCP did prove useful in discovering new populations of all the target species and thus they fulfilled their purpose. From an accuracy standpoint, though, the models were over-predictive (as expected) as can be seen where only 56 out of the 547 total sites had target species found. It must be kept in mind though, that due to cost limitations only a small percentage of the County that was predicted as habitat was actually surveyed. Surveys were also not conducted within a 4 hectare buffer around previously known species occurrence records. In addition, although a target species was not observed, the habitat predicted by the model may have still been present. This was the purpose of requiring ICF to collect information on the soil types and any indicator species present at each location.

One of the main weaknesses of the draft models was that there were no geologic maps that covered the entire County at a large enough scale to be useful (1:150k or lower). Therefore, the models were composed of portions of several different geology maps with different mapping schemes and resolutions. This was especially true for the southeastern portion of the County where the 1:250k statewide geology map was used because no larger scale maps were available. This caused that area of the County to be greatly over-predictive for sand habitat.

In addition, there are three separate SSURGO soil maps for the County and no soil survey has been completed for the northwestern portion of the County. The soil surveys were completed at different times and thus have differences in the quality of the mapping and the amount of specific soil information available. The draft predictive sand models did use SSURGO sand data as a factor in the models but because no data was available for the northwest corner, this area was less detailed than other areas of the County.

For the gypsum models, the lack of cohesive geologic maps for the County also created issues. Spring deposits were determined to be an important geologic unit for the gypsum plants but this unit has not been mapped at the same level of detail in the geologic maps used. Due to the coarse scales used for some of the maps, many smaller gypsiferous geologic units were not mapped. In addition, many large geologic units were categorized as gypsiferous when only a small portion of the entire unit may actually be gypsiferous. This results in an over-prediction of the amount of gypsiferous soils in some areas.

The original models also did not incorporate the known elevation range of the species as is typical in habitat modeling exercises (Aitken et al, 2007; Boykin et al, 2008). The elevation range was purposely left out to not bias the survey efforts to only the previously known elevation ranges.

3.3 Model Refinement Options

3.3.1 Maxent

One of the common habitat modeling software tools in use today is Maxent (Phillips et al, 2006). With Maxent, the user provides a table of known occurrence locations along with gridded data that can be either continuous (e.g., elevation, climate) or categorical (e.g., vegetation). Maxent overlays the points on the background layers to produce the habitat model. The software uses all the input data layers provided to create the model but does report on the effect of each layer on the model. Thus the models tend to be run iteratively by the user, eliminating layers in each subsequent run that do not have much influence on the model. The decision about a background layers importance, though, is up to the user and can be highly subjective.

For this modeling exercise, the use of categorical data in Maxent can produce over-predictive models due to issues with the resolution and accuracy of the GIS data along with the accuracy of the occurrence data. As an example, assume that a species is known to exist only on soil type A and not on soil type B. Mapping resolution or errors in the GIS soil map or species occurrence records can lead to some occurrence records falling within areas mapped as soil type B. Maxent does not know these are errors and thus will produce a habitat model that includes both soil types. Therefore, Maxent was determined to not be useful to refine the soils based habitat models.

3.3.2 Other Modeling Options

Because Maxent alone was not a desirable modeling option for this exercise, it was decided to instead refine the models through manual selection of suitable polygons based on their attributes for a variety of factors similar to how the original predictive habitat models were created. Since the creation of the original predictive habitat models, a 1:150k surficial geology map for the County has been created by the University of Nevada Reno (UNR) under contract with the DCP (House et al, 2010). We reviewed the use of the new geologic map in conjunction with the original predictive habitat models to see if it would help refine the original models. Unfortunately it did not provide much improvement to the original models by itself. The SSURGO soil maps were also reinvestigated to see if they could be used to improve the models but were found to still be difficult to use. The weaknesses in the original predictive habitat models were still present.

One option was to refine the original sandy and gypsiferous soil models developed by TerraSpectra. The field survey results would provide new training locations and absence information that could be used to refine the original ASTER classifications. In addition, the refined ASTER classifications could be combined with the new UNR geologic map instead of the multiple maps used originally. Because this appeared to be the best option, the decision was made by the DCP to contract with TerraSpectra again to refine the original sandy and gypsiferous soil models.

3.4 Refinement of Sand and Gypsum Models

The following sections summarize the methods used by TerraSpectra to refine the gypsum and sand soil models that were used in the creation of the EPG refined habitat models. For a more detailed description of their methods, refer to the metadata provided with the sand and gypsum models (TerraSpectra, 2011).

3.4.1 Sandy Soil Model

3.4.1.1 Geology Refinement

As mentioned earlier, the recent completion of the UNR 1:150k surficial geologic map for the County was the impetus to refine the sand model produced by TerraSpectra. It was determined that the eolian class on the new geologic map was highly spatially consistent with known locations of "sandy" plants. Therefore, the geologic map was attributed with two classes: Non-eolian = 0; Eolian = 5. Sand from these eolian deposits can be blown up onto neighboring non-eolian geologic units. This un-mapped blown sand can provide habitat for the species of interest. Therefore, it was decided to create a 100 meter buffer around the eolian geologic units that was assigned a code of 2. In addition, the known locations of "sandy" plants were found to be highly negatively spatially correlated with bedrock, hillslope deposits, and the Lake Mead Reservoir (2009 level). Thus an additional field was added with the following codes: Bedrock, Hillslope, or 2009 Lake Mead = 0; All other geologic units = 1. Unlike the first sand model where multiple geologic units were considered “sandy”, for the refined model, only the eolian geologic unit type and the 100 meter buffer around these units were considered “sandy” based on the geologic map.

3.4.1.2 SSURGO Refinement

Similar to the first sand model, SSURGO was used as one of the inputs but this time with a different coding scheme. The area weighted percent sand within the top one foot of soil was again used as the metric. Based on this metric, the SSURGO maps were recoded into 3 classes: Less than 80 percent sand = 0; Between 80 and 90 percent sand = 1; Greater than 90 percent sand = 2.

3.4.1.3 ASTER Refinement

ASTER imagery was again used to identify quartz as a surrogate for sand. The same 14 ASTER images used in the creation of the gypsum model were used for the new sand model. Similar to the methods used in producing the original sand model, PCA was performed on the VNIR and SWIR image bands and the results were used in a supervised classification. Known locations of sand plants and eolian geologic units were used to help determine threshold levels for the classification. The classification resulted in three codes: Strong quartz presence likely = 2; Moderate quartz presence likely = 1; Insufficient quartz presence likely = 0. Refer to the metadata provided by TerraSpectra for a more detailed description of how the ASTER imagery was analyzed (TerraSpectra, 2011).

3.4.1.4 Combined Sand Soil Model

The recoded geologic, SSURGO, and ASTER classification maps were then combined into a single map. The codes for each of the maps were then totaled to produce an overall "sand" score ranging from 0 (not sandy) to 9 (very high potential for sand). The resulting total was then multiplied by the bedrock code from the geologic map. This had the effect of converting the "sand" score to zero in areas of bedrock, hillslope, and the 2009 Lake Mead basin regardless of the SSURGO or ASTER code. Because the ASTER map is actually a map of quartz, areas of sandstone receive a code of 1 or 2 even though they are actually solid rock. The bedrock code of 0 therefore has the effect of converting the sandstone units back into being non-sand units.

3.4.2 Gypsiferous Soil Model

3.4.2.1 Geology Refinement

Because the presence of gypsum in the soil is largely a function of the bedrock geology of an area, a surficial geologic map such as that produced by UNR of the County is not as helpful in mapping gypsiferous units. Unfortunately, no medium or large scale bedrock geologic maps exist for the entire County. Thus, similar to the original gypsum model produced by TerraSpectra, a combination of existing geologic maps was used to identify potentially gypsiferous geologic units. For this refinement, they also used large scale (1:24k) scale geology maps where available in GIS format. Therefore, for the refined model, six total medium (1:100k) and small (1:250k) scale geologic maps along with thirty large (1:24k) scale geologic maps were used to identify potentially gypsiferous geologic units throughout the County. For a list of geologic maps used, refer to the metadata provided by TerraSpectra in the gypsum soil model. The maps were merged with priority given first to the large scale (1:24k) maps, followed by the medium scale (1:100k) maps, and lastly by the small scale (1:250k) Geologic Map of Nevada (refer to the metadata provided by TerraSpectra for specific map citations).

The geologic map units for the compiled map were then recoded into the following classes: gypsiferous (unit was described as gypsiferous); gypsiferous, partially or locally (unit was described as partially or locally gypsiferous); spring deposit (spring or ground-water discharge deposit not described as gypsiferous); spring deposit, gypsiferous (spring or ground-water discharge deposit that was described as gypsiferous); spring deposit, gypsiferous, partially or locally (spring or ground-water discharge deposit that was described as partially or locally gypsiferous); other (unit not described as gypsiferous). TerraSpectra then assigned a numeric gypsum score to each map unit as follows: Gypsiferous = 2; Gypsiferous, Partially or Locally = 1; Spring Deposit = 1; Spring Deposit, Gypsiferous = 2; Spring Deposit, Gypsiferous, Partially or Locally = 1; Other = 0.

3.4.2.2 ASTER Refinement

The same ASTER imagery that was used in the creation of the initial gypsum model was used for the refined version. A mosaic of 14 ASTER images from 2004 was used to create a decorrelation enhancement image. This decorrelation enhancement image was used as the input for a supervised classification to identify potential gypsiferous areas. The compiled geologic map, the location of past and present gypsum mines from the USGS MRDS (accessed online 9/23/08), and two field trips conducted during the creation of the initial gypsum model were used to identify thirty gypsum training sites. The large scale (1:24k) geologic maps were used to help verify and control the classification. The result was an image classified into three possible values: Non gypsiferous = 0; Somewhat likely gypsiferous = 1; More likely gypsiferous = 2.

3.4.2.3 SSURGO Refinement

Even though SSURGO was used as a tool in the initial gypsum model for identification of potential gypsum habitat, it was not used as an input in the first gypsum model produced by TerraSpectra. Because SSURGO may identify areas with high gypsum that are not identified in the geologic maps or identified by the ASTER imagery, it was decided to add SSURGO as an input into the new refined gypsum model. The combined SSURGO maps were recoded into the following classes based on the area weighted amount of gypsum within the top one foot of the soil: Gypsum greater than or equal to 23 percent = 2; Gypsum greater than or equal to 5 percent and less than 23 percent = 1; Gypsum less than 5 percent = 0.

3.4.2.4 Combined Gypsiferous Soil Model

The recoded geologic, ASTER, and SSURGO maps were then combined to form the final refined gypsum model. The codes from each of the maps were totaled to produce an overall “gypsum” score ranging from 0 (non-gypsiferous unit) to 6 (highly likely gypsiferous with a score of 2 from each input layer). Playas and playa fringes in the County are often gypsiferous but are not considered potential habitat for any of the plant species of interest in this study. Therefore, playas and playa fringes were given a code of -1 and were considered non-gypsiferous in the subsequent habitat modeling.

3.5 Refinement of Individual Species Habitat Models

Once TerraSpectra refined the sandy and gypsiferous soil models, the next step was to use those new models to create refined predictive habitat models for each of the rare plant species.

3.5.1 Sand Species Models

For the sand species, each model was first refined using the known or expected elevation range within Nevada for each species. This information was gathered from the Nevada Natural Heritage Program (NNHP) Nevada Rare Plant Atlas (NNHP, 2001) and the LERPCMS. The elevation ranges were compared with the available occurrence data including the new data from the ICF 2009-10 field surveys to verify that the elevation range would encompass all the occurrences within the County. A slightly broader range was used by rounding up or down the elevations. Elevation was not used to clip the models. Instead, polygons were only removed if the entire polygon was outside of the elevation range. If the elevation range of the polygon overlapped with a species elevation range, it was retained in the model even if most of the polygon was outside of the elevation range. Clipping the polygons with elevation would have resulted in many sliver polygons at the edges of the elevation range which was undesirable.

The occurrence data for each species was then intersected with the sand model to analyze any patterns that could be used to help further refine the models. This included identifying any geologic, ASTER, or SSURGO score or combination of scores that did not have any known occurrences for a species. These categories were assumed to not be suitable habitat and thus removed from the individual species predictive habitat model.

Some categories or combinations only had a few occurrences. The occurrences in these categories were investigated further to analyze whether to retain these categories in the model or not. Some of these occurrence records were at the edge of a polygon, next to a polygon type with several other occurrence records. If so, it was concluded that these edge occurrences were most likely in the habitat type of the neighboring polygons and that the category could be removed from the model. Some of the occurrence records were of questionable accuracy and were dropped which may have led to the category being removed from the model also. If the occurrence data could not be explained by edge effects or location errors, the categories were retained in the model even if they only contained a few occurrences. Thus professional judgment was used in determining which categories from the sand model could be removed to further refine the individual species predictive habitat models.

Besides elevation and the sand model coding, other SSURGO soil attributes and geologic unit codes were investigated for their usefulness in refining the models. For SSURGO, the use of soil series name, percent of surface fragments, and percent of rock within the soil was analyzed. Except where discussed below, the use of these other factors did not prove useful in refining the models. This was due to the species occurring on a wide range of the possible types.

3.5.1.1 White Margined Penstemon

The elevation range used to refine the white margined penstemon model was 800 to 1100 meters. The model was then refined by removing all polygons that did not cross a SSURGO mapunit having either a Bluepoint, Birdspring, Commski, or Prisonear soil series type within the unit. These soil series tend to have a low surface rock fragment cover and a sand content greater than 80 percent. The majority of known occurrences in the County occur on Bluepoint, Birdspring, or Prisonear soil types with other locations occurring just on the edge of these soil types. The occurrence data available for the Nye County population are on Bluepoint and Commski soils. Finally, any polygon that had only an ASTER quartz signature and was not indicated by geology or soils to be sandy was removed. The known occurrences were only on areas indicated by the geologic or SSURGO classifications to be high sand.

3.5.1.2 Threecorner Milkvetch

The threecorner milkvetch model was initially refined using an elevation range of 325 to 750 meters. The model was then refined by removing polygons with the following attributes: areas with sand between 80-90% but an ASTER signature below 2; eolian areas with sand below 80% and no ASTER quartz signature. When only part of an eolian unit met the above criteria but another part had an ASTER quartz signature of 2, the whole unit was still retained. Essentially these cuts removed areas that are lower in sand content, as determined by SSURGO or ASTER classifications, than the areas with known occurrences.

3.5.1.3 Pahrump Valley Buckwheat

The Pahrump Valley buckwheat model was refined with an elevation range of 700 to 860 meters. The model was further refined by eliminating all non-eolian areas and then further eliminating those eolian areas with an ASTER signature and areas mapped by SSURGO as having sand content above 80%. It was then further refined by removing areas with greater than 23% surface fragment cover. Essentially, while Pahrump Valley buckwheat occurs on sandy soils, those soils appear to be not as sandy as the soils for the other sand species. It also appears it may be on sand that isn't dominated by quartz and does not have a strong ASTER quartz signature. The difficulty in modeling this species is that the majority of its occurrences are in Pahrump and Stewart valleys in Nye County, Nevada, and neighboring Inyo County, California, outside of the Clark County study area.

3.5.1.4 Sticky Buckwheat

The sticky buckwheat predictive habitat model was initially refined using an elevation range of 360 to 715 meters. The ICF surveys found a potential occurrence of the species along the upper reaches of the Muddy River at a higher elevation than the published range. Thus, the elevation range for this species was expanded to include this potential occurrence. This occurrence has not been verified and may be a misidentification. The model could be refined using a narrower elevation range in the future if this occurrence turns out to be a misidentification. Similar to threecorner milkvetch, the sticky buckwheat model was further refined by removing areas with sand between 80-90% but ASTER below 2 and eolian areas without an ASTER signature. Again, this removes areas lower in sand content as determined by SSURGO or ASTER. Finally, areas with a geologic type of Avxk (as defined in House et al, 2010), or sand sheets over calcrete, were removed. This Avxk unit only occurs on top of Mormon Mesa and was not included in the original model. Sticky buckwheat seems to occur in sandy areas mostly along large drainages and is possibly gaining habitat along the margins of Lake Mead as the water level recedes.

3.5.1.5 Beaver Dam Breadroot

The Beaver Dam breadroot model was first refined using an elevation range of 390 to 750 meters. The NNHP reported range includes a record for breadroot at 1524 meters on the north slope of the Spring Mountains that was collected in 1964. This record is questionable due to its disjunct nature from the rest of the known occurrences. There is another species of *Pediomelum* in the Spring Mountains (Niles and Leary, 2007) and this record may be an individual of that species which was misidentified. This occurrence and its elevation were not used as part of the analysis. Instead, the upper elevation was determined using the other known occurrences. The model was then refined to remove those areas without an ASTER signature and non-eolian areas with an ASTER signature but a sand content of 80-90%. Thus the model depicts areas with high levels of sand as indicated by the ASTER signature and the SSURGO sand content.

3.5.2 <u>Gypsum Species Models</u>

Like the sand species models, the first level of refinement with the gypsum species models was with elevation range. Unlike the sand species models, the use of SSURGO gypsum codes or the ASTER gypsum signature codes were not helpful in refining the models further. In addition, because the gypsum soil model includes geologic units from multiple sources, the use of the geologic unit codes to refine the models is problematic because the map sources did not use a standardized code system. Thus, similar geologic units that occur on multiple source maps may have different codes. Other than the Las Vegas buckwheat model, the other models were not further refined using the geology, ASTER, or SSURGO layers. The exception to this is the small polygons in the southeastern portion of the County that had only an ASTER gypsum signature and were not near any gypsiferous geology or soils. These small polygons were removed from all of the models.

3.5.2.1 Las Vegas Buckwheat

The Las Vegas buckwheat model was first refined using an elevation range of 570 to 1180 meters. The model was further refined by eliminating polygons with a gypsiferous geologic score of 2. The majority of the known Las Vegas Buckwheat occurrences are on spring deposit geologic units with the others on geology with a gypsum score of 0 or 1. Based on observations, many of the occurrences are in the run-off from higher gypsum content areas indicating that the species may not be able to tolerate high gypsum contents. There may also be some other soil chemistry reason the species prefers spring deposits and that condition may only be present in a few non-spring deposit units.

3.5.2.2 Las Vegas Bearpoppy, Sticky Ringstem, White Bearpoppy

The Las Vegas bearpoppy model was refined using an elevation range of 300 to 1120 meters. The sticky ringstem model was refined using an elevation range of 360 to 725 meters. The white bearpoppy model was refined using an elevation range of 600 to 1920 meters. The white bearpoppy model, as discussed in more detail below, does not capture most of the species locations in the County which was true for the initial predictive habitat models also.

3.6 Model Discussion

Overall, elevation was the biggest factor in helping refine the models from the general soil models. The initial models purposely did not include elevation as a factor so as to not bias the models to only those elevations where surveys had been conducted previously. Elevation, although not directly used in the creation of the initial models, was incorporated as a filter for most of the geographic units to remove high elevation locations. One occurrence of sticky buckwheat, pending positive identification, was found outside of its published elevation range. The rest of the occurrences during the ICF 2009-10 surveys were within the previously published ranges.

There may be other factors that could have been used to help refine the models. For example, there may be other soil parameters within the SSURGO database that could be used to refine the models for specific species. The sand models used an area weighted average of the amount of sand within the first foot of soil. Depending on rooting depth, the first foot of soil may not be the best determinate for each species. Some may only require 6 inches or less of sand on the surface due to shallow roots while other may require even deeper sand due to deep roots. The investigation of additional SSURGO soil factors or other factors was out of the scope of this model refinement contract.

As with any GIS based habitat modeling, the quality of the models is heavily dependent on the accuracy and quality of the input GIS data used to create the model. Some of the over or under predictions in each of the models may be due to errors or scale issues in the geologic or soils maps used as input layers in creating and refining the models. The accuracy of the occurrence data can also lead to problems when refining the models and interpreting their quality.

3.6.1 Sand Plant Models

The use of the UNR surficial geologic map was an improvement over the initial draft models use of multiple geologic maps. The use of SSURGO soils and ASTER information in addition to the geology helps to identify other potentially sandy areas not mapped as eolian. There may still be some issues with using SSURGO because there are three separate soil series that cover only part of the County. Some of the refinements may be artifacts of differences in the soil surveys and not true differences. In addition, the lack of a soil survey for the northwestern portion of the County does affect some of the models. For white-margined penstemon, the model had suitable habitat in the northwestern corner due to the lack of soil survey information that could be used to refine the model based on soil type. NRCS is currently conducting a soil survey over a portion of the Sheep Range which could be incorporated into the model to help refine the soil models in the future.

ASTER indicates stronger quartz signatures in the northern portion of the County versus the southern portion. It is unknown whether the sands in the northern portion of the County are higher in quartz content than those in the southern portion or whether this difference is due to other factors such as vegetative cover, overall reflectance of the soils, ruggedness of the areas, barren rock outcrops versus soil exposure, etc. The causes of the differences in the ASTER signature, in addition to elevation, may be a clue to differences in the distribution of the species.

3.6.1.1 Gypsum Plant Models

Because gypsum is a function of the bedrock type, the UNR surficial geologic map of the County was not useful in refining the model except for the consistently mapped spring deposits. The use of multiple sources for geologic maps still leads to the inability to refine the models based on actual geologic unit types which would be ideal. For example, there are several geologic units west of Las Vegas valley that are coded as gypsiferous and indeed, there are gypsum mines in some of these areas. These hills, though, may be a different geologic unit type than the gypsiferous units where the species occur since the gypsum on these units is mostly under the surface and not exposed. Since the coding between geologic maps is not uniform, these units cannot be excluded by geologic code. It would probably take significant effort to standardize the coding between all the maps. The use of 1:24k geology maps improved the models over the original models by adding more detail. Unfortunately, only a subset of the County has digital 1:24k geology maps available even though some of these maps cover areas of the County with a large area of gypsum habitat such as Rainbow Gardens. Ideally, as additional 1:24k geology maps become available, they could be incorporated into the gypsum model.

Even though the EPG models do better than the original predictive habitat models, the new models still miss some of the known occurrences for each of the species. These species, especially the bearpoppies, can occur in very small patches on geologic units that are too small to be mapped at the scale of even 1:24k. Some of the patches are also small enough to not be picked up by the ASTER imagery. The species are also known to be found on units that are not necessarily considered gypsiferous but gypsum is likely present in the soil due to erosion from the neighboring gypsiferous units. This can be seen by the population of Las Vegas buckwheat in Gold Butte that does not occur on what the model calls gypsum but is surrounded by gypsiferous units that are probably eroding down into the wash. If desired, this could be dealt with by buffering the models to help capture these additional erosion based areas.

Similar to the draft model, the new model for white bearpoppy is still weak and misses the majority of the known occurrences. This species can occur on very small outcrops and can occur on not only gypsum but also carbonate soils (NNHP, 2001). Thus a strictly gypsum based model is probably not the best approach. In addition, this species range extends into Lincoln and Nye County, Nevada and Inyo and San Bernardino County, California. Thus only a small portion of its overall range is being modeled.

3.7 Addition of Survey Attributes

Once the models were completed, each model was attributed with information on whether a polygon had a known occurrence within the polygon including the results of the ICF rare plant surveys, whether there was a known occurrence within 10 meters of the polygon, whether a ICF rare plant survey was done in part of the polygon and the species was not found, or whether a polygon had no known occurrences and no known survey has been conducted within the polygon. These attributes should not be assumed to mean that the species is present throughout the whole polygon or that the species is definitely absent from the polygon. In addition, the ICF rare plant surveys may have only surveyed a very small portion of the polygon and not the entire polygon. The presence data is only as good as the input occurrence data and therefore should not be construed to mean definitive presence. Because of the unknown positional accuracy of some of the occurrence records and the fact that some points actually represent a population of the plants over a larger area, the “occurrence within 10 meters” was added to identify areas near occurrence records that are likely to have the species also.

These attributes could be useful for the users of the models because it provides information on where in the model the species is known to occur and what areas of the model may warrant additional survey efforts before management decisions are made. As additional surveys are performed, new occurrence records or “absence” information should be added to the models to keep the models current.

3.8 Assessment of Model Accuracy

The habitat models are designed to depict predicted suitable habitat for each of the species and not necessarily depict only the actual occurrence of a species. As such, the models are expected to depict more area than just where a species occurs. In addition, the models do not necessarily identify all areas where a species occurs as being suitable habitat due to issues with the scales of the maps and any errors that might be present in the maps used to create the sand and gypsum models. These issues should be kept in mind when trying to discuss the accuracy of the habitat models.

The attribution of the models with whether there are known occurrences or whether partial surveys have been performed without finding the species does give some idea as to how much the models may overpredict the habitat for a species. The absence information should be used with caution, because rarely has the whole polygon been surveyed and absence during one survey does not definitively indicate that the species is not present in other years (i.e., annual species may not be found in dry years, but the seeds are still present in the soil). Also, absence does not mean that the area is not suitable habitat. The area may be suitable habitat but the species is not present due to other reasons such as the seeds not having reached the area.

3.9 Climate Based Models

Some of the sand and gypsum predictive based models still appeared to overpredict the geographic extent of possible suitable habitat for some species and the white bearpoppy model is a poor predictor of the species occurrence in general. In addition, one of the original species of interest, two-tone beardtongue, is neither gypsum nor a sand species. Instead, it occurs mostly in rocky and gravelly drainages and along roadsides, therefore a gypsiferous or sandy soil-based model was not created.

Another form of habitat model is one created using climate data to model the climatic ranges of the species occurrence. These models are often created in Maxent using climate variables modeled for the whole world at around an 800 meter pixel resolution (Worldclim Bioclim data – Hijmans et al, 2005). Since the models are relatively quick to run, a climate model was created for each species, including two-tone beardtongue, using Maxent. See Appendix C for the outputs from the Maxent runs. The grid outputs from Maxent were recoded into 1 (suitable) or null (not suitable) using the Maxent computed threshold of equal specificity and sensitivity (Boykin et al, 2008). The recoded grids were then converted into polygons representing the modeled climate based distribution.

The climate models included in Appendix C were run using all the available occurrence data for each of the species, including occurrence data outside of the County. This occurrence data, though, may still not represent the full range of the species occurrence and thus the climate models may not reflect the full range of a species' climatic limits. The drawback in using climate data is that using points from a subset of a species range versus its entire range can produce very different results. The Maxent results from using just the County occurrences for white-margined penstemon are provided at the end of Appendix C. As one can see, the predicted suitable habitat using just County data is much smaller than the predicted suitable habitat using all the data. Thus, not using data over the entire distribution does not accurately capture the species full climatic range.

For the species with sandy or gypsiferous soil-based models, the climate based models could be used to clip those models to produce a model representing both the soil based and climate-based suitable habitat. Most of the soil-based models already match fairly well with the climate models with only some outlying polygons. This is understandable because elevation was used as a filter in the soil based models and climate is highly correlated with elevation at a local scale. The species with larger differences between the soil and climate based models tended to have much wider distributions than just in the County. For white bearpoppy and two-tone beardtongue, the climate based model by itself may be the best model choice due to reasons stated previously. A better soils and geology based model may be able to be created for white bearpoppy by expanding the geographic scope of the model and also looking at carbonate soil types. It is unknown if a soils or geology based model could be created for two-tone beardtongue.

4.0 GENERAL DISCUSSION

There is a great deal of uncertainty in resource management. Resource managers are faced with making landscape level decisions everyday with limited information, data, and funding. Models are used as way of reducing the uncertainty when making decisions. As is well known models never reflect reality. George Box wrote “All models are wrong, but some are useful.” They are meant to be used as a tool to provide understanding of the complexities of the topic of interest. In that mode, a predictive habitat model provides managers with a tool to better survey for, manage, and make decisions about a species for which the actual, complete distribution is not known. A habitat model also provides a starting point when studying the natural history of a species.

The original draft models developed by the DCP and TerraSpectra were useful in finding new populations of the target species, and were an improvement over the TNC buffered populations. The way the models were developed showed that using soils and geological information from multiple sources was helpful and that no one single input layer was sufficient in modeling the plants habitat. The weaknesses in the original and updated habitat models do point out the need for things like a really good geologic map or really good soils maps. The soils and geologic maps must also map the soil types of interest. SSURGO maps are mostly focused on mapping soils suitable for agriculture and thus in some cases do not capture the soil characteristics needed for mapping rare plant habitat such as percent gypsum. In addition, a detailed surficial geologic may be useful for identifying sandy soils but is not as useful in identifying soil types more influenced by the bedrock such as gypsum. The reverse is true for trying to use detailed bedrock geology maps to identify surficial soil types like sand. The modeling efforts also showed the difficulties in trying to model species habitat within a political boundary without the ability to take the full species distribution into account.

The modeling efforts were based upon the idea that the species, except for two-tone beardtongue, either occurred on gypsiferous or sandy soil. Describing plants as gypsum or sand plants is an oversimplification of the natural history of the plants, which adds difficulty in the. The best example is Pahrump Valley buckwheat where it is considered a sand plant (occurring primarily in sand) but may require some level of clay content also, and not just pure sand. For two-tone beardtongue, creating a soil based habitat model would present its own challenges and would likely be very overpredictive. As mentioned, two-tone beardtongue is found on rocky and gravelly soils. In the County, these soils are very common and can probably be found in the majority of the SSURGO soil or geology map units.

The importance of absence data was also highlighted during the project. Historically surveys were conducted to only collect data on species presence. It is important to not only collect presence data but also absence data. Documentation of a species absence provides a greater understanding of the species full distribution and natural history. The Las Vegas buckwheat has historically been characterized as a gypsum plant, but through the modeling effort, more of the locations were found to on spring deposits and the species was found to be absent from many non-spring deposit areas with gypsiferous soils. Documented absence data or a full species list for an area can provide guidance to management when making decisions about an area thought to have suitable habitat for a given species. Absence data also provides documentation that an area has been surveyed instead of just relying on institutional knowledge about where a species does or does not occur. This institutional knowledge is easily lost as employees change jobs or retire.

The surveys performed as part of the DCP project did provide documentation of where the surveys were conducted and whether a species was observed or not. This type of documentation is not necessarily an easy task. Most species occurrence data is documented in GIS as point dataset but absence data is not point based. Documenting absence requires documenting the area where the surveys were conducted, preferably as a polygon, including documenting all species present. In addition, absence data does not automatically mean a site should not be surveyed again. In the case of annuals (and some perennials), a species may be found one year and not the next due to rainfall events, drought, and survey timing compared to plant phenology. We can and should do a better job of tracking absence data.

This project also highlights the need for reliable occurrence data where the positional accuracy of the data and the methods used to collect the data are known. This project used occurrence data from multiple sources with different levels of quality. The data sets from multiple projects were used because no single dataset represented the full distribution of any of the target species. Some of the datasets contain highly accurate location information for individual plants. Other datasets contain single points representing entire populations. Some datasets, though, contain points for which it is not known whether they represent individual plants or populations. The positional accuracy of these datasets can also be coarse or even unknown at times. In other words, is the point accurate within a mile, 5 meters, or 2 meters? For example, NNHP creates polygons around the points or populations which are buffered by the positional uncertainty. This makes developing habitat models with the use of the occurrence data challenging at times. A few occurrence points for some of the species were not used in the analysis due to their questionable location. The use of inaccurate occurrence location data can lead to the inclusion or removal of habitat types in the model that can lead to the over or under prediction of potential habitat.

4.1 Potential Future Improvements to Models

There are multiple options for improving the habitat models in the future if desired. The gypsum habitat models are still composed of geologic maps from multiple sources and scales. As more 1:24k geology maps become available, these could be incorporated into the model. These maps provide more detail than the smaller scale maps and also provide more information on potentially gypsiferous geologic units. Another option for improving the gypsum models in the future would be to try to standardize the geologic coding schemes used in the different maps and then identify the potentially gypsiferous units that are also potential habitat for the species of interest. Not all gypsiferous geologic units are necessarily potential habitat for a “gypsum-loving” species. The gypsum may not be at the surface such as the gypsiferous geologic units west of Las Vegas valley where there are gypsum mines but most of the gypsum is not at the surface of the soil. Other gypsiferous units may be too high in gypsum for some species. The coding of the gypsiferous units needs to be standardized to ensure similar geologic units are being compared between maps.

For the sand models, it may be possible to identify other soil parameters from the SSURGO surveys that could be used to further refine the models. In addition, when the soil survey for the Sheep Range is completed, the map could be incorporated into the model to help fill in where SSRUGO soils information is missing.

As mentioned earlier, the models should be updated as new occurrence or absence data becomes available. This will ensure that management decisions that are made using the models are using the most current data available. The new occurrence and absence data will also allow for continued improvement of knowledge about a species distribution and natural history.

This page intentionally left blank.

5.0 LITERATURE CITED

Aitken, M., D.W. Roberts, and L.M. Shultz. 2007. Modeling Distributions of Rare Plants in the Great Basin, Western North America. Western North American Naturalist 67:26-38.

Boykin, Kenneth G., David F. Bradford, and William G. Kepner. 2008. Habitat Distribution Models for 37 Vertebrate Species in the Mojave Desert Ecoregion of Nevada, Arizona, and Utah. U.S. Environmental Protection Agency, Office of Research and Development (EPA/600/R-08/117, 142 pp.).

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very High Resolution Interpolated Climate Surfaces for Global Land Areas. International Journal of Climatology 25:1965-1978. Data available online at http://www.worldclim.org/

House, P. Kyle, Heather L. Green, Abbey K. Grimmer, and Rachel E. Wearne. 2010. 1:150,000 Surficial Geologic Map of Clark County, Nevada. Prepared for the Desert Conservation Program, Clark County Multiple Species Habitat Conservation Plan, Las Vegas, NV.

ICF Jones & Stokes. 2009a. Desert Conservation Program Rare Plant Inventories -2005-JS-497B-Work Plan. Prepared for the Desert Conservation Program, Clark County Multiple Species Habitat Conservation Plan, Las Vegas, NV.

ICF Jones & Stokes. 2009b. Desert Conservation Program Rare Plant Inventories -2005-JS-497B-Data Management Plan. Prepared for the Desert Conservation Program, Clark County Multiple Species Habitat Conservation Plan, Las Vegas, NV.

ICF Jones & Stokes. 2010. Desert Conservation Program Rare Plant Inventories -2005-JS-497B-Final Project Report. Prepared for the Desert Conservation Program, Clark County Multiple Species Habitat Conservation Plan, Las Vegas, NV.

Kokos, Sonja R., Larry R. Tinney, David W. Brickey, Mathew E. Hamilton, Analie R. Barnett, and Robert D. Sutter. 2009. Clark County Rare Plant Modeling and Inventory. Poster presented at the 5th Southwest Rare Plant Conference. Salt Lake City, UT.

Nevada Natural Heritage Program (NNHP). 2001. Nevada Rare Plant Atlas. Available online at http://heritage.nv.gov/atlas/atlas.html

Niles, Wesley E. and Patrick J. Leary. 2007. Annotated Checklist of the Vascular Plants of the Spring Mountains Clark and Nye Counties, Nevada. Mentzelia, Number 8, pg 38.

Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190:231-259.

TerraSpectra Geomatics (TerraSpectra). 2011. Revised Gypsiferous and Sandy Soil Models. Prepared for the Desert Conservation Program, Clark County Multiple Species Habitat Conservation Plan, Las Vegas, NV.

The Nature Conservancy (TNC). 2007. A Conservation Management Strategy for Nine Low Elevation Rare Plants in Clark County, Nevada. Prepared for the Desert Conservation Program, Clark County Multiple Species Habitat Conservation Plan, Las Vegas, NV.

U.S. Department of Agriculture, Natural Resources Conservation Service (NRCS). 2007a. NV755 – SSURGO database for Clark County Area, Nevada. Available online at http://SoilDataMart.nrcs.usda.gov/

U.S. Department of Agriculture, Natural Resources Conservation Service (NRCS). 2007b. NV788 – SSURGO database for Las Vegas Valley Area, Nevada, Part of Clark County. Available online at http://SoilDataMart.nrcs.usda.gov/

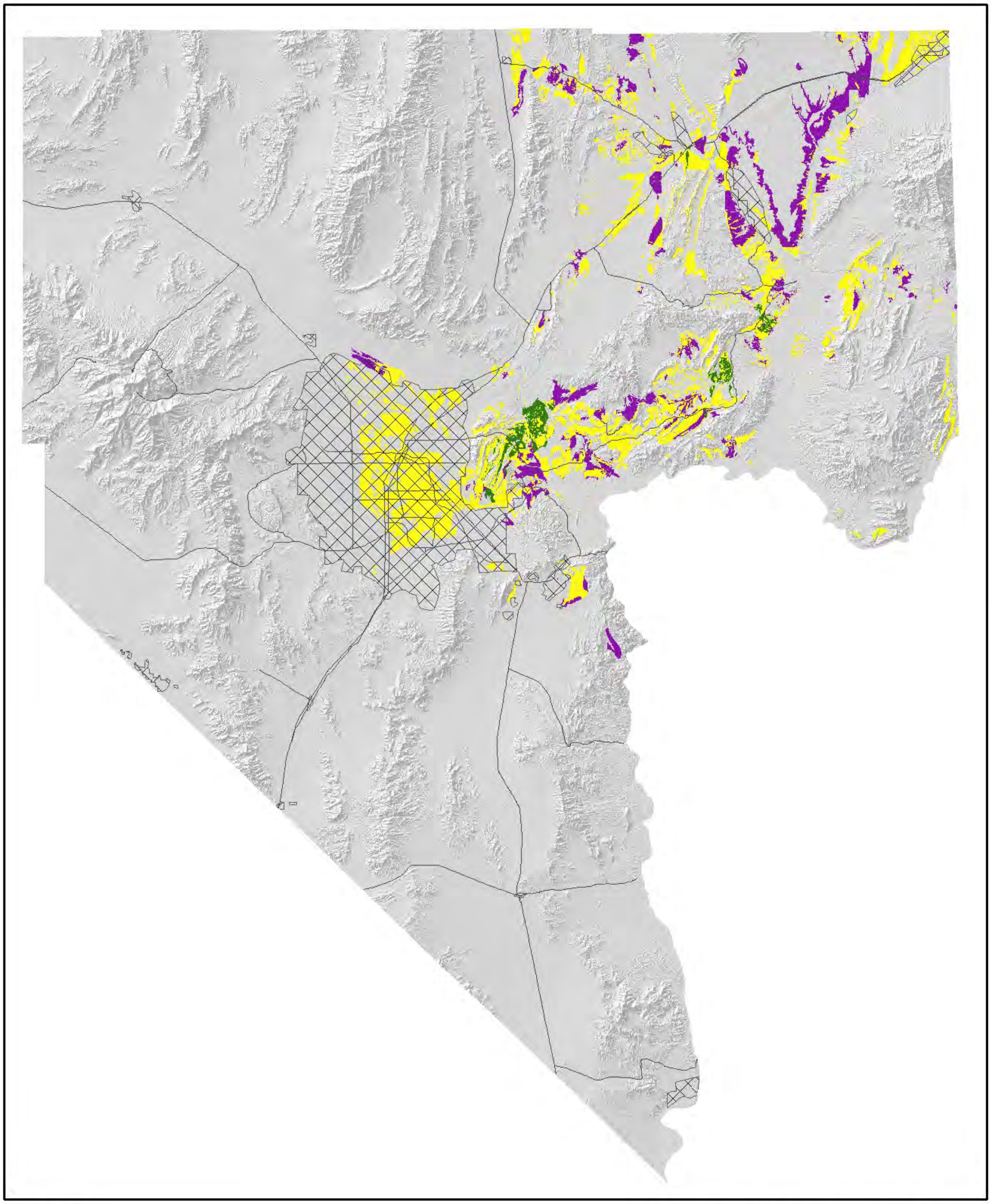
U.S. Department of Agriculture, Natural Resources Conservation Service (NRCS). 2009. NV608 – SSURGO database for Virgin River Area Nevada and Arizona, Parts of Clark and Lincoln Counties Nevada, and Mohave County, Arizona. Available online at http://SoilDataMart.nrcs.usda.gov/

U.S. Geological Survey (USGS). 2005. Mineral Resources Data System. Available online at http://tin.er.usgs.gov/mrds/

Appendix A

Refined Habitat Models

This page intentionally left blank.



Legend

Developed Area

Major Road

Habitat Suitability Model

Known occurrence within 10 meters

Known occurrence within polygon

No known surveys performed

Partial survey, species not found

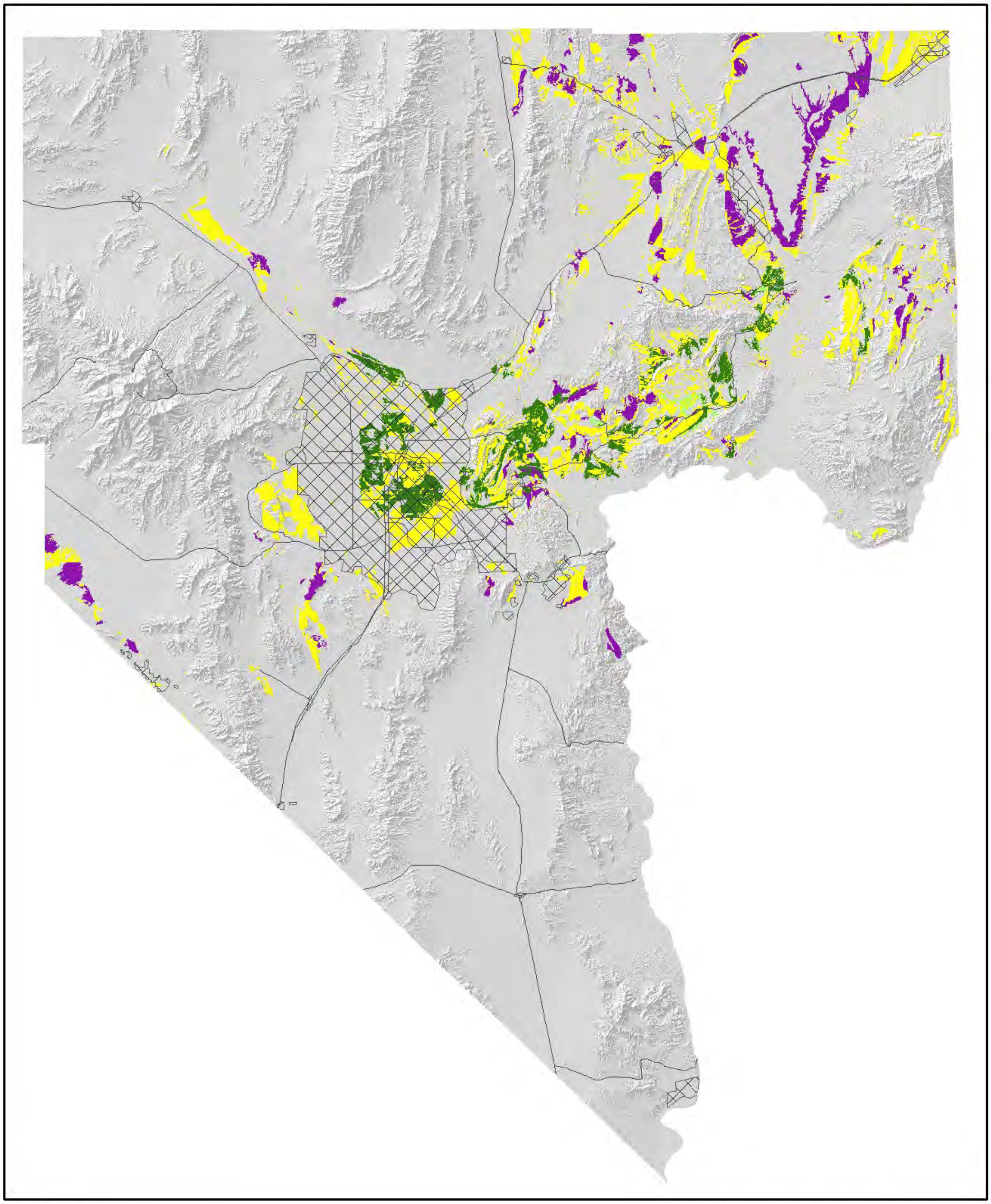
Anulocaulis leiosolenus

Sticky Ringstem

N

Miles

0 5 10 20 30 40



Legend

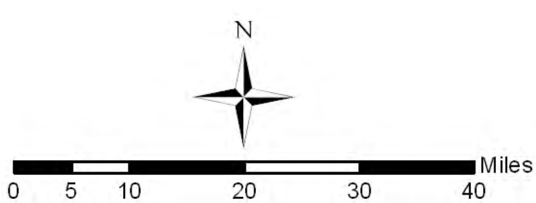
Developed Area
Major Road

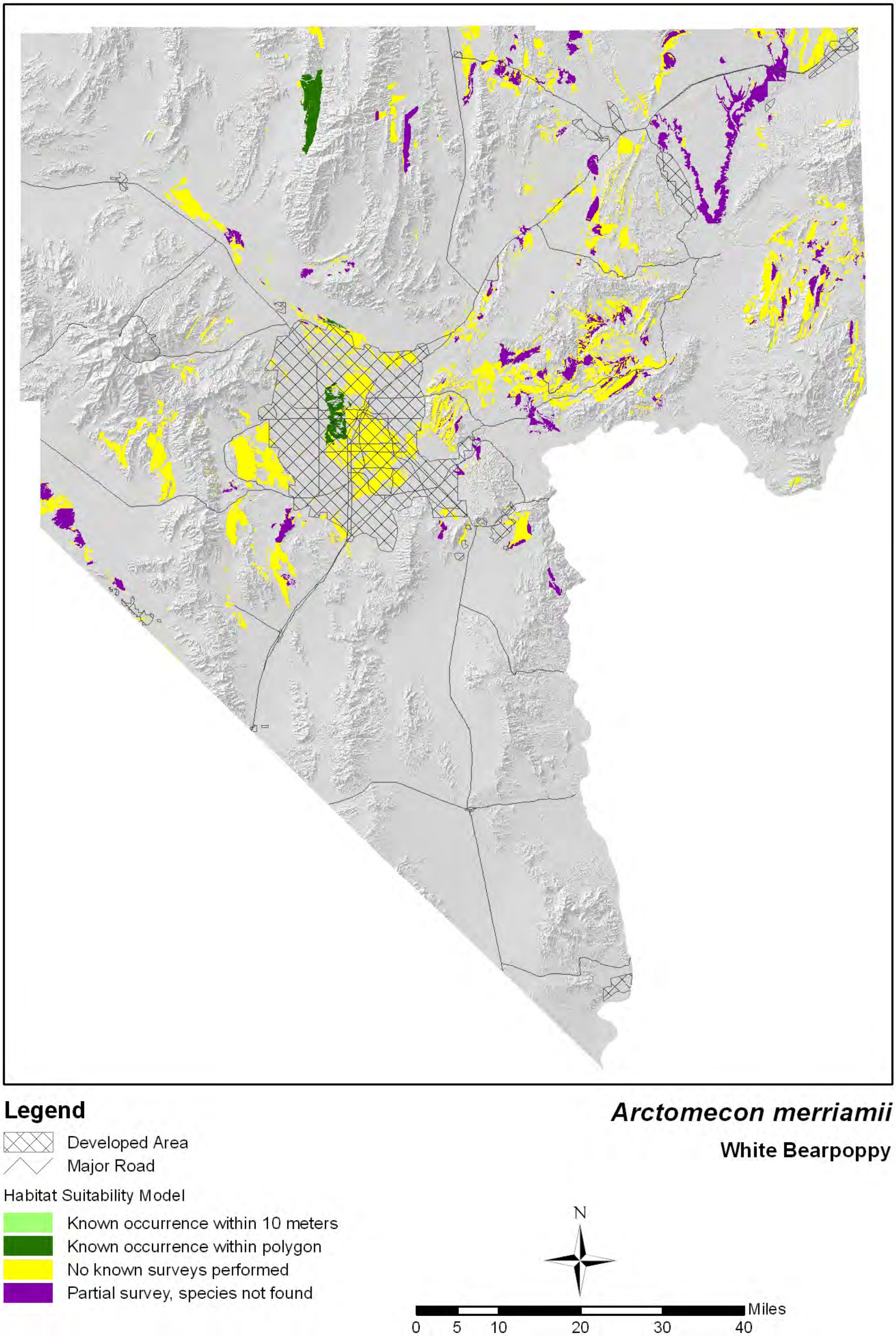
Habitat Suitability Model

Known occurrence within 10 meters
Known occurrence within polygon
No known surveys performed
Partial survey, species not found

Arctomecon californica

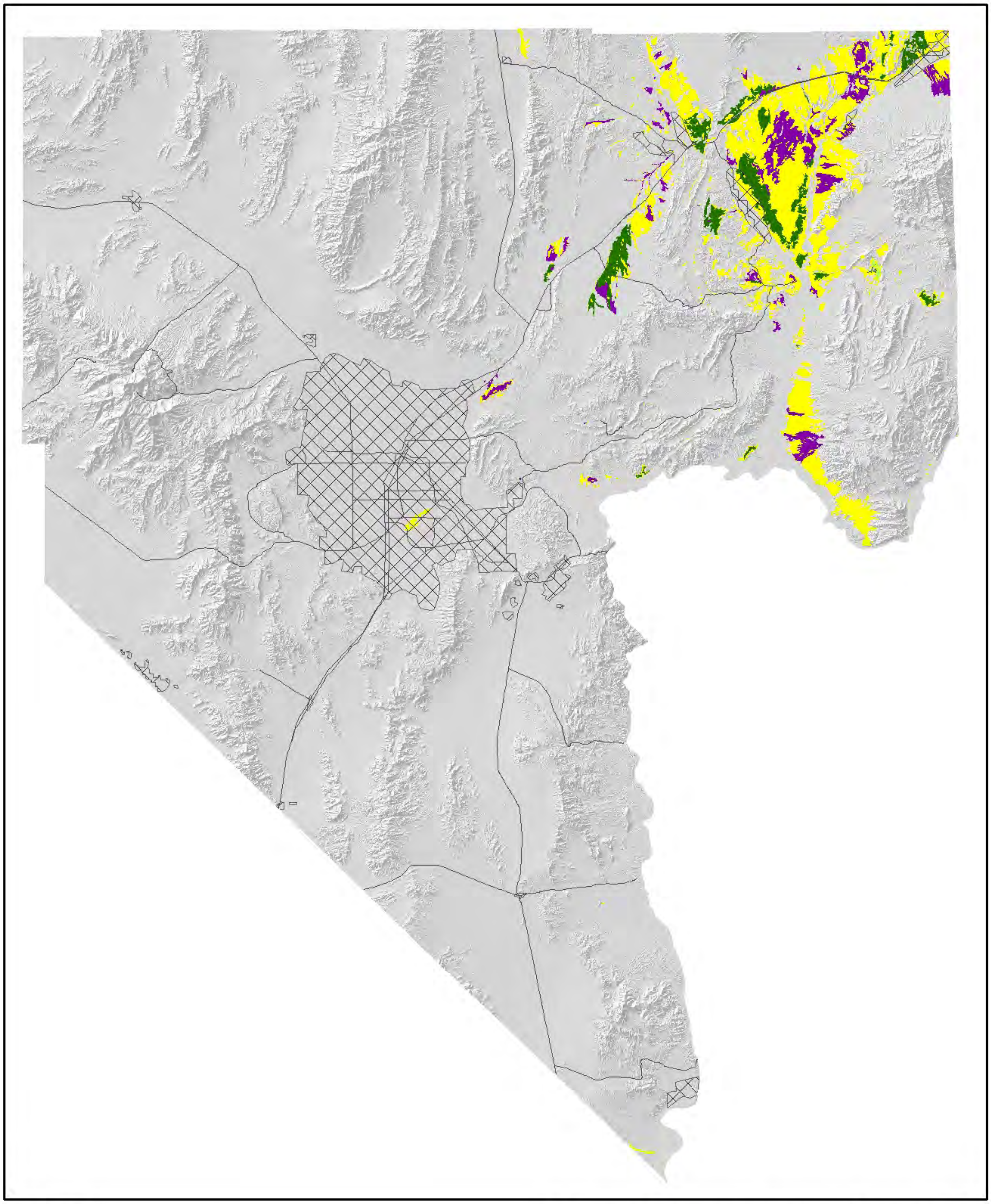
Las Vegas Bearpoppy





Arctomecon merriamii

White Bearpoppy



Legend

Developed Area

Major Road

Habitat Suitability Model

Known occurrence within 10 meters

Known occurrence within polygon

No known surveys performed

Partial survey, species not found

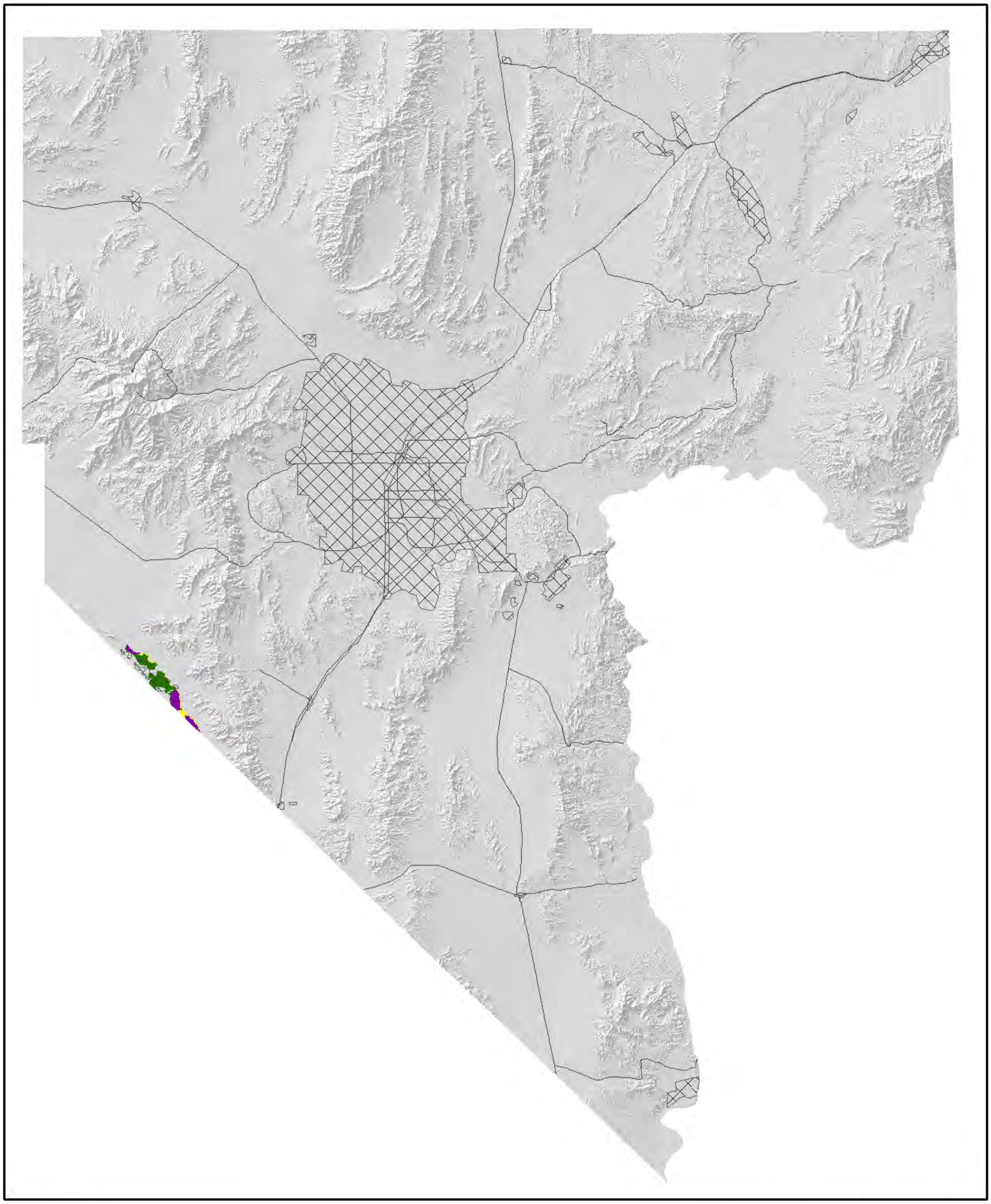
Astragalus geyeri var. triquetrus

Threecorner Milkvetch

N

Miles

0 5 10 20 30 40



Legend

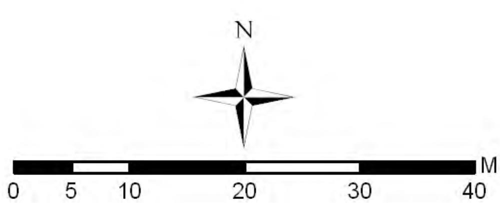
Developed Area
Major Road

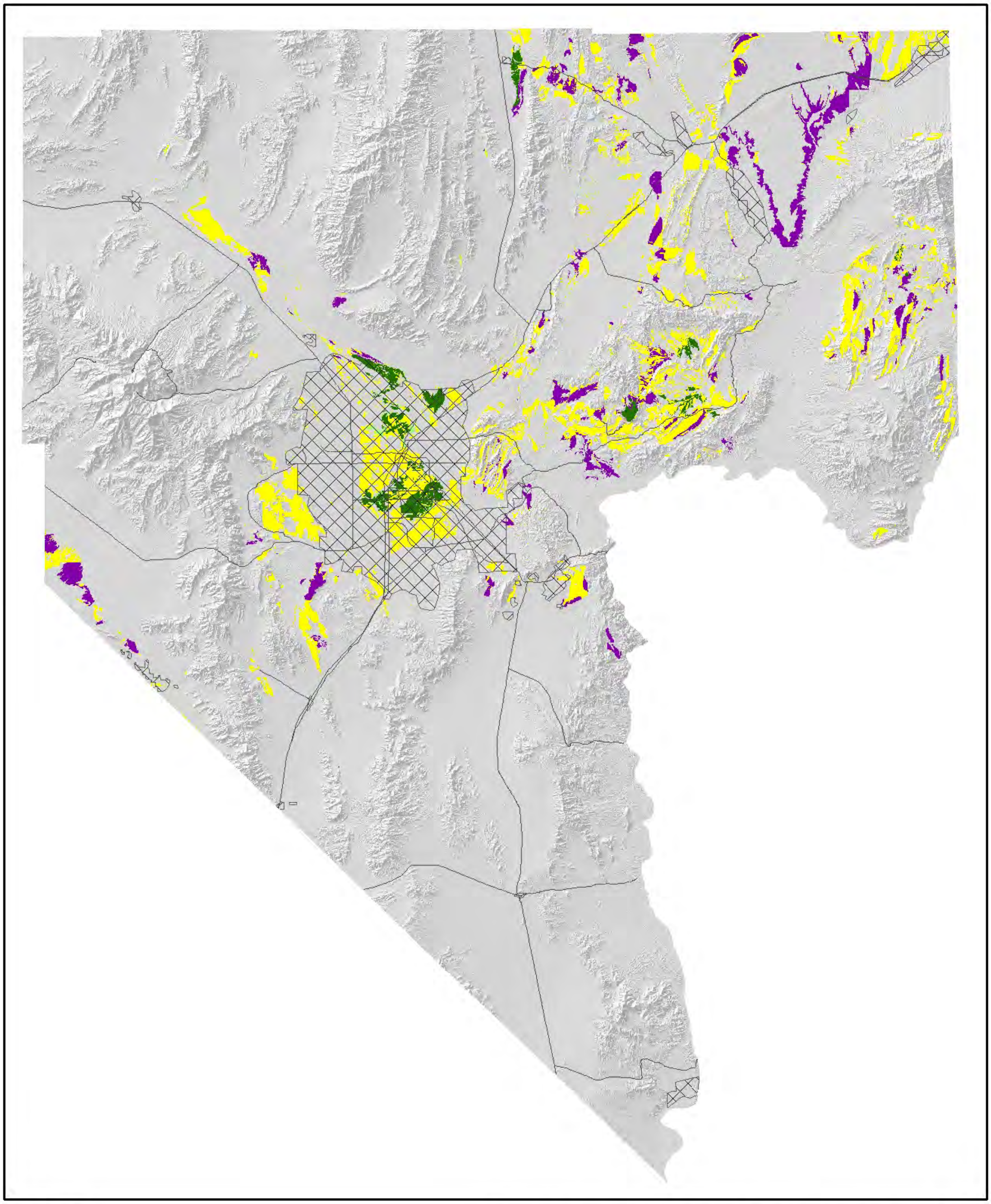
Habitat Suitability Model

Known occurrence within 10 meters
Known occurrence within polygon
No known surveys performed
Partial survey, species not found

Eriogonum bifurcatum

Pahrump Valley Buckwheat





Eriogonum corymbosum var. nilesii

Las Vegas Buckwheat

Legend

- Developed Area
- Major Road

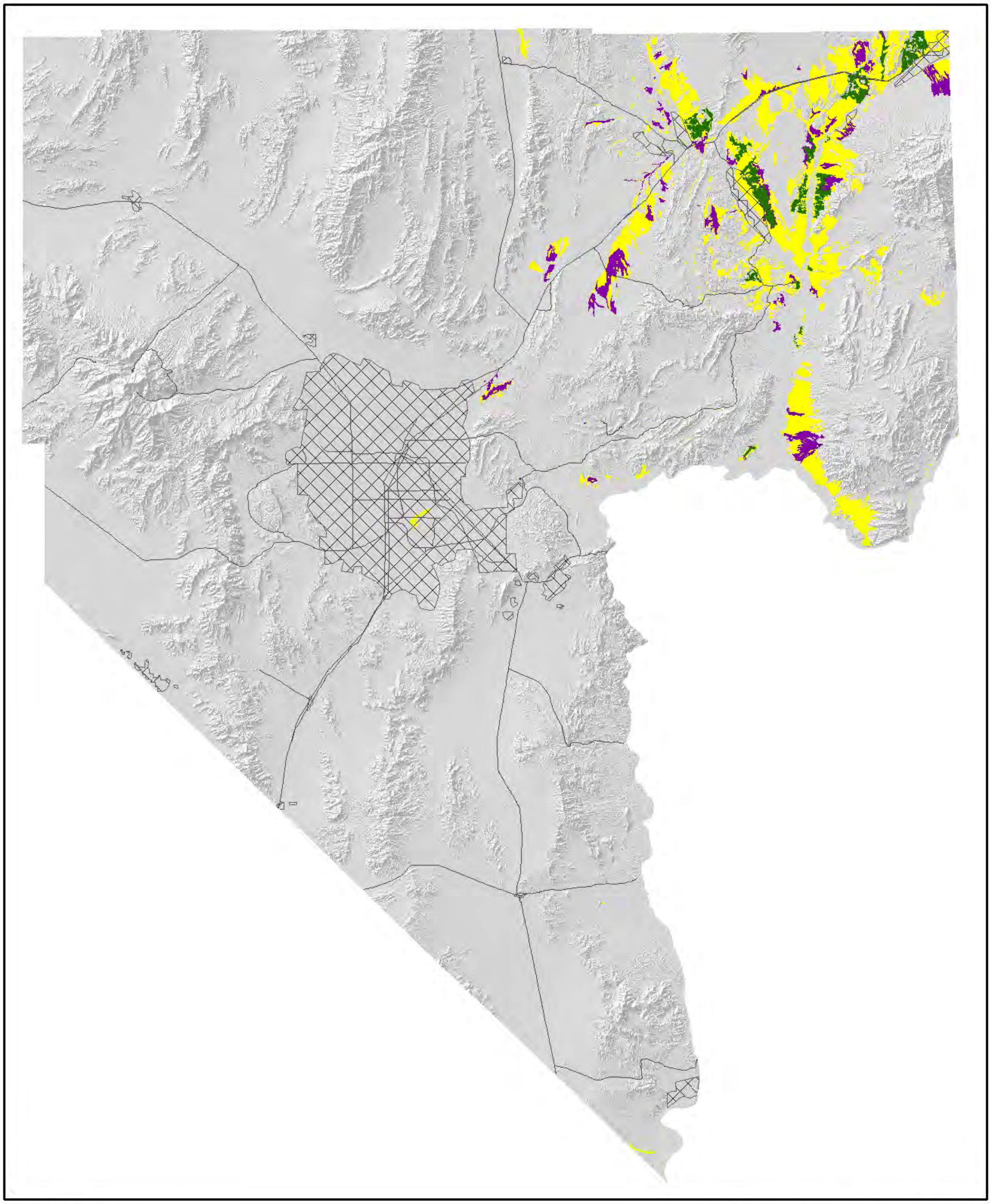
Habitat Suitability Model

- Known occurrence within 10 meters
- Known occurrence within polygon
- No known surveys performed
- Partial survey, species not found

N

Miles

0 5 10 20 30 40



Legend

Developed Area
Major Road

Habitat Suitability Model

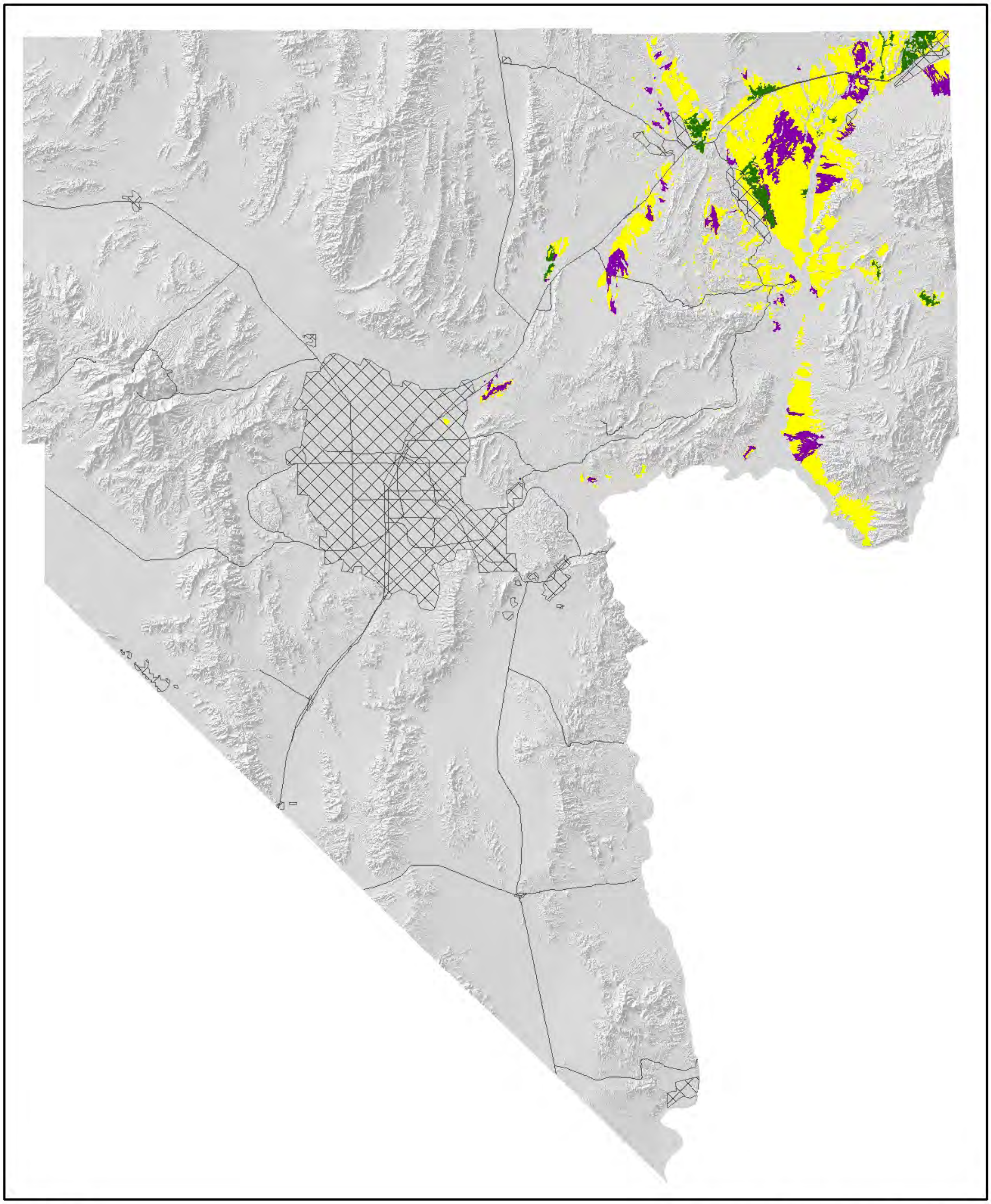
- Known occurrence within 10 meters
- Known occurrence within polygon
- No known surveys performed
- Partial survey, species not found

Eriogonum viscidulum

Sticky Buckwheat

N

0 5 10 20 30 40 Miles



Legend

Developed Area

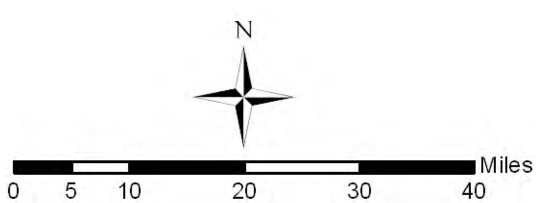
Major Road

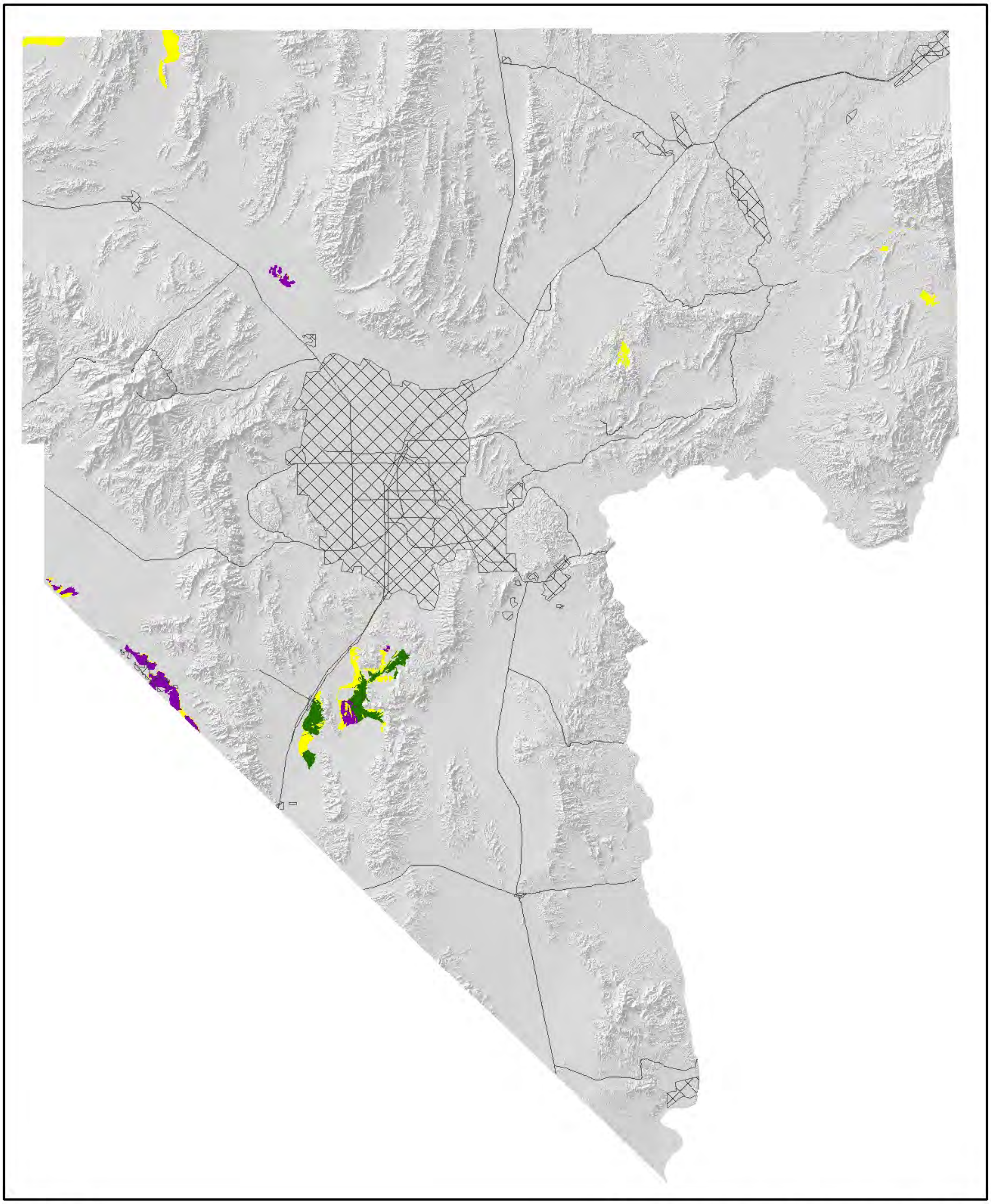
Habitat Suitability Model

- Known occurrence within 10 meters
- Known occurrence within polygon
- No known surveys performed
- Partial survey, species not found

Pediomelum castoreum

Beaver Dam Breadroot





Legend

Developed Area
Major Road

Habitat Suitability Model

Known occurrence within 10 meters
Known occurrence within polygon
No known surveys performed
Partial survey, species not found

Penstemon albomarginatus

White-margined Penstemon

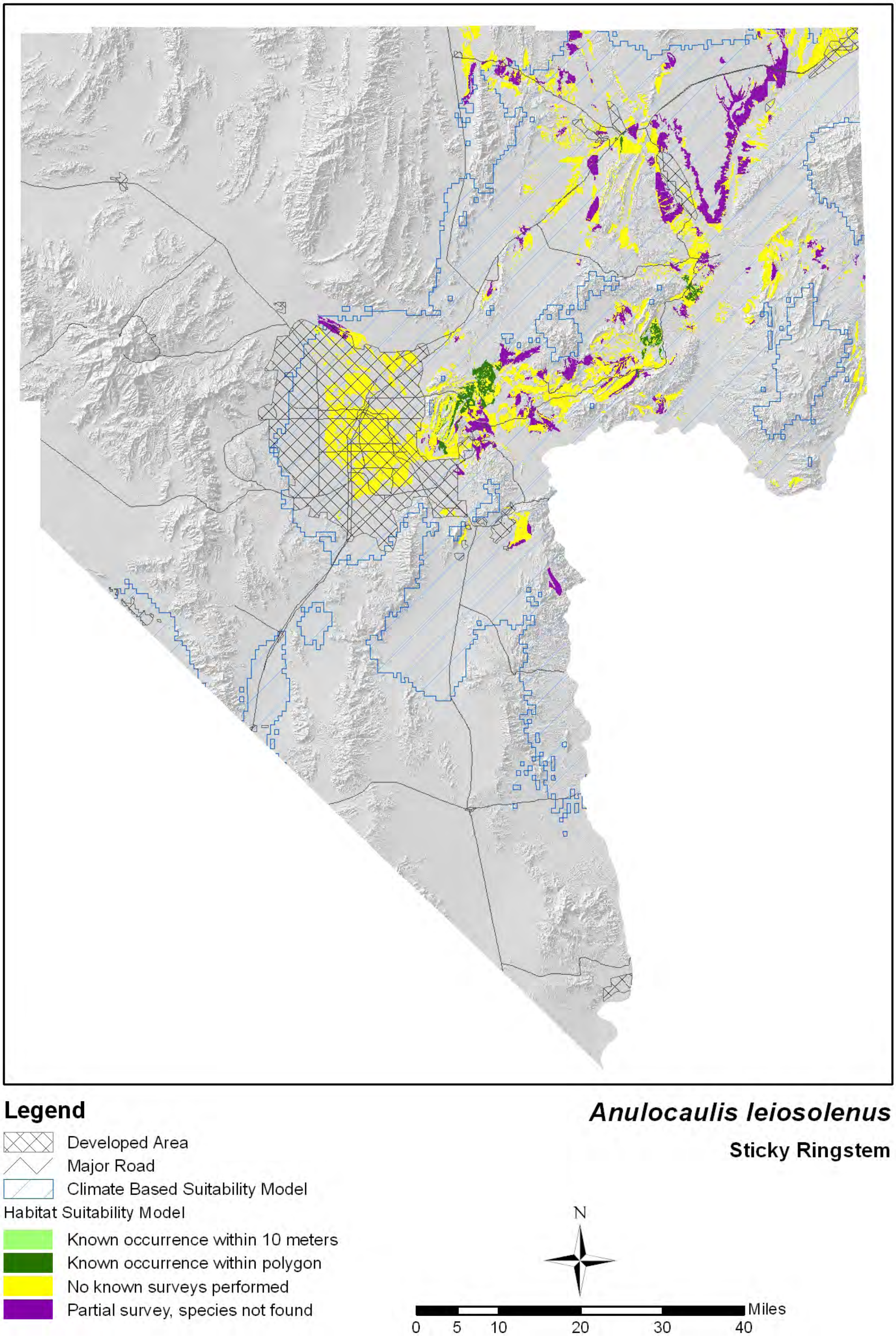
N

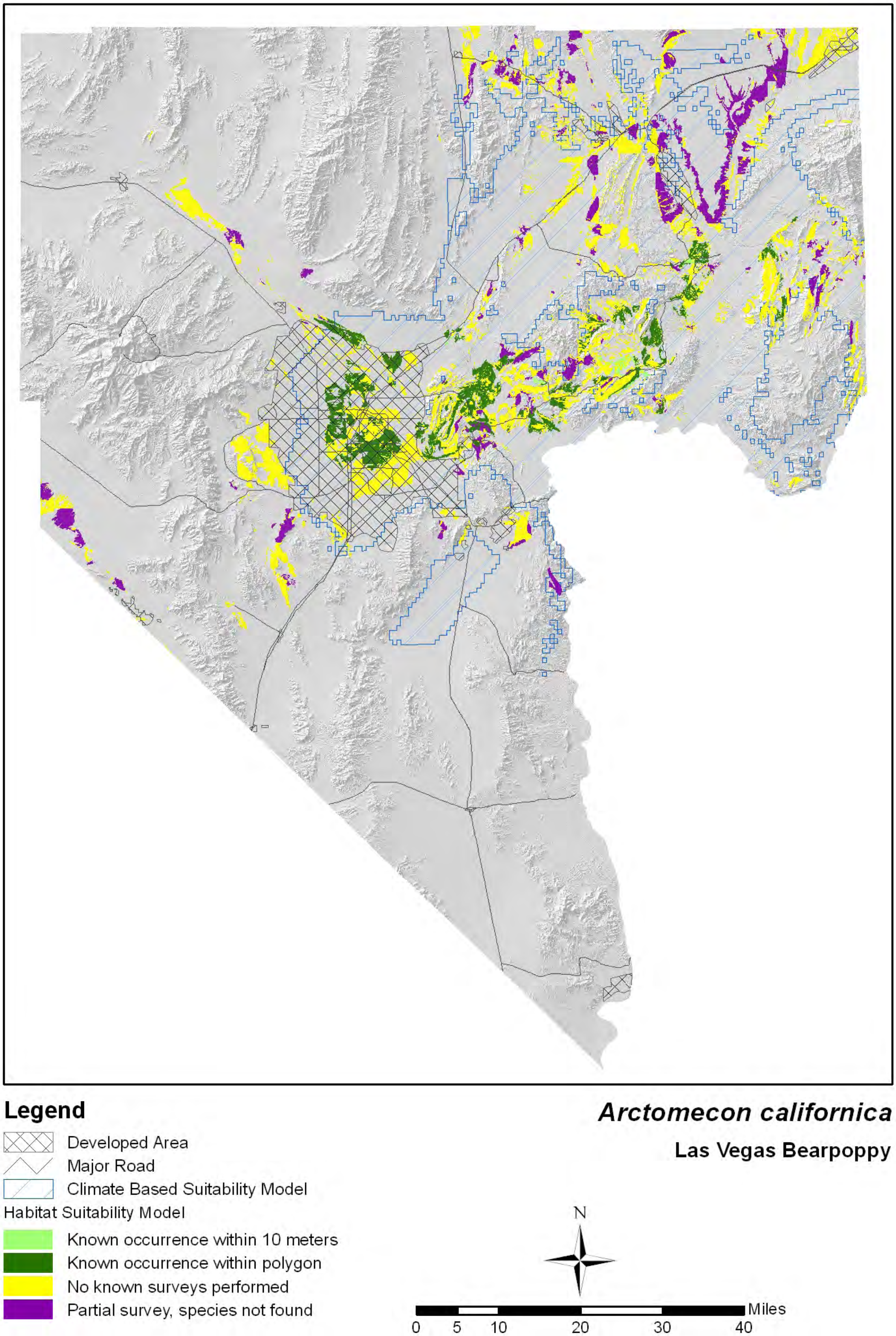
0 5 10 20 30 40 Miles

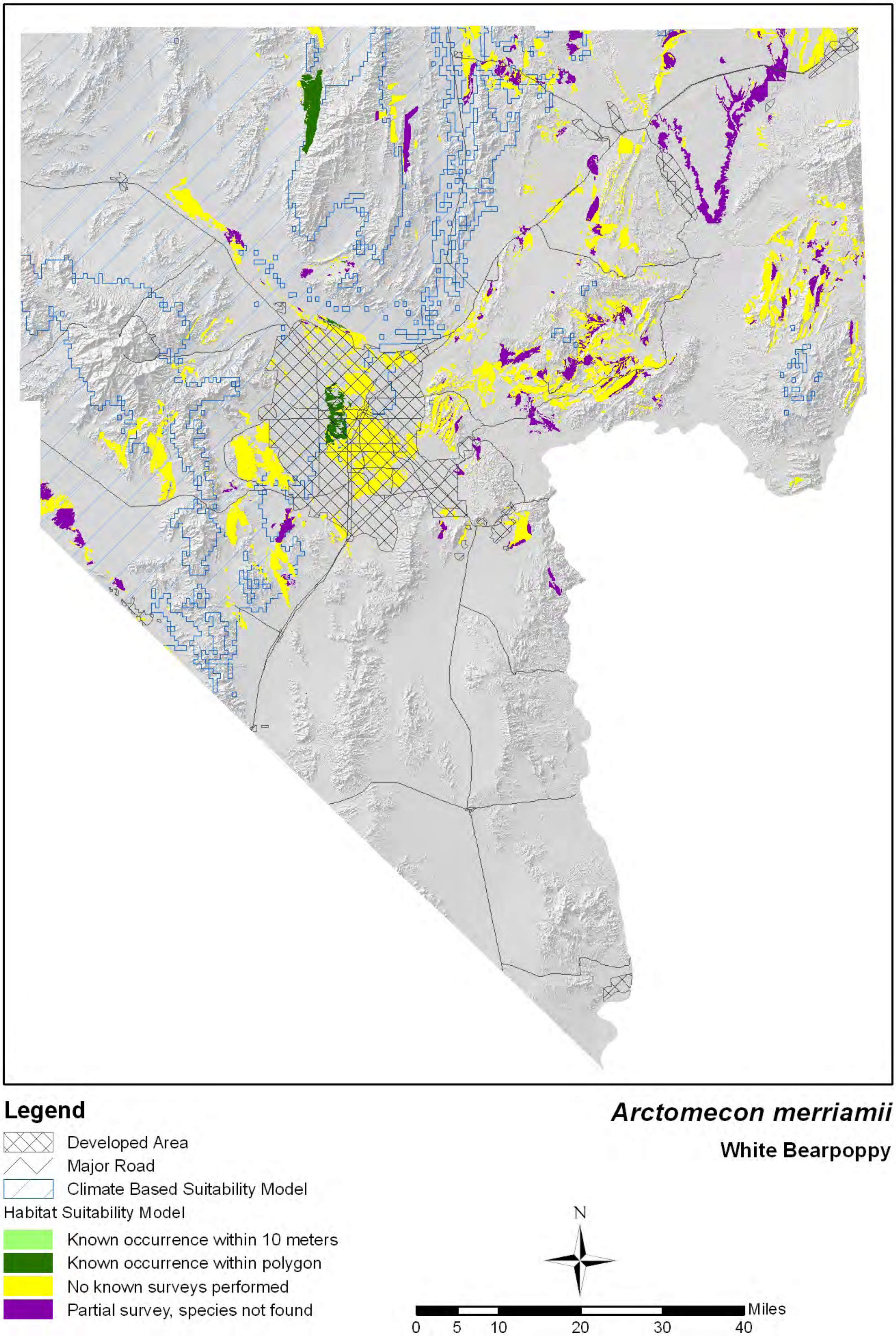
Appendix B

Climate Based Models

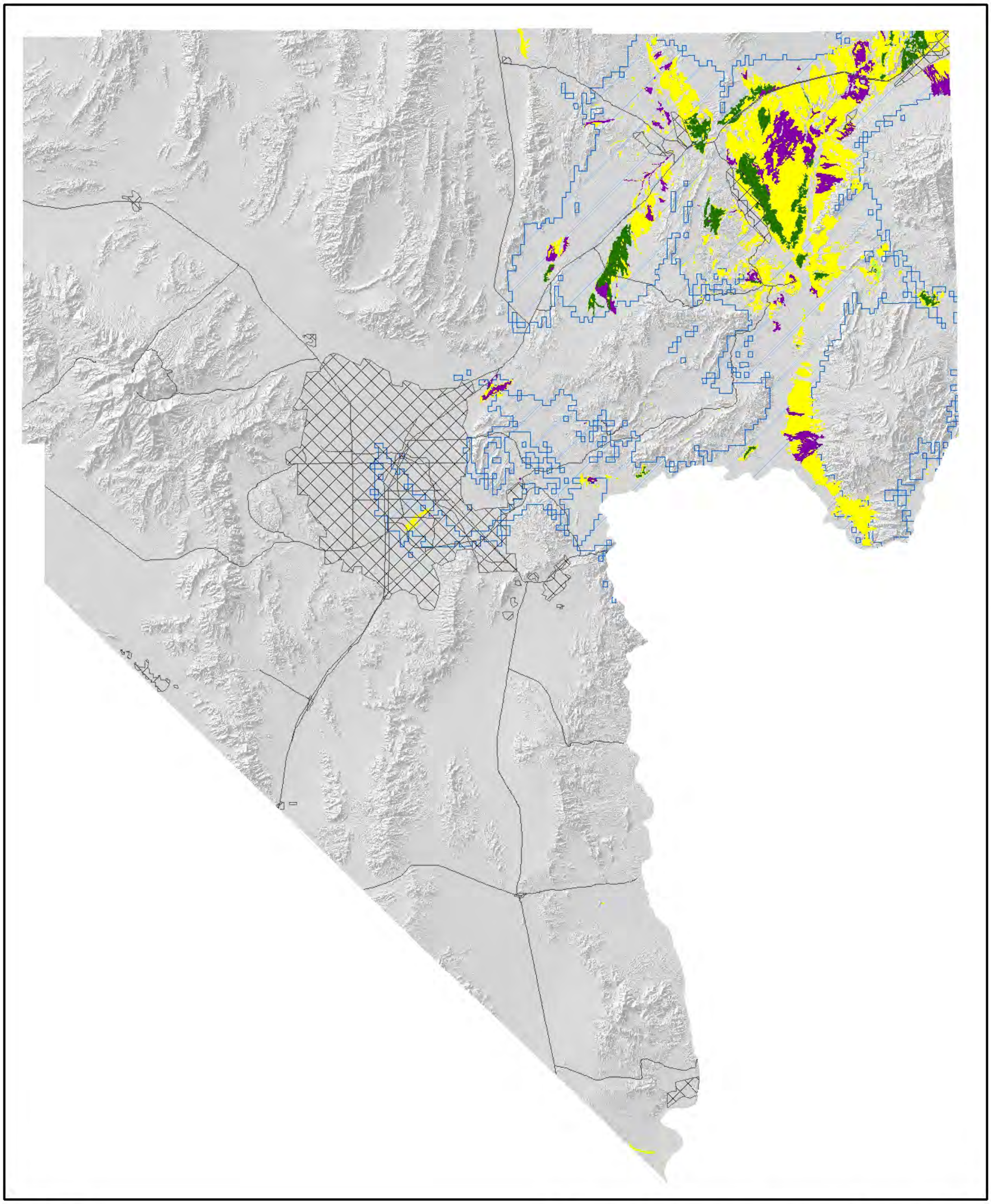
This page intentionally left blank.


Legend
Developed Area
Major Road
Climate Based Suitability Model
Habitat Suitability Model
Known occurrence within 10 meters
Known occurrence within polygon
No known surveys performed
Partial survey, species not found
Anulocaulis leiosolenus
Sticky Ringstem
N
Miles
0 5 10 20 30 40


Legend
Developed Area
Major Road
Climate Based Suitability Model
Habitat Suitability Model
Known occurrence within 10 meters
Known occurrence within polygon
No known surveys performed
Partial survey, species not found
Arctomecon californica
Las Vegas Bearpoppy
N
Miles
0
5
10
20
30
40



Legend
Developed Area
Major Road
Climate Based Suitability Model
Habitat Suitability Model
Known occurrence within 10 meters
Known occurrence within polygon
No known surveys performed
Partial survey, species not found
Arctomecon merriamii
White Bearpoppy
N
0 5 10 20 30 40
Miles



Legend

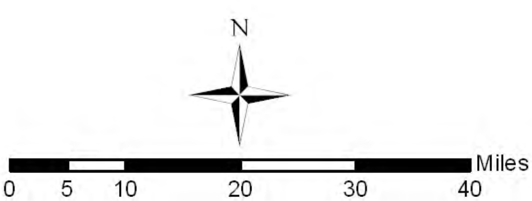
	Developed Area
	Major Road
	Climate Based Suitability Model

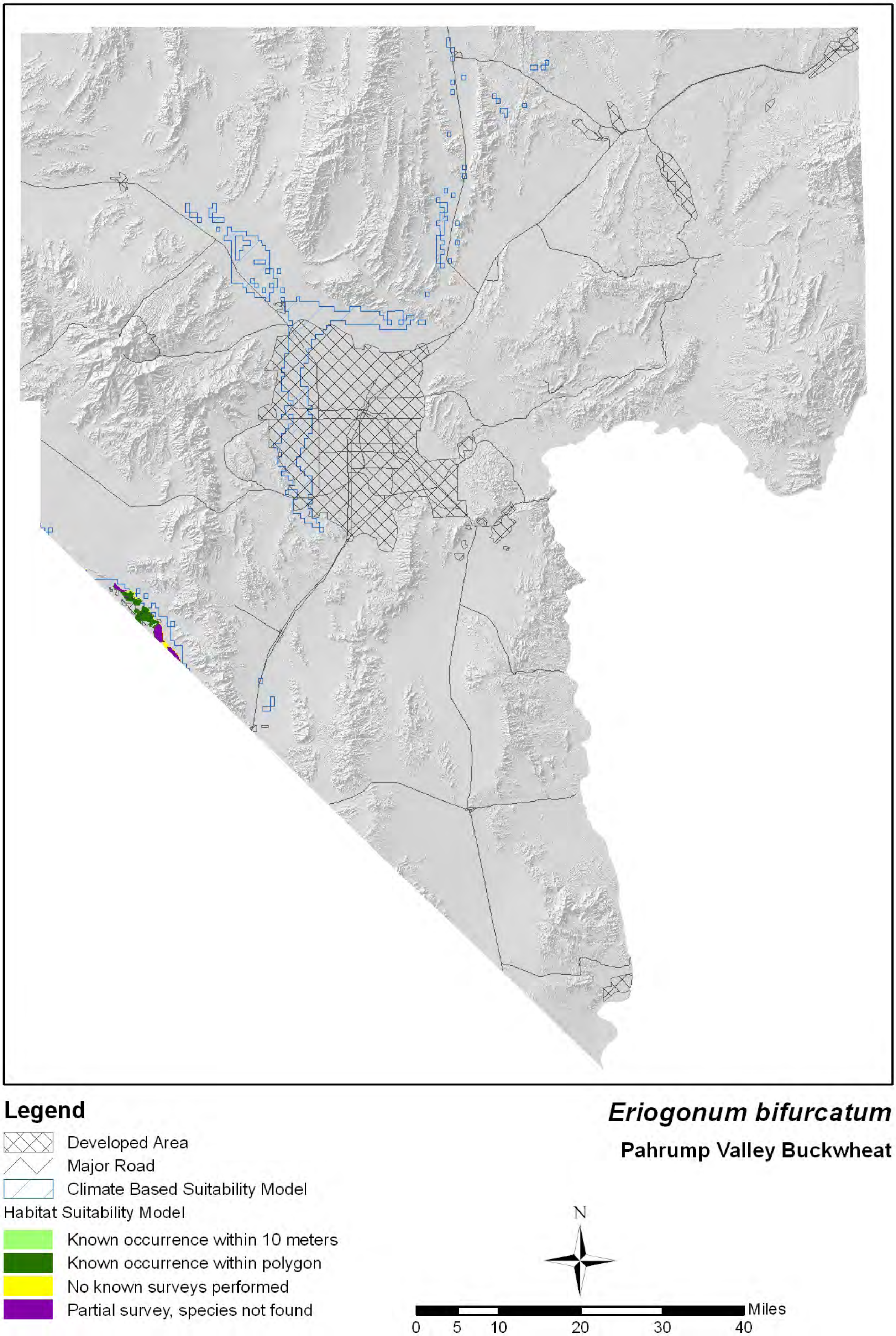
Habitat Suitability Model

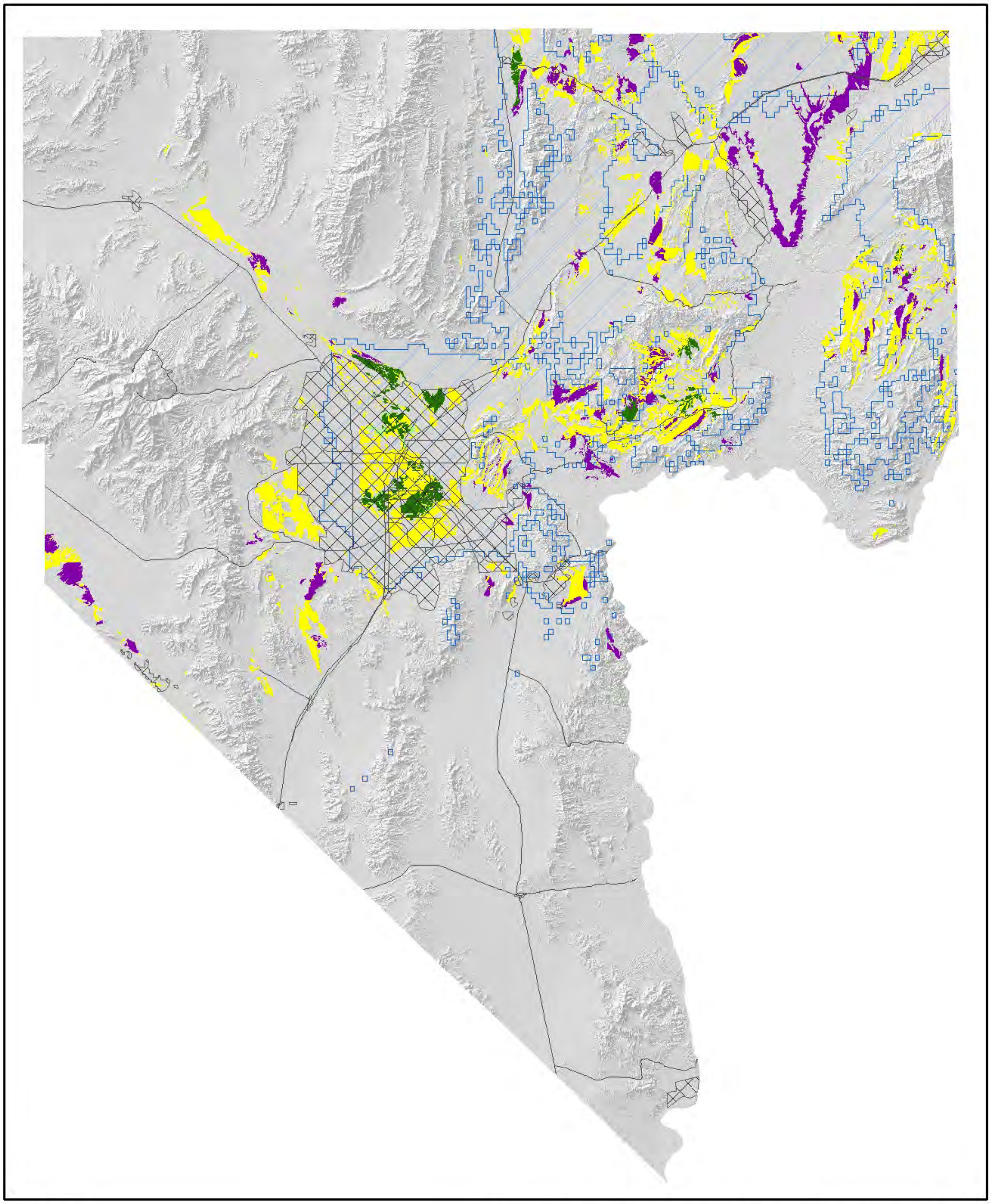
	Known occurrence within 10 meters
	Known occurrence within polygon
	No known surveys performed
	Partial survey, species not found

Astragalus geyeri var. triquetrus

Threecorner Milkvetch




Legend
Developed Area
Major Road
Climate Based Suitability Model
Habitat Suitability Model
Known occurrence within 10 meters
Known occurrence within polygon
No known surveys performed
Partial survey, species not found
Eriogonum bifurcatum
Pahrump Valley Buckwheat
N
Miles
0
5
10
20
30
40

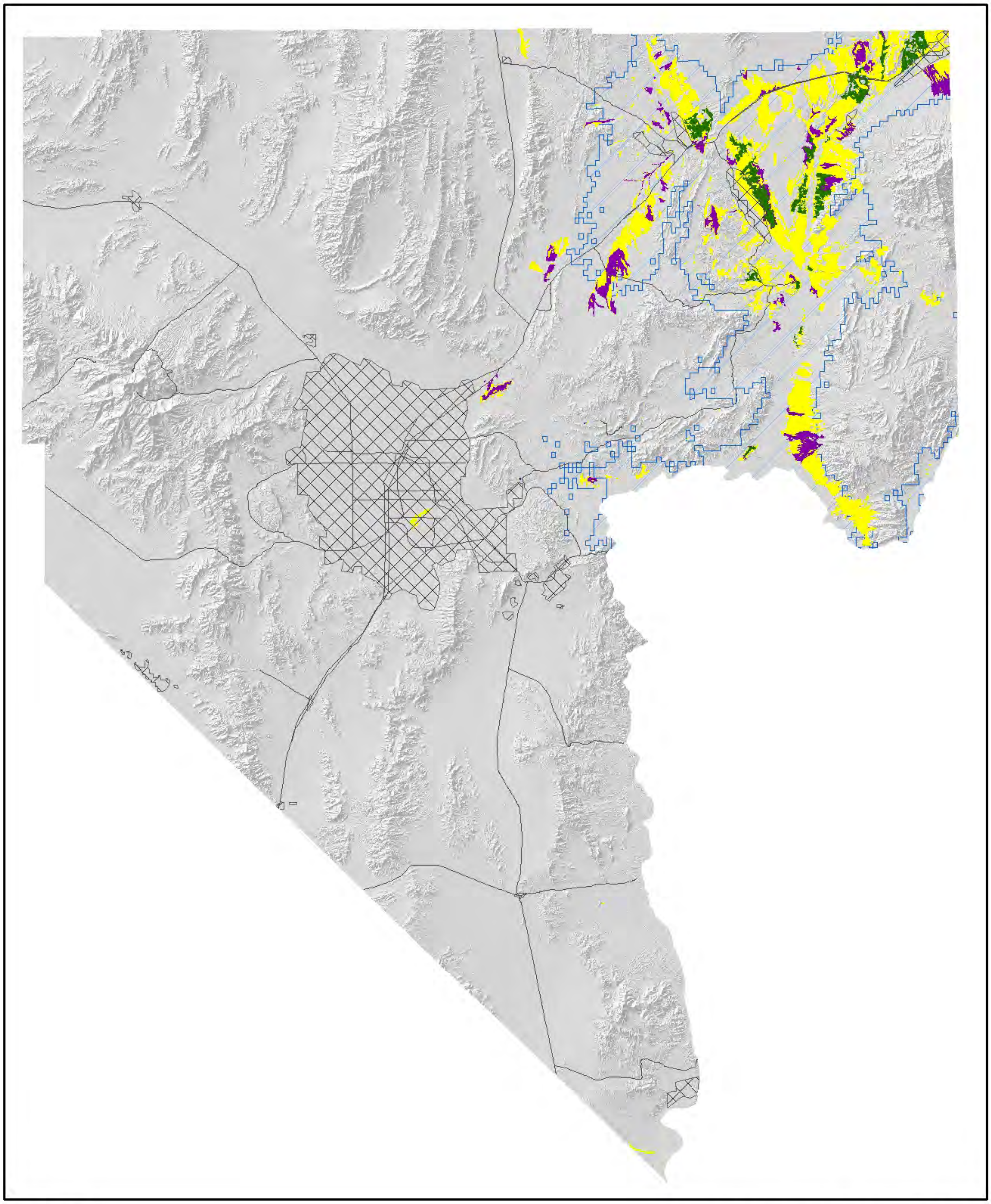


Legend

Developed Area
Major Road
Climate Based Suitability Model
Habitat Suitability Model
Known occurrence within 10 meters
Known occurrence within polygon
No known surveys performed
Partial survey, species not found

Eriogonum corymbosum var. nilesii

Las Vegas Buckwheat



Legend

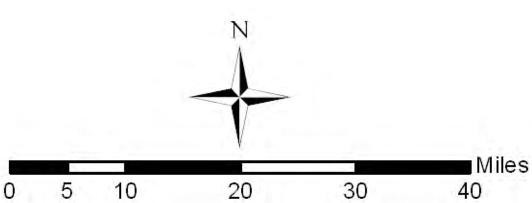
- Developed Area
- Major Road
- Climate Based Suitability Model

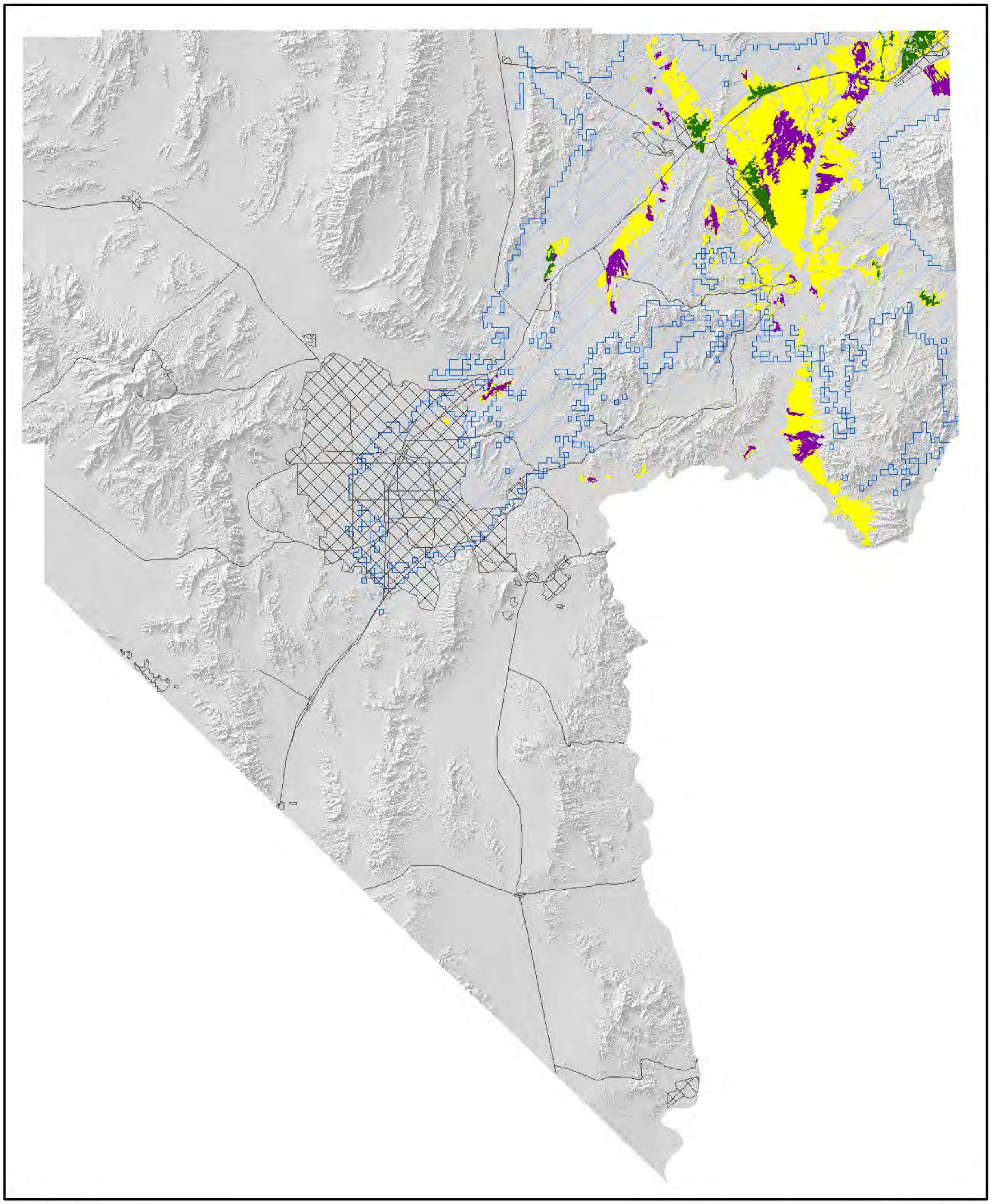
Habitat Suitability Model

- Known occurrence within 10 meters
- Known occurrence within polygon
- No known surveys performed
- Partial survey, species not found

Eriogonum viscidulum

Sticky Buckwheat



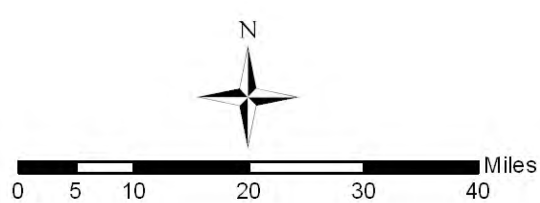


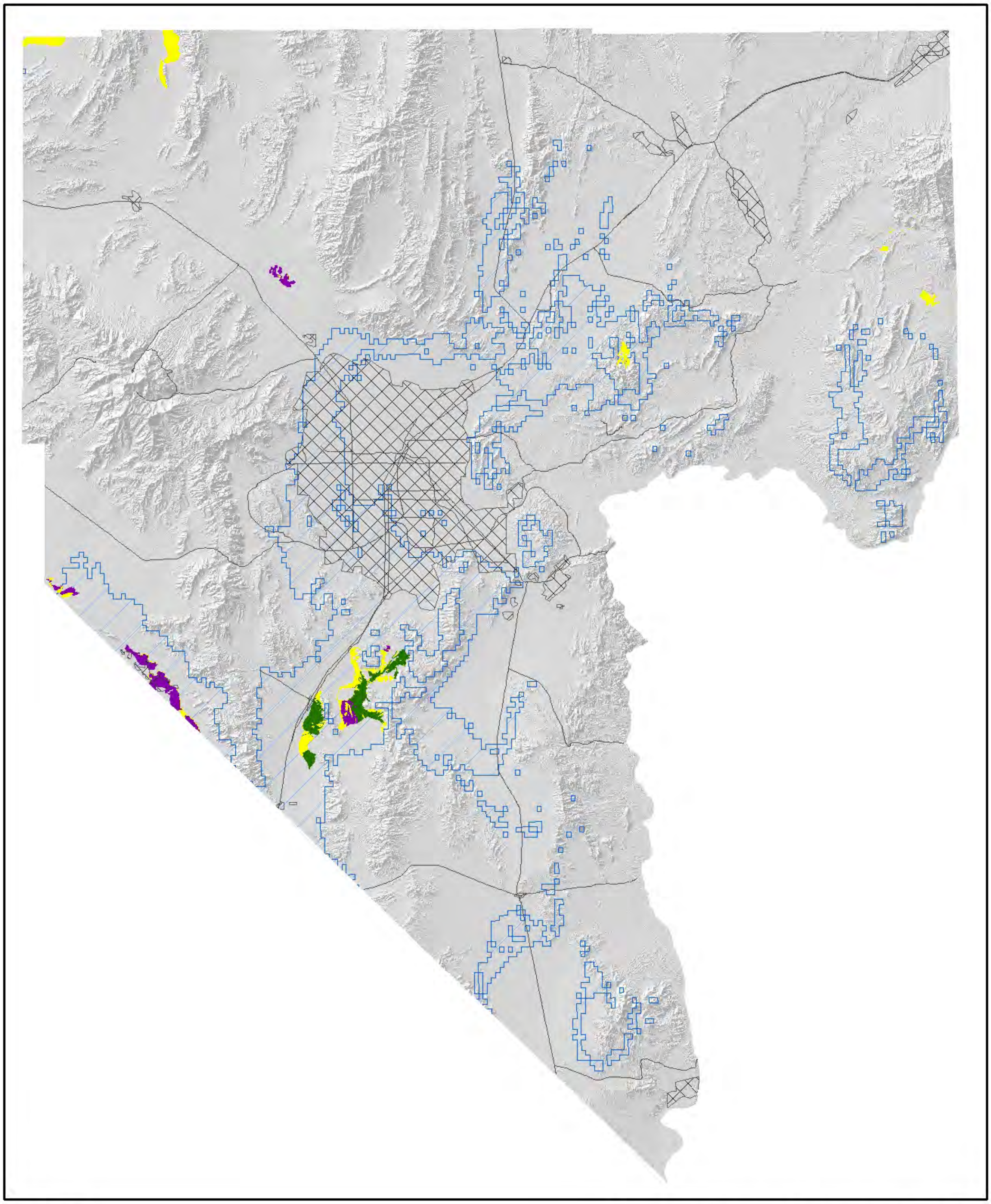
Pediomelum castoreum

Beaver Dam Breadroot

Legend

Developed Area
Major Road
Climate Based Suitability Model
Habitat Suitability Model
Known occurrence within 10 meters
Known occurrence within polygon
No known surveys performed
Partial survey, species not found





Legend

Developed Area

Major Road

Climate Based Suitability Model

Habitat Suitability Model

Known occurrence within 10 meters

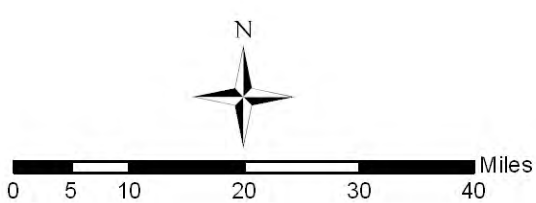
Known occurrence within polygon

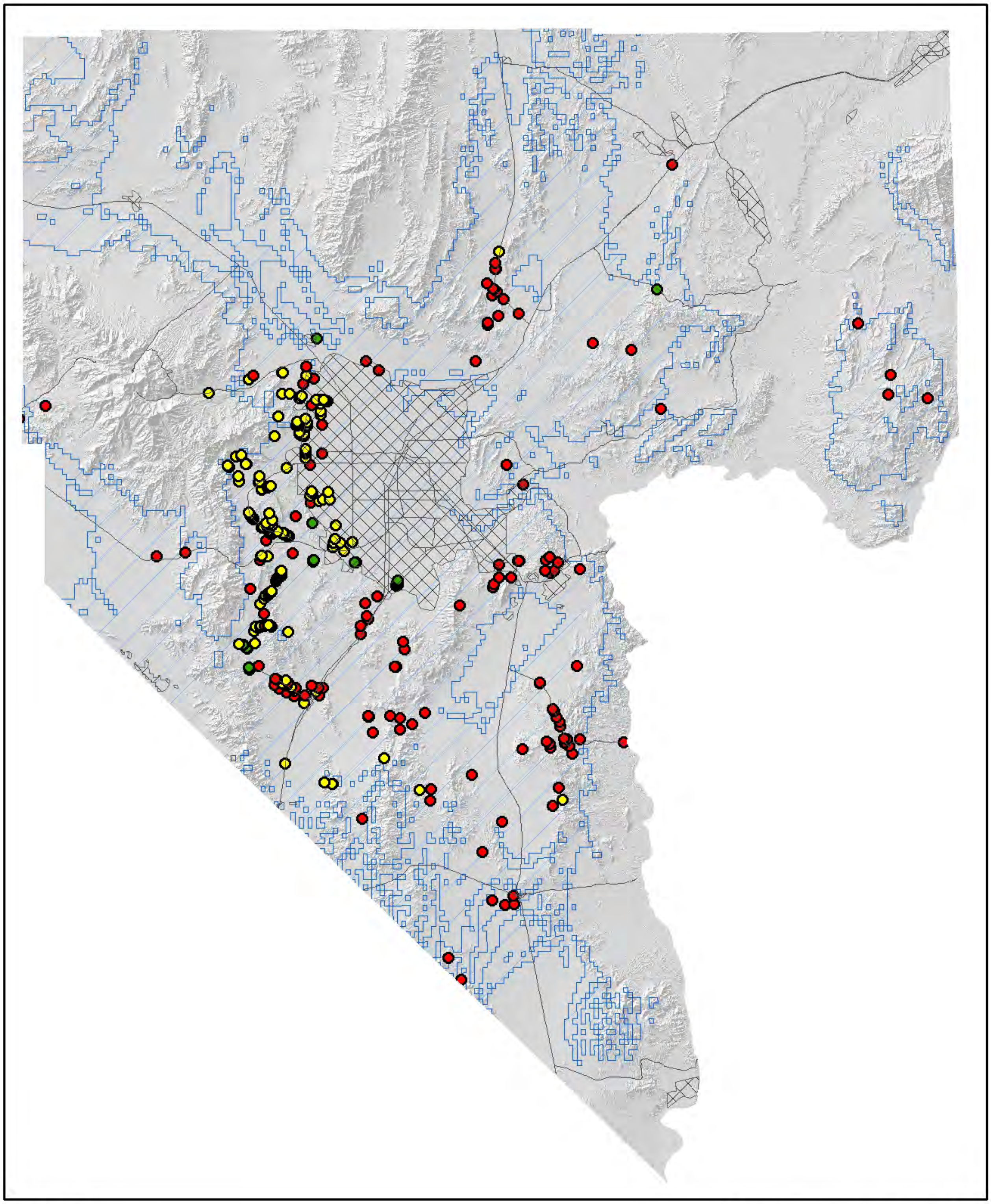
No known surveys performed

Partial survey, species not found

Penstemon albomarginatus

White-margined Penstemon





Legend

Developed Area

Major Road

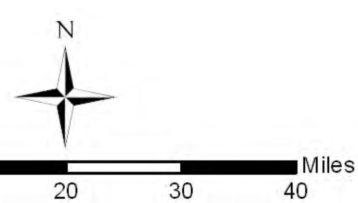
Climate Based Suitability Model

Occurrence Record

- Penstemon bicolor
- Penstemon bicolor ssp bicolor
- Penstemon bicolor ssp roseus

Penstemon bicolor

Two-tone Beardtongue



Appendix C

Maxent Results

This page intentionally left blank.

This appendix includes the outputs from the Maxent climate based model run for each of the species. The models were run using the WorldClim Bioclim climatic data sets (www.worldclim.org/bioclim - accessed February 2011). The datasets used are as follows:

BIO1 = Annual Mean Temperature

BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))

BIO3 = Isothermality (BIO2/BIO7) (* 100)

BIO4 = Temperature Seasonality (standard deviation *100)

BIO5 = Max Temperature of Warmest Month

BIO6 = Min Temperature of Coldest Month

BIO7 = Temperature Annual Range (BIO5-BIO6)

BIO8 = Mean Temperature of Wettest Quarter

BIO9 = Mean Temperature of Driest Quarter

BIO10 = Mean Temperature of Warmest Quarter

BIO11 = Mean Temperature of Coldest Quarter

BIO12 = Annual Precipitation

BIO13 = Precipitation of Wettest Month

BIO14 = Precipitation of Driest Month

BIO15 = Precipitation Seasonality (Coefficient of Variation)

BIO16 = Precipitation of Wettest Quarter

BIO17 = Precipitation of Driest Quarter

BIO18 = Precipitation of Warmest Quarter

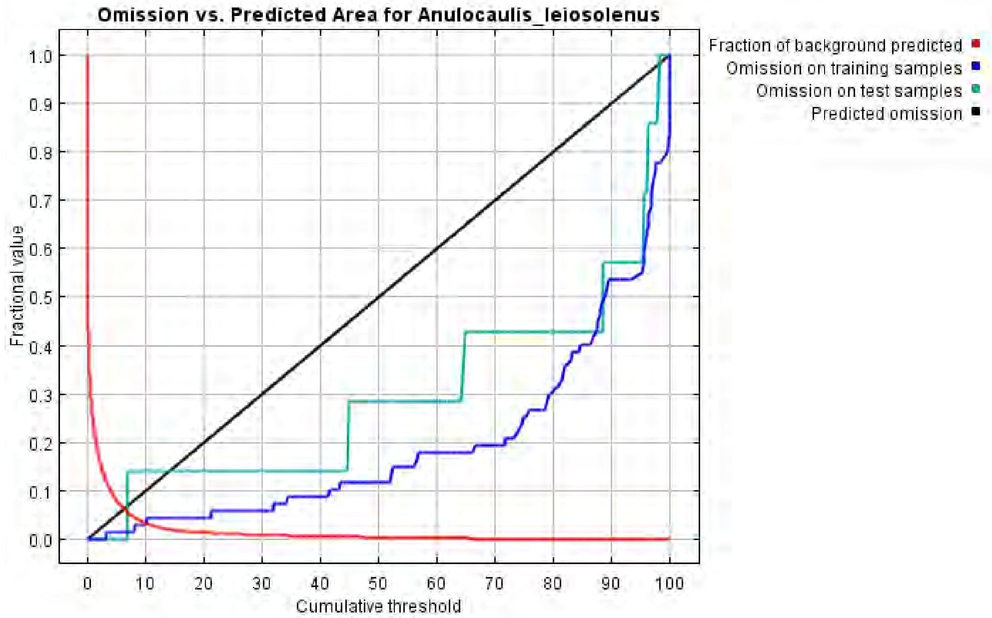
BIO19 = Precipitation of Coldest Quarter

This page intentionally left blank.

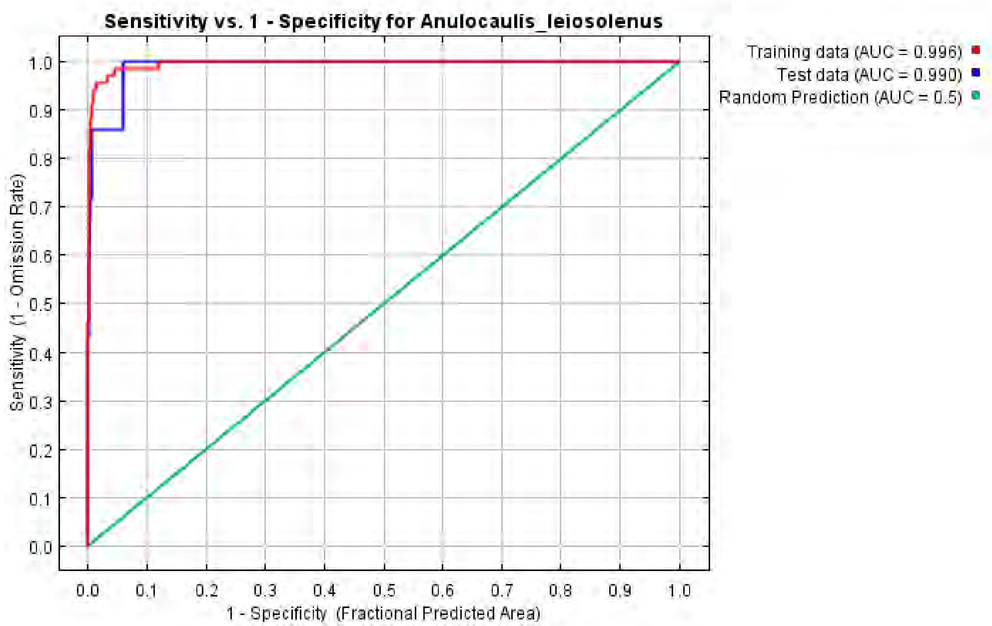
Maxent model for Anulocaulis_leiosolenus

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.978 rather than 1; in practice the test AUC may exceed this bound.

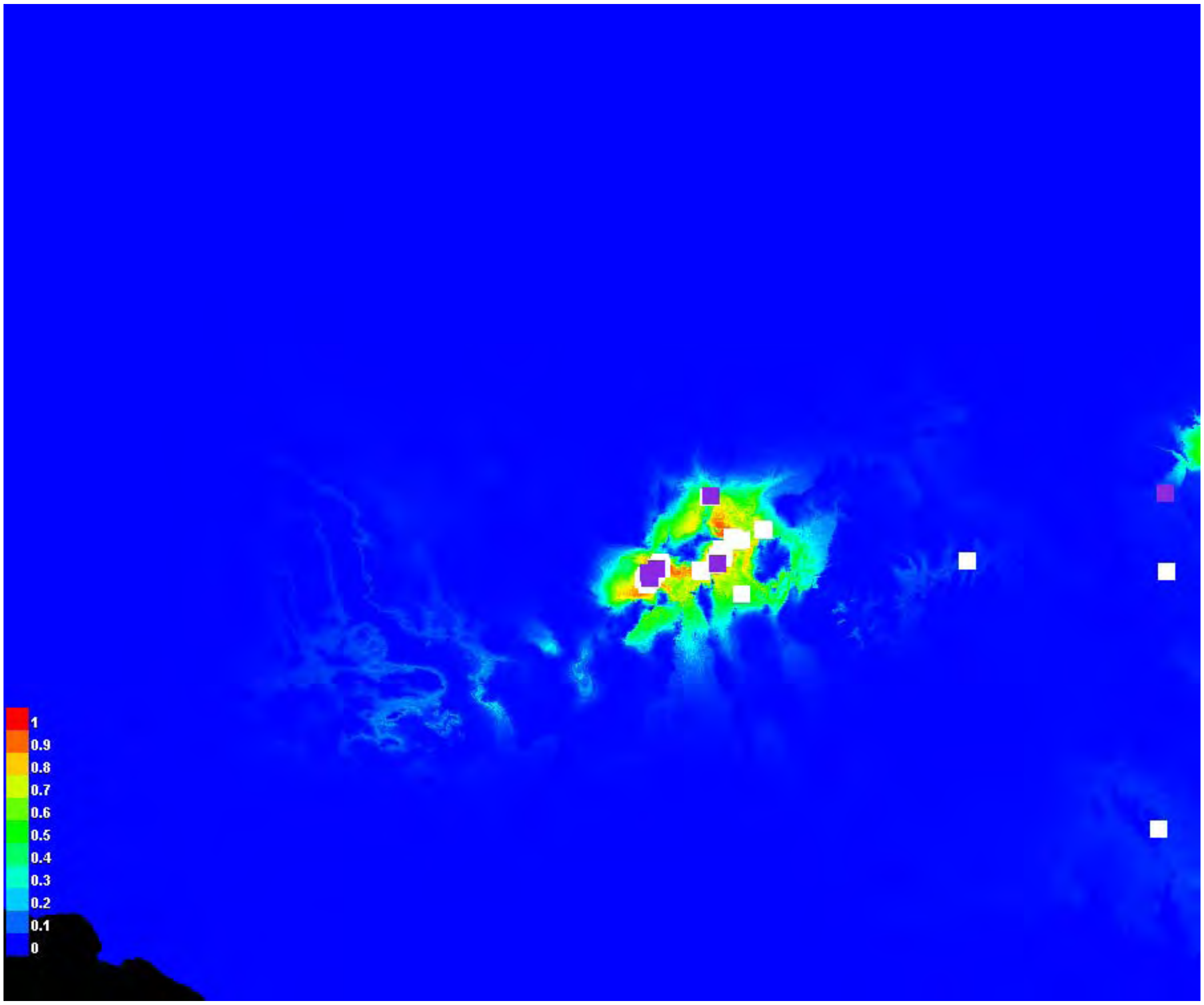


Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.003	Fixed cumulative value	0.240	0.000	0.000	4.533E-5
5.000	0.018	Fixed cumulative value	0.081	0.015	0.000	2.229E-8
10.000	0.056	Fixed cumulative value	0.034	0.030	0.143	1.014E-8
3.235	0.010	Minimum training presence	0.120	0.000	0.000	3.542E-7
41.329	0.566	10 percentile training presence	0.007	0.090	0.143	6.296E-13
10.121	0.058	Equal training sensitivity and specificity	0.033	0.030	0.143	9.107E-9
21.228	0.312	Maximum training sensitivity plus specificity	0.014	0.045	0.143	5.912E-11
6.723	0.028	Equal test sensitivity and specificity	0.058	0.015	0.000	2.345E-9
6.723	0.028	Maximum test sensitivity plus specificity	0.058	0.015	0.000	2.345E-9
3.235	0.010	Balance training omission, predicted area and threshold value	0.120	0.000	0.000	3.542E-7
10.888	0.072	Equate entropy of thresholded and non-thresholded distributions	0.030	0.045	0.143	4.874E-9

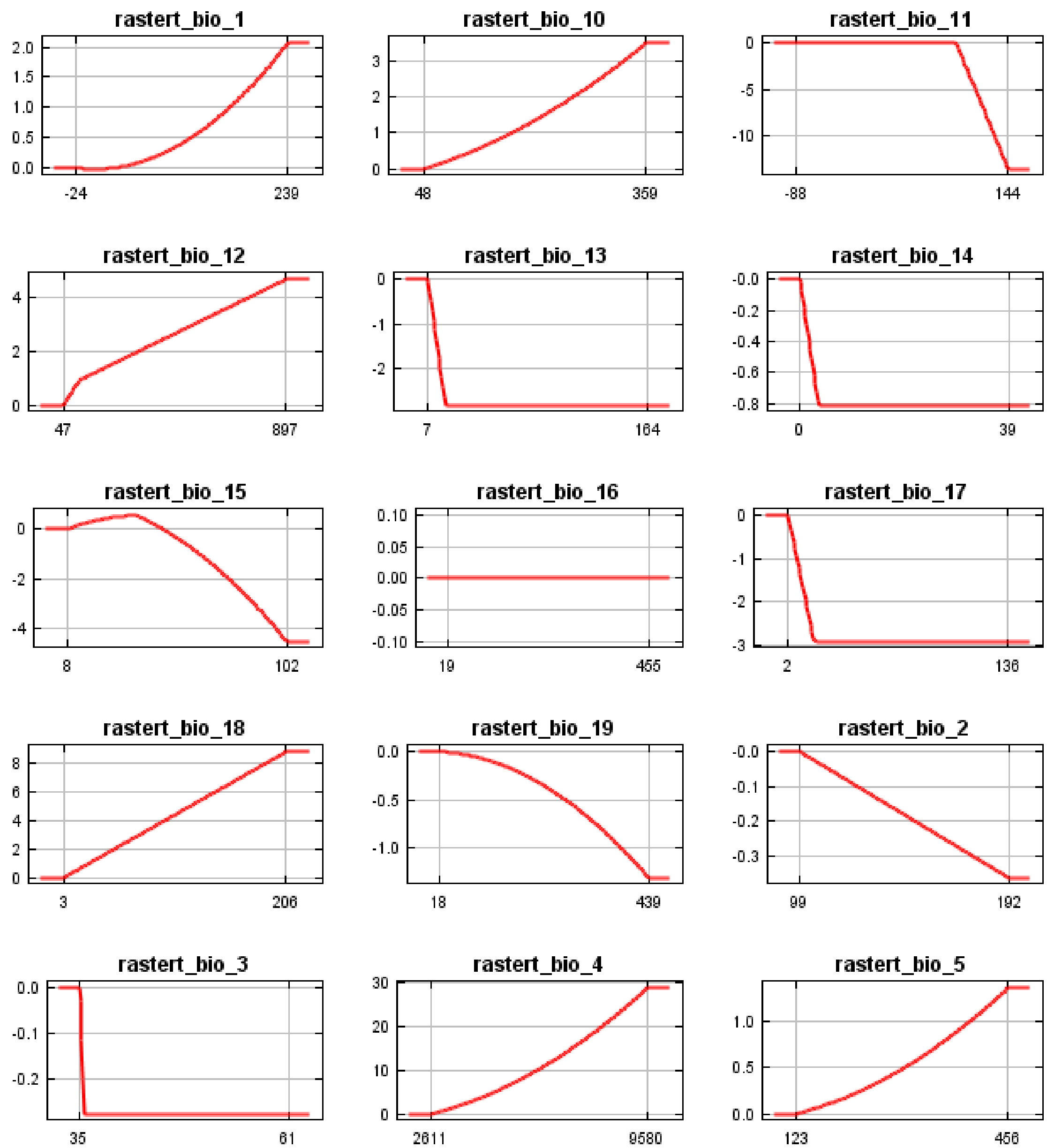
Pictures of the model

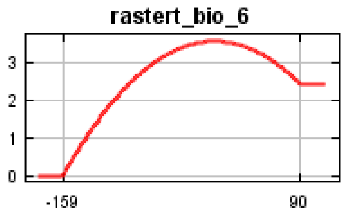
This is a representation of the Maxent model for Anulocaulis_leiosolenus. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

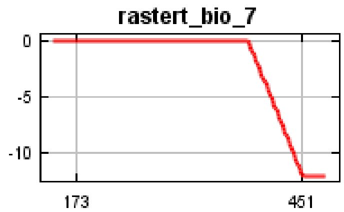


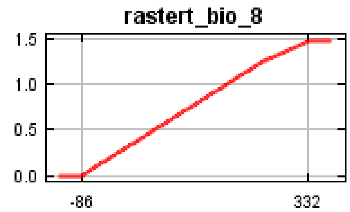
Response curves

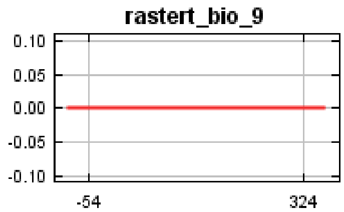
These curves show how each environmental variable affects the Maxent prediction. The (raw) Maxent model has the form exp(...)/constant, and the curves show how the exponent changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version.









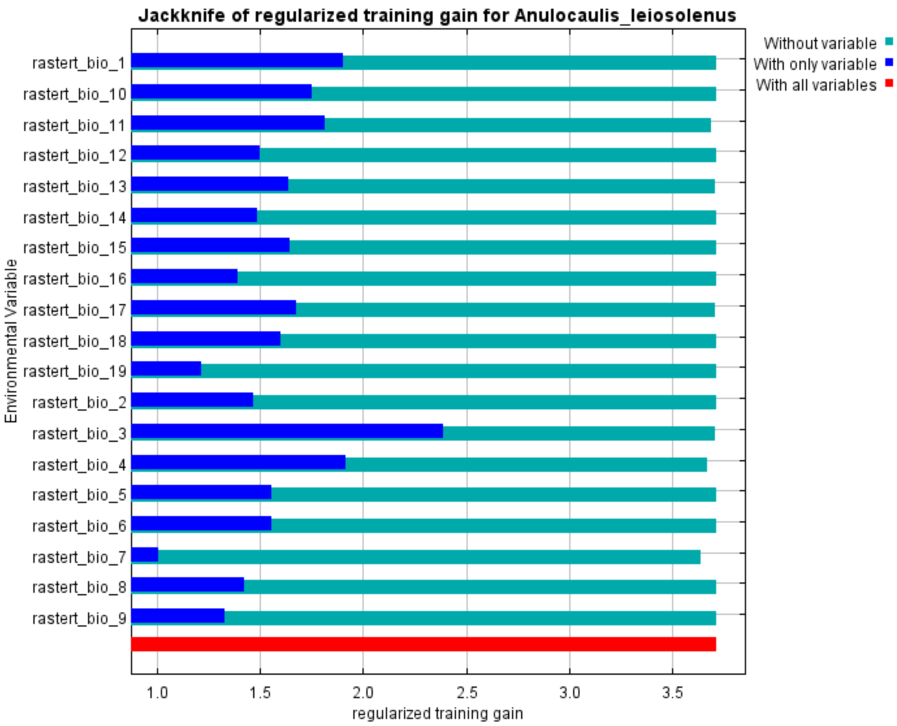


Analysis of variable contributions

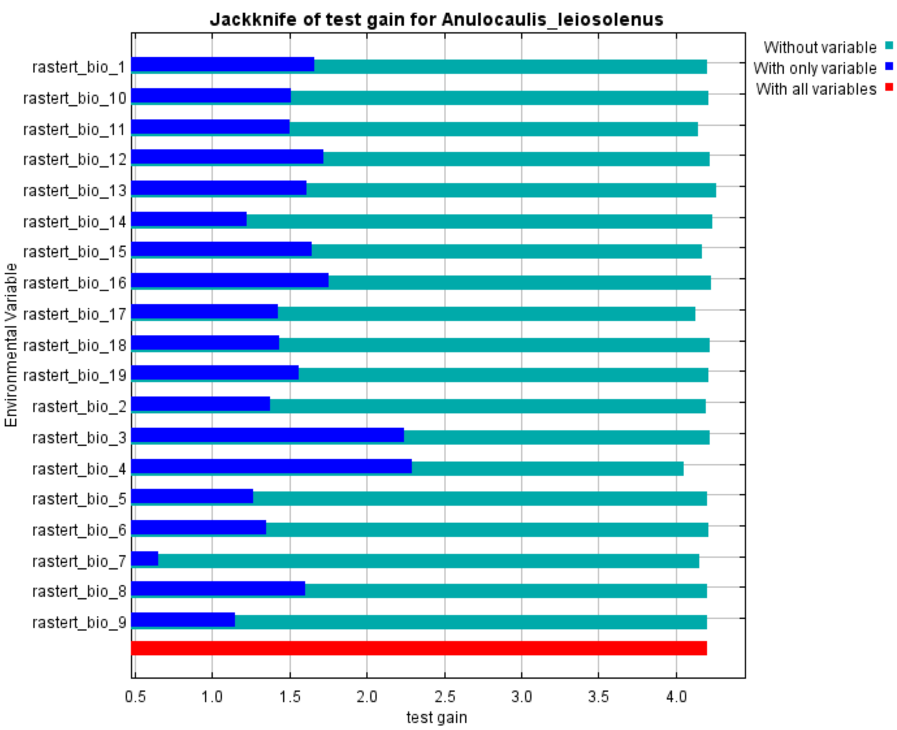
The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution
rastert_bio_3	58.4
rastert_bio_13	19.1
rastert_bio_6	6.2
rastert_bio_11	2.9
rastert_bio_1	2.7
rastert_bio_7	1.9
rastert_bio_5	1.7
rastert_bio_4	1.6
rastert_bio_8	1.6
rastert_bio_2	1.3
rastert_bio_17	0.8
rastert_bio_18	0.7
rastert_bio_12	0.5
rastert_bio_15	0.4
rastert_bio_10	0.2
rastert_bio_14	0
rastert_bio_19	0
rastert_bio_16	0
rastert_bio_9	0

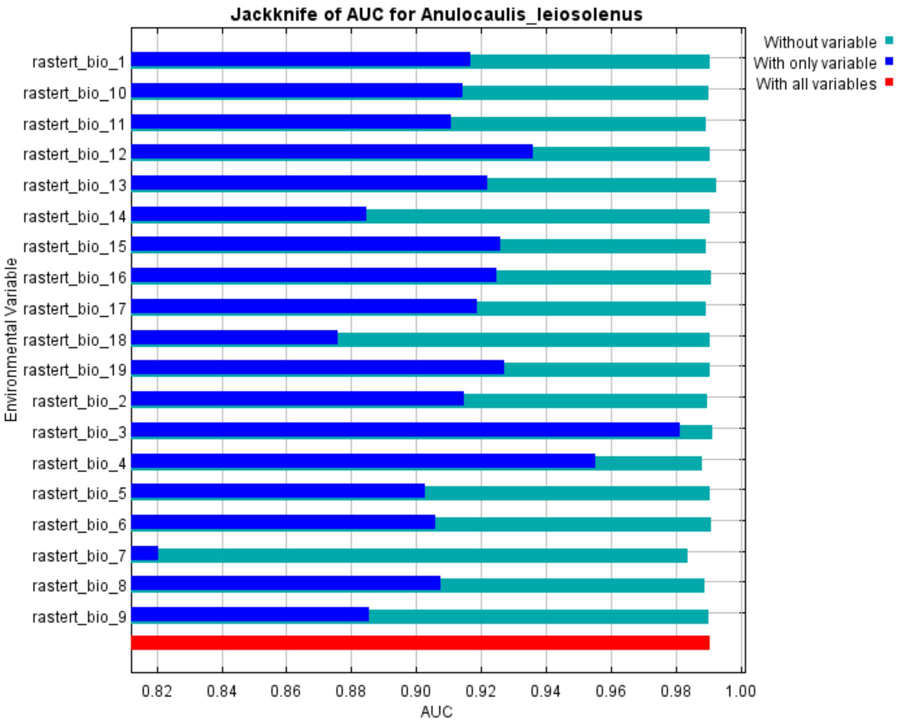
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is rastert_bio_3, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is rastert_bio_7, which therefore appears to have the most information that isn't present in the other variables.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Raw data outputs and control parameters

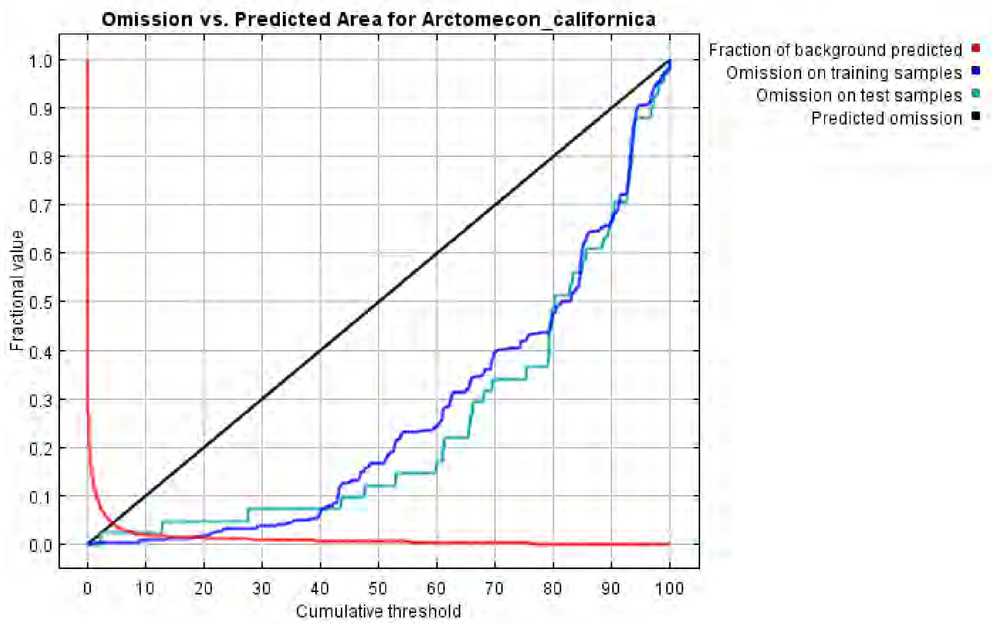
Regularized training gain is 3.716, training AUC is 0.996, unregularized training gain is 4.039.
Unregularized test gain is 4.203. Test AUC is 0.990, standard deviation is 0.008 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm terminated after 500 iterations (13 seconds).

The follow parameters and settings were used during the run:
67 presence records used for training, 7 for testing.
10067 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used (all continuous): rastert_bio_1 rastert_bio_10 rastert_bio_11 rastert_bio_12 rastert_bio_13 rastert_bio_14 rastert_bio_15 rastert_bio_16 rastert_bio_17 rastert_bio_18 rastert_bio_19 rastert_bio_2 rastert_bio_3 rastert_bio_4 rastert_bio_5 rastert_bio_6 rastert_bio_7 rastert_bio_8 rastert_bio_9
Command line:
Feature types used: Linear Quadratic Hinge
Regularization multiplier is 1.0
Regularization values: linear/quadratic/product: 0.144 categorical: 0.250 hinge: 0.500
Species file is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all\gyp_plants_dd.csv
Environmental variables from I:\rare_plant_data\habitat_modeling_maxent\worldclim\bio_30s_esri\bio_ascii_sw
Output directory is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all
Output format is Logistic
Output file type is .asc
Maximum iterations is 500
Convergence threshold is 1.0E-5
Random test percentage is 10
Jackknife selected
Remove duplicates selected
Make pictures selected
Create response curves selected

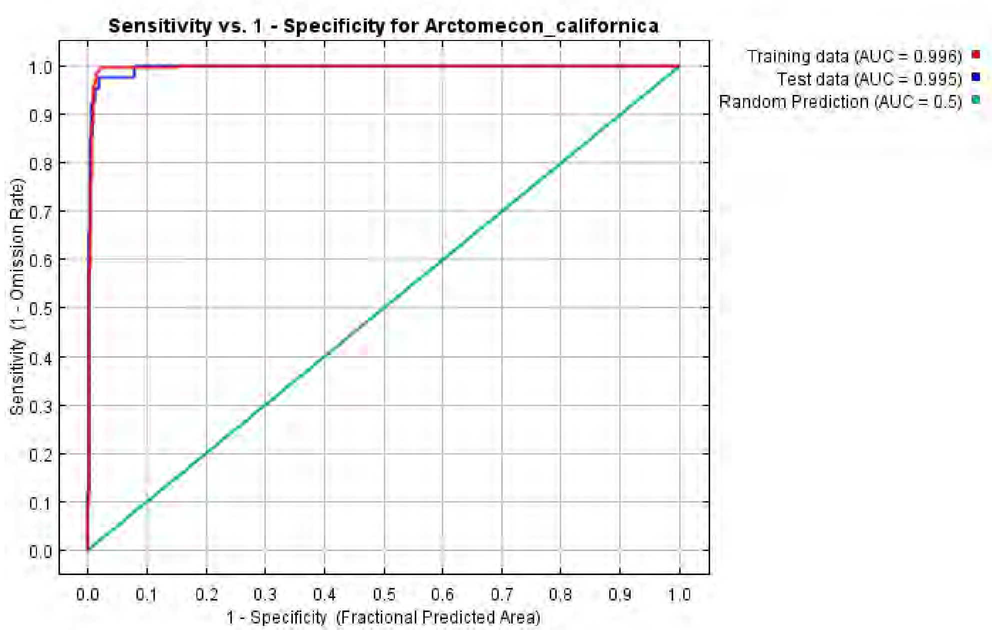
Maxent model for Arctomecon_californica

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.987 rather than 1; in practice the test AUC may exceed this bound.

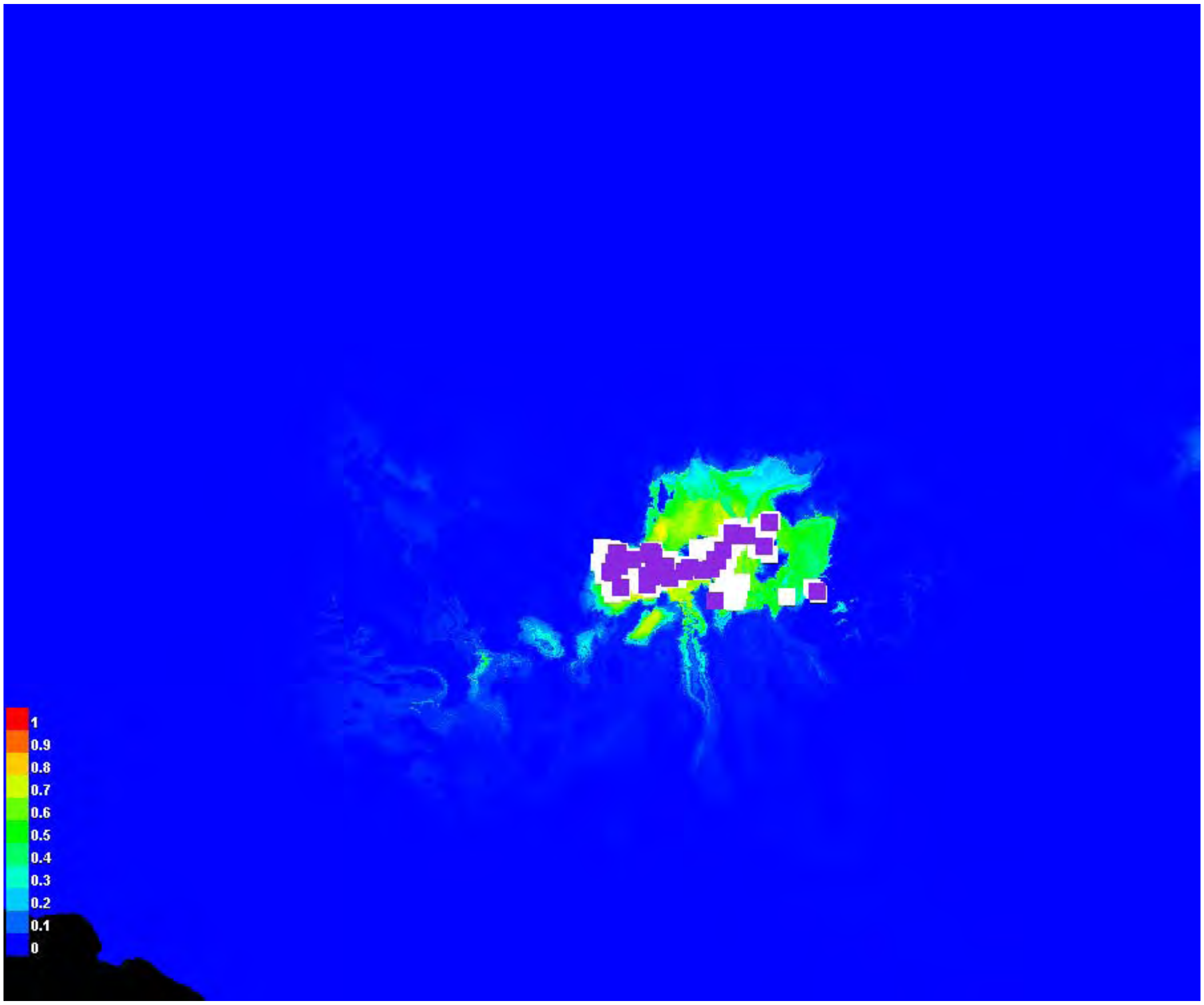


Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.004	Fixed cumulative value	0.128	0.003	0.000	6.01E-63
5.000	0.032	Fixed cumulative value	0.038	0.003	0.024	0E0
10.000	0.208	Fixed cumulative value	0.020	0.011	0.024	0E0
0.690	0.003	Minimum training presence	0.153	0.000	0.000	1.773E-51
42.975	0.582	10 percentile training presence	0.008	0.099	0.073	0E0
18.172	0.379	Equal training sensitivity and specificity	0.015	0.016	0.049	0E0
8.912	0.125	Maximum training sensitivity plus specificity	0.022	0.005	0.024	0E0
7.845	0.096	Equal test sensitivity and specificity	0.024	0.005	0.024	0E0
12.782	0.284	Maximum test sensitivity plus specificity	0.018	0.011	0.024	0E0
2.294	0.011	Balance training omission, predicted area and threshold value	0.076	0.003	0.024	0E0
6.808	0.063	Equate entropy of thresholded and non-thresholded distributions	0.028	0.005	0.024	0E0

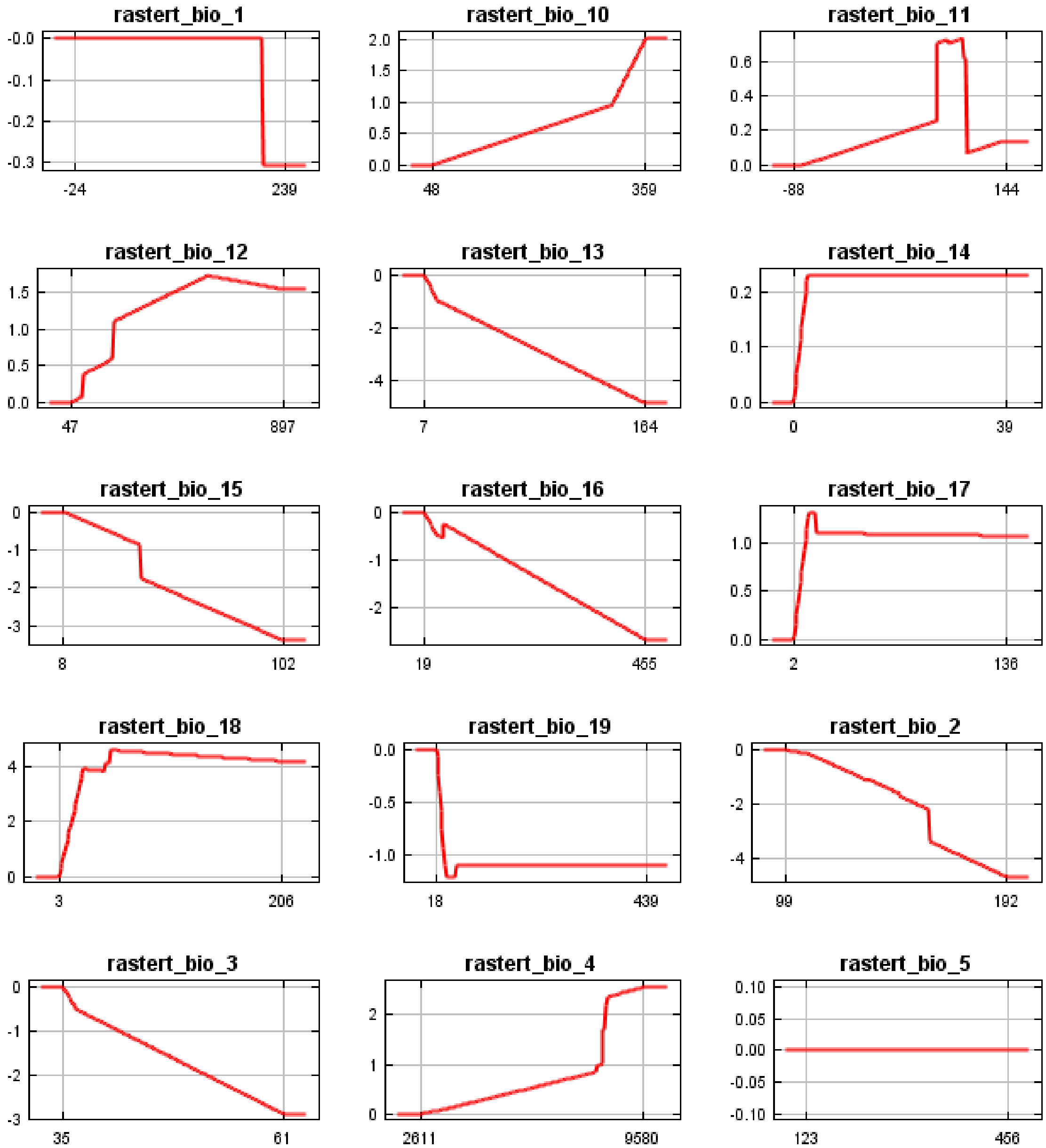
Pictures of the model

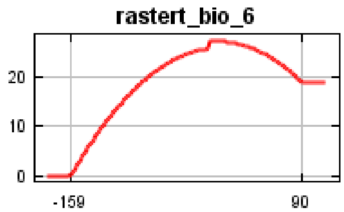
This is a representation of the Maxent model for Arctomecon_californica. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

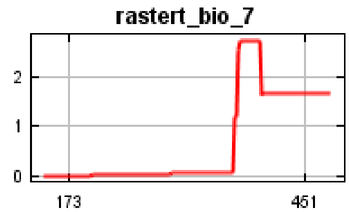


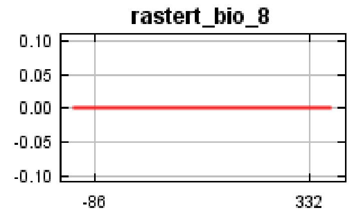
Response curves

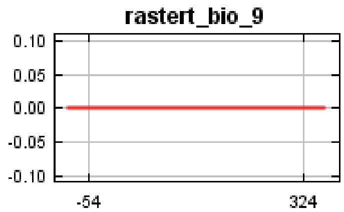
These curves show how each environmental variable affects the Maxent prediction. The (raw) Maxent model has the form exp(...)/constant, and the curves show how the exponent changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version.









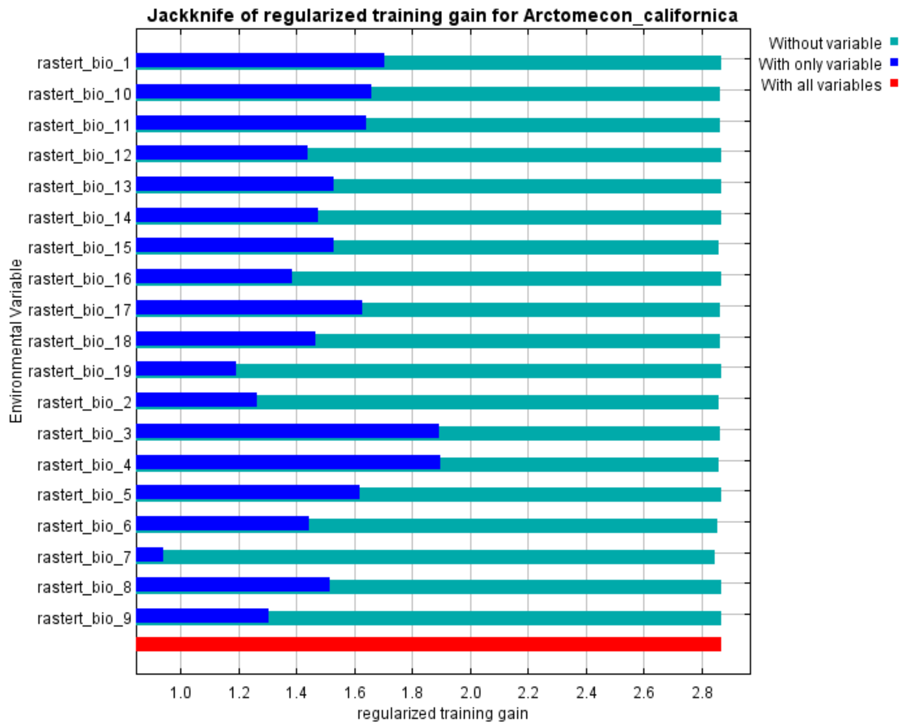


Analysis of variable contributions

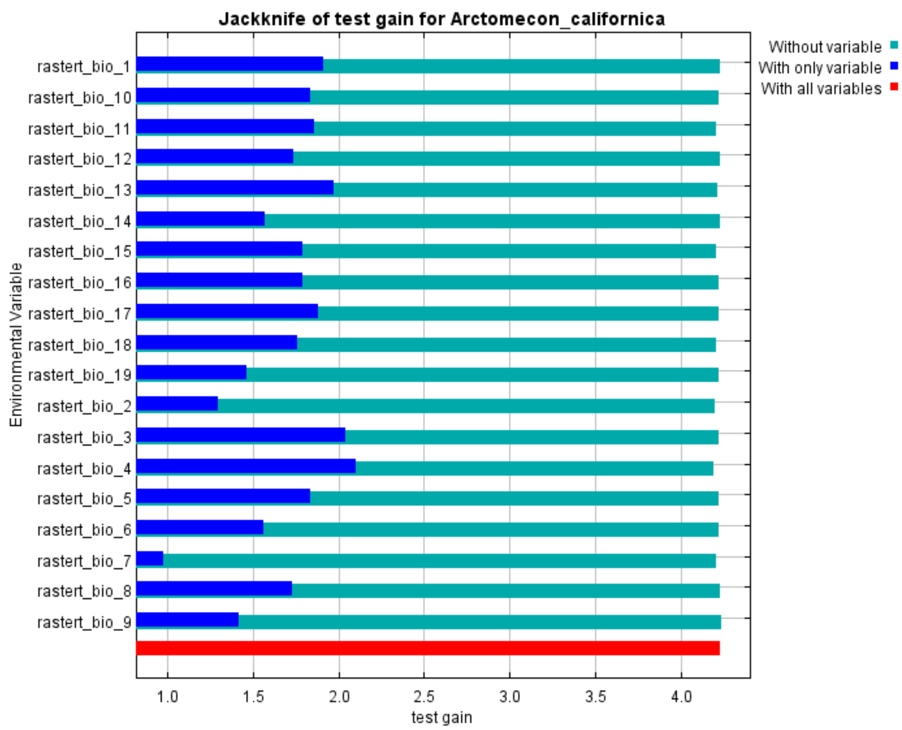
The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution
rastert_bio_4	34.2
rastert_bio_10	22.4
rastert_bio_3	12.3
rastert_bio_2	10.9
rastert_bio_6	7.3
rastert_bio_15	5.5
rastert_bio_7	3.2
rastert_bio_11	2.1
rastert_bio_1	0.5
rastert_bio_13	0.5
rastert_bio_16	0.4
rastert_bio_18	0.2
rastert_bio_19	0.2
rastert_bio_12	0.1
rastert_bio_17	0.1
rastert_bio_14	0
rastert_bio_9	0
rastert_bio_5	0
rastert_bio_8	0

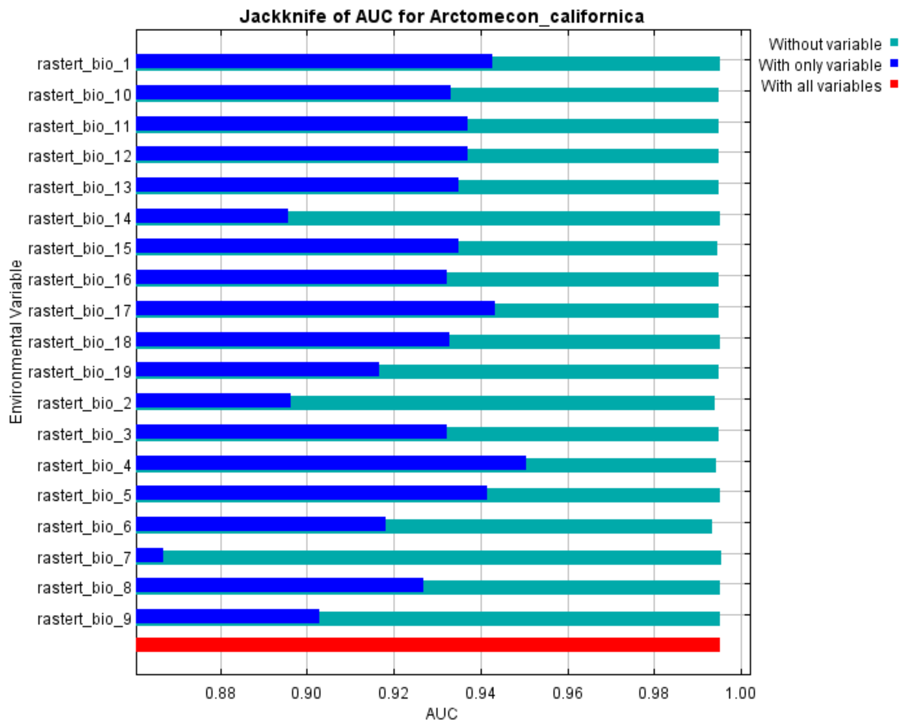
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is rastert_bio_4, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is rastert_bio_7, which therefore appears to have the most information that isn't present in the other variables.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Raw data outputs and control parameters

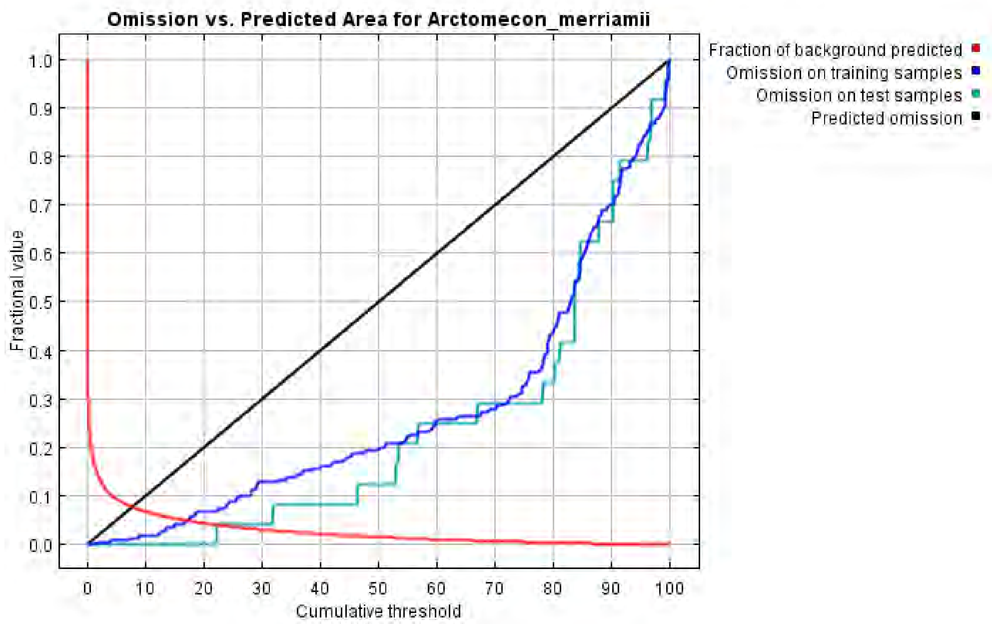
Regularized training gain is 2.868, training AUC is 0.996, unregularized training gain is 2.958.
Unregularized test gain is 4.233. Test AUC is 0.995, standard deviation is 0.002 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm terminated after 500 iterations (13 seconds).

The follow parameters and settings were used during the run:
374 presence records used for training, 41 for testing.
10374 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used (all continuous): rastert_bio_1 rastert_bio_10 rastert_bio_11 rastert_bio_12 rastert_bio_13 rastert_bio_14 rastert_bio_15 rastert_bio_16 rastert_bio_17 rastert_bio_18 rastert_bio_19 rastert_bio_2 rastert_bio_3 rastert_bio_4 rastert_bio_5 rastert_bio_6 rastert_bio_7 rastert_bio_8 rastert_bio_9
Command line:
Feature types used: Linear Quadratic Product Threshold Hinge
Regularization multiplier is 1.0
Regularization values: linear/quadratic/product: 0.050 categorical: 0.250 threshold: 1.000 hinge: 0.500
Species file is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all\gyp_plants_dd.csv
Environmental variables from I:\rare_plant_data\habitat_modeling_maxent\worldclim\bio_30s_esri\bio_ascii_sw
Output directory is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all
Output format is Logistic
Output file type is .asc
Maximum iterations is 500
Convergence threshold is 1.0E-5
Random test percentage is 10
Jackknife selected
Remove duplicates selected
Make pictures selected
Create response curves selected

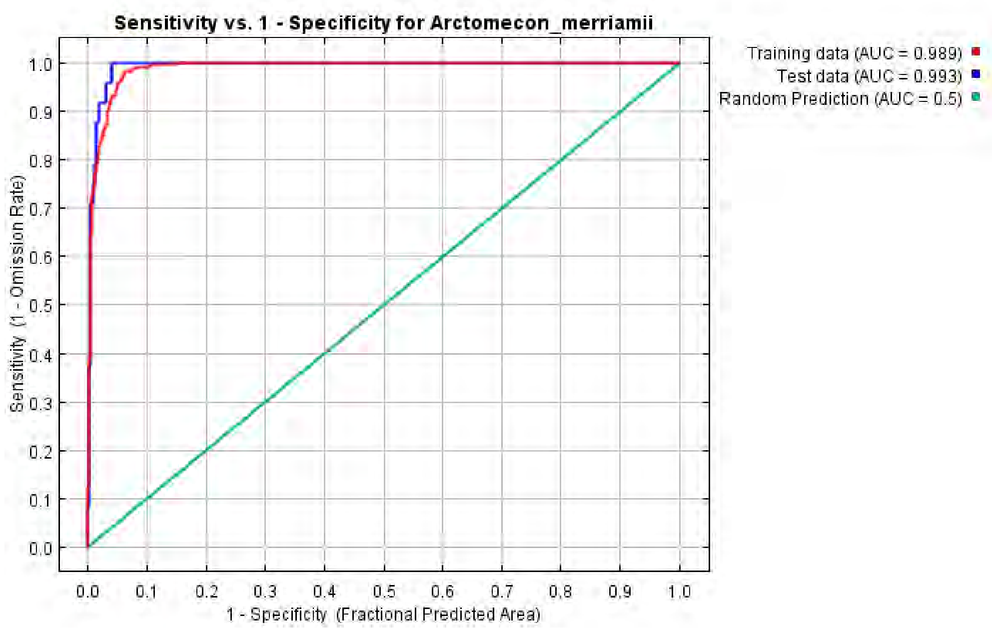
Maxent model for Arctomecon_merriamii

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.971 rather than 1; in practice the test AUC may exceed this bound.

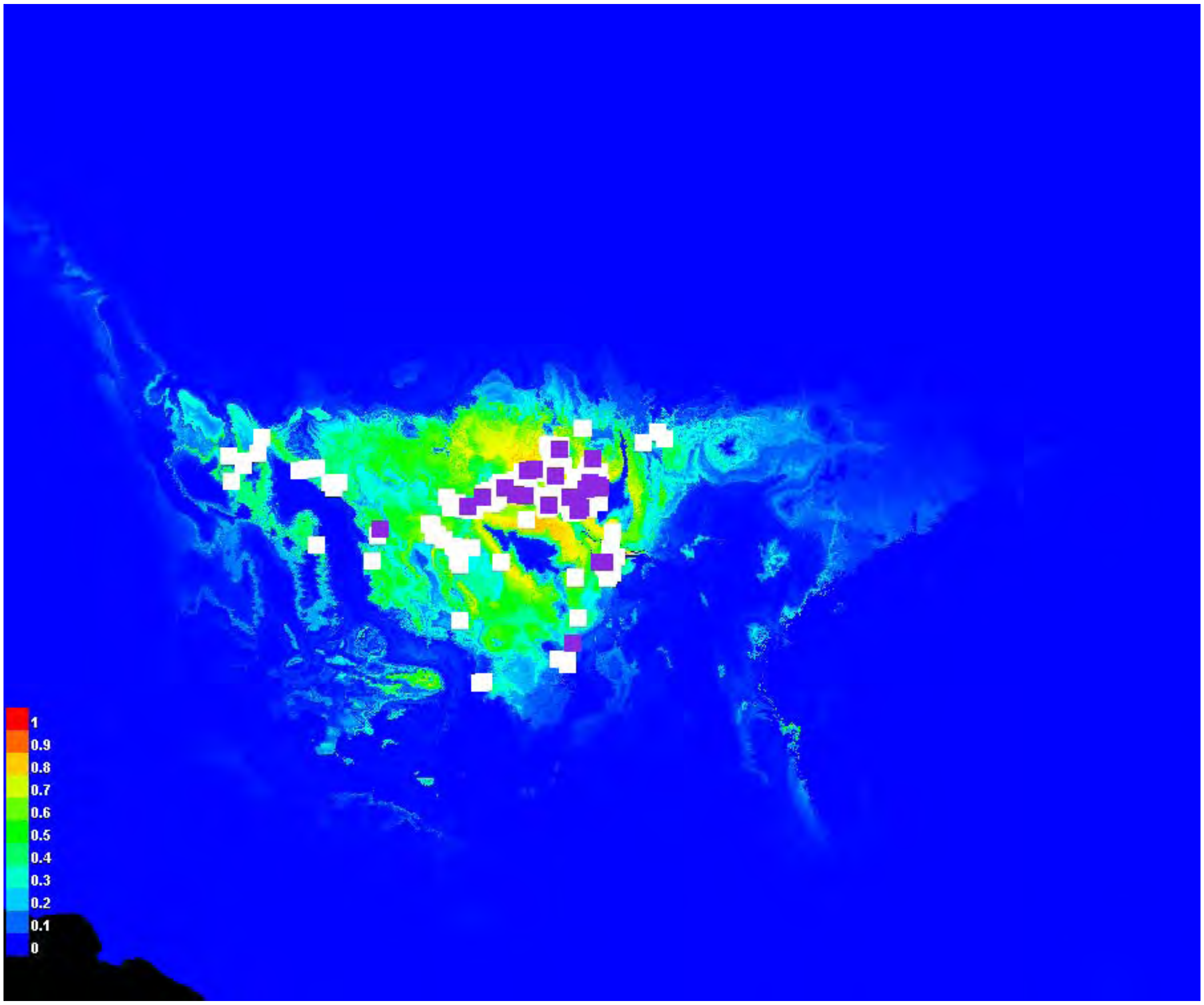


Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.013	Fixed cumulative value	0.174	0.000	0.000	6.355E-19
5.000	0.082	Fixed cumulative value	0.095	0.009	0.000	2.639E-25
10.000	0.171	Fixed cumulative value	0.068	0.019	0.000	9.555E-29
1.513	0.021	Minimum training presence	0.151	0.000	0.000	2.071E-20
26.185	0.371	10 percentile training presence	0.035	0.097	0.042	1.048E-32
16.952	0.264	Equal training sensitivity and specificity	0.050	0.051	0.000	5.681E-32
12.238	0.207	Maximum training sensitivity plus specificity	0.061	0.019	0.000	7.046E-30
22.257	0.337	Equal test sensitivity and specificity	0.041	0.069	0.042	2.559E-31
22.242	0.337	Maximum test sensitivity plus specificity	0.041	0.069	0.000	4.527E-34
1.513	0.021	Balance training omission, predicted area and threshold value	0.151	0.000	0.000	2.071E-20
8.618	0.148	Equate entropy of thresholded and non-thresholded distributions	0.073	0.014	0.000	5.98E-28

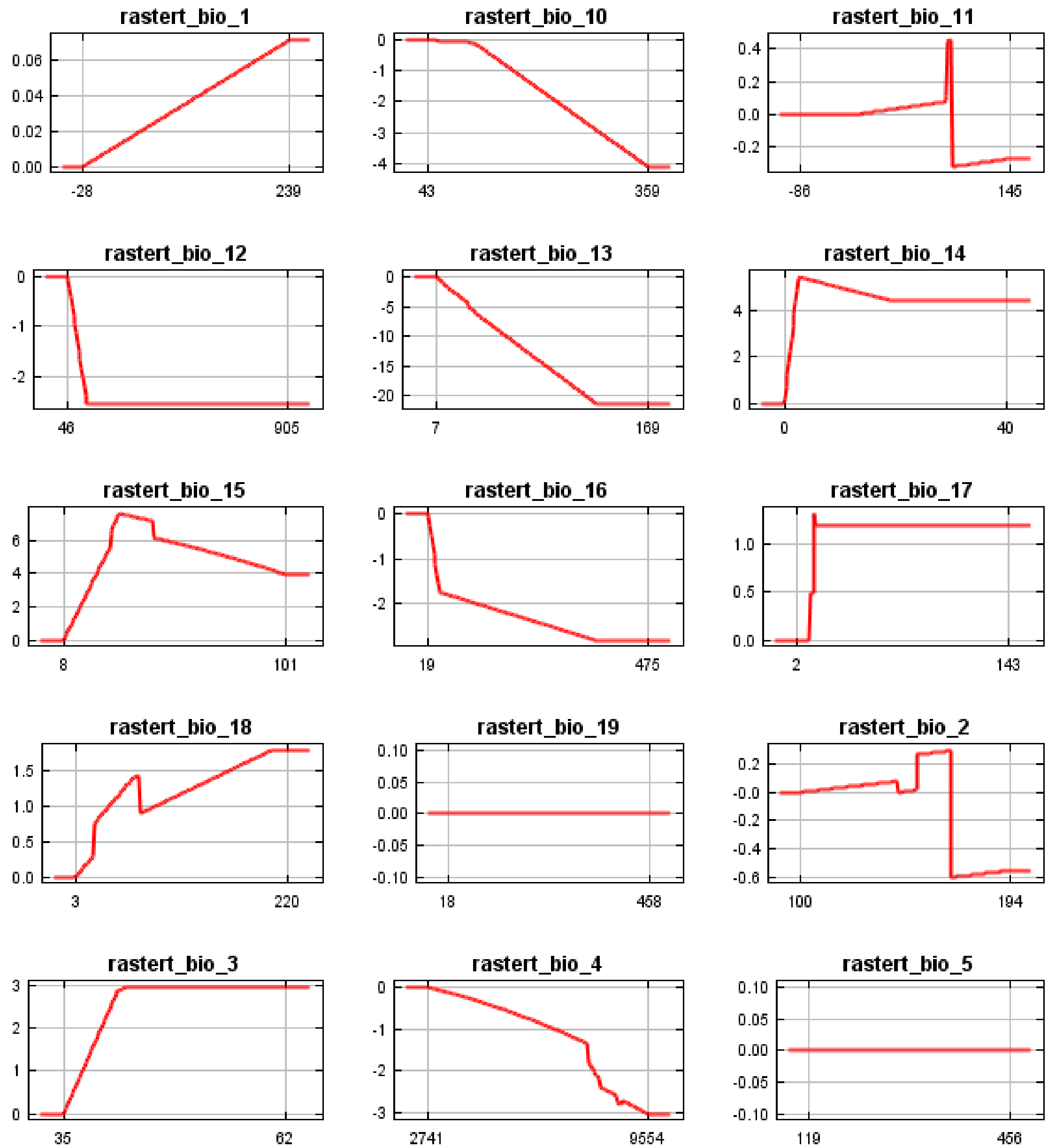
Pictures of the model

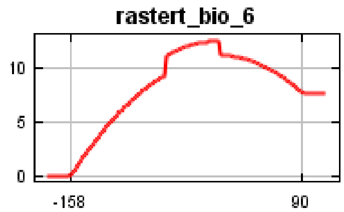
This is a representation of the Maxent model for Arctomecon_merriamii. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

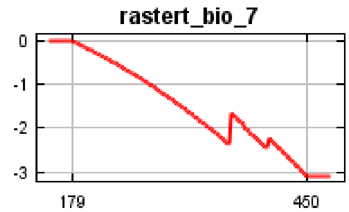


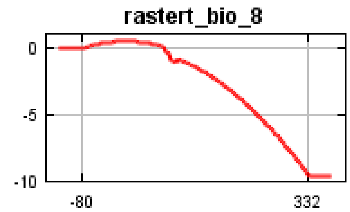
Response curves

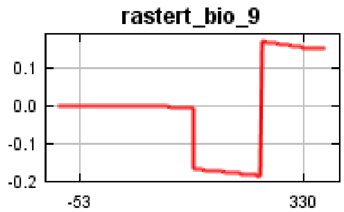
These curves show how each environmental variable affects the Maxent prediction. The (raw) Maxent model has the form exp(...)/constant, and the curves show how the exponent changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version.









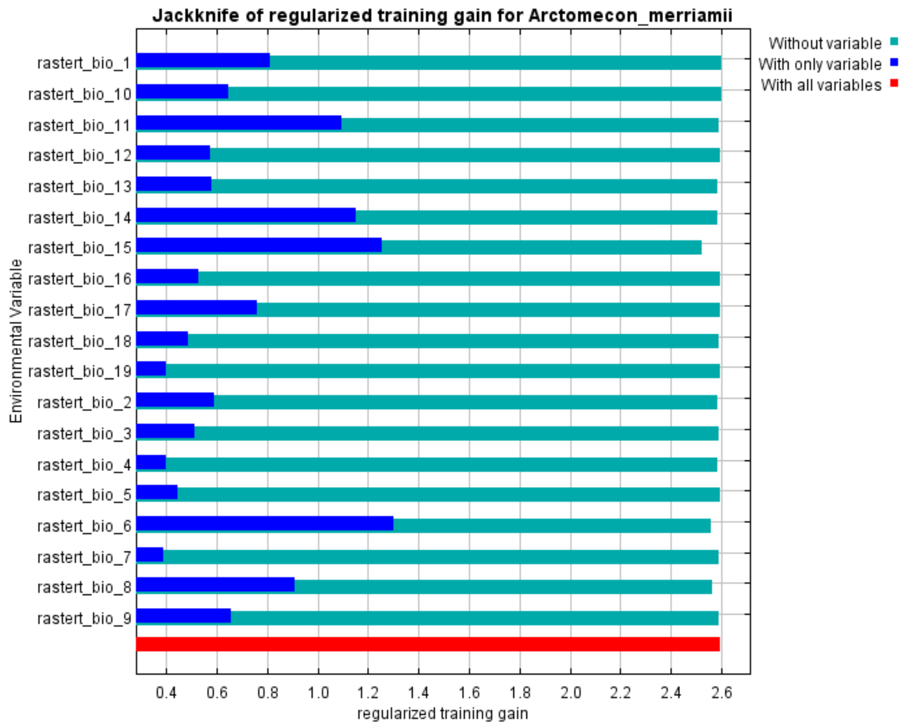


Analysis of variable contributions

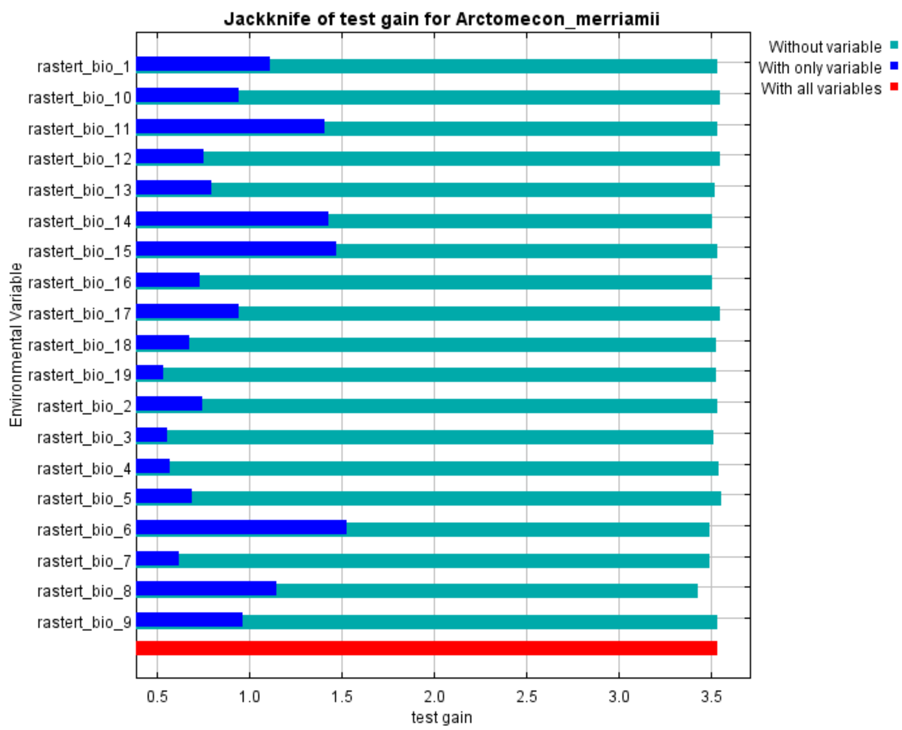
The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution
rastert_bio_6	28.4
rastert_bio_15	23.8
rastert_bio_8	18.8
rastert_bio_13	16.2
rastert_bio_2	3.4
rastert_bio_3	3
rastert_bio_4	2.2
rastert_bio_14	0.9
rastert_bio_17	0.8
rastert_bio_12	0.6
rastert_bio_7	0.6
rastert_bio_10	0.5
rastert_bio_18	0.4
rastert_bio_16	0.4
rastert_bio_11	0.1
rastert_bio_19	0
rastert_bio_9	0
rastert_bio_1	0
rastert_bio_5	0

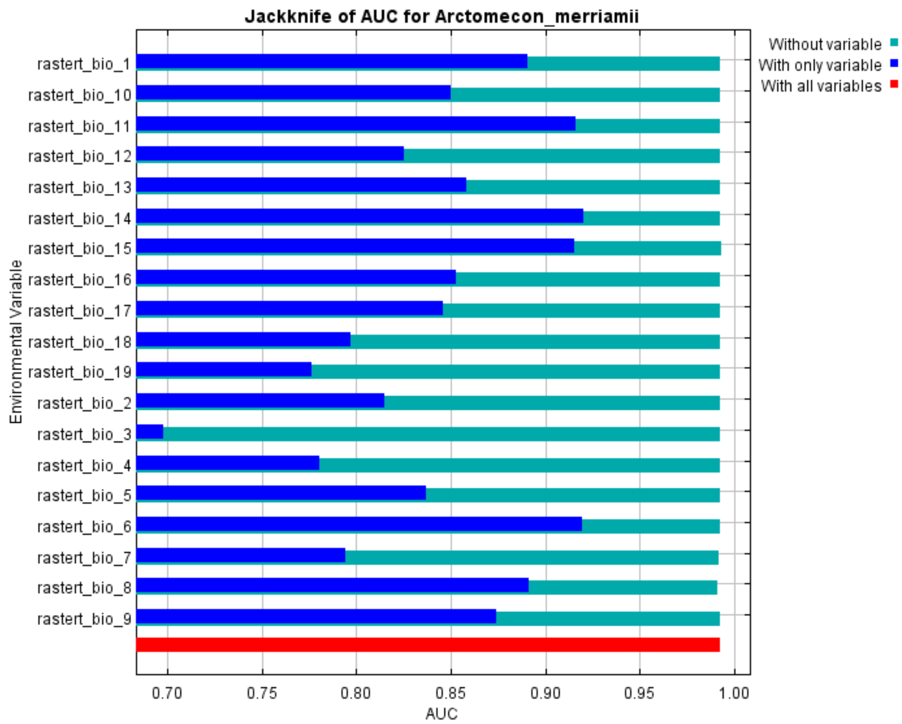
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is rastert_bio_6, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is rastert_bio_15, which therefore appears to have the most information that isn't present in the other variables.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Raw data outputs and control parameters

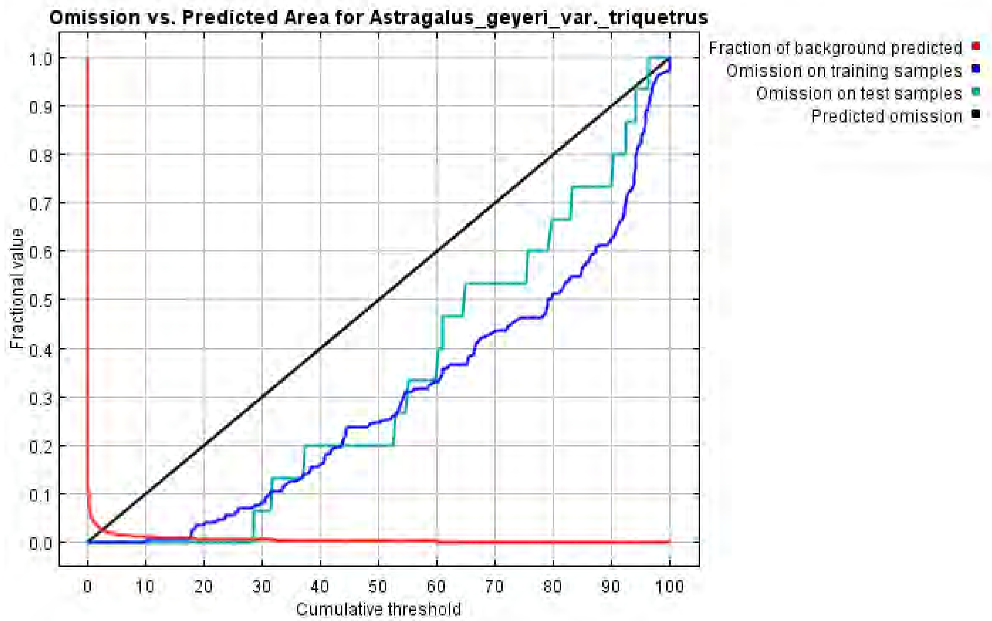
Regularized training gain is 2.598, training AUC is 0.989, unregularized training gain is 2.823.
Unregularized test gain is 3.533. Test AUC is 0.993, standard deviation is 0.002 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm terminated after 500 iterations (13 seconds).

The follow parameters and settings were used during the run:
216 presence records used for training, 24 for testing.
10216 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used (all continuous): rastert_bio_1 rastert_bio_10 rastert_bio_11 rastert_bio_12 rastert_bio_13 rastert_bio_14 rastert_bio_15 rastert_bio_16 rastert_bio_17 rastert_bio_18 rastert_bio_19 rastert_bio_2 rastert_bio_3 rastert_bio_4 rastert_bio_5 rastert_bio_6 rastert_bio_7 rastert_bio_8 rastert_bio_9
Command line:
Feature types used: Linear Quadratic Product Threshold Hinge
Regularization multiplier is 1.0
Regularization values: linear/quadratic/product: 0.050 categorical: 0.250 threshold: 1.000 hinge: 0.500
Species file is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all\gyp_plants_dd.csv
Environmental variables from I:\rare_plant_data\habitat_modeling_maxent\worldclim\bio_30s_esri\bio_ascii_sw
Output directory is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all
Output format is Logistic
Output file type is .asc
Maximum iterations is 500
Convergence threshold is 1.0E-5
Random test percentage is 10
Jackknife selected
Remove duplicates selected
Make pictures selected
Create response curves selected

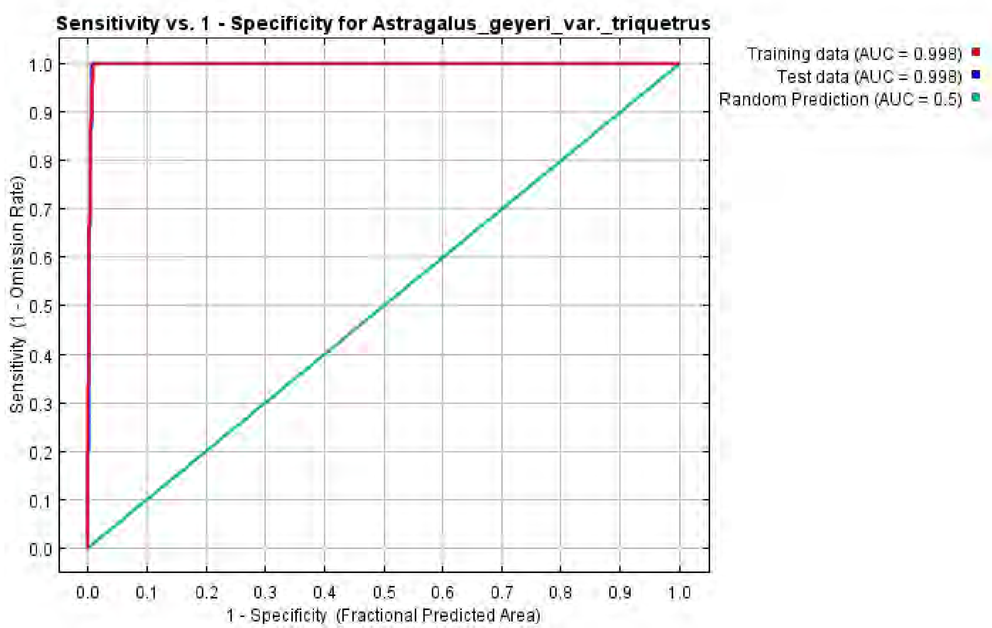
Maxent model for Astragalus_geyeri_var._triquetrus

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.993 rather than 1; in practice the test AUC may exceed this bound.

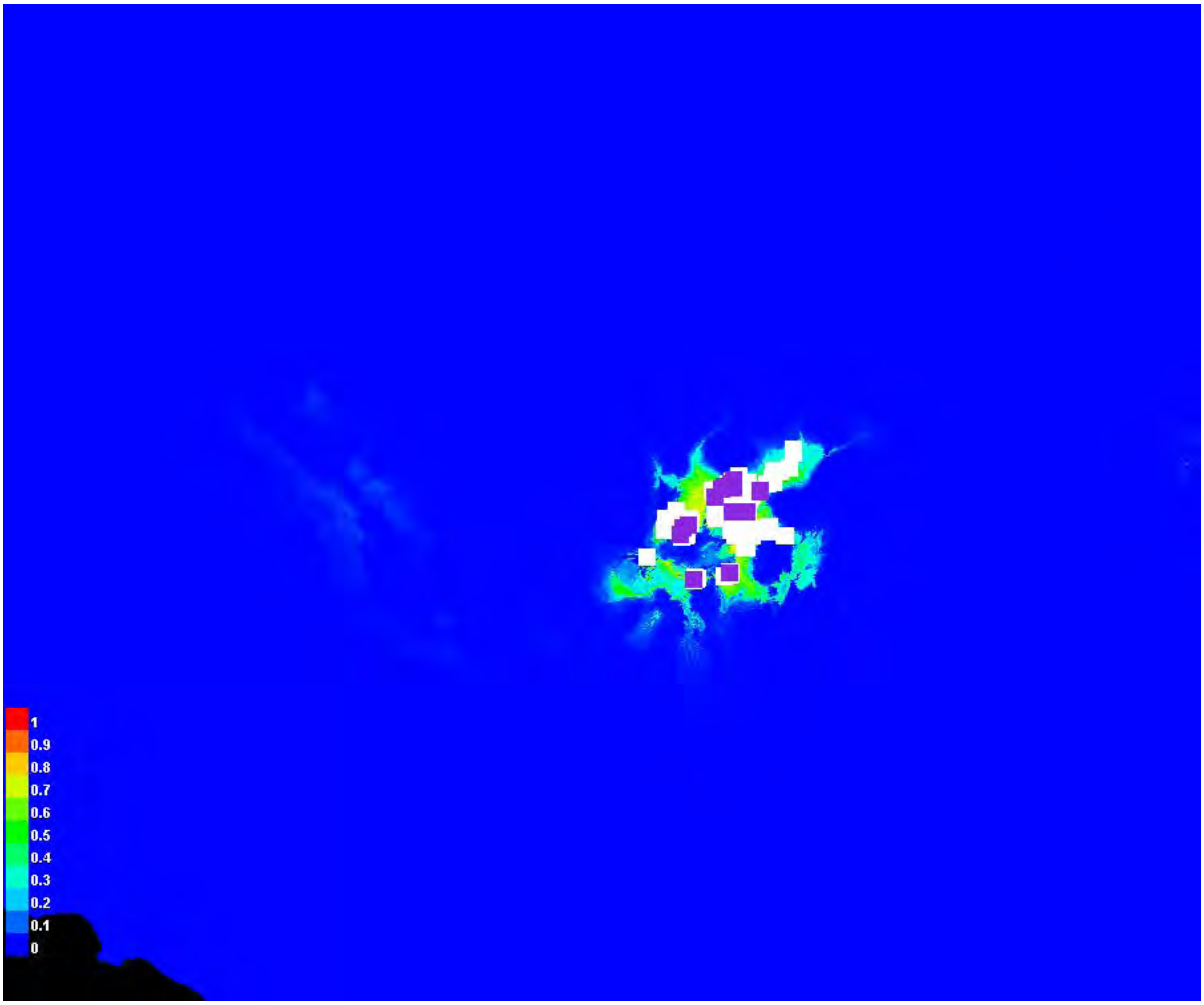


Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.007	Fixed cumulative value	0.045	0.000	0.000	6.942E-21
5.000	0.076	Fixed cumulative value	0.017	0.000	0.000	4.057E-27
10.000	0.188	Fixed cumulative value	0.012	0.000	0.000	2.231E-29
10.248	0.192	Minimum training presence	0.012	0.000	0.000	1.974E-29
31.179	0.506	10 percentile training presence	0.006	0.099	0.067	4.451E-31
17.661	0.301	Equal training sensitivity and specificity	0.009	0.007	0.000	1.238E-31
10.248	0.192	Maximum training sensitivity plus specificity	0.012	0.000	0.000	1.974E-29
28.423	0.463	Equal test sensitivity and specificity	0.006	0.070	0.000	6.025E-34
28.423	0.463	Maximum test sensitivity plus specificity	0.006	0.070	0.000	6.025E-34
0.893	0.006	Balance training omission, predicted area and threshold value	0.048	0.000	0.000	1.603E-20
6.931	0.125	Equate entropy of thresholded and non-thresholded distributions	0.015	0.000	0.000	3.58E-28

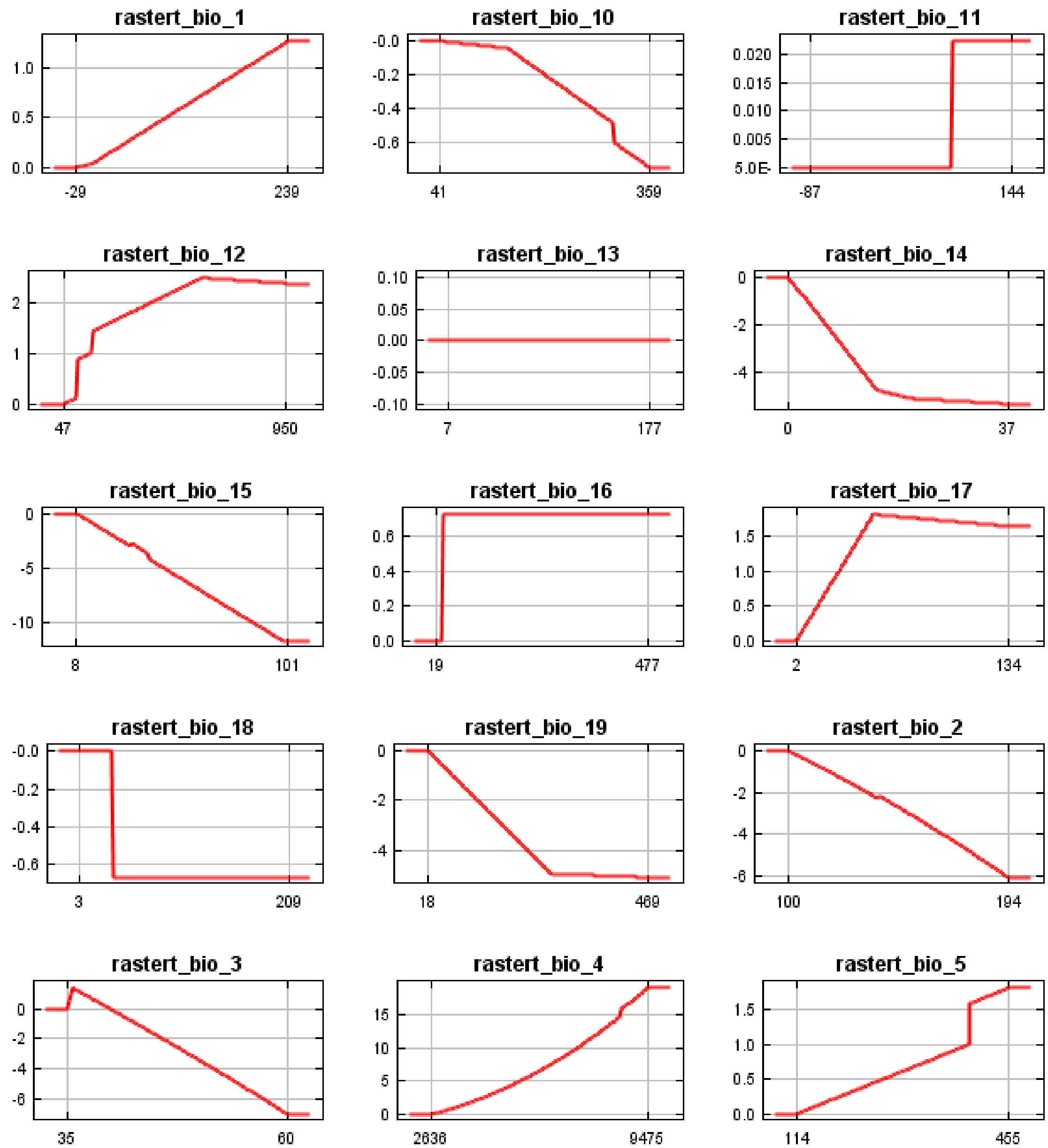
Pictures of the model

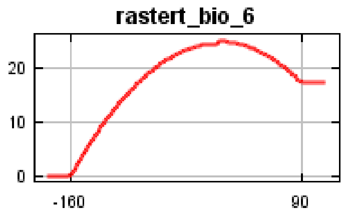
This is a representation of the Maxent model for Astragalus_geyeri_var._triquetrus. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

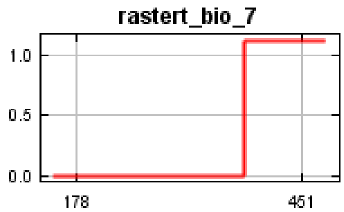


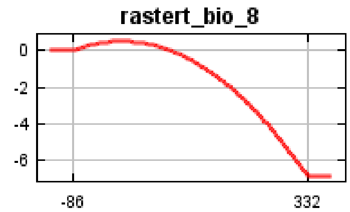
Response curves

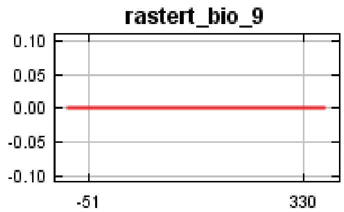
These curves show how each environmental variable affects the Maxent prediction. The (raw) Maxent model has the form exp(...)/constant, and the curves show how the exponent changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version.









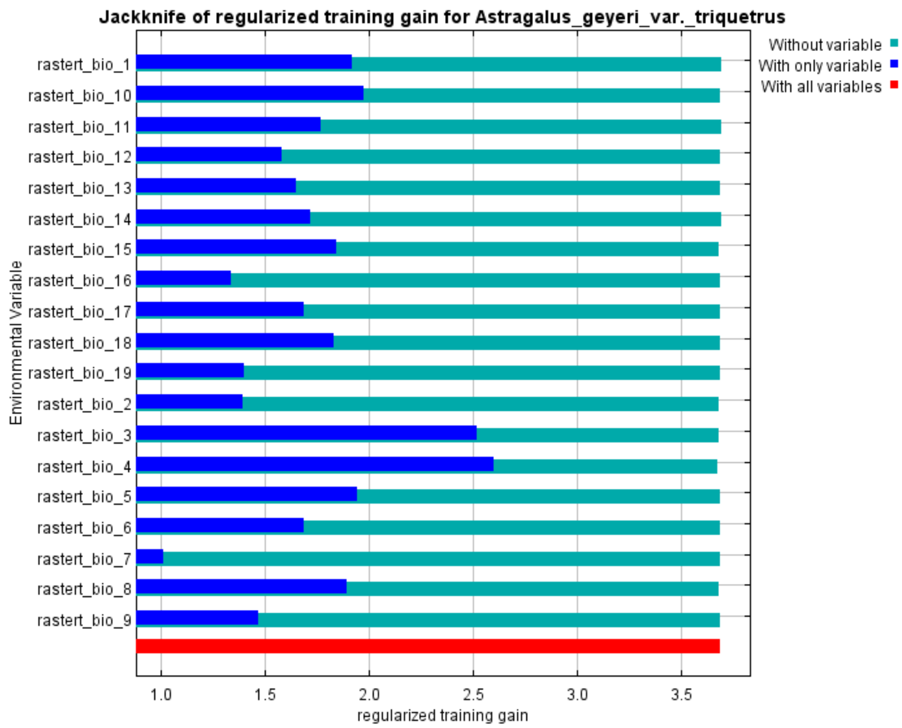


Analysis of variable contributions

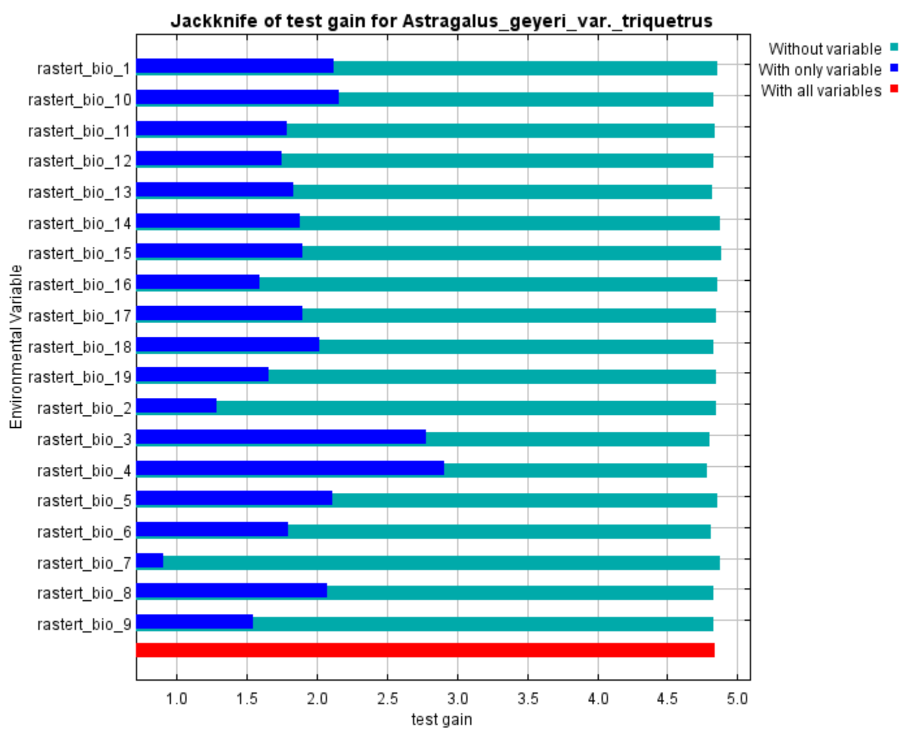
The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution
rastert_bio_4	36.7
rastert_bio_5	22.7
rastert_bio_3	17
rastert_bio_15	7.3
rastert_bio_10	6.9
rastert_bio_2	2.9
rastert_bio_12	2
rastert_bio_7	1.4
rastert_bio_8	1
rastert_bio_6	0.8
rastert_bio_18	0.5
rastert_bio_14	0.3
rastert_bio_1	0.3
rastert_bio_16	0.2
rastert_bio_17	0
rastert_bio_19	0
rastert_bio_11	0
rastert_bio_13	0
rastert_bio_9	0

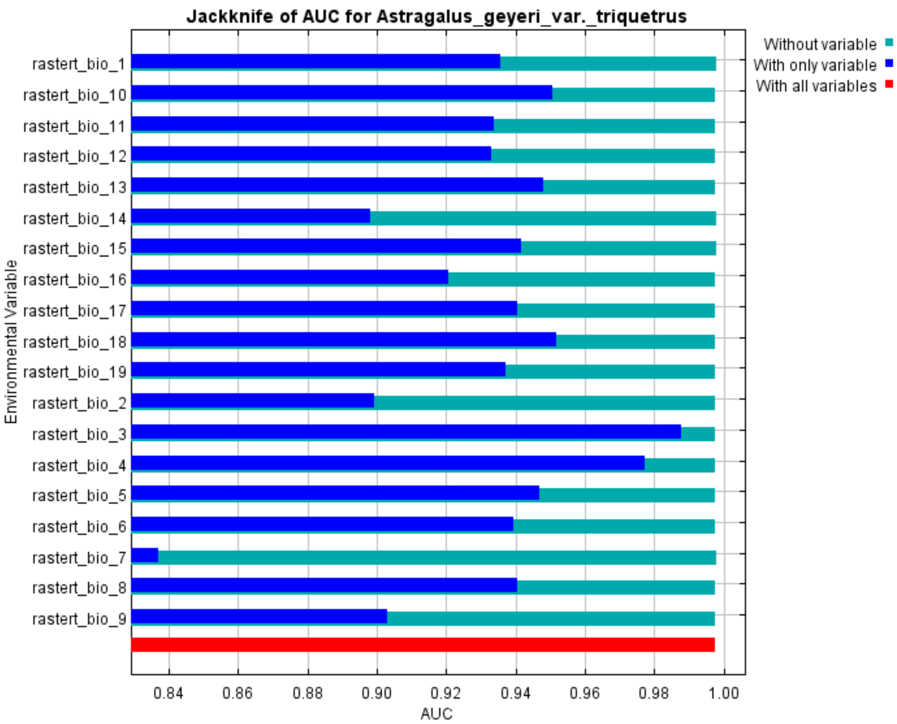
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is rastert_bio_4, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is rastert_bio_4, which therefore appears to have the most information that isn't present in the other variables.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Raw data outputs and control parameters

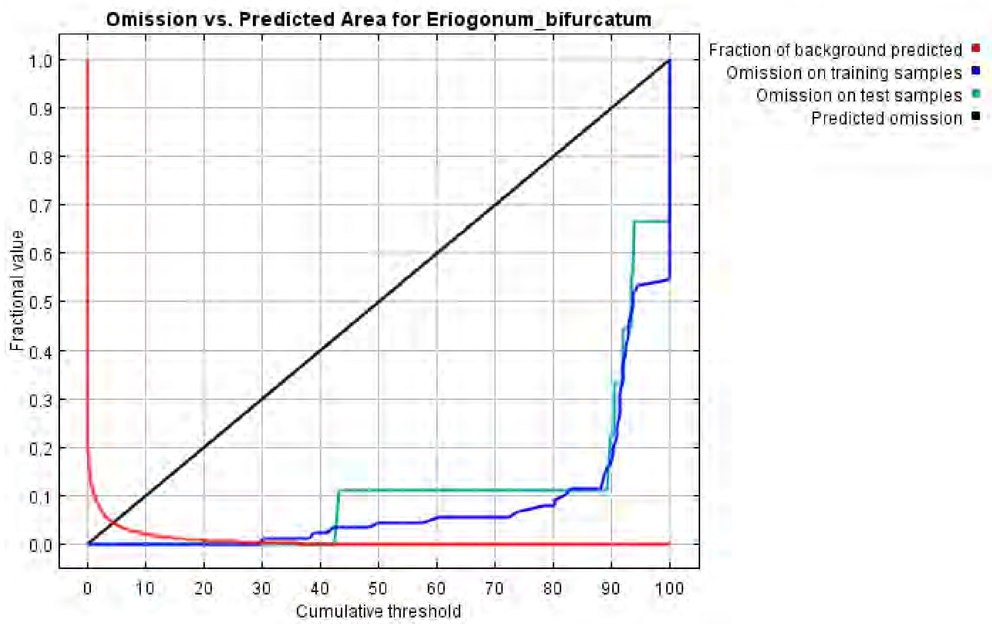
Regularized training gain is 3.686, training AUC is 0.998, unregularized training gain is 3.768.
Unregularized test gain is 4.839. Test AUC is 0.998, standard deviation is 0.001 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm terminated after 500 iterations (12 seconds).

The follow parameters and settings were used during the run:
142 presence records used for training, 15 for testing.
10142 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used (all continuous): rastert_bio_1 rastert_bio_10 rastert_bio_11 rastert_bio_12 rastert_bio_13 rastert_bio_14 rastert_bio_15 rastert_bio_16 rastert_bio_17 rastert_bio_18 rastert_bio_19 rastert_bio_2 rastert_bio_3 rastert_bio_4 rastert_bio_5 rastert_bio_6 rastert_bio_7 rastert_bio_8 rastert_bio_9
Command line:
Feature types used: Linear Quadratic Product Threshold Hinge
Regularization multiplier is 1.0
Regularization values: linear/quadratic/product: 0.050 categorical: 0.250 threshold: 1.000 hinge: 0.500
Species file is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all\sand_plants_dd.csv
Environmental variables from I:\rare_plant_data\habitat_modeling_maxent\worldclim\bio_30s_esri\bio_ascii_sw
Output directory is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all
Output format is Logistic
Output file type is .asc
Maximum iterations is 500
Convergence threshold is 1.0E-5
Random test percentage is 10
Jackknife selected
Remove duplicates selected
Make pictures selected
Create response curves selected

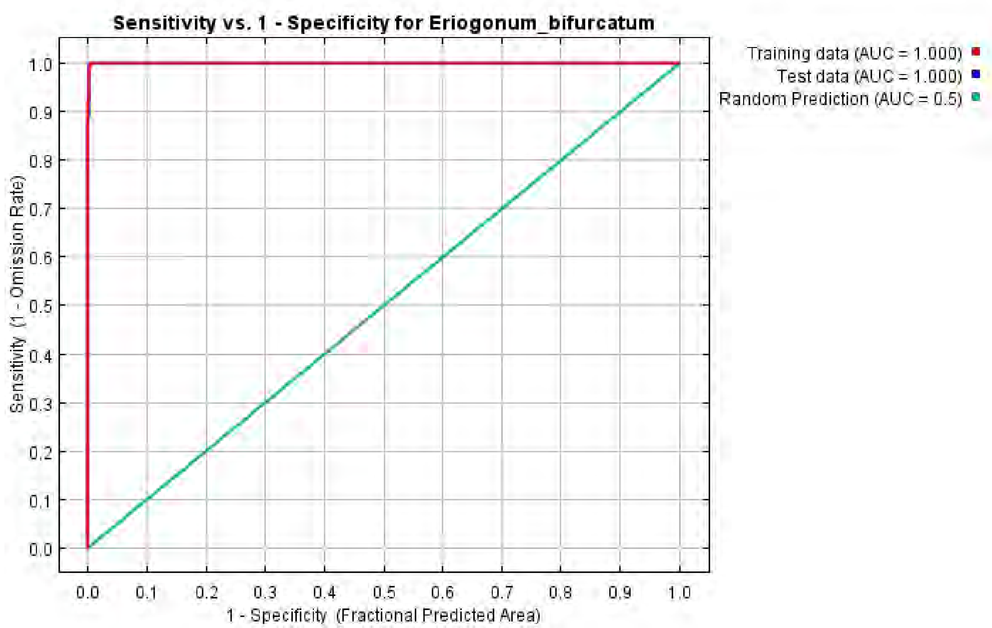
Maxent model for Eriogonum_bifurcatum

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.987 rather than 1; in practice the test AUC may exceed this bound.

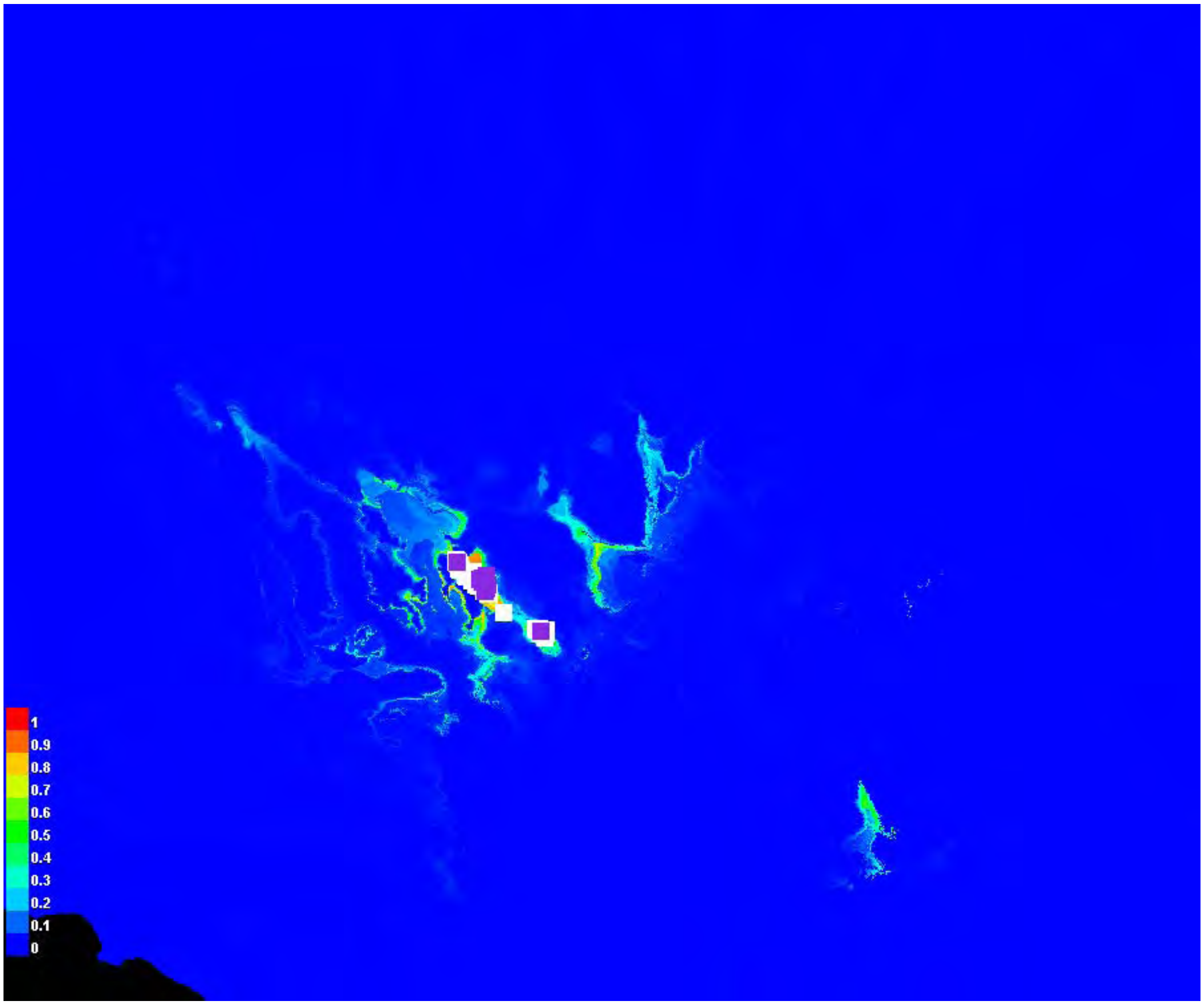


Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.003	Fixed cumulative value	0.108	0.000	0.000	2.016E-9
5.000	0.017	Fixed cumulative value	0.041	0.000	0.000	3.496E-13
10.000	0.048	Fixed cumulative value	0.022	0.000	0.000	1.207E-15
29.890	0.266	Minimum training presence	0.004	0.000	0.000	2.621E-22
80.304	0.844	10 percentile training presence	0.000	0.093	0.111	5.896E-27
29.890	0.266	Equal training sensitivity and specificity	0.004	0.000	0.000	2.621E-22
29.890	0.266	Maximum training sensitivity plus specificity	0.004	0.000	0.000	2.621E-22
42.441	0.556	Equal test sensitivity and specificity	0.002	0.035	0.000	1.984E-25
42.441	0.556	Maximum test sensitivity plus specificity	0.002	0.035	0.000	1.984E-25
1.728	0.005	Balance training omission, predicted area and threshold value	0.084	0.000	0.000	2.173E-10
16.409	0.103	Equate entropy of thresholded and non-thresholded distributions	0.012	0.000	0.000	4.785E-18

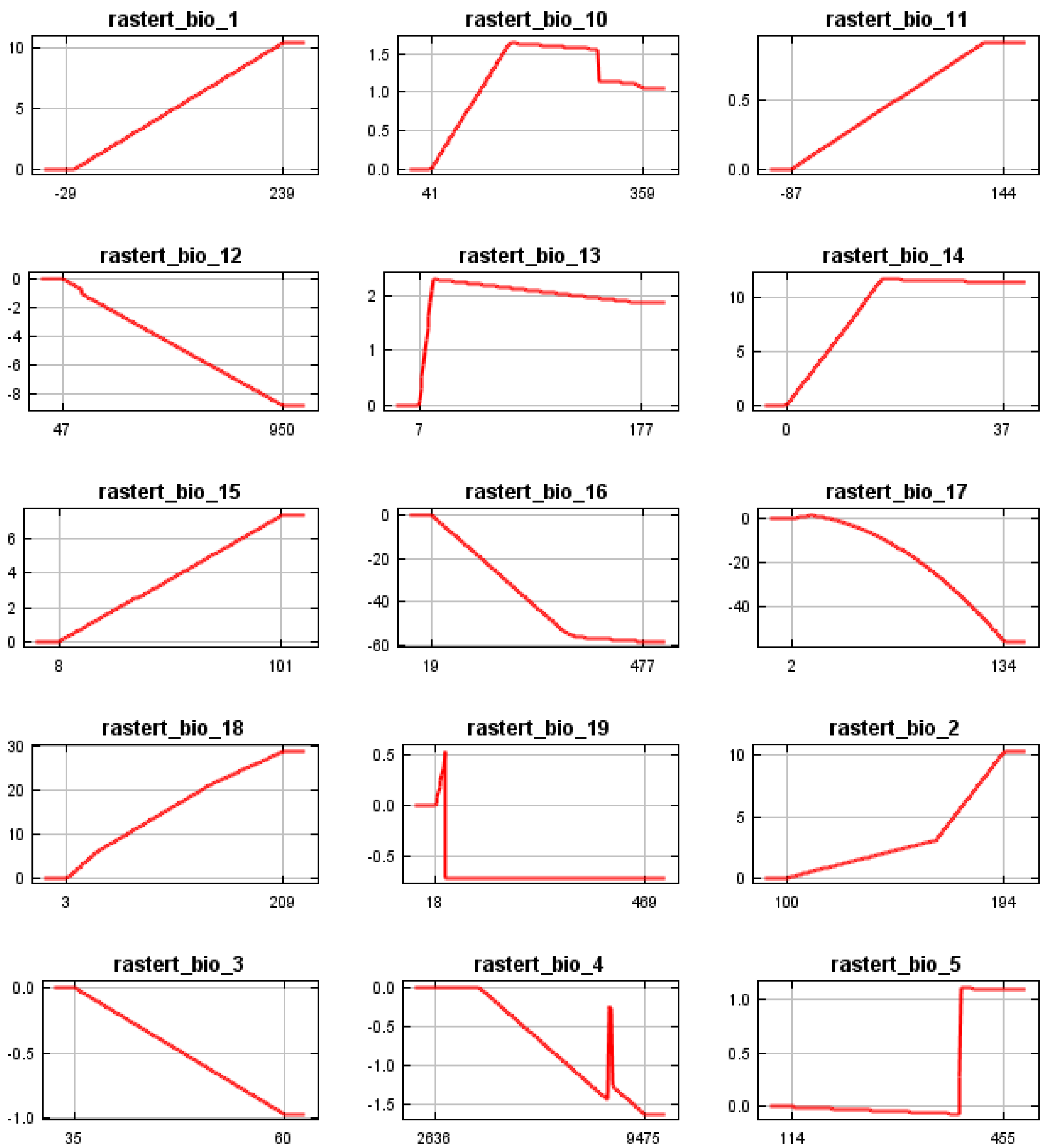
Pictures of the model

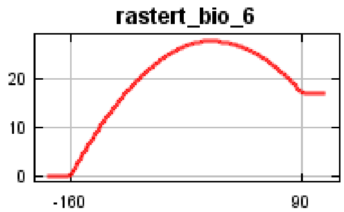
This is a representation of the Maxent model for Eriogonum_bifurcatum. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

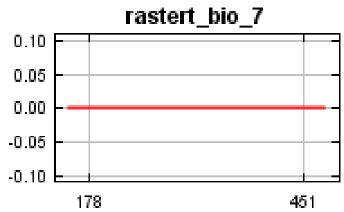


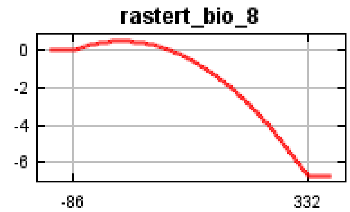
Response curves

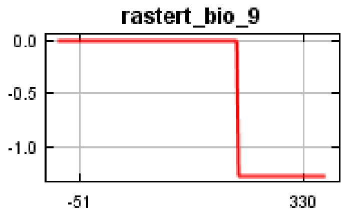
These curves show how each environmental variable affects the Maxent prediction. The (raw) Maxent model has the form exp(...)/constant, and the curves show how the exponent changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version.









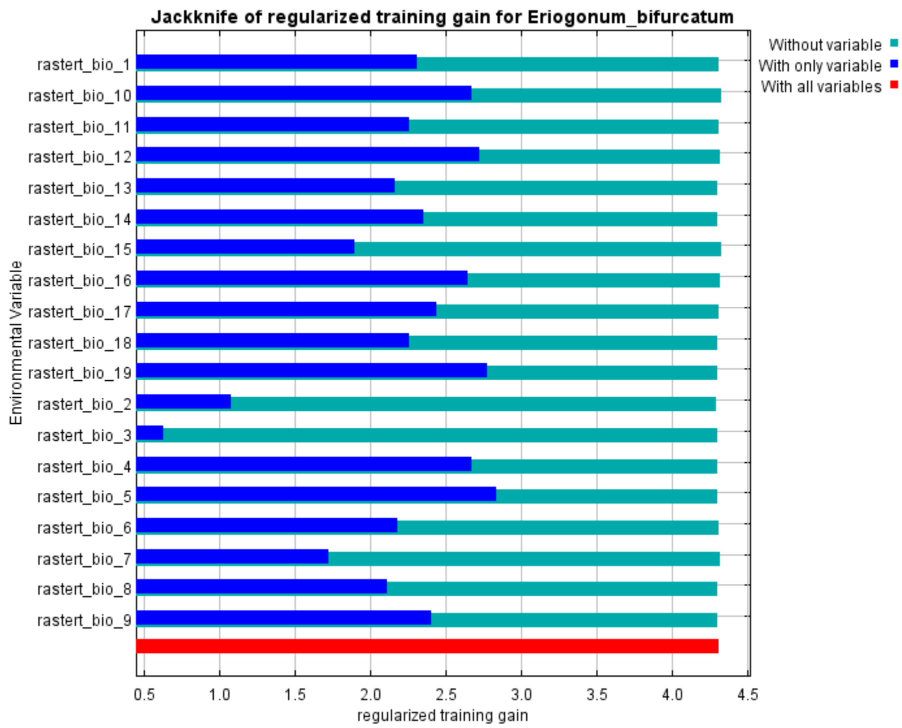


Analysis of variable contributions

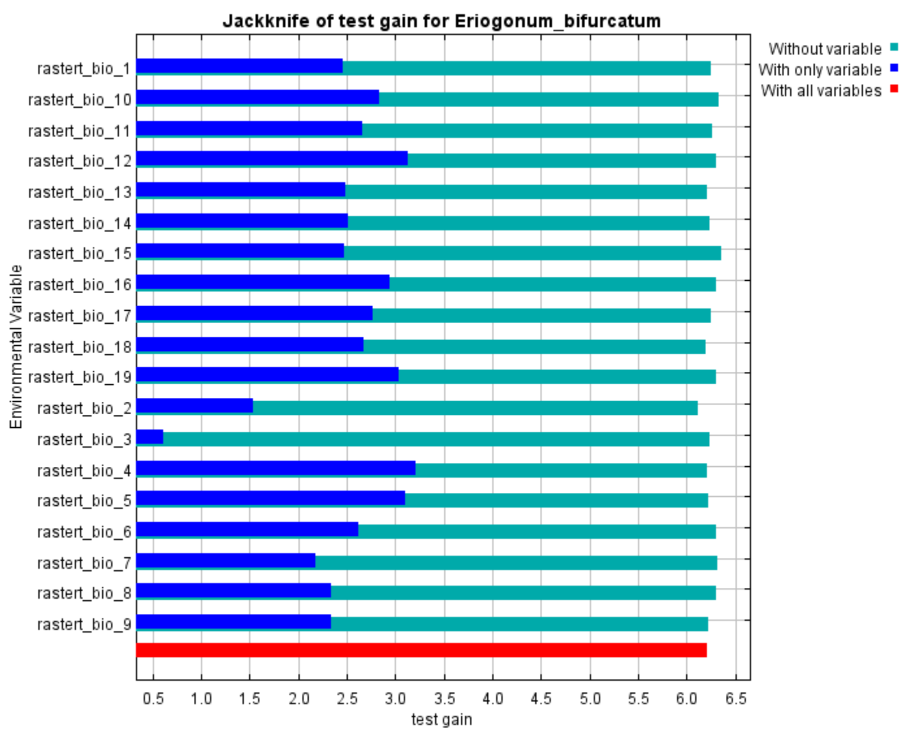
The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution
rastert_bio_12	19.7
rastert_bio_4	19.2
rastert_bio_5	14.7
rastert_bio_9	13.3
rastert_bio_8	10.3
rastert_bio_16	8.9
rastert_bio_19	8.5
rastert_bio_2	2.8
rastert_bio_17	0.7
rastert_bio_15	0.6
rastert_bio_14	0.5
rastert_bio_6	0.3
rastert_bio_3	0.2
rastert_bio_18	0.1
rastert_bio_10	0.1
rastert_bio_13	0
rastert_bio_1	0
rastert_bio_11	0
rastert_bio_7	0

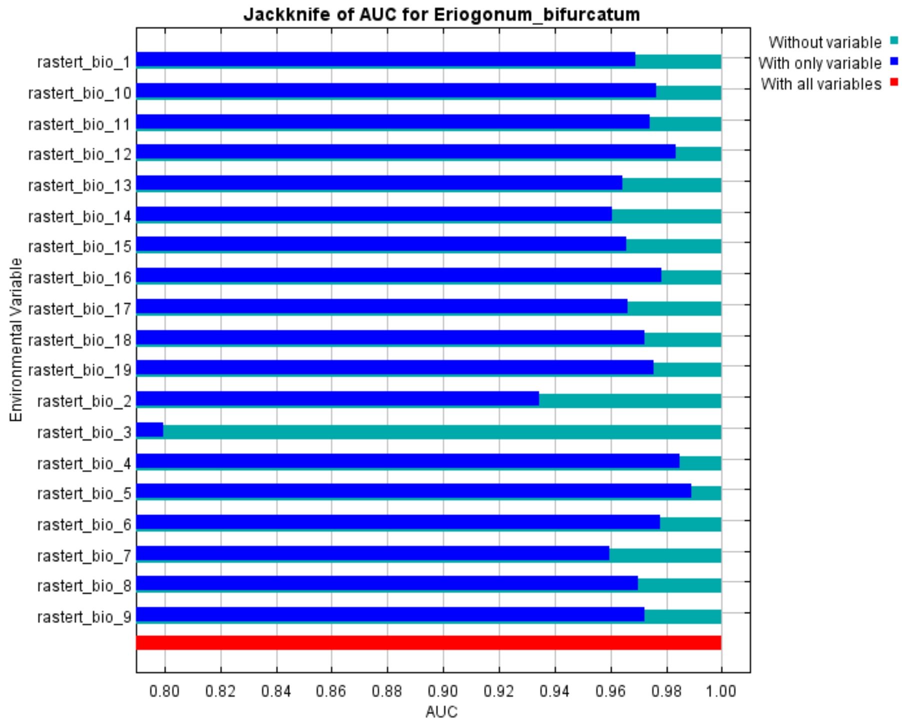
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is rastert_bio_5, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is rastert_bio_2, which therefore appears to have the most information that isn't present in the other variables.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Raw data outputs and control parameters

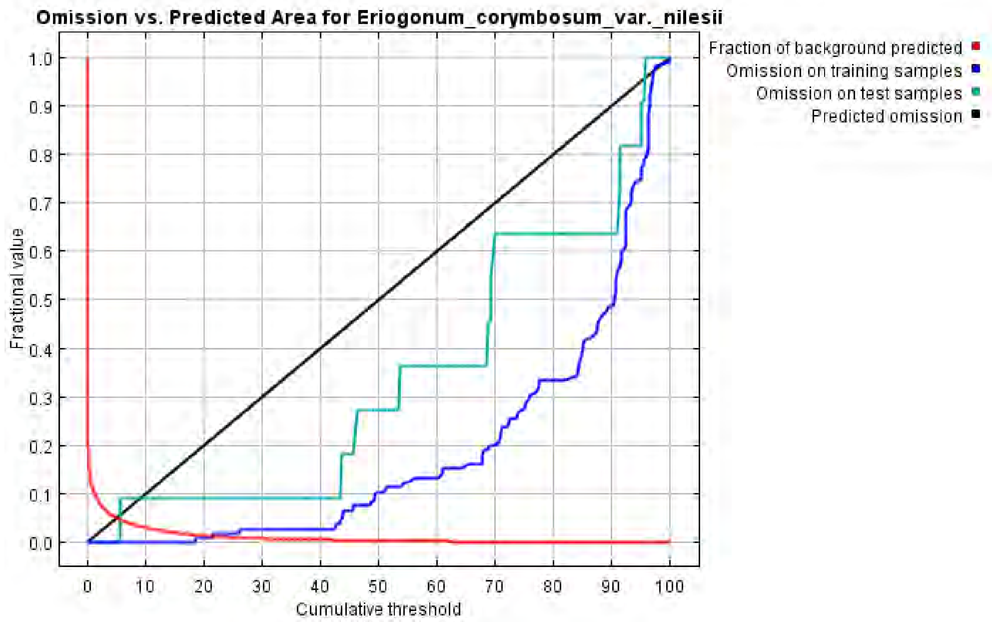
Regularized training gain is 4.316, training AUC is 1.000, unregularized training gain is 4.510.
Unregularized test gain is 6.206. Test AUC is 1.000, standard deviation is 0.000 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm terminated after 500 iterations (12 seconds).

The follow parameters and settings were used during the run:
86 presence records used for training, 9 for testing.
10086 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used (all continuous): rastert_bio_1 rastert_bio_10 rastert_bio_11 rastert_bio_12 rastert_bio_13 rastert_bio_14 rastert_bio_15 rastert_bio_16 rastert_bio_17 rastert_bio_18 rastert_bio_19 rastert_bio_2 rastert_bio_3 rastert_bio_4 rastert_bio_5 rastert_bio_6 rastert_bio_7 rastert_bio_8 rastert_bio_9
Command line:
Feature types used: Linear Quadratic Product Threshold Hinge
Regularization multiplier is 1.0
Regularization values: linear/quadratic/product: 0.150 categorical: 0.250 threshold: 1.140 hinge: 0.500
Species file is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all\sand_plants_dd.csv
Environmental variables from I:\rare_plant_data\habitat_modeling_maxent\worldclim\bio_30s_esri\bio_ascii_sw
Output directory is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all
Output format is Logistic
Output file type is .asc
Maximum iterations is 500
Convergence threshold is 1.0E-5
Random test percentage is 10
Jackknife selected
Remove duplicates selected
Make pictures selected
Create response curves selected

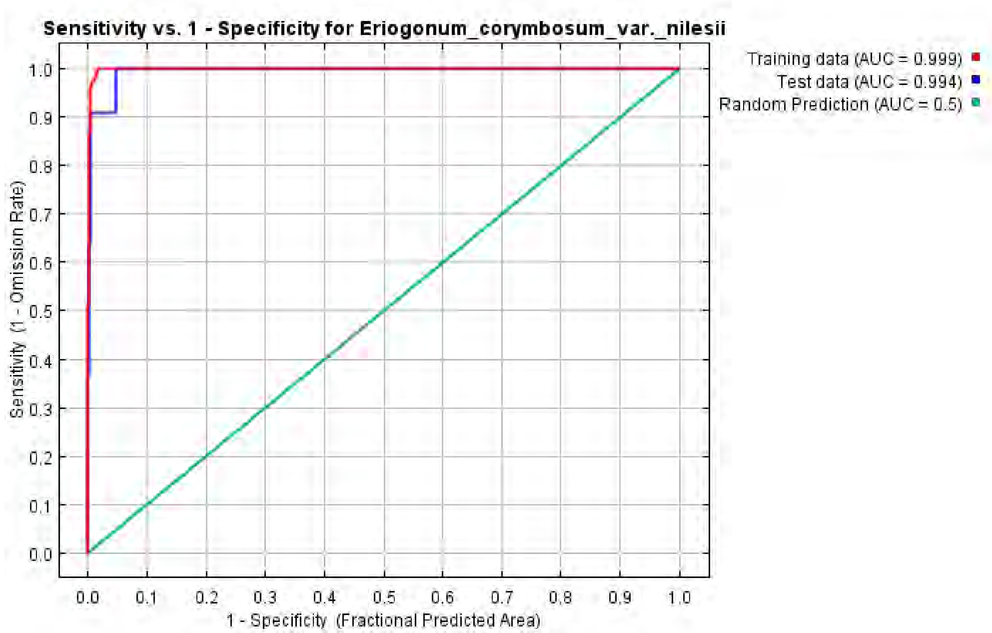
Maxent model for Eriogonum_corymbosum_var._nilesii

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.986 rather than 1; in practice the test AUC may exceed this bound.

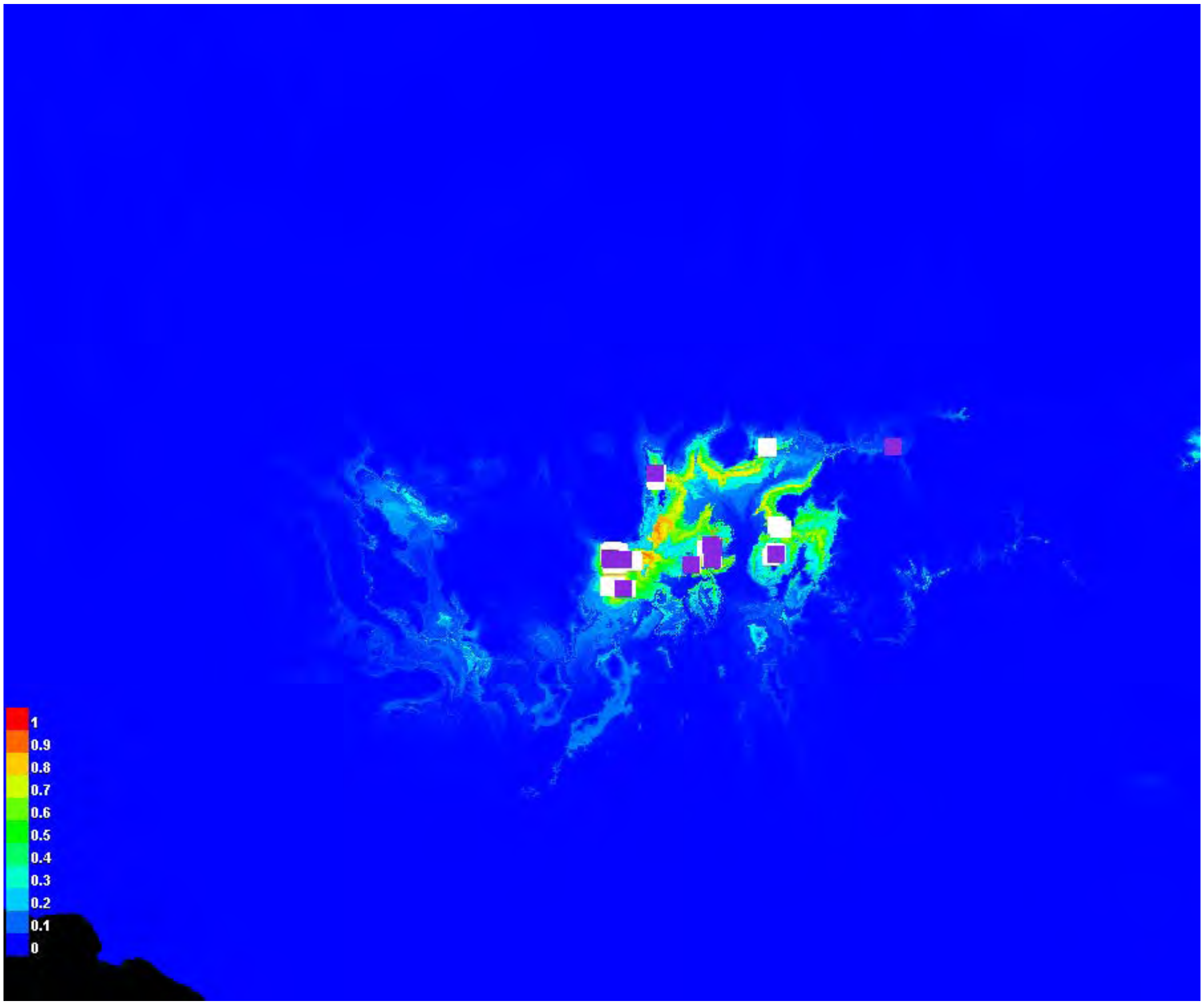


Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.007	Fixed cumulative value	0.107	0.000	0.000	2.193E-11
5.000	0.039	Fixed cumulative value	0.051	0.000	0.000	6.071E-15
10.000	0.083	Fixed cumulative value	0.031	0.000	0.091	8.762E-15
18.464	0.188	Minimum training presence	0.016	0.000	0.091	1.349E-17
49.182	0.619	10 percentile training presence	0.004	0.095	0.273	1.907E-17
21.445	0.229	Equal training sensitivity and specificity	0.013	0.010	0.091	2.028E-18
18.464	0.188	Maximum training sensitivity plus specificity	0.016	0.000	0.091	1.349E-17
5.594	0.045	Equal test sensitivity and specificity	0.048	0.000	0.091	6.417E-13
5.592	0.045	Maximum test sensitivity plus specificity	0.048	0.000	0.000	2.909E-15
1.444	0.010	Balance training omission, predicted area and threshold value	0.094	0.000	0.000	5.123E-12
12.793	0.111	Equate entropy of thresholded and non-thresholded distributions	0.025	0.000	0.091	8.728E-16

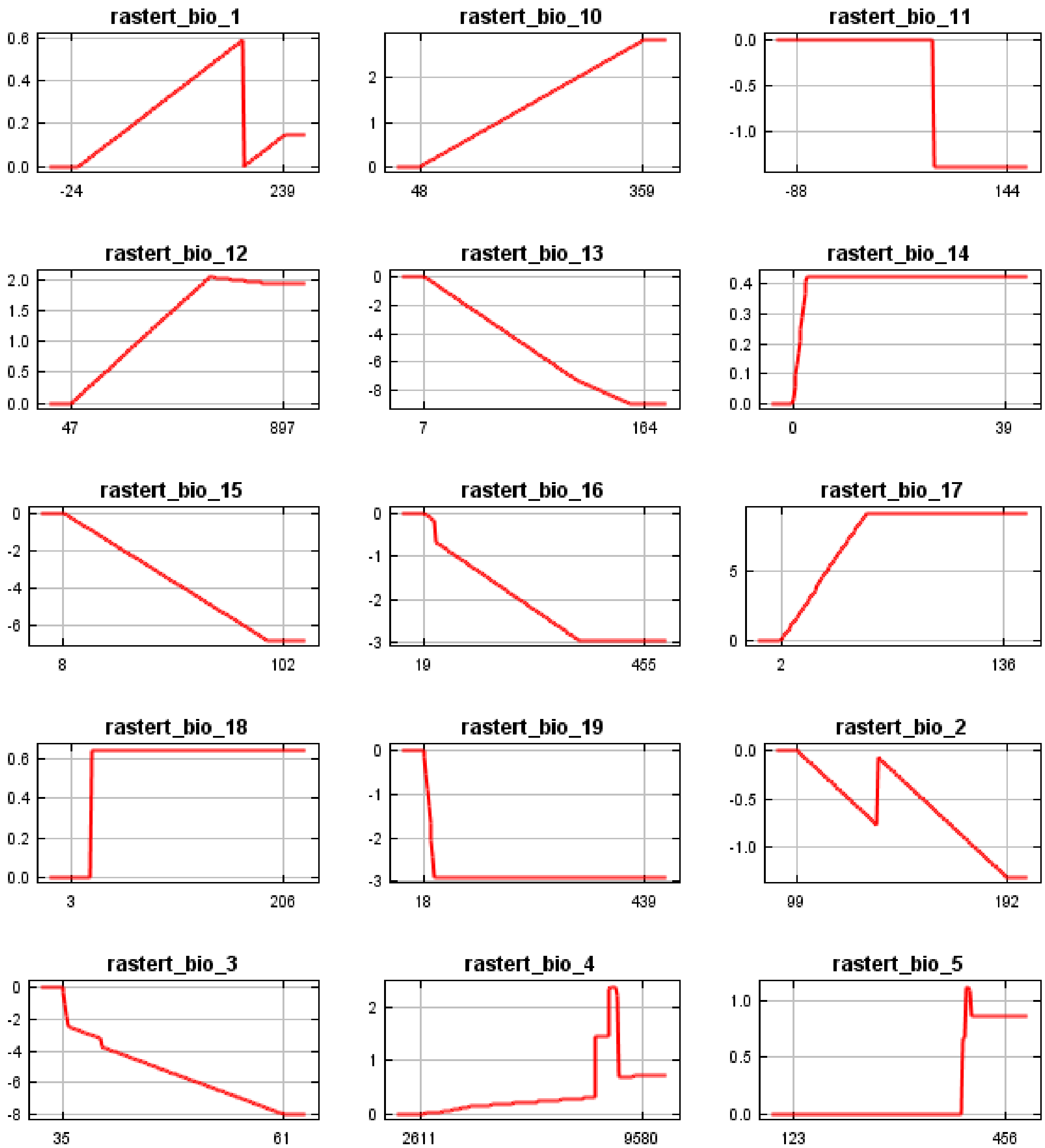
Pictures of the model

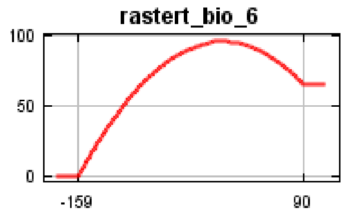
This is a representation of the Maxent model for Eriogonum_corymbosum_var._nilesii. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

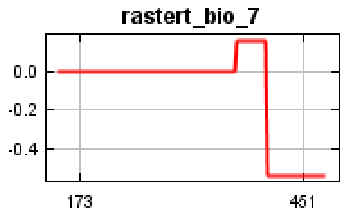


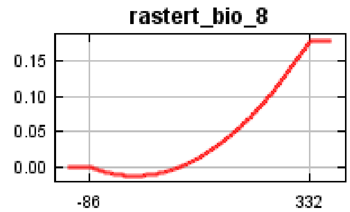
Response curves

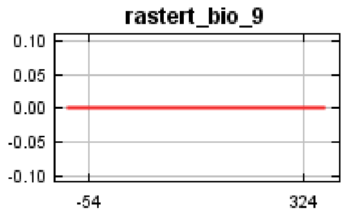
These curves show how each environmental variable affects the Maxent prediction. The (raw) Maxent model has the form exp(...)/constant, and the curves show how the exponent changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version.









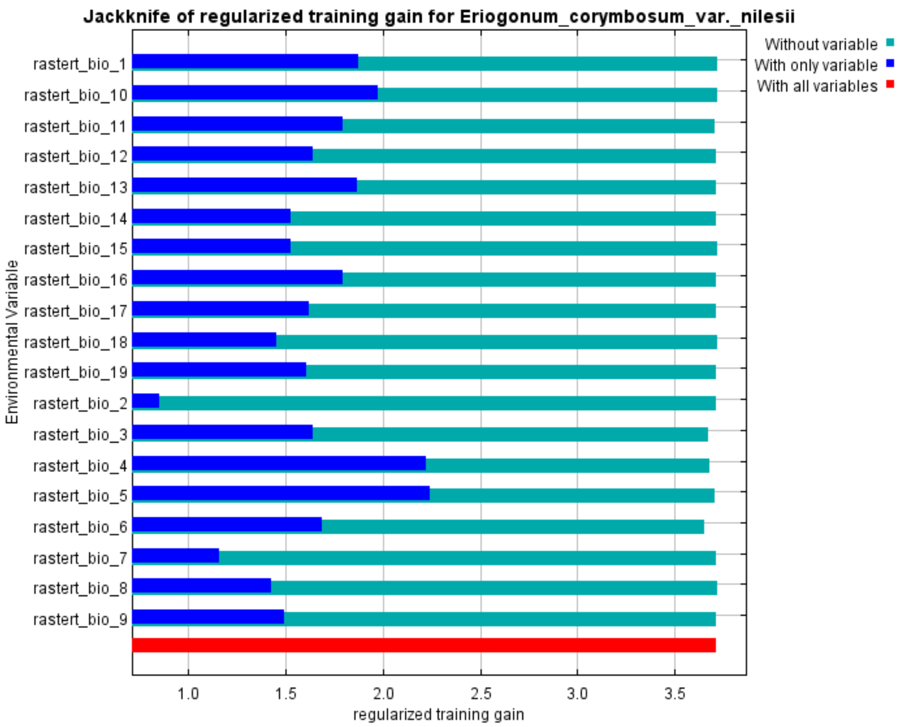


Analysis of variable contributions

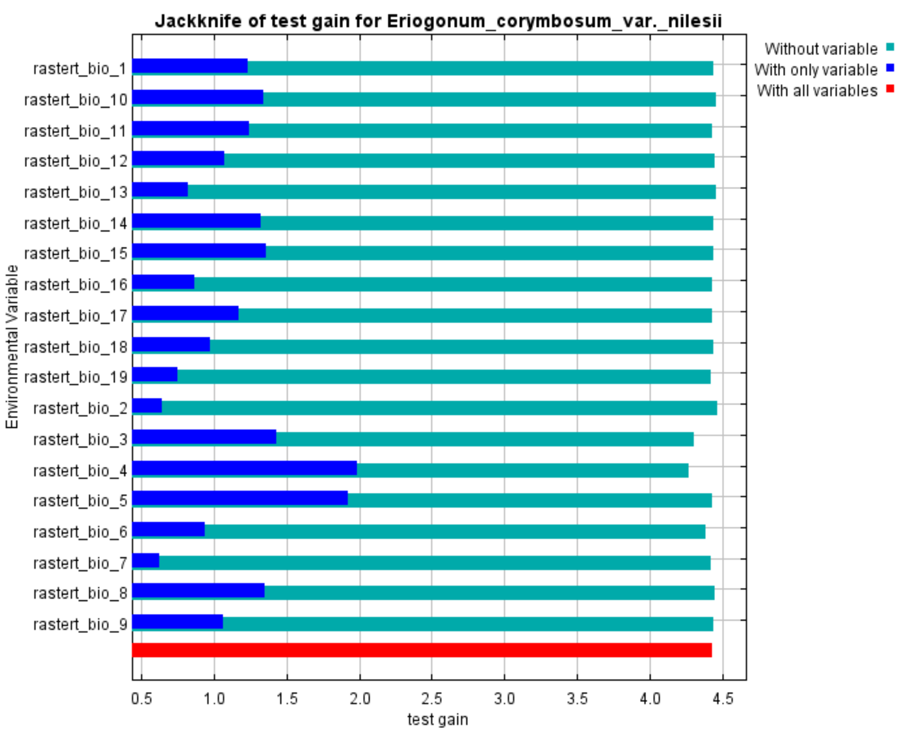
The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution
rastert_bio_4	23.3
rastert_bio_3	18.3
rastert_bio_5	17.5
rastert_bio_6	16.1
rastert_bio_15	9.6
rastert_bio_13	4.9
rastert_bio_19	3.6
rastert_bio_7	2.7
rastert_bio_1	1.2
rastert_bio_8	0.6
rastert_bio_18	0.6
rastert_bio_11	0.5
rastert_bio_16	0.4
rastert_bio_2	0.4
rastert_bio_14	0.3
rastert_bio_17	0.1
rastert_bio_10	0.1
rastert_bio_12	0
rastert_bio_9	0

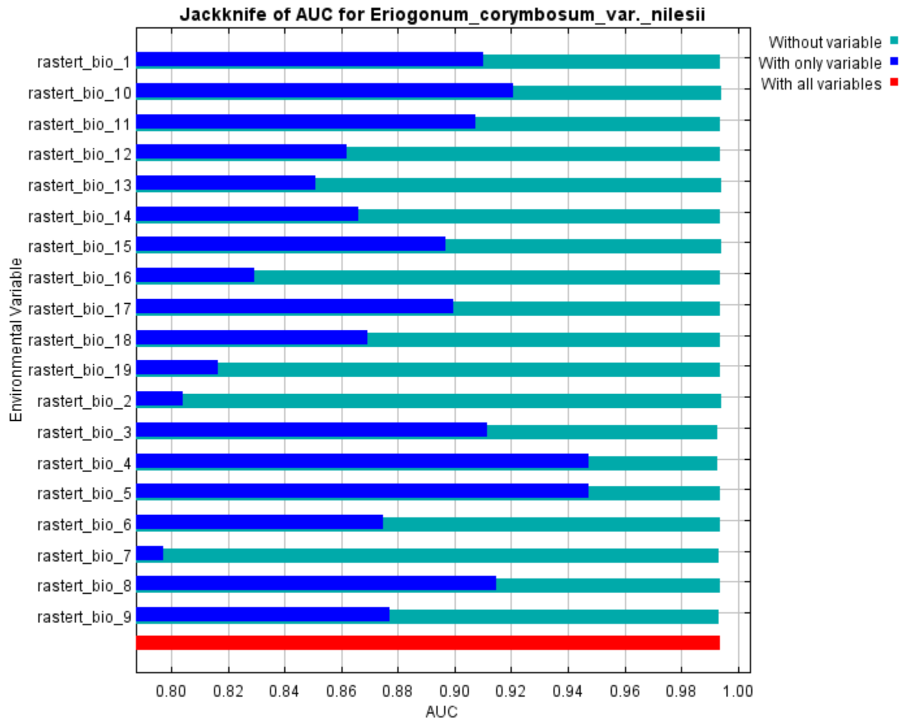
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is rastert_bio_5, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is rastert_bio_6, which therefore appears to have the most information that isn't present in the other variables.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Raw data outputs and control parameters

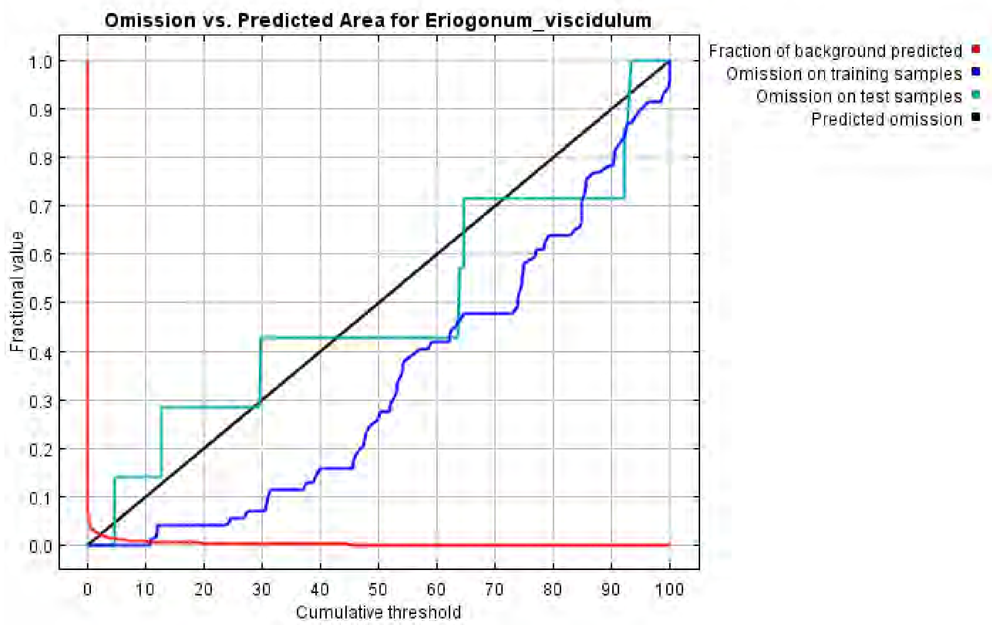
Regularized training gain is 3.713, training AUC is 0.999, unregularized training gain is 4.005.
Unregularized test gain is 4.433. Test AUC is 0.994, standard deviation is 0.004 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm terminated after 500 iterations (12 seconds).

The follow parameters and settings were used during the run:
105 presence records used for training, 11 for testing.
10105 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used (all continuous): rastert_bio_1 rastert_bio_10 rastert_bio_11 rastert_bio_12 rastert_bio_13 rastert_bio_14 rastert_bio_15 rastert_bio_16 rastert_bio_17 rastert_bio_18 rastert_bio_19 rastert_bio_2 rastert_bio_3 rastert_bio_4 rastert_bio_5 rastert_bio_6 rastert_bio_7 rastert_bio_8 rastert_bio_9
Command line:
Feature types used: Linear Quadratic Product Threshold Hinge
Regularization multiplier is 1.0
Regularization values: linear/quadratic/product: 0.050 categorical: 0.250 threshold: 1.000 hinge: 0.500
Species file is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all\gyp_plants_dd.csv
Environmental variables from I:\rare_plant_data\habitat_modeling_maxent\worldclim\bio_30s_esri\bio_ascii_sw
Output directory is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all
Output format is Logistic
Output file type is .asc
Maximum iterations is 500
Convergence threshold is 1.0E-5
Random test percentage is 10
Jackknife selected
Remove duplicates selected
Make pictures selected
Create response curves selected

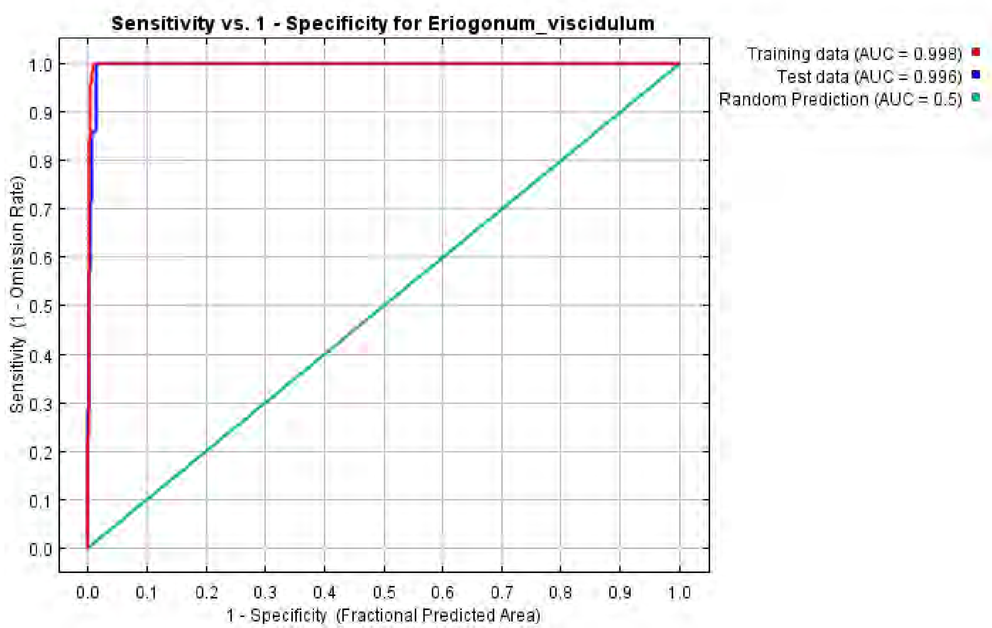
Maxent model for Eriogonum_viscidulum

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.995 rather than 1; in practice the test AUC may exceed this bound.

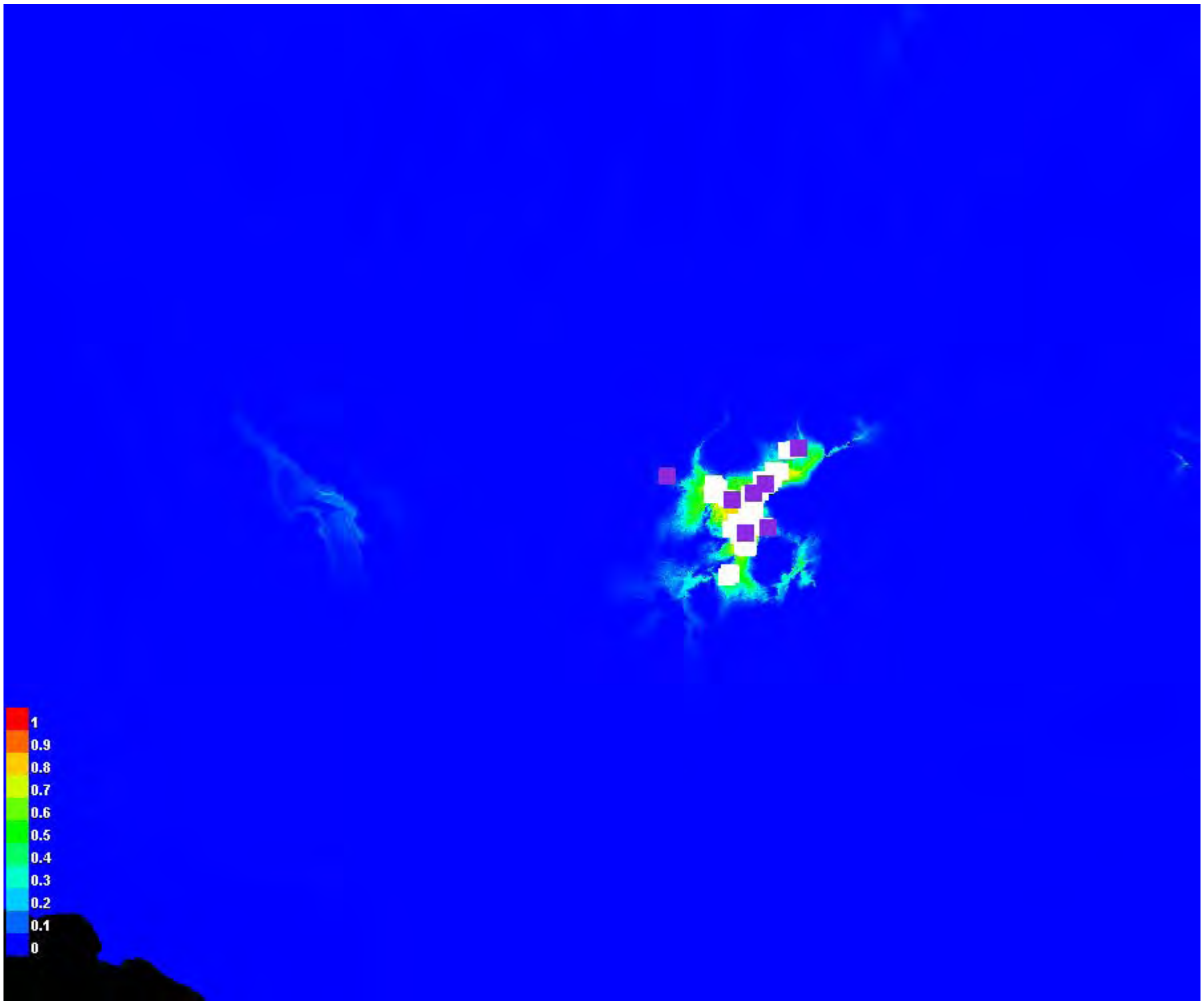


Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.009	Fixed cumulative value	0.031	0.000	0.000	2.943E-11
5.000	0.051	Fixed cumulative value	0.014	0.000	0.143	4.989E-11
10.000	0.146	Fixed cumulative value	0.009	0.000	0.143	2.811E-12
10.785	0.153	Minimum training presence	0.008	0.000	0.143	1.963E-12
30.808	0.459	10 percentile training presence	0.004	0.087	0.429	7.232E-9
10.950	0.155	Equal training sensitivity and specificity	0.008	0.014	0.143	1.963E-12
10.785	0.153	Maximum training sensitivity plus specificity	0.008	0.000	0.143	1.963E-12
4.649	0.046	Equal test sensitivity and specificity	0.014	0.000	0.000	1.348E-13
4.649	0.046	Maximum test sensitivity plus specificity	0.014	0.000	0.000	1.348E-13
0.539	0.004	Balance training omission, predicted area and threshold value	0.038	0.000	0.000	1.254E-10
8.002	0.109	Equate entropy of thresholded and non-thresholded distributions	0.010	0.000	0.143	6.94E-12

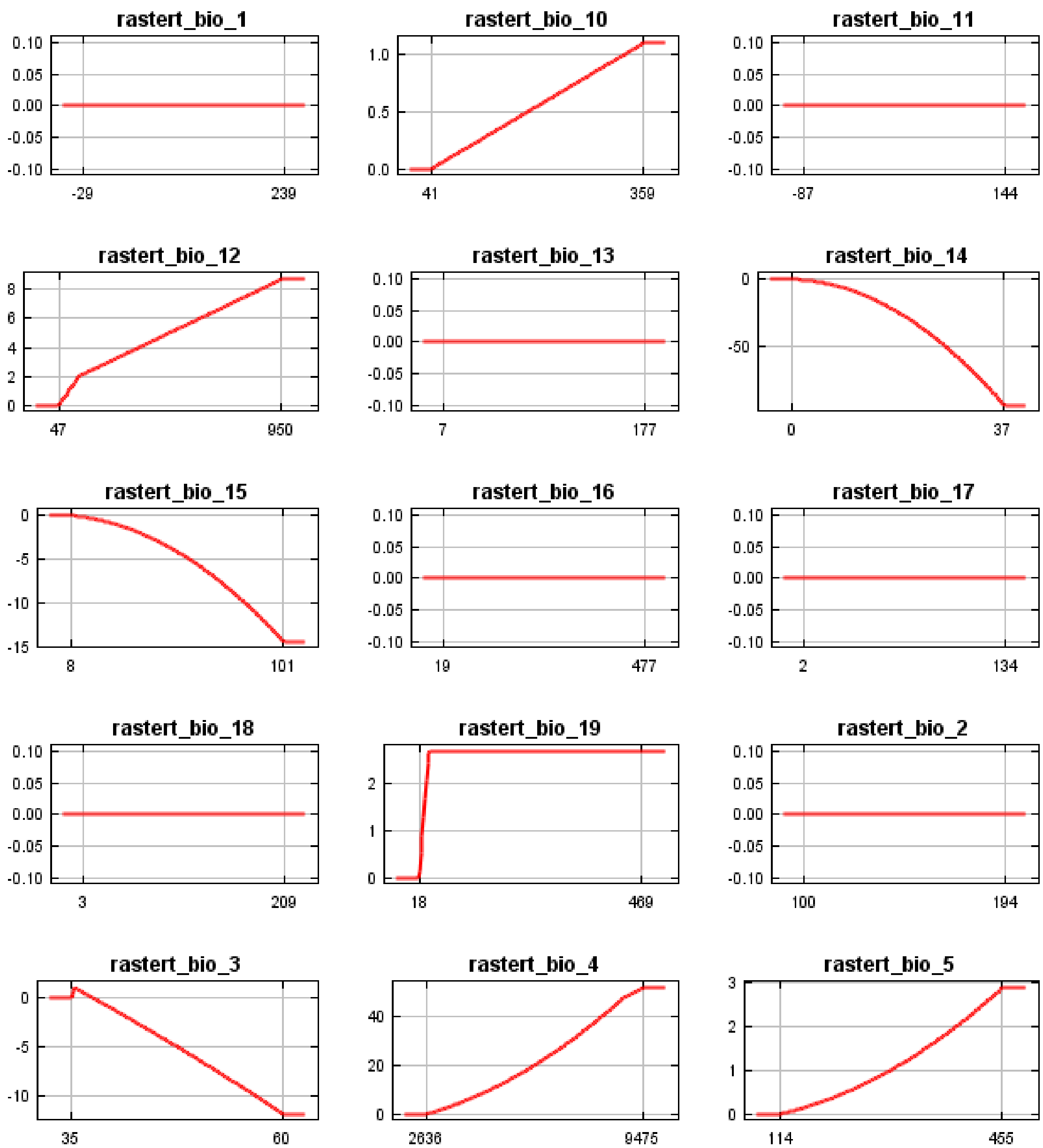
Pictures of the model

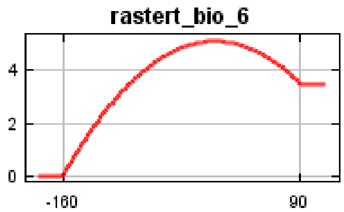
This is a representation of the Maxent model for Eriogonum_viscidulum. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

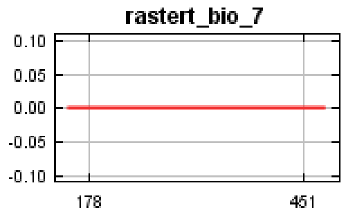


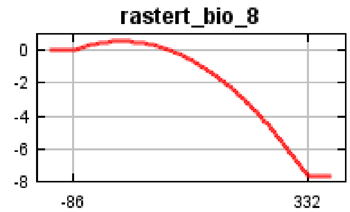
Response curves

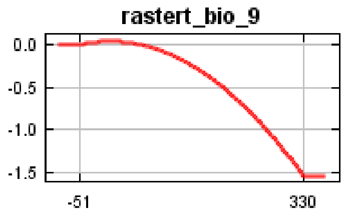
These curves show how each environmental variable affects the Maxent prediction. The (raw) Maxent model has the form exp(...)/constant, and the curves show how the exponent changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version.









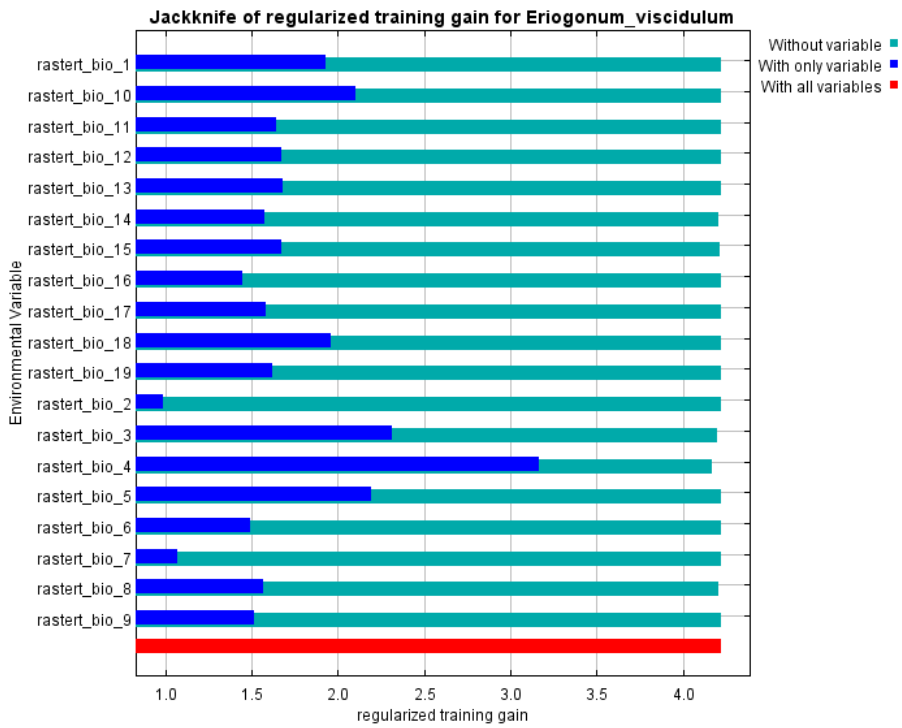


Analysis of variable contributions

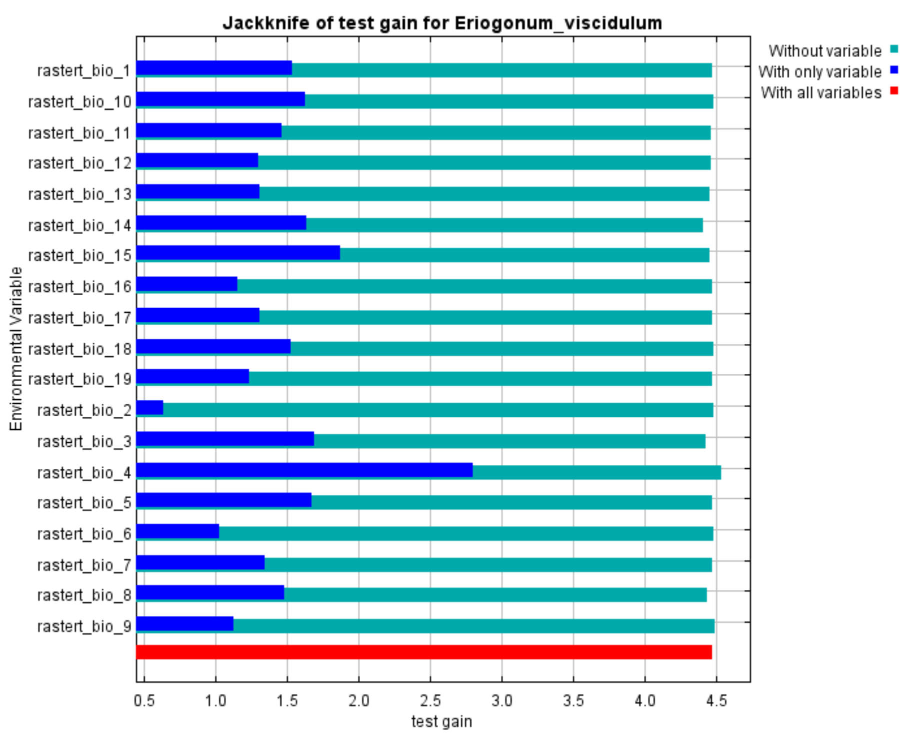
The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution
rastert_bio_4	38.3
rastert_bio_3	22.9
rastert_bio_10	17.5
rastert_bio_6	9.1
rastert_bio_12	8.4
rastert_bio_1	1.7
rastert_bio_2	0.6
rastert_bio_19	0.5
rastert_bio_8	0.4
rastert_bio_17	0.2
rastert_bio_7	0.2
rastert_bio_14	0.1
rastert_bio_15	0.1
rastert_bio_5	0
rastert_bio_9	0
rastert_bio_18	0
rastert_bio_13	0
rastert_bio_11	0
rastert_bio_16	0

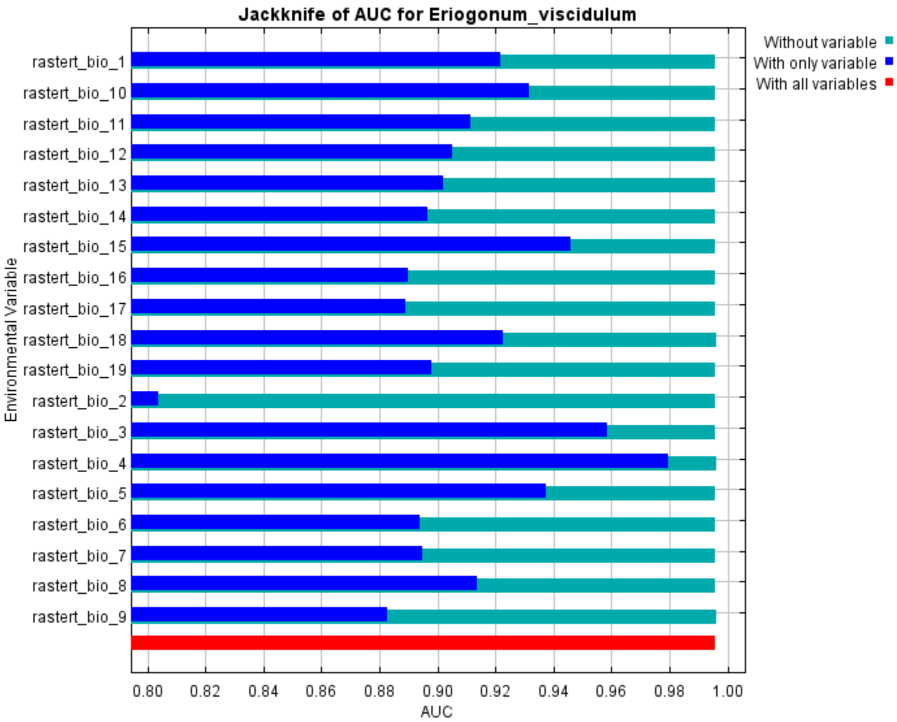
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is rastert_bio_4, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is rastert_bio_4, which therefore appears to have the most information that isn't present in the other variables.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Raw data outputs and control parameters

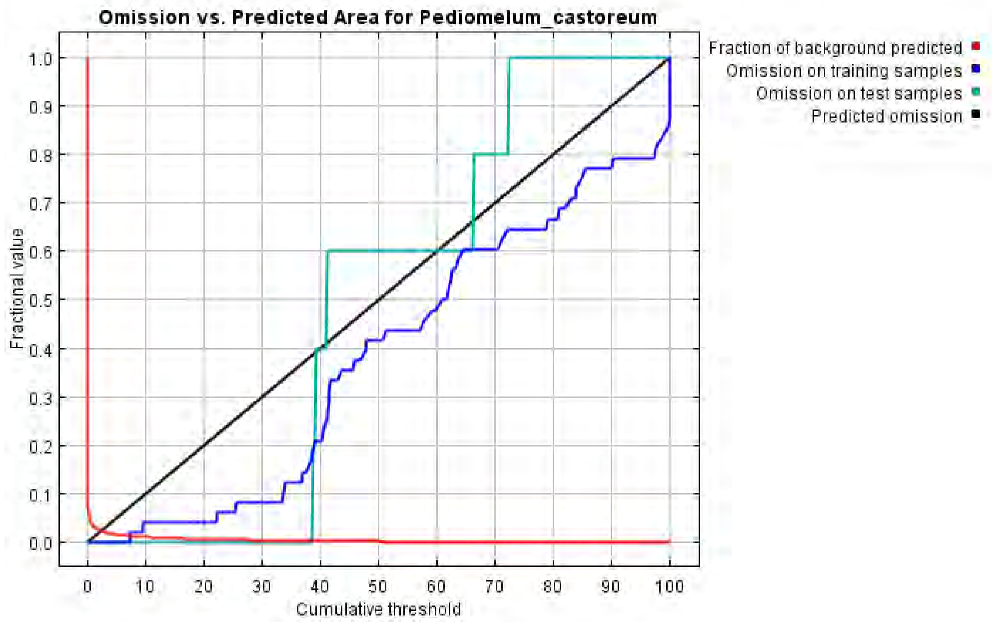
Regularized training gain is 4.220, training AUC is 0.998, unregularized training gain is 4.320.
Unregularized test gain is 4.473. Test AUC is 0.996, standard deviation is 0.002 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm terminated after 500 iterations (13 seconds).

The follow parameters and settings were used during the run:
69 presence records used for training, 7 for testing.
10069 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used (all continuous): rastert_bio_1 rastert_bio_10 rastert_bio_11 rastert_bio_12 rastert_bio_13 rastert_bio_14 rastert_bio_15 rastert_bio_16 rastert_bio_17 rastert_bio_18 rastert_bio_19 rastert_bio_2 rastert_bio_3 rastert_bio_4 rastert_bio_5 rastert_bio_6 rastert_bio_7 rastert_bio_8 rastert_bio_9
Command line:
Feature types used: Linear Quadratic Hinge
Regularization multiplier is 1.0
Regularization values: linear/quadratic/product: 0.139 categorical: 0.250 hinge: 0.500
Species file is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all\sand_plants_dd.csv
Environmental variables from I:\rare_plant_data\habitat_modeling_maxent\worldclim\bio_30s_esri\bio_ascii_sw
Output directory is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all
Output format is Logistic
Output file type is .asc
Maximum iterations is 500
Convergence threshold is 1.0E-5
Random test percentage is 10
Jackknife selected
Remove duplicates selected
Make pictures selected
Create response curves selected

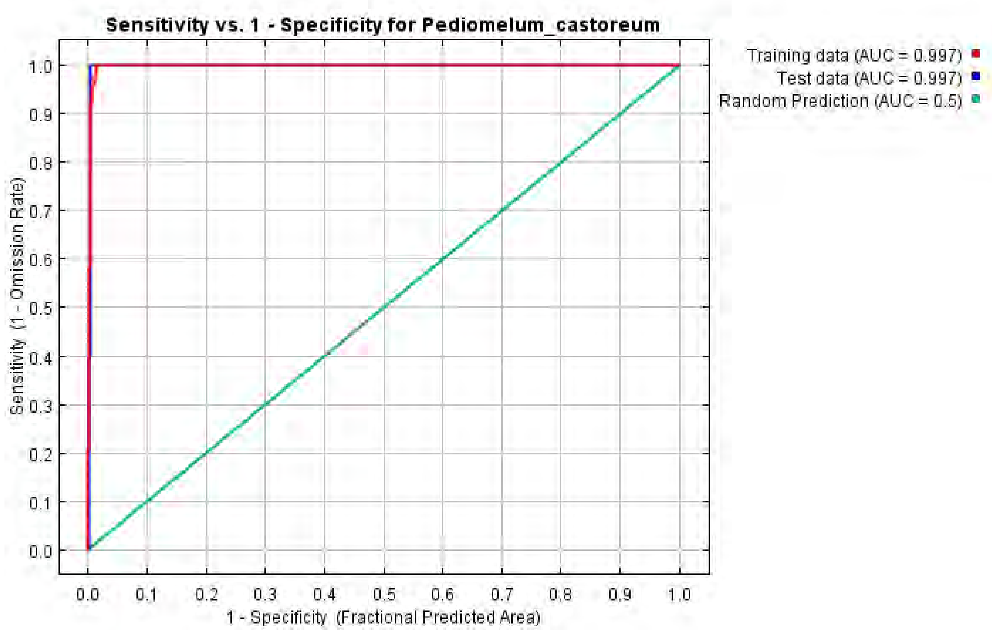
Maxent model for Pediomelum_castoreum

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.994 rather than 1; in practice the test AUC may exceed this bound.

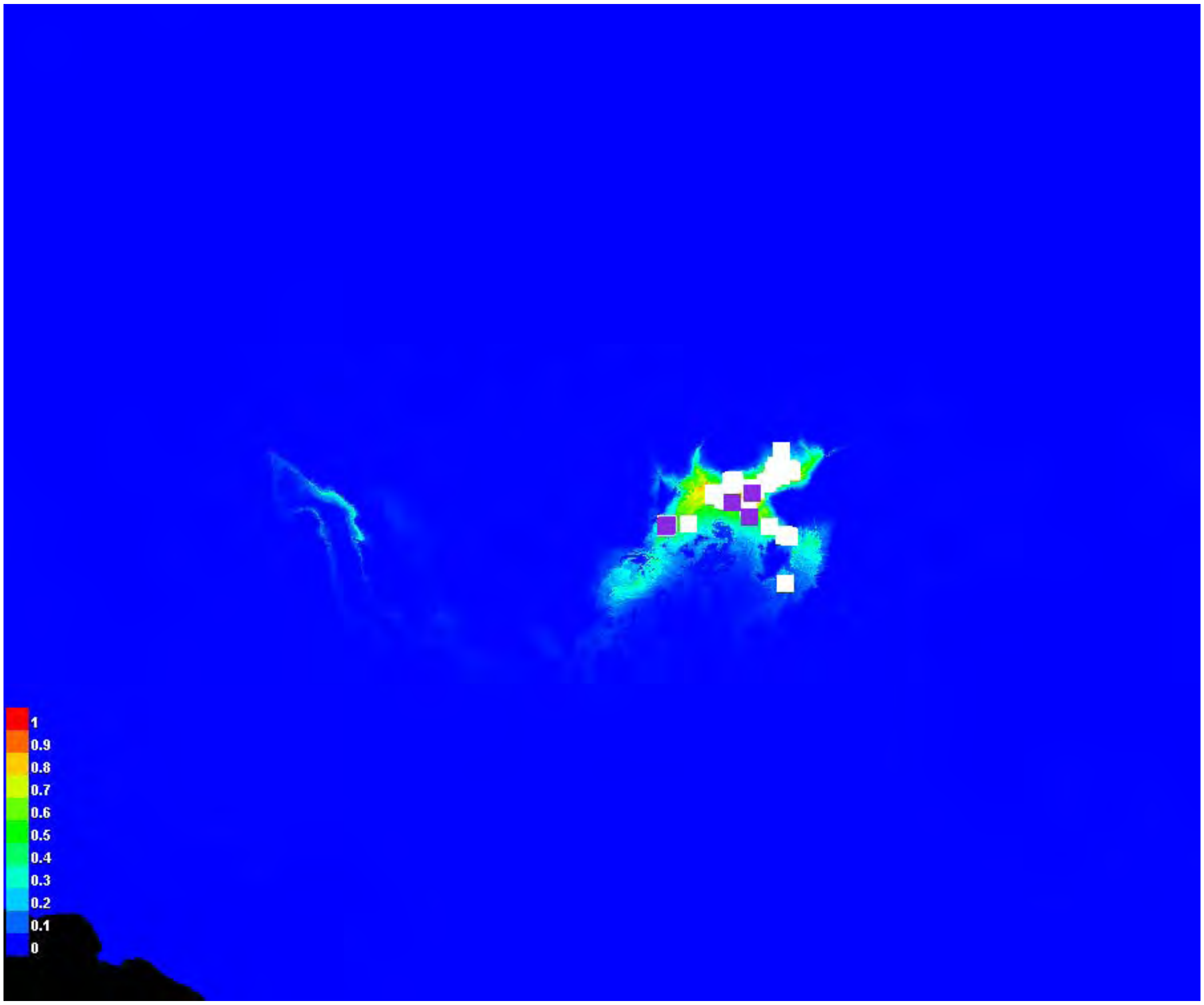


Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.009	Fixed cumulative value	0.035	0.000	0.000	5.481E-8
5.000	0.073	Fixed cumulative value	0.017	0.000	0.000	1.379E-9
10.000	0.150	Fixed cumulative value	0.012	0.042	0.000	2.386E-10
7.411	0.125	Minimum training presence	0.014	0.000	0.000	5.378E-10
33.294	0.456	10 percentile training presence	0.004	0.083	0.000	1.845E-12
7.420	0.126	Equal training sensitivity and specificity	0.014	0.021	0.000	5.378E-10
7.411	0.125	Maximum training sensitivity plus specificity	0.014	0.000	0.000	5.378E-10
38.492	0.502	Equal test sensitivity and specificity	0.004	0.167	0.000	7.924E-13
38.492	0.502	Maximum test sensitivity plus specificity	0.004	0.167	0.000	7.924E-13
0.642	0.005	Balance training omission, predicted area and threshold value	0.042	0.000	0.000	1.323E-7
8.955	0.138	Equate entropy of thresholded and non-thresholded distributions	0.013	0.021	0.000	3.304E-10

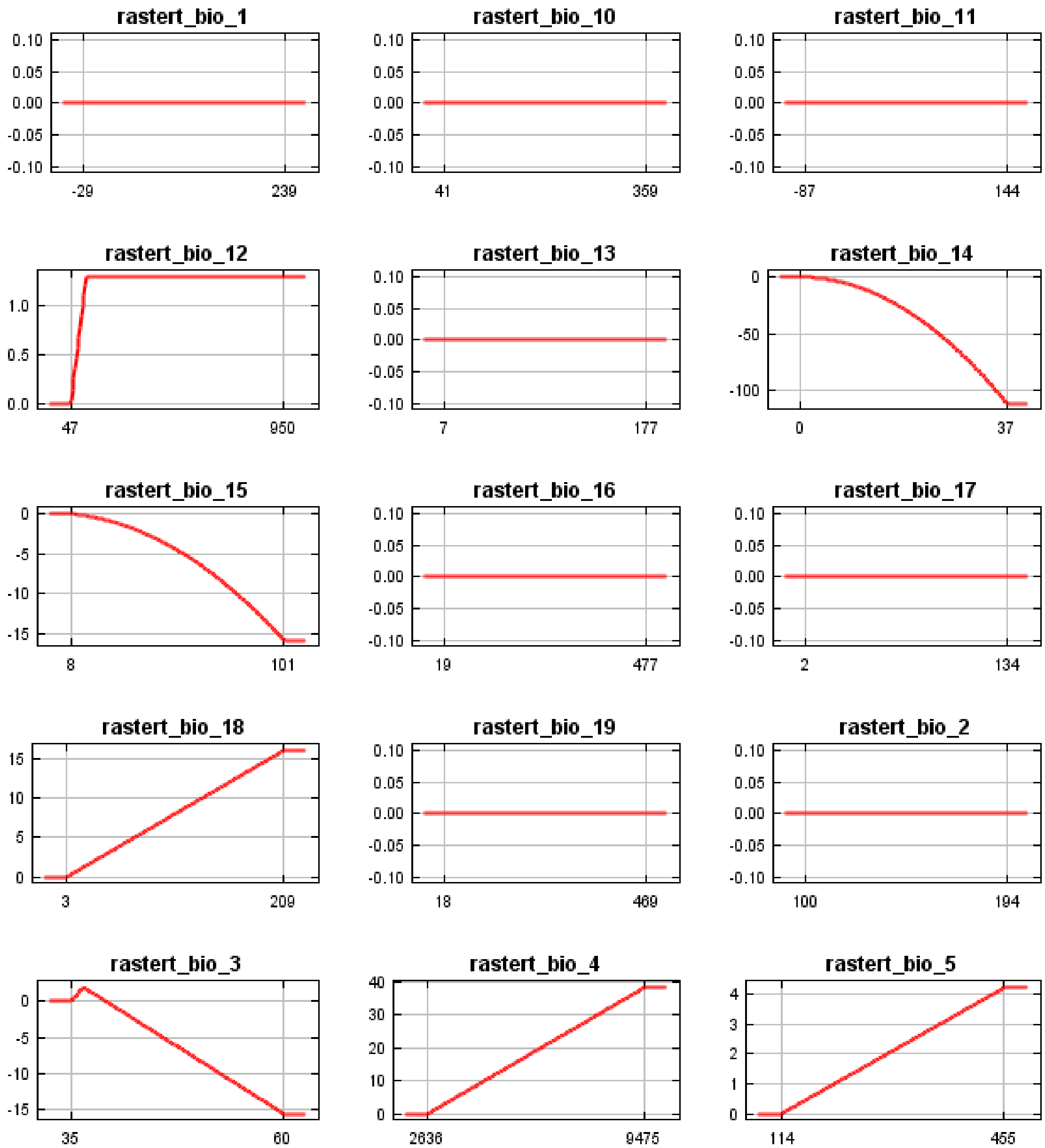
Pictures of the model

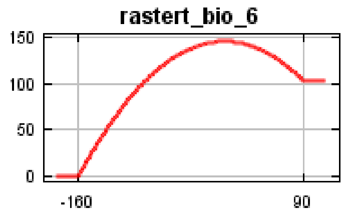
This is a representation of the Maxent model for Pediomelum_castoreum. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

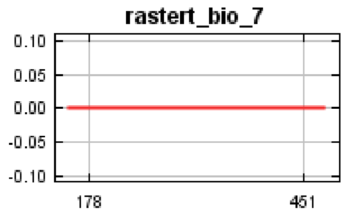


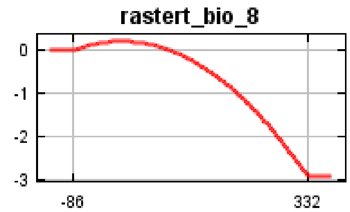
Response curves

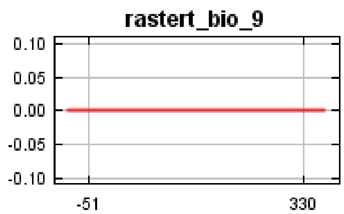
These curves show how each environmental variable affects the Maxent prediction. The (raw) Maxent model has the form exp(...)/constant, and the curves show how the exponent changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version.









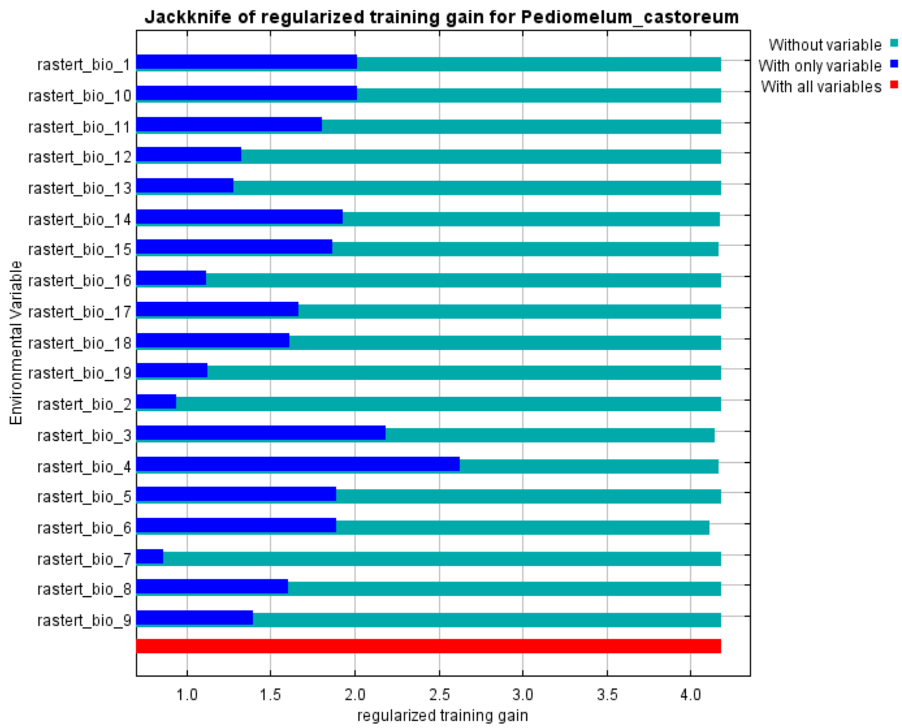


Analysis of variable contributions

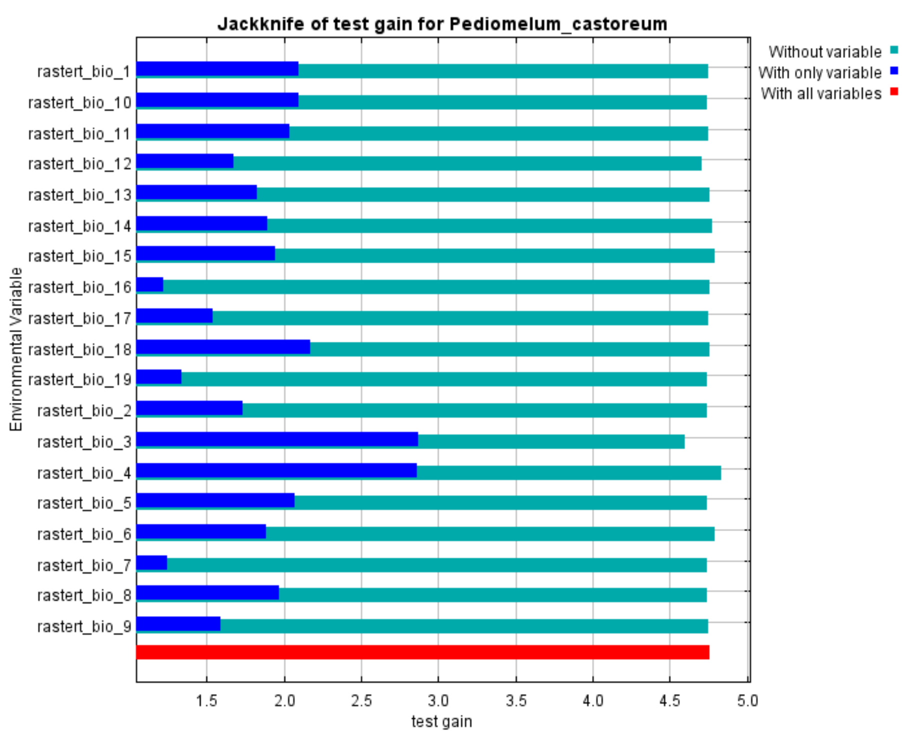
The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution
rastert_bio_4	37.2
rastert_bio_6	23.8
rastert_bio_3	14.5
rastert_bio_17	8.5
rastert_bio_10	8.4
rastert_bio_12	2.2
rastert_bio_8	1.7
rastert_bio_5	1.3
rastert_bio_1	0.9
rastert_bio_18	0.6
rastert_bio_19	0.5
rastert_bio_15	0.2
rastert_bio_14	0.1
rastert_bio_11	0
rastert_bio_16	0
rastert_bio_13	0
rastert_bio_2	0
rastert_bio_7	0
rastert_bio_9	0

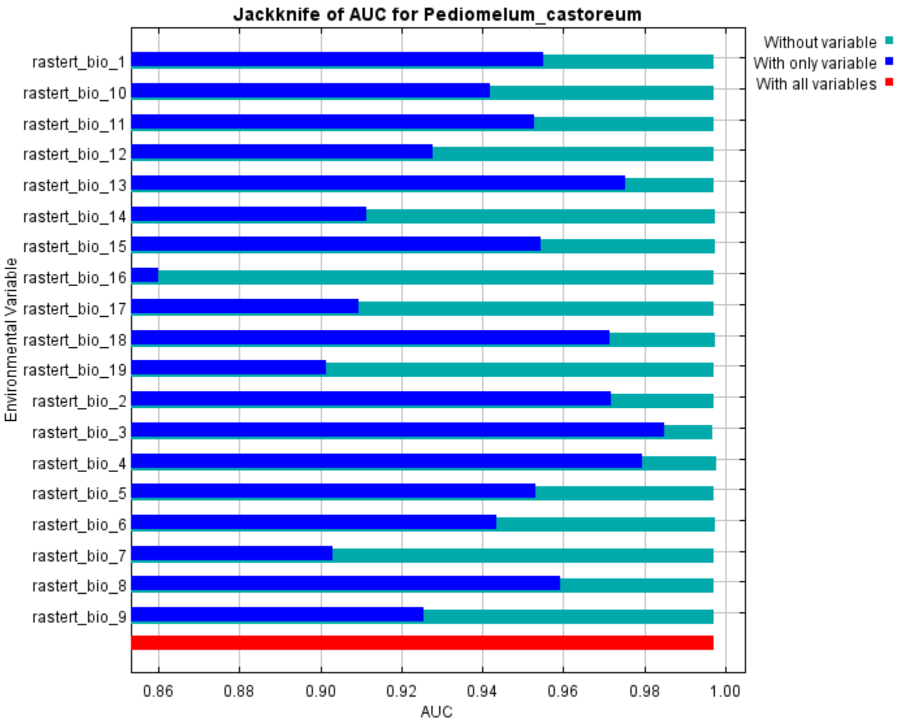
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is rastert_bio_4, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is rastert_bio_6, which therefore appears to have the most information that isn't present in the other variables.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Raw data outputs and control parameters

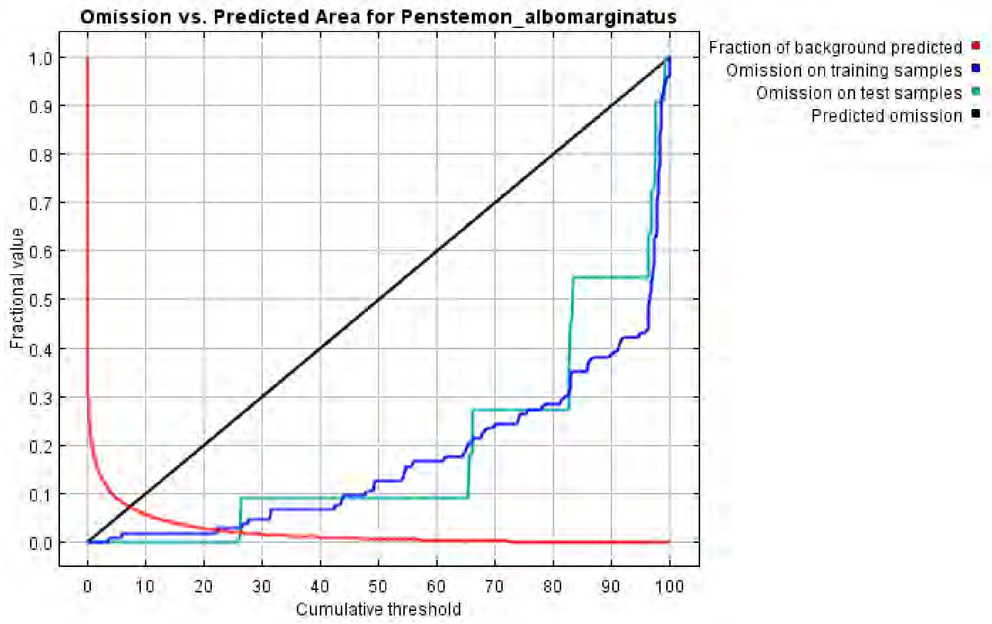
Regularized training gain is 4.185, training AUC is 0.997, unregularized training gain is 4.294.
Unregularized test gain is 4.756. Test AUC is 0.997, standard deviation is 0.001 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm terminated after 500 iterations (11 seconds).

The follow parameters and settings were used during the run:
48 presence records used for training, 5 for testing.
10048 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used (all continuous): rastert_bio_1 rastert_bio_10 rastert_bio_11 rastert_bio_12 rastert_bio_13 rastert_bio_14 rastert_bio_15 rastert_bio_16 rastert_bio_17 rastert_bio_18 rastert_bio_19 rastert_bio_2 rastert_bio_3 rastert_bio_4 rastert_bio_5 rastert_bio_6 rastert_bio_7 rastert_bio_8 rastert_bio_9
Command line:
Feature types used: Linear Quadratic Hinge
Regularization multiplier is 1.0
Regularization values: linear/quadratic/product: 0.199 categorical: 0.250 hinge: 0.500
Species file is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all\sand_plants_dd.csv
Environmental variables from I:\rare_plant_data\habitat_modeling_maxent\worldclim\bio_30s_esri\bio_ascii_sw
Output directory is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all
Output format is Logistic
Output file type is .asc
Maximum iterations is 500
Convergence threshold is 1.0E-5
Random test percentage is 10
Jackknife selected
Remove duplicates selected
Make pictures selected
Create response curves selected

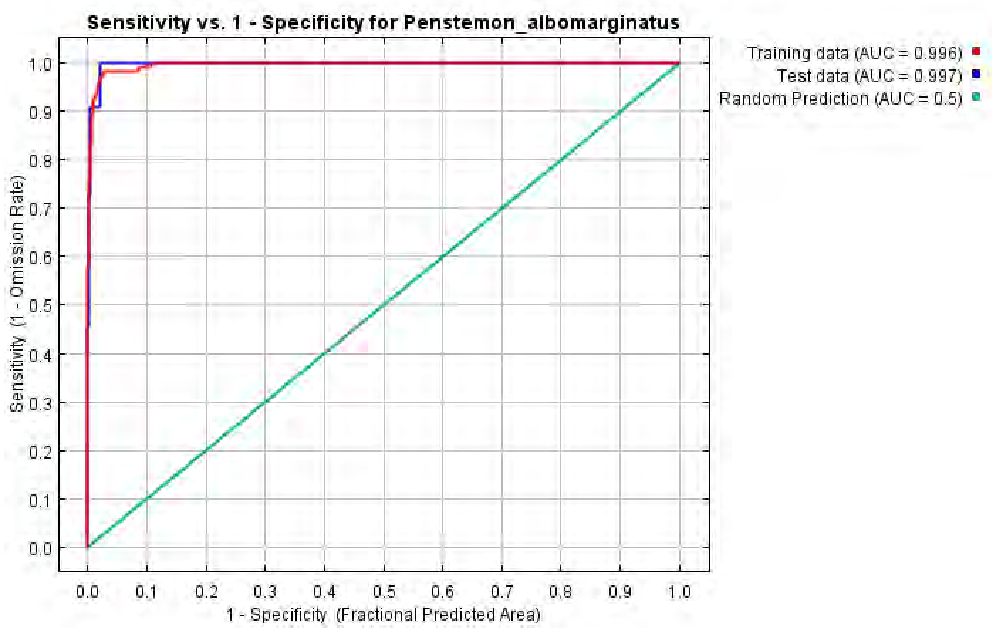
Maxent model for Penstemon_albomarginatus

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.977 rather than 1; in practice the test AUC may exceed this bound.

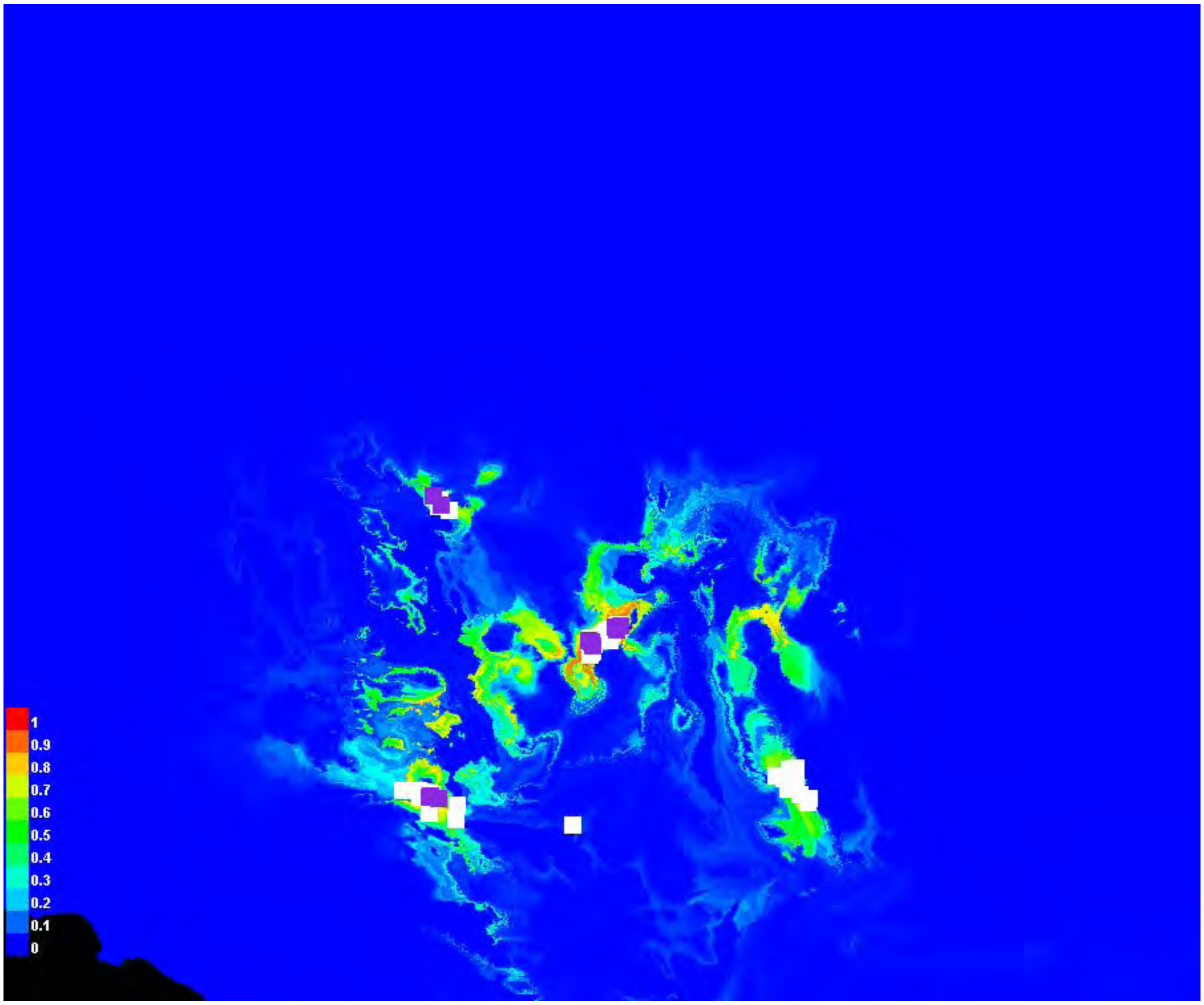


Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.007	Fixed cumulative value	0.185	0.000	0.000	8.739E-9
5.000	0.041	Fixed cumulative value	0.093	0.010	0.000	4.344E-12
10.000	0.090	Fixed cumulative value	0.058	0.020	0.000	2.406E-14
3.750	0.030	Minimum training presence	0.108	0.000	0.000	2.404E-11
47.463	0.573	10 percentile training presence	0.008	0.098	0.091	1.327E-20
22.175	0.237	Equal training sensitivity and specificity	0.026	0.029	0.000	3.518E-18
22.113	0.236	Maximum training sensitivity plus specificity	0.026	0.020	0.000	3.518E-18
26.216	0.300	Equal test sensitivity and specificity	0.021	0.029	0.000	3.153E-19
26.216	0.300	Maximum test sensitivity plus specificity	0.021	0.029	0.000	3.153E-19
2.246	0.018	Balance training omission, predicted area and threshold value	0.137	0.000	0.000	3.165E-10
13.245	0.125	Equate entropy of thresholded and non-thresholded distributions	0.045	0.020	0.000	1.73E-15

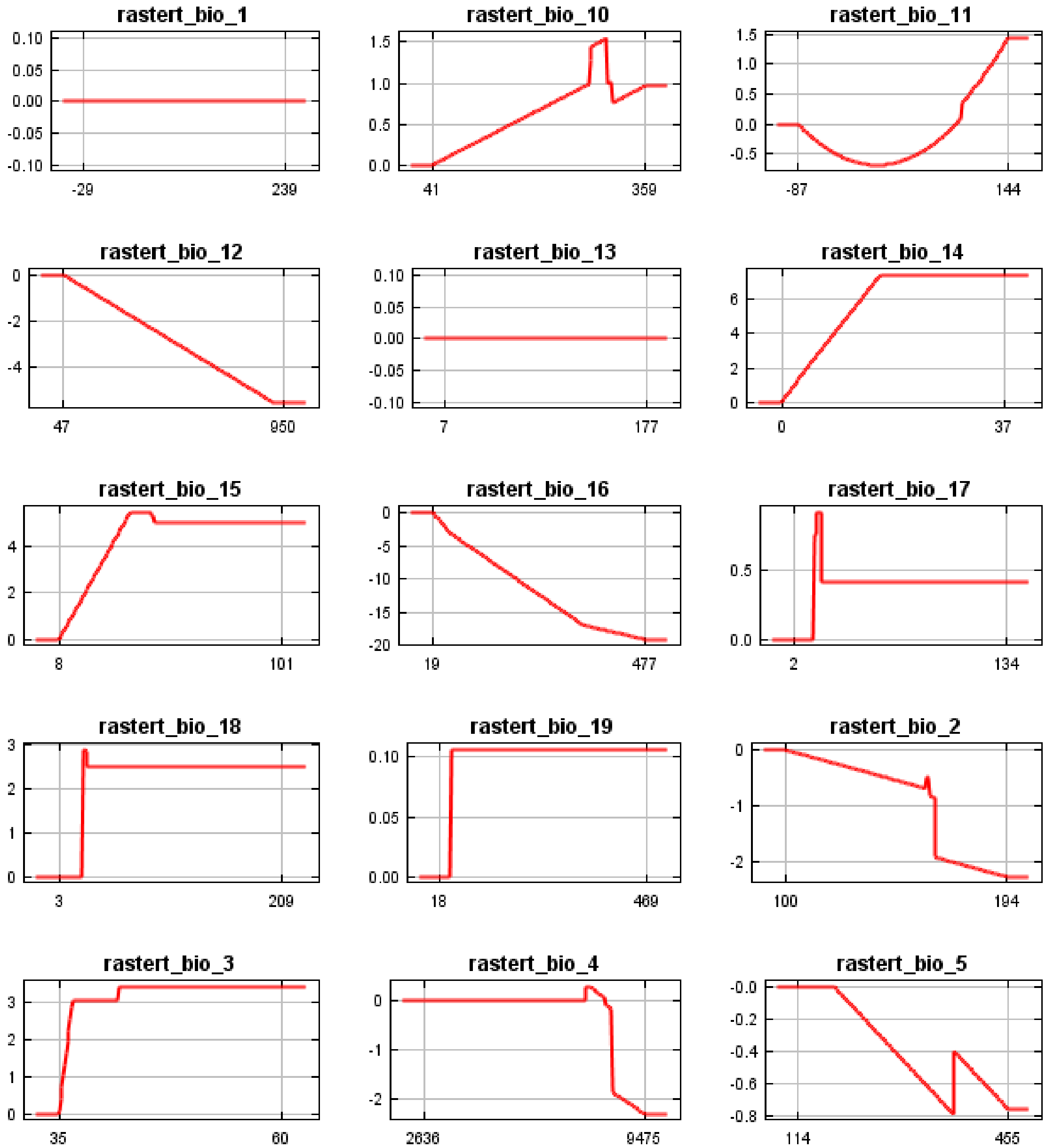
Pictures of the model

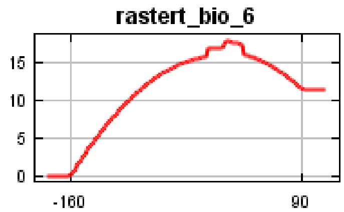
This is a representation of the Maxent model for Penstemon_albomarginatus. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

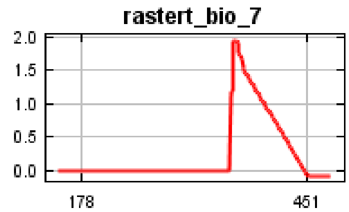


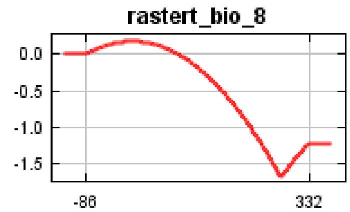
Response curves

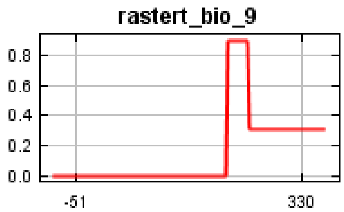
These curves show how each environmental variable affects the Maxent prediction. The (raw) Maxent model has the form exp(...)/constant, and the curves show how the exponent changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version.









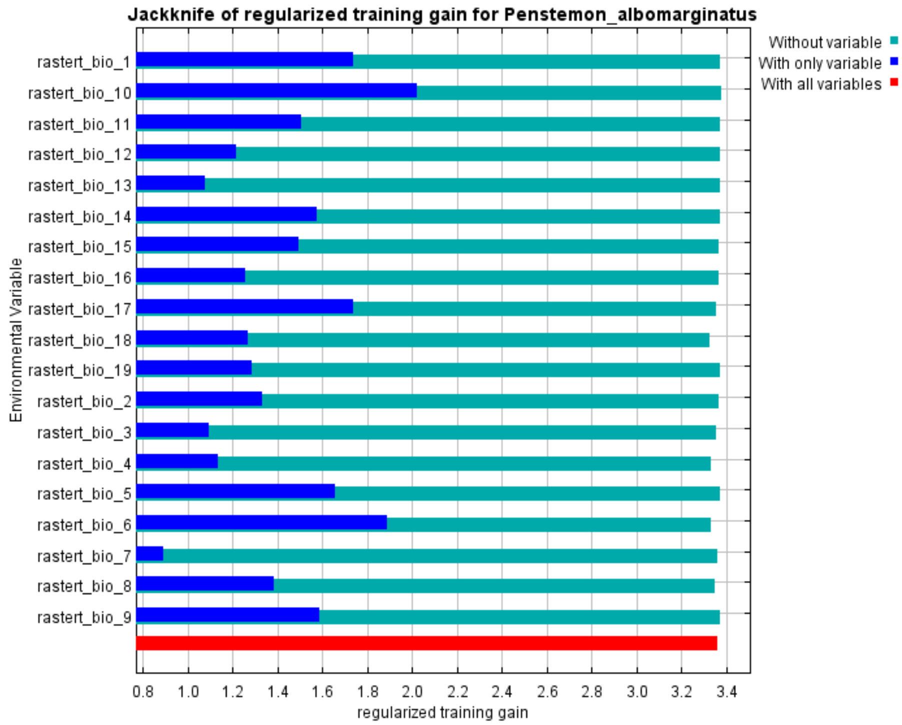


Analysis of variable contributions

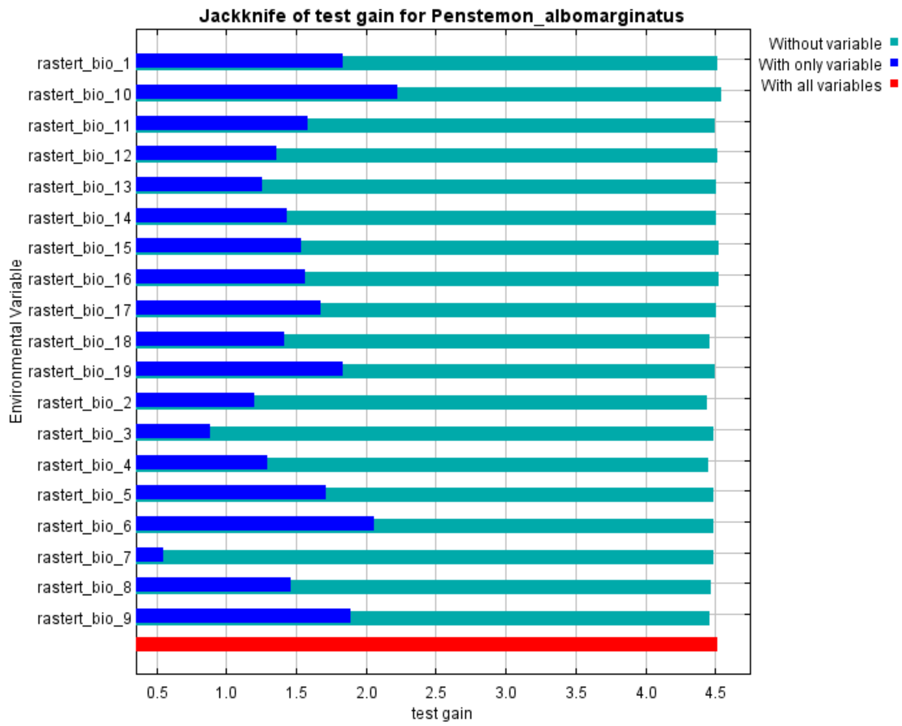
The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution
rastert_bio_10	28.6
rastert_bio_6	12.4
rastert_bio_2	8
rastert_bio_9	7.7
rastert_bio_8	7
rastert_bio_17	6.9
rastert_bio_4	5.6
rastert_bio_16	5.4
rastert_bio_7	4.2
rastert_bio_15	3.9
rastert_bio_18	3.6
rastert_bio_3	3
rastert_bio_5	2.6
rastert_bio_14	0.6
rastert_bio_12	0.2
rastert_bio_19	0.1
rastert_bio_11	0.1
rastert_bio_13	0
rastert_bio_1	0

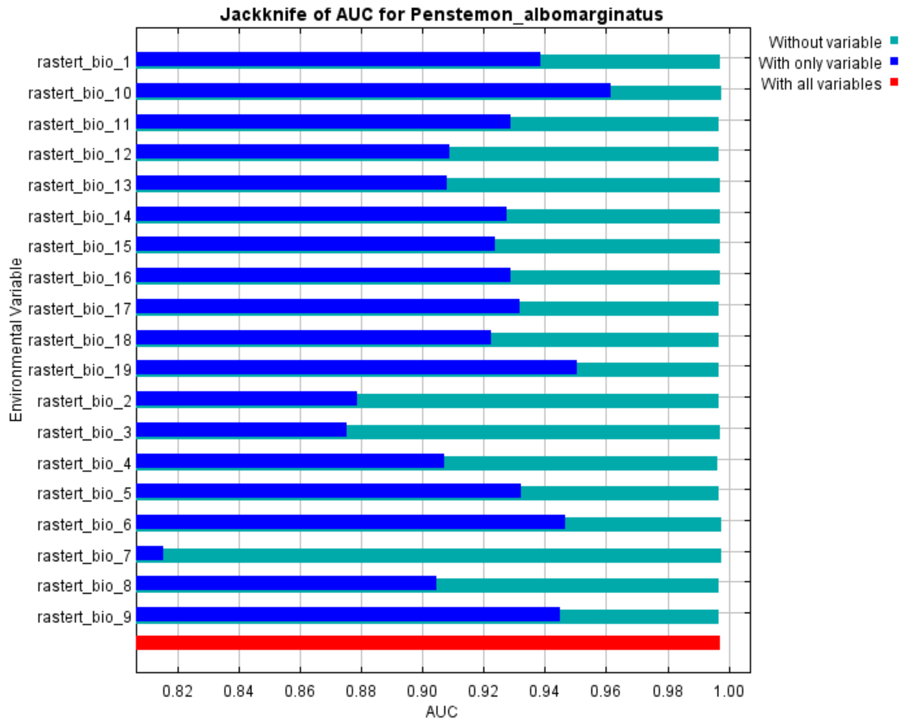
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is rastert_bio_10, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is rastert_bio_18, which therefore appears to have the most information that isn't present in the other variables.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Raw data outputs and control parameters

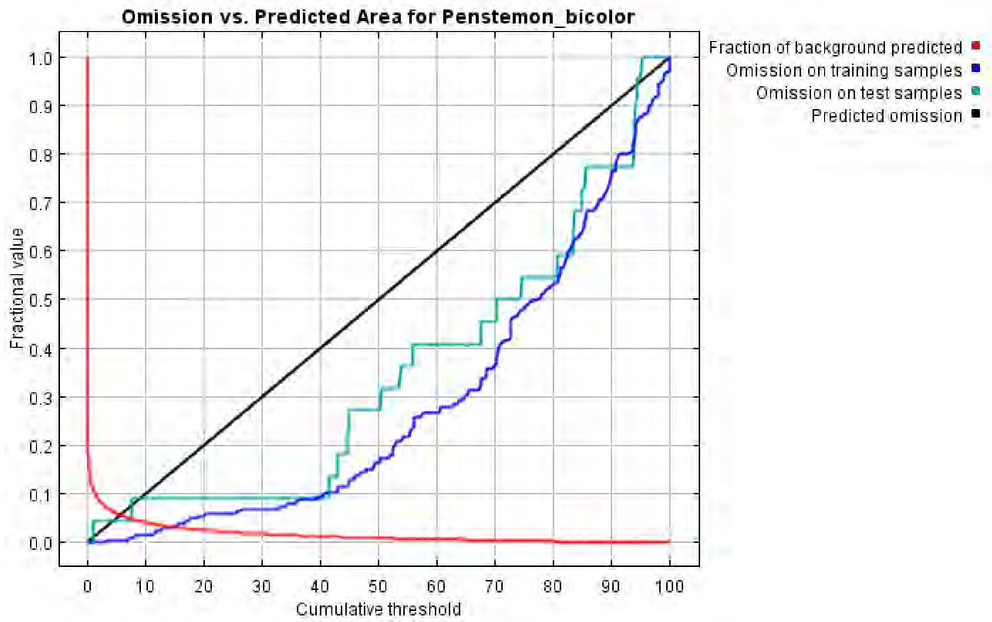
Regularized training gain is 3.363, training AUC is 0.996, unregularized training gain is 3.726.
Unregularized test gain is 4.513. Test AUC is 0.997, standard deviation is 0.002 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm terminated after 500 iterations (13 seconds).

The follow parameters and settings were used during the run:
102 presence records used for training, 11 for testing.
10102 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used (all continuous): rastert_bio_1 rastert_bio_10 rastert_bio_11 rastert_bio_12 rastert_bio_13 rastert_bio_14 rastert_bio_15 rastert_bio_16 rastert_bio_17 rastert_bio_18 rastert_bio_19 rastert_bio_2 rastert_bio_3 rastert_bio_4 rastert_bio_5 rastert_bio_6 rastert_bio_7 rastert_bio_8 rastert_bio_9
Command line:
Feature types used: Linear Quadratic Product Threshold Hinge
Regularization multiplier is 1.0
Regularization values: linear/quadratic/product: 0.050 categorical: 0.250 threshold: 1.000 hinge: 0.500
Species file is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all\sand_plants_dd.csv
Environmental variables from I:\rare_plant_data\habitat_modeling_maxent\worldclim\bio_30s_esri\bio_ascii_sw
Output directory is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all
Output format is Logistic
Output file type is .asc
Maximum iterations is 500
Convergence threshold is 1.0E-5
Random test percentage is 10
Jackknife selected
Remove duplicates selected
Make pictures selected
Create response curves selected

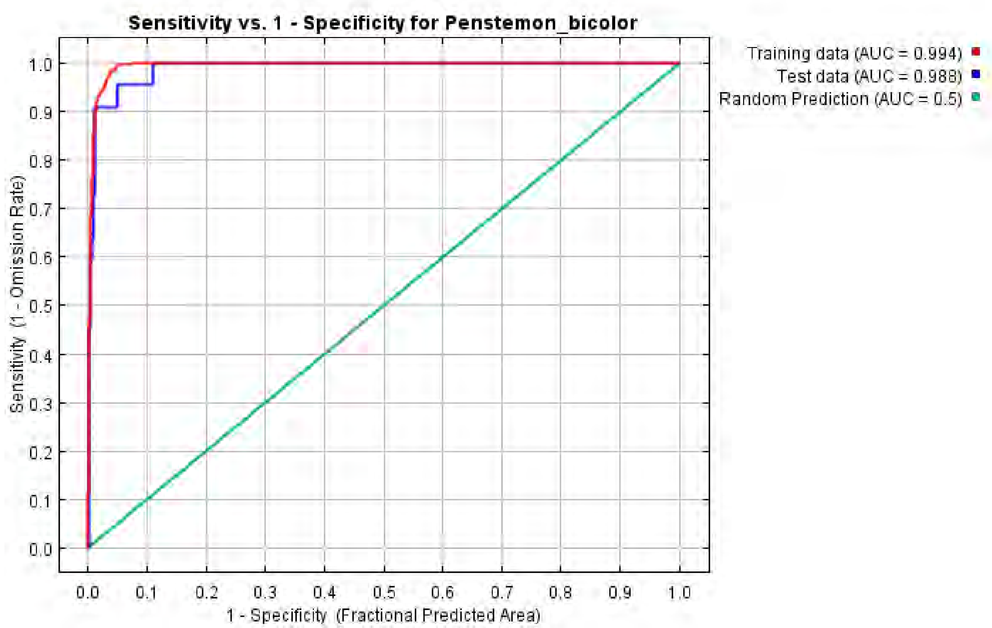
Maxent model for Penstemon_bicolor

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.982 rather than 1; in practice the test AUC may exceed this bound.

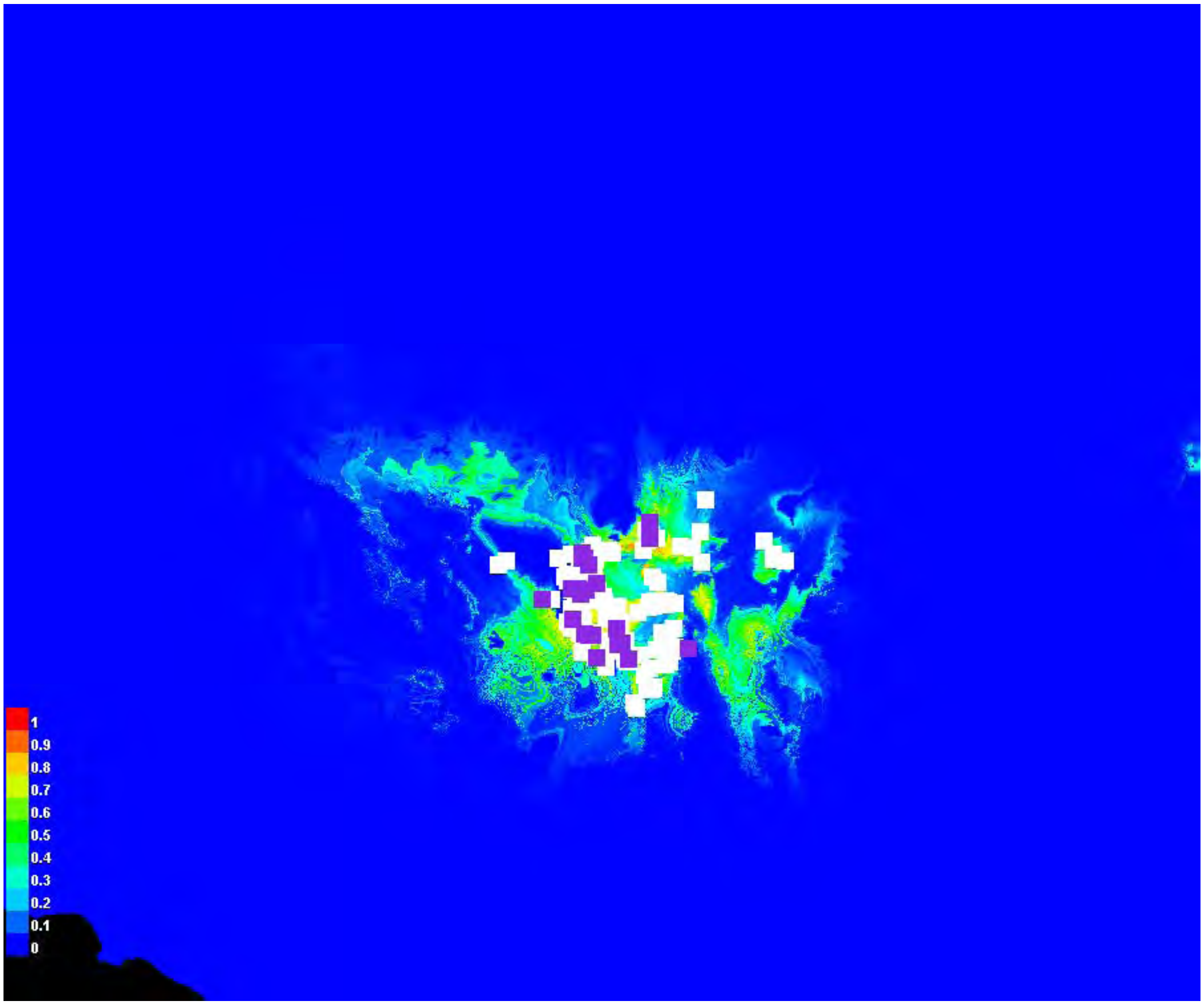


Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.013	Fixed cumulative value	0.109	0.000	0.045	1.159E-19
5.000	0.075	Fixed cumulative value	0.060	0.005	0.045	4.878E-25
10.000	0.147	Fixed cumulative value	0.042	0.015	0.091	5.133E-26
3.237	0.047	Minimum training presence	0.073	0.000	0.045	2.39E-23
40.809	0.522	10 percentile training presence	0.012	0.100	0.091	1.666E-36
14.939	0.220	Equal training sensitivity and specificity	0.032	0.030	0.091	2.586E-28
11.216	0.170	Maximum training sensitivity plus specificity	0.039	0.015	0.091	1.217E-26
7.538	0.109	Equal test sensitivity and specificity	0.049	0.010	0.045	6.268E-27
7.538	0.109	Maximum test sensitivity plus specificity	0.049	0.010	0.045	6.268E-27
1.276	0.017	Balance training omission, predicted area and threshold value	0.101	0.000	0.045	2.35E-20
9.117	0.133	Equate entropy of thresholded and non-thresholded distributions	0.044	0.015	0.091	1.569E-25

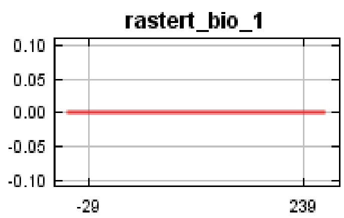
Pictures of the model

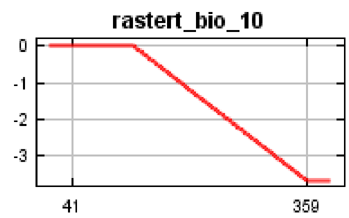
This is a representation of the Maxent model for Penstemon_bicolor. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

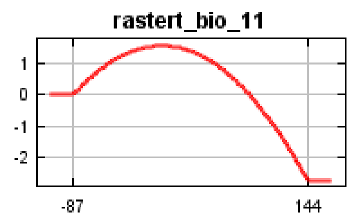


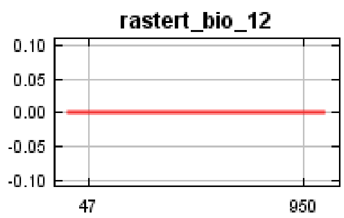
Response curves

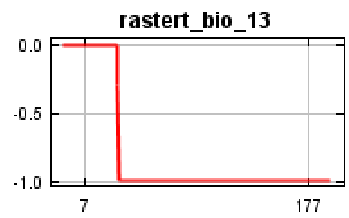
These curves show how each environmental variable affects the Maxent prediction. The (raw) Maxent model has the form exp(...)/constant, and the curves show how the exponent changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version.

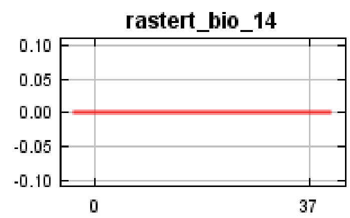


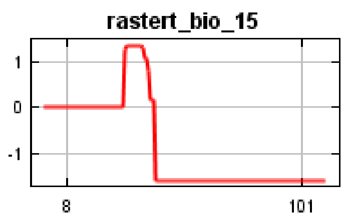


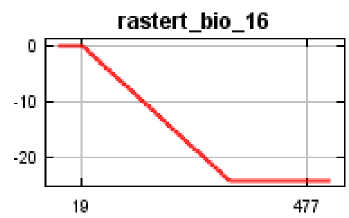


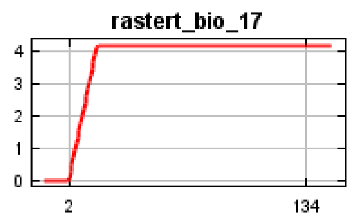


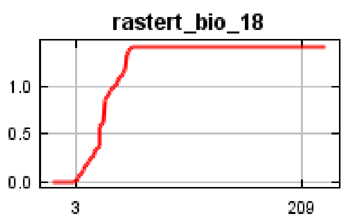


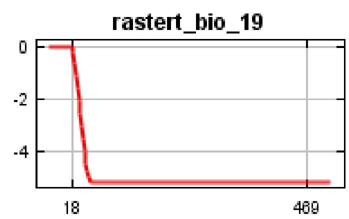


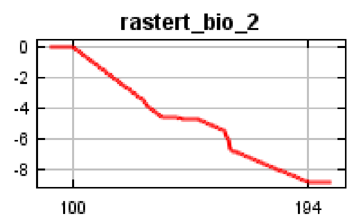


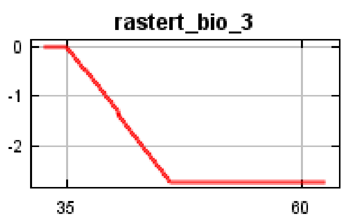


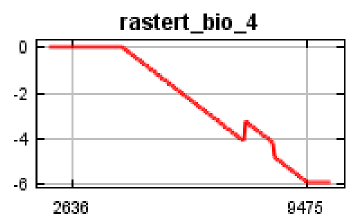


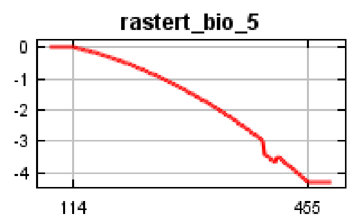


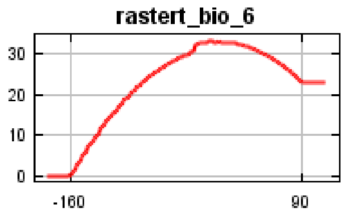


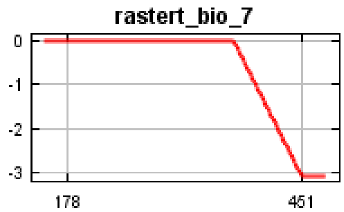


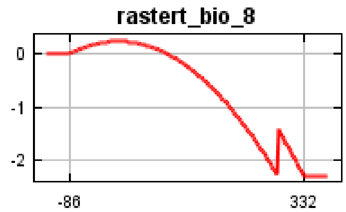


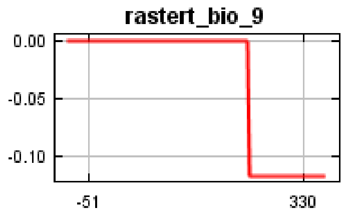










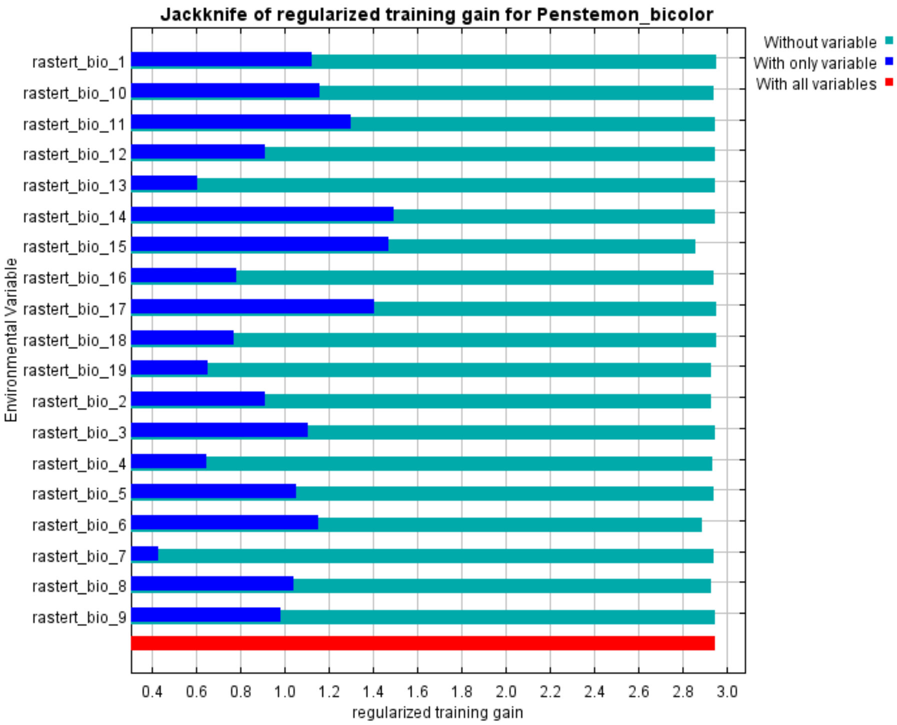


Analysis of variable contributions

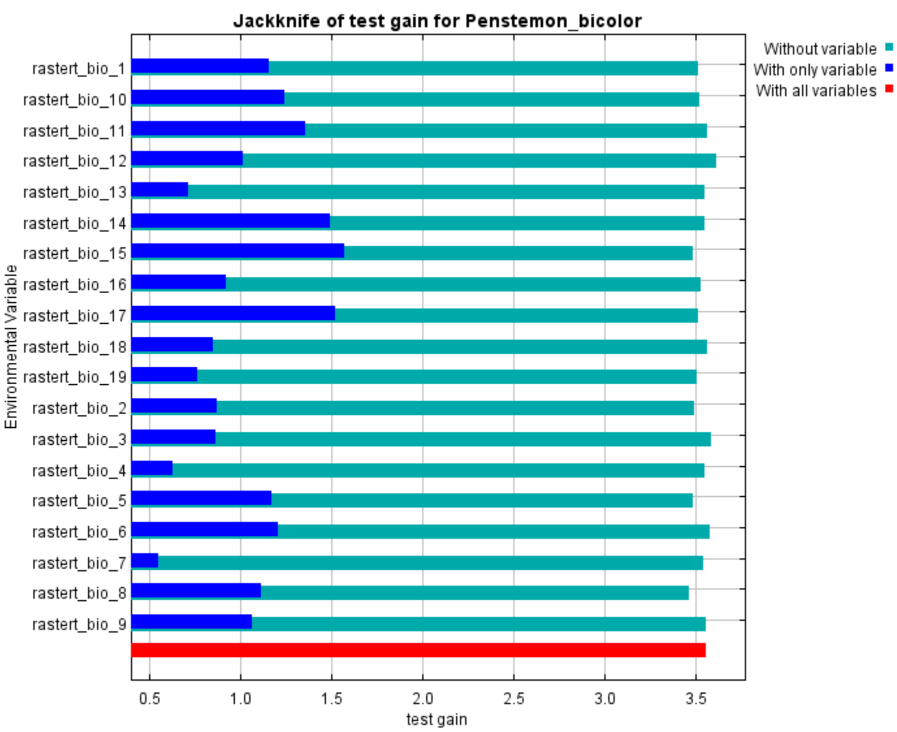
The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution
rastert_bio_15	24
rastert_bio_6	22.7
rastert_bio_3	15.8
rastert_bio_8	9.6
rastert_bio_10	7.8
rastert_bio_2	5.6
rastert_bio_7	3.8
rastert_bio_4	3.1
rastert_bio_12	2.3
rastert_bio_19	1.6
rastert_bio_5	1.5
rastert_bio_17	0.8
rastert_bio_13	0.6
rastert_bio_16	0.3
rastert_bio_11	0.2
rastert_bio_14	0.1
rastert_bio_18	0.1
rastert_bio_9	0
rastert_bio_1	0

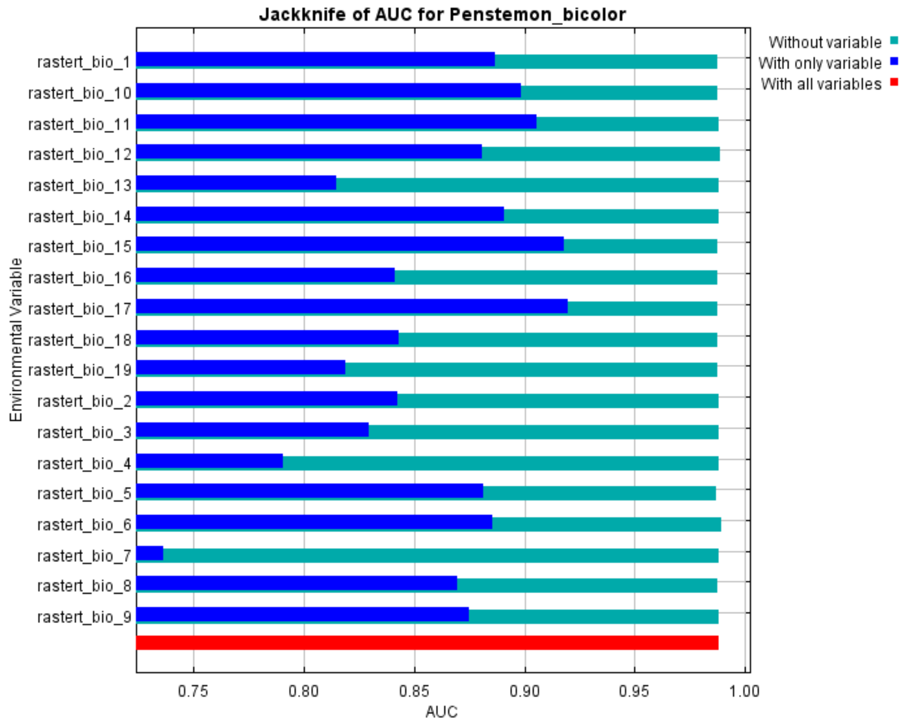
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is rastert_bio_14, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is rastert_bio_15, which therefore appears to have the most information that isn't present in the other variables.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Raw data outputs and control parameters

Regularized training gain is 2.944, training AUC is 0.994, unregularized training gain is 3.138.
Unregularized test gain is 3.561. Test AUC is 0.988, standard deviation is 0.005 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm terminated after 500 iterations (13 seconds).

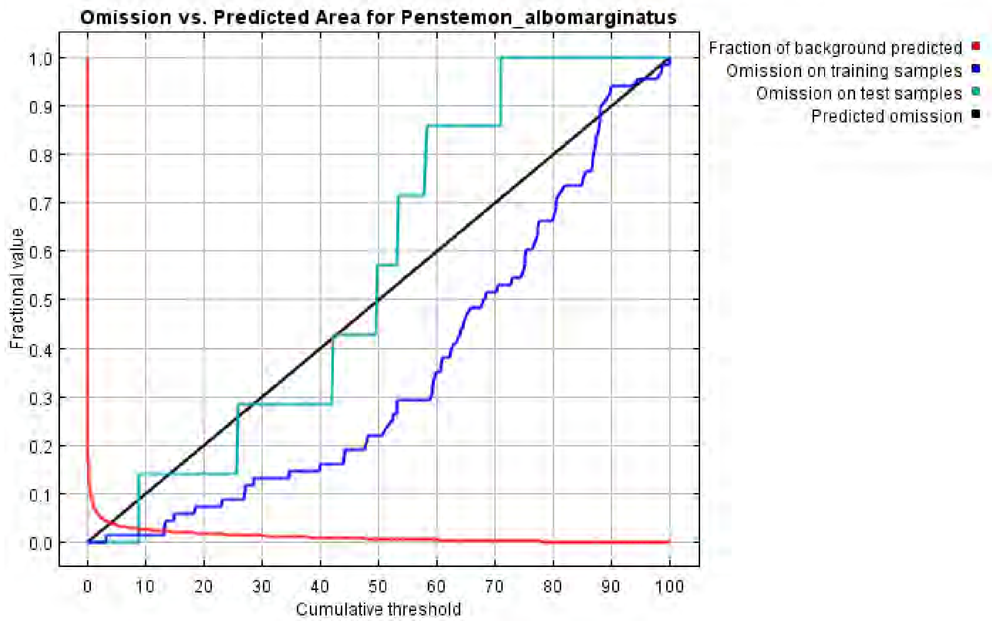
The follow parameters and settings were used during the run:
201 presence records used for training, 22 for testing.
10201 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used (all continuous): rastert_bio_1 rastert_bio_10 rastert_bio_11 rastert_bio_12 rastert_bio_13 rastert_bio_14 rastert_bio_15 rastert_bio_16 rastert_bio_17 rastert_bio_18 rastert_bio_19 rastert_bio_2 rastert_bio_3 rastert_bio_4 rastert_bio_5 rastert_bio_6 rastert_bio_7 rastert_bio_8 rastert_bio_9
Command line:
Feature types used: Linear Quadratic Product Threshold Hinge
Regularization multiplier is 1.0
Regularization values: linear/quadratic/product: 0.050 categorical: 0.250 threshold: 1.000 hinge: 0.500
Species file is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all\sand_plants_dd.csv
Environmental variables from I:\rare_plant_data\habitat_modeling_maxent\worldclim\bio_30s_esri\bio_ascii_sw
Output directory is I:\rare_plant_data\habitat_modeling_maxent\worldclim\models_all
Output format is Logistic
Output file type is .asc
Maximum iterations is 500
Convergence threshold is 1.0E-5
Random test percentage is 10
Jackknife selected
Remove duplicates selected
Make pictures selected
Create response curves selected

The following is a Maxent output for Penstemon albomarginatus run just using occurrence data for Clark County and not the species full distribution. It is included for information purposes only to allow one to compare the results of running a Maxent model for just a subset of a species distribution with the output above which was run using the species full distribution. The output from the following run was not used in the results.

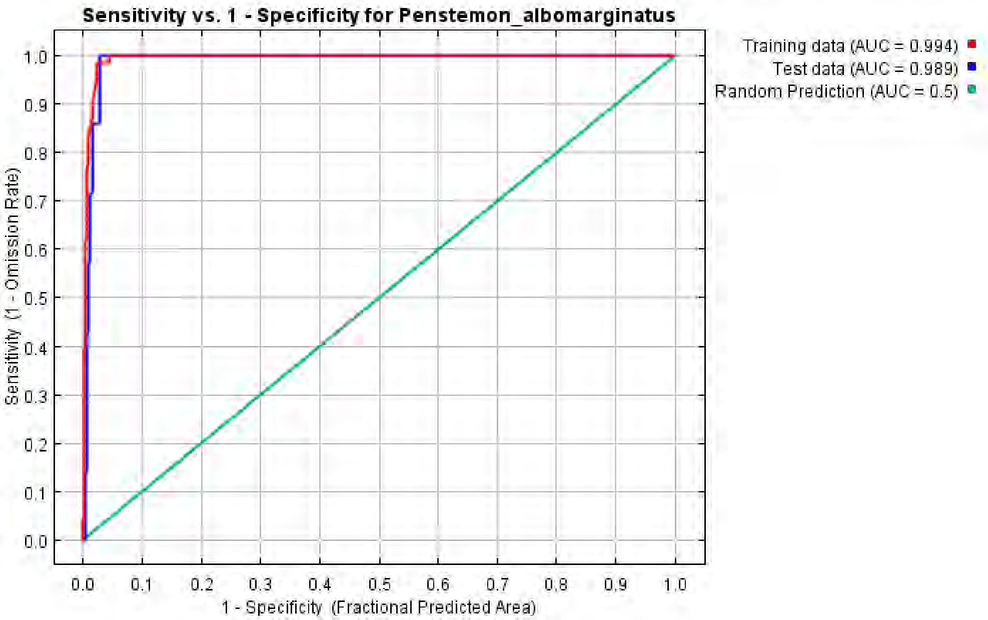
Maxent model for Penstemon_albomarginatus

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.



The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.987 rather than 1; in practice the test AUC may exceed this bound.

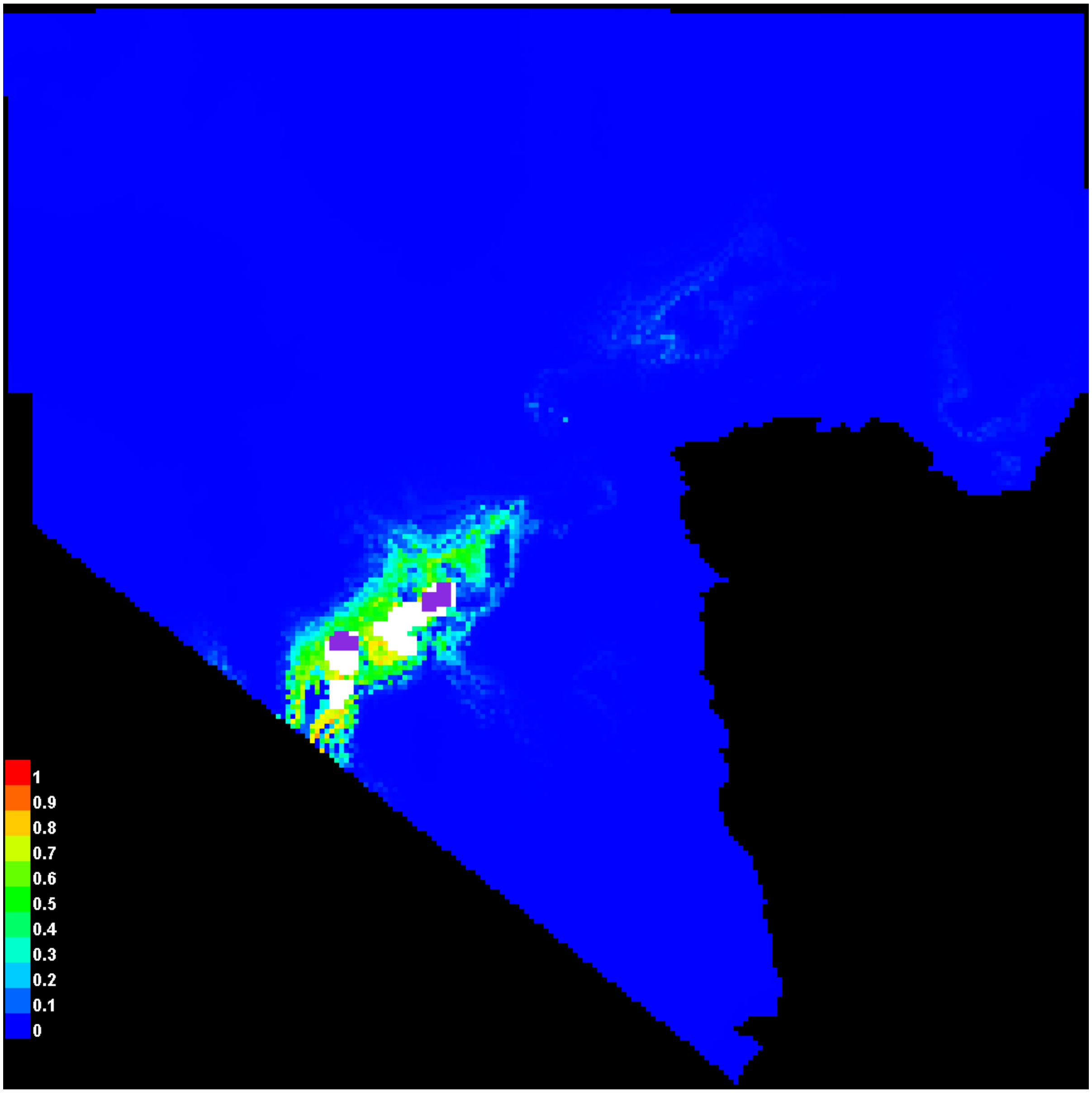


Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.008	Fixed cumulative value	0.078	0.000	0.000	1.741E-8
5.000	0.094	Fixed cumulative value	0.036	0.015	0.000	8.146E-11
10.000	0.224	Fixed cumulative value	0.027	0.015	0.143	2.708E-9
3.125	0.046	Minimum training presence	0.045	0.000	0.000	3.679E-10
26.889	0.424	10 percentile training presence	0.016	0.088	0.286	1.951E-8
13.280	0.280	Equal training sensitivity and specificity	0.024	0.029	0.143	1.278E-9
13.208	0.273	Maximum training sensitivity plus specificity	0.024	0.015	0.143	1.278E-9
8.859	0.196	Equal test sensitivity and specificity	0.028	0.015	0.000	1.527E-11
8.859	0.196	Maximum test sensitivity plus specificity	0.028	0.015	0.000	1.527E-11
1.282	0.013	Balance training omission, predicted area and threshold value	0.069	0.000	0.000	7.676E-9
6.295	0.133	Equate entropy of thresholded and non-thresholded distributions	0.033	0.015	0.000	4.084E-11

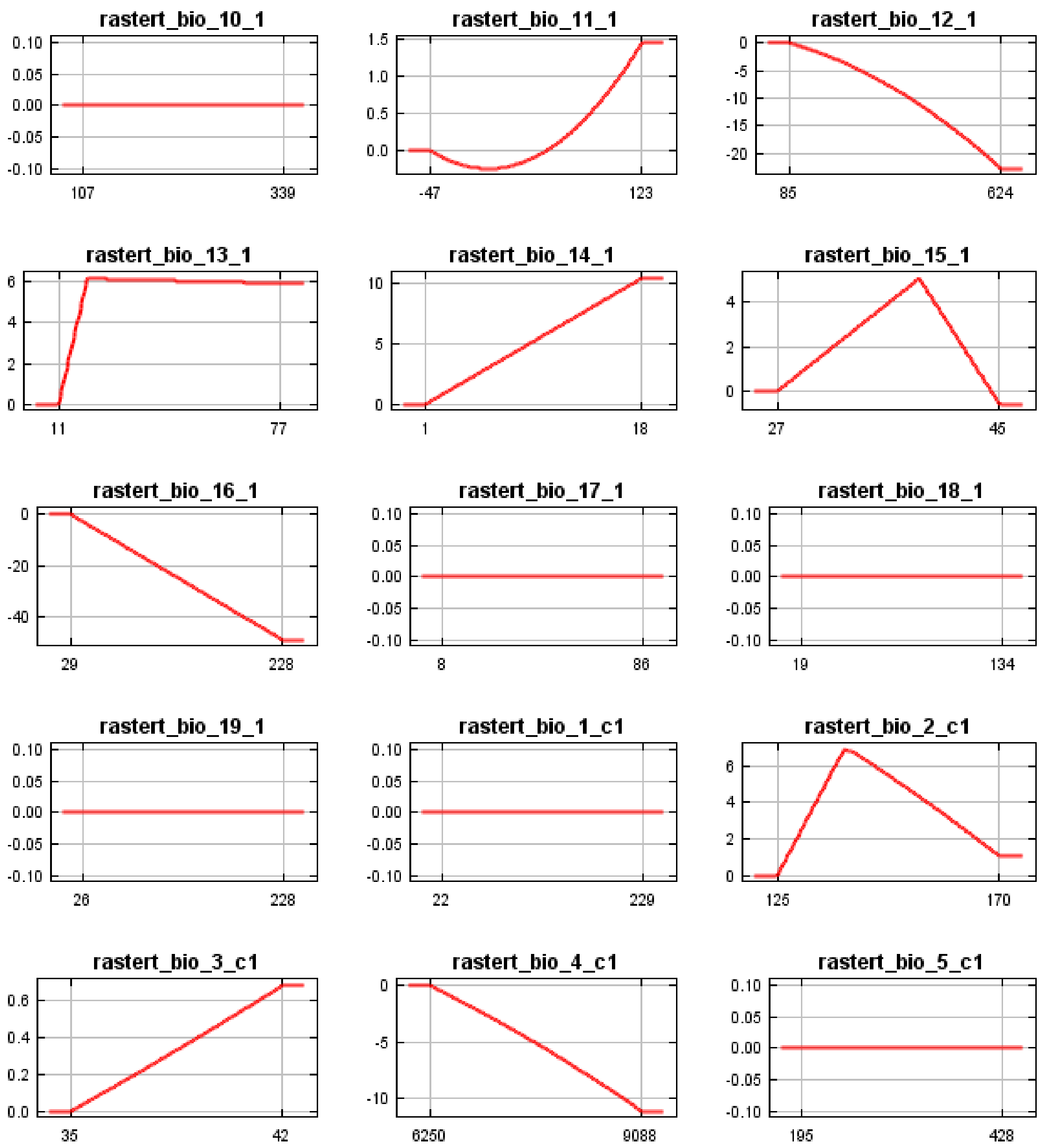
Pictures of the model

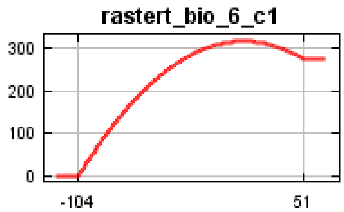
This is a representation of the Maxent model for Penstemon_albomarginatus. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

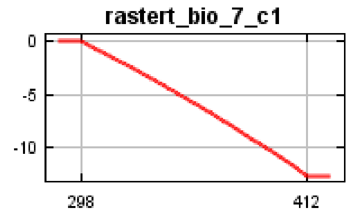


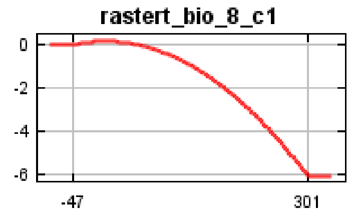
Response curves

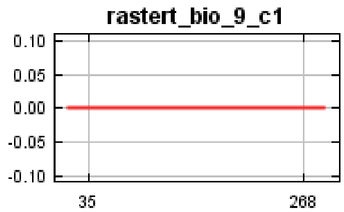
These curves show how each environmental variable affects the Maxent prediction. The (raw) Maxent model has the form exp(...)/constant, and the curves show how the exponent changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version.









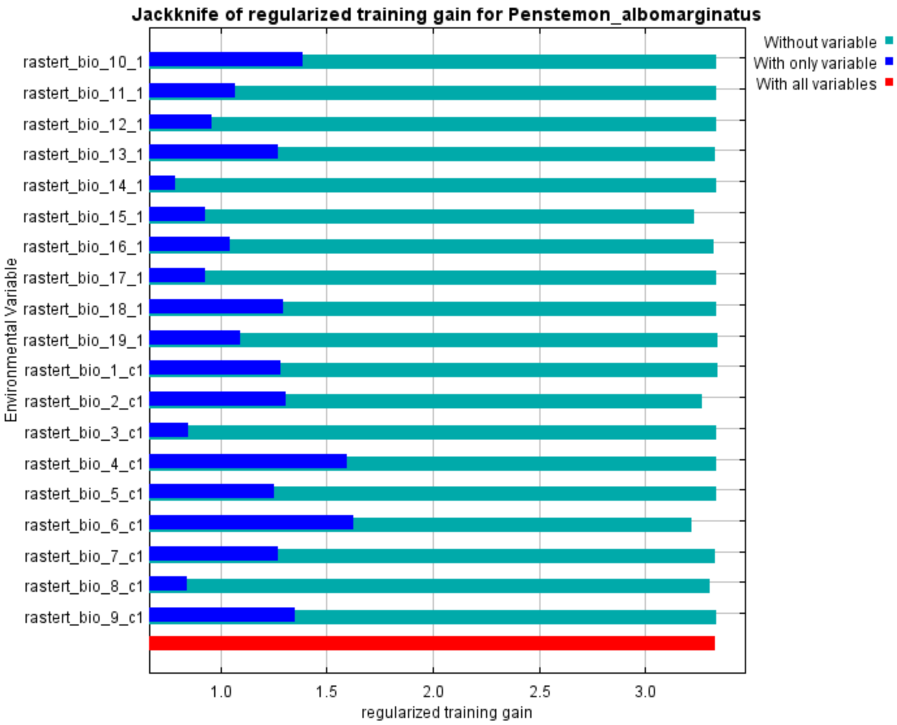


Analysis of variable contributions

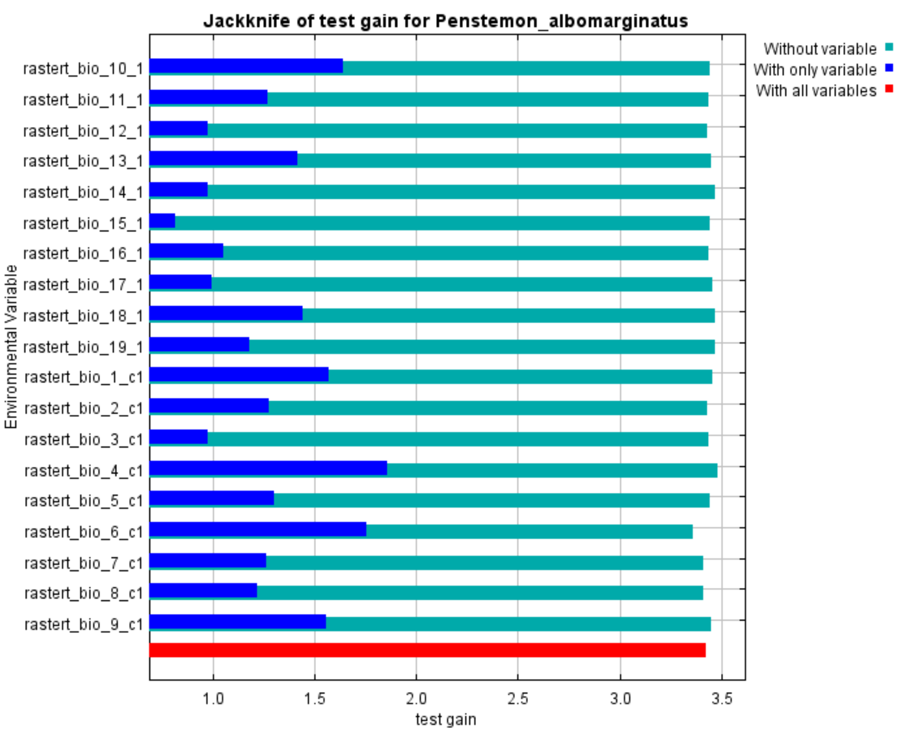
The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution
rastert_bio_2_c1	23.4
rastert_bio_4_c1	21.7
rastert_bio_3_c1	18.7
rastert_bio_6_c1	14.1
rastert_bio_16_1	9.1
rastert_bio_15_1	6.8
rastert_bio_13_1	3.3
rastert_bio_7_c1	2.2
rastert_bio_8_c1	0.7
rastert_bio_14_1	0
rastert_bio_18_1	0
rastert_bio_11_1	0
rastert_bio_12_1	0
rastert_bio_9_c1	0
rastert_bio_5_c1	0
rastert_bio_1_c1	0
rastert_bio_19_1	0
rastert_bio_17_1	0
rastert_bio_10_1	0

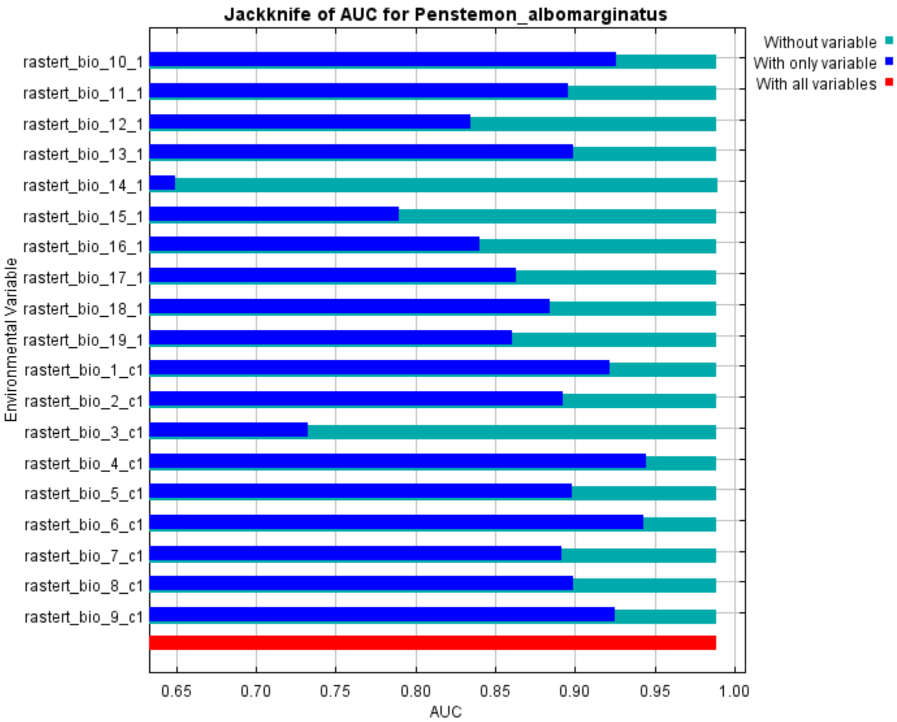
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is rastert_bio_6_c1, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is rastert_bio_6_c1, which therefore appears to have the most information that isn't present in the other variables.



The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.



Lastly, we have the same jackknife test, using AUC on test data.



Raw data outputs and control parameters

Regularized training gain is 3.331, training AUC is 0.994, unregularized training gain is 3.583.
Unregularized test gain is 3.423. Test AUC is 0.989, standard deviation is 0.003 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm terminated after 500 iterations (8 seconds).

The follow parameters and settings were used during the run:
68 presence records used for training, 7 for testing.
10068 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used (all continuous): rastert_bio_10_1 rastert_bio_11_1 rastert_bio_12_1 rastert_bio_13_1 rastert_bio_14_1 rastert_bio_15_1 rastert_bio_16_1 rastert_bio_17_1 rastert_bio_18_1 rastert_bio_19_1 rastert_bio_1_c1 rastert_bio_2_c1 rastert_bio_3_c1 rastert_bio_4_c1 rastert_bio_5_c1 rastert_bio_6_c1 rastert_bio_7_c1 rastert_bio_8_c1 rastert_bio_9_c1
Command line:
Feature types used: Linear Quadratic Hinge
Regularization multiplier is 1.0
Regularization values: linear/quadratic/product: 0.141 categorical: 0.250 hinge: 0.500
Species file is E:\rare_plant_data\habitat_modeling_maxent\worldclim\Pen_albo_dd.csv
Environmental variables from E:\rare_plant_data\habitat_modeling_maxent\worldclim\bio_30s_esri\bio_ascii
Output directory is E:\rare_plant_data\habitat_modeling_maxent\worldclim\test
Output format is Logistic
Output file type is .asc
Maximum iterations is 500
Convergence threshold is 1.0E-5
Random test percentage is 10
Jackknife selected
Remove duplicates selected
Make pictures selected
Create response curves selected