

PREDICTING HYDROMANTES SHASTAE
OCCURRENCES IN SHASTA COUNTY, CALIFORNIA

by

Teri Lorraine Mooney

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

December 2010

Copyright 2010

Teri Lorraine Mooney

Acknowledgements

I would like to express my sincere gratitude to my thesis advisor, Dr. John P. Wilson, for his knowledge, support and guidance; to the members of my thesis committee, Dr. Travis R. Longcore and Dr. Jennifer N. Swift, for their interest and contribution; and to my academic advisor, Katherine Kelsey, for her exceptional and dependable assistance.

Table of Contents

Acknowledgements	ii
List of Tables	iv
List of Figures	v
Abstract	vi
Chapter 1: Introduction	1
1.1 Species Distribution Modeling	1
1.2 Description of Species	1
1.3 Description of Study Area	4
Chapter 2: Literature Review	6
2.1 Species Distribution Models	6
2.2 Correlative and Mechanistic Modeling Approaches	6
2.3 Presence/Absence or Presence-Only Data	7
2.4 Model Selection Factors	8
2.4.1 Data Quality	8
2.4.2 Sample Size	8
2.4.3 Data Type	10
2.4.4 Data Scale	11
2.4.5 Data Relevance/Predictor Selection	11
2.4.6 Linear and Curvilinear Responses	12
2.5 Model Performance Measures	13
Chapter 3: Data and Methods	15
3.1 Biological Data	15
3.2 Environmental Variables	16
3.3 Species Distribution Model	19
3.4 Model Performance Measures	21
Chapter 4: Results	23
4.1 Species Distribution Map	23
4.2 Model Performance	25
4.3 Response Curves	26
4.4 Predictor Variable Importance	29
Chapter 5: Discussion and Conclusions	31
5.1 Model Strengths	31
5.2 Model Challenges	31
5.3 Model Viability for <i>Hydromantes shastae</i>	33
5.4 Future Uses of Research	34
References	35

List of Tables

1. Types and sources of biological and environmental data used in Maxent modeling of <i>Hydromantes shastae</i> distribution in Shasta County, CA	16
2. Descriptions of geology classes	18
3. Description of land cover classes	18
4. Regularization parameters and applied constraints	21
5. AUC classification	22
6. Variable contribution	30

List of Figures

1. Flattened body (a) and webbed toes (b) assist in climbing sheer and slippery surfaces; characteristics include a short tail (c) and juveniles resembling adults (d)	3
2. Location of study area	5
3. Conceptual summary of modeling method	20
4. Maxent representation of predictive <i>Hydromantes shastae</i> distribution. Red and orange colors represent more accurate predictions and blues represent lower probability of occupancy. Dots represent training (white) and testing (purple) data	23
5. Maxent representation of predictive <i>Hydromantes shastae</i> distribution with study area boundary and an alternative palette for the habitat suitability values	24
6. Binary model results from threshold, representing suitable and unsuitable habitat. All of the test samples fall within suitable habitat areas	25
7. AUC curve for the model summarizing model performance	26
8. Maxent response curves for the five predictor variables: (a) elevation, (b) geology, (c) land cover, (d) precipitation, and (e) soils	27
9. Jackknife results for (a) training gain, (b) test gain, and (c) AUC	30

Abstract

Species distribution models use species occurrence data and environmental variables to estimate species-habitat relationships and predict potentially suitable habitat. This research analyzes the usefulness of a maximum entropy model, Maxent, for estimating species occurrences and environmental predictor variables for *Hydromantes shastae*, a rare species of salamander with a small geographic extent and limited occurrence records. Environmental variables included elevation, geology, land cover, precipitation, and soils. Seventy-five percent of the presence data was used to train the model and the remaining 25% was used for testing. Model performance was measured by area under the Receiver Operating Characteristics (ROC) curve (AUC). The AUC of 0.879 indicated that the model performed substantially better than a random prediction. The log loss plot indicated that soils contributed most to model fit. These results indicate Maxent's effectiveness for identifying potentially suitable habitat for *H. shastae* and predicting potential species occurrences. This model can be used to support species impact analyses and conservation efforts. Further, this model could be enhanced to focus surveys for populations in new areas and predict species responses to altered environmental conditions.

Chapter 1:

Introduction

1.1 Species Distribution Modeling

For rare species with limited distributions, such as the salamander *H. shastae*, occurrence data and knowledge of habitat requirements can be scarce. Species distribution modeling (SDM) provides an alternative method for identifying suitable habitat and predicting potential species distribution when thorough species data are not available. Models that perform well can inform species impact studies and conservation efforts. This study used Maxent, a maximum entropy machine learning model (Phillips 2010). Unlike many other SDMs, Maxent can be run without true absence data and with a small sample size and still achieve high predictive accuracy (Phillips and Dudik 2008). The maximum entropy approach can be effective even with incomplete data because it uses what is known and avoids assuming anything about what is unknown (Jaynes 1990). The species input data to be used in this research are observed occurrences of *Hydromantes shastae* recorded in the California Natural Diversity Database (CNDDB), a continually updated inventory of rare plants and animals in California, maintained by the California Department of Fish and Game (California Department of Fish and Game 2010). CNDDB data is a common biological input for species distribution modeling (Meentemeyer et al. 2004; Hernandez et al. 2006; Thorne et al. 2006; Williams et al. 2009). Environmental variables included elevation, precipitation, geology, land cover and soils. This work analyses the usefulness of Maxent to predict suitable habitat and potential occurrences of the *H. shastae*.

1.2 Description of Species

The Shasta salamander (*H. shastae*) is endemic to California. It is found in the headwaters of the Shasta Reservoir drainage in Shasta County, California, in an area less than 35 km across its greatest dimension (Wake and Papenfuss 2005). Most of its range is within the Shasta-Trinity

National Forest (Hansen and Papenfuss 1994), but surveys in other potential habitat areas have not been conducted. The Shasta salamander has a discontinuous distribution within its range (Hansen and Papenfuss 1994). There are presently 64 known occurrence sites (California Department of Fish and Game 2010).

The primary habitat of the Shasta salamander is defined by limestone rock outcrops and the slopes surrounding these outcrops (Olsen and Lewendal 1999). It is known to occur in the Kennett Formation, McCloud Limestone and Hosselkus Limestone, but not the Pit formation (Lewendal 1995). Although one population has been observed on a volcanic rock outcrop (Papenfuss and Cross 1980), and it may occur in other non-limestone areas (Lewendal 1995; Lindstrand 2000), occurrences in non-limestone habitats have seldom been documented. The species occurs in elevations ranging from 330 to 773 m. Land cover types include hardwood-conifer, ponderosa pine, and mixed conifer habitats.

The Shasta salamander exhibits an entirely terrestrial life cycle (Olsen and Lewendal 1999). The species is sensitive to temperature and moisture, and occurs in cool, moist microhabitats (Olsen and Lewendal 1999). As a member of the family Plethodontidae, it is a lungless salamander that respire through its thin, moist skin (Olsen and Lewendal 1999). It has a flattened body (Figure 1a), webbed feet (Figure 1b), a short tail, and a long, projectile tongue (Gorman and Camp 1953). It ranges from 7.5 to 11 cm in total length (Gorman and Camp 1953). Juvenile members resemble adults (Figure 1c) (Gorman 1956). Genetic distinctions may be present since populations do not appear to migrate between separated outcrops (Hansen and Papenfuss 1994).



(a)



(b)



(c)



(d)

Figure 1: Flattened body (a) and webbed toes (b) assist in climbing sheer and slippery surfaces; characteristics include a short tail (c) and juveniles resembling adults (d) (Photos reproduced with permission of Gary Nafis and www.californiaherps.com)

The Shasta salamander is most active in the evening during rainy periods in the fall, winter and spring, and is typically found at the surface in limestone areas and under limestone boulders and logs (Hansen and Papenfuss 1994; Stebbins 1985). It is known to be an adept climber, using its webbed feet and short tail to climb over sheer and slippery rock surfaces (Gorman and Camp 1953). During periods of low surface moisture, the Shasta salamander retreats to subsurface cave refuges of rocks, downed wood, vegetative litter, and substrata (Gorman 1956; Keen 1982; Lindstrand 2000).

The 1949 filling of Shasta Lake submerged some of the Shasta salamander's historical range, and current human activities including recreation, mining, logging, and water management activities around Shasta Lake continue to impact its habitat (Hansen and Papenfuss 1994). In 1979, the United States Forest Service initiated a special management plan for the Shasta salamander designed to both prevent habitat disturbances and re-vegetate habitat to promote

lower temperatures (Papenfuss and Brouha 1979). In addition, a 300 foot (91 m) habitat protection buffer zone was proposed (Hansen and Papenfuss 1994). In 1999, an investigation was implemented by the United States Department of the Interior, Bureau of Reclamation to study the potential impacts of raising Shasta Dam (United States Department of the Interior, Bureau of Reclamation 1999), which would flood salamander habitat and thereby threaten the genetic diversity and overall survival of the species (Hansen and Papenfuss 1994).

Current conservation status listings for the Shasta salamander include “threatened” by the California Department of Fish and Game, “category 2” candidate for listing as an endangered species by the United States Fish and Wildlife Service, “sensitive” by the United States Bureau of Land Management and the United States Forest Service, “critically impaired” by NatureServe Global Conservation Status Ranks, and “vulnerable” by the World Conservation Union (California Department of Fish and Game 2009).

1.3 Description of Study Area

The study area encompasses the known Shasta salamander habitat range in Shasta County, California (Figure 1) (Lindstrand 2000). It measures 2,320 km² (573,239 acres), with elevation ranging from 153 to 1711 m and average annual precipitation values ranging from 125 to 250 cm. The major hydrologic feature within the study area is Shasta Lake, the largest freshwater reservoir in the state of California, covering 121 km² (30,000 acres). The United States Geological Survey Gap Analysis database suggests that approximately 84% of the study area is covered with forest and woodland systems, with 6% shrubland, steppe and savanna systems, 5% aquatic, 2% riparian and wetland, 1% each recently disturbed or modified, human land use, and sparse and barren, and less than 1% grassland systems (United States Geological Survey 2010b). Primary geologic bedrock includes lava flows, tuff beds, volcanic rocks and limestone outcrops, including four limestone belts: the Kennett Formation, McCloud Limestone, Hosselkus, and Pit formations (Lewendal 1995). Jepson bioregions in the study area include the Klamath Ranges,

High Cascade Range, Cascade Range Foothills, Modoc Plateau, and North Coast Ranges (Jepson and Hickman 1993).

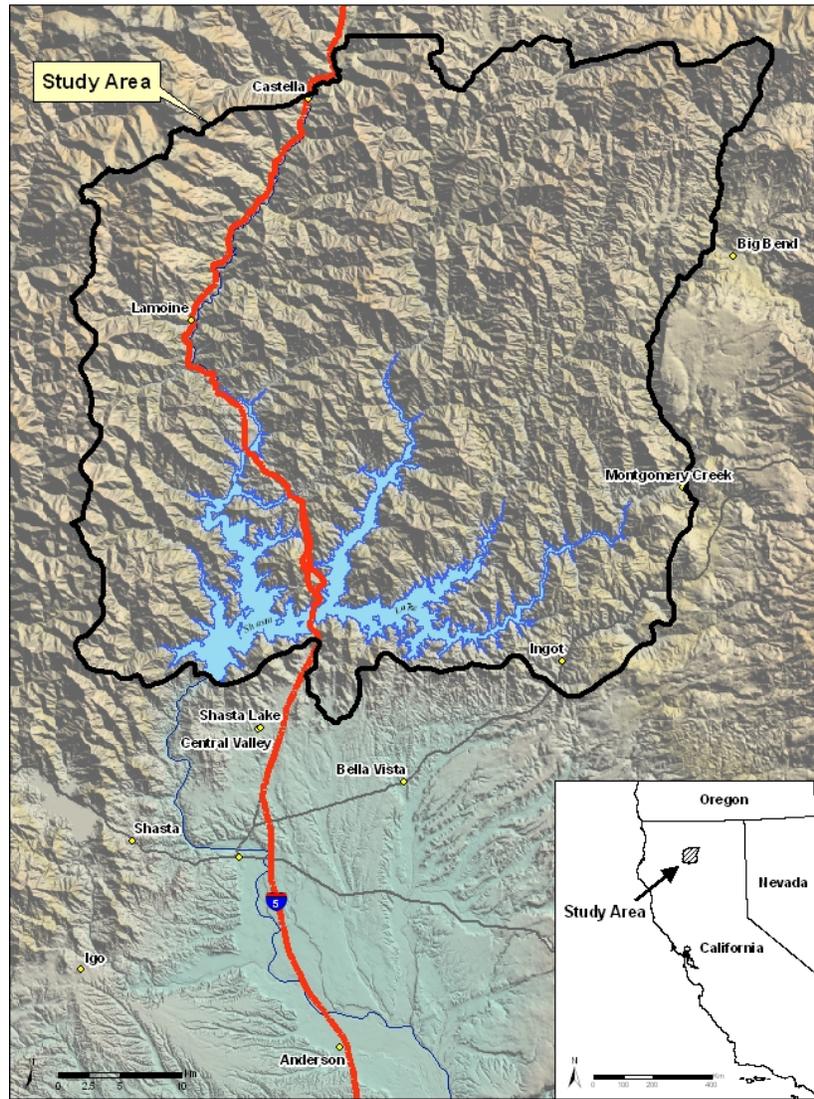


Figure 2: Location of study area

Ownership of the study area is approximately 60% United States Forest Service, 38% private land, and 2% United States Department of the Interior Bureau of Land Management and California State-owned lands. Human activities in the study area include mining, logging, water resource management (including damming, storage, power generation, provisioning, and recreation), and outdoor recreation.

Chapter 2:

Literature Review

2.1 Species Distribution Models

Species distribution models are used for examining and predicting the natural distribution of species. They use data to describe empirical correlations between biological data of species occurrence and the environmental variables that define the species' physical environment (Pearson 2007). Resulting predictions identify the species niche and suitable habitat to support the species. These data are useful for environmental research, resource management, and conservation efforts.

Species distribution models use a variety of approaches. Statistical models include generalized linear models (GLMs), generalized additive models (GAMs), boosted regression trees (BRT), and multivariate adaptive regression splines (MARS). Machine learning techniques include artificial neural network (ANN; Pearson et al. 2002), genetic algorithm (GA; Stockwell and Peters 1999), ecological niche factor analysis (ENFA; Hirzel et al. 2002), and maximum entropy (Maxent; Phillips et al. 2006) models. There are also heuristic models, such as BIOCLIM, which model sequential BIOSphere systems under CLIMate change (Beaumont and Hughes 2002) and combinatorial optimization methods such as the genetic algorithm for rule-set production (GARP; Fitzpatrick et al. 2007).

2.2 Correlative and Mechanistic Modeling Approaches

Two distinct modeling approaches, correlative and mechanistic, are used to make species distribution predictions (Pearson 2007). Correlative SDMs use two types of input data: biological data, which is the observed or sampled data that characterize the known species distribution, and environmental data that represents environmental factors that characterize the species' physical

environment. Mechanistic SDMs do not use observed biological distribution sample data but instead predict distribution based on species physiological and environmental limitations.

The types of available input data will determine whether a correlative or a mechanistic SDM should be used. A correlative method can be used when species occurrence data are available. To produce reliable results, a correlative method should only be used when at least the minimum number of reliable occurrence records necessary for the model to perform accurately (as identified by the literature) is available (Elith and Leathwick 2009; Stockwell and Peterson 2002; Wisz et al. 2008). A mechanistic approach can be utilized when no species occurrence data is available, or when the sample size is too small to provide accurate results in a correlative model. Mechanistic models require an increased knowledge of the species' biology as well as eco-physiological requirements (Robertson et al. 2003). Mechanistic models may be more appropriate than correlative methods at finer spatial scales (Guisan and Zimmermann 2000; Helmuth et al. 2005; Richardson et al. 2004). A mechanistic approach should be used only when the knowledge of species can be relied upon to determine the environmental predictors that influence the species distribution. Because the mechanistic approach uses few or no known observations, they may be more difficult to validate than the correlative methods (Pearson 2007).

2.3 Presence/Absence or Presence-Only Data

All correlative models use species presence data (which defines where the species has been observed) as the primary biological input. However, some correlative models also require species absence data (which defines where the species has been observed to be absent). Common models that require absence data include the GLM, GAM, BRT, MARS, GA, and ANN methods. While Brotons et al. (2004) showed that absence data may improve model performance, absence records are often times unavailable or include false absences, making them unreliable (Pearson 2007). When absence data are not available, models that require absence data can use sampled pseudo-

absence data which can be randomly selected or generated using weighted criteria (Engler et al. 2004).

2.4 Model Selection Factors

Typical factors considered when selecting the best fitting species distribution model include data characteristics, data relevance/predictor selection, and linear vs. curvilinear responses.

2.4.1 Data Quality

Data used for model inputs must be accurate in order for model results to be accurate (Johnson and Gillingham 2005). In an SDM comparison study, data error and bias were found to be the primary contributors to inaccurate model results (Wisiz et al. 2008). Data error can come from many sources which might include inaccuracy and imprecision in data capture methods. Data bias might come from collection efforts conducted in areas that are more easily accessible and therefore not a true sample. Biased results might come from data that was collected for a specific purpose other than to be used as SDM input (Austin 2007; Elith and Leathwick 2009). For example, if the observed species data set is acquired from a museum, it may well have been collected with the goals of preserving unusual or rare species, which would prompt different collection methods and acquired data than would a collection effort focused on collecting species data for a distribution study. One way to address this bias is to supplement the existing data with new survey data (Austin 2007) or compare data from different sources to examine the accuracy of species' range representations (Hernandez et al. 2006).

2.4.2 Sample Size

Sample size of required data is an important factor influencing model results. Since the amount of observed data available to be used for model input affects the results, sample size must be considered in the model selection process. Research has demonstrated significant impact of the biological data sample size on model performance (Elith and Leathwick 2009; Johnson and

Gillingham 2008; Pearson 2007; Robertson et al. 2003; Stockwell and Peterson 2002; Wisz et al. 2008; Zuo et al. 2008). In certain instances, model accuracy is improved with increases in environmental data (Zuo et al. 2008) and reduced with decreases in sample size (Elith and Leathwick 2009). Results of a study comparing 12 different species distribution models (at sample sizes of 10, 30, and 100) showed that increasing sample size leads to increases in model performance and decreases in the variability of predictive accuracy (Wisz et al. 2008). Because none of the models tested predicted well with sample sizes less than 30, the researchers concluded that small sample sizes should be used for exploratory purposes only.

Wisz et al. (2008) concluded that methods that performed best with large sample sizes, like GBM, MARS-INT and BRUTO, did not perform well with smaller sample sizes; Maxent, a maximum entropy model often cited as being an accurate predictor of species distribution, performed well with both small and large sample sizes; and DOMAIN (a distance metric algorithm) and OM-GARP (a genetic algorithm) performed among the best with small sample sizes, but only moderately with larger sample sizes.

Stockwell and Peterson (2002) compared three models for effects of sample size on model performance. Results showed that a machine learning method, genetic algorithm for rule-set production (GARP), and a coarse surrogate method (CSM) had accuracy rates within 90% of the maximum obtained with 10 sample points and achieved near maximal accuracy with 50 sample points, while a fine surrogate method (FSM) and a logistic regression (LR) model required 100 data points to achieve near maximal accuracy (Stockwell and Peterson 2002).

Hernandez et al. (2006) found the machine learning Maxent model more capable than three other modeling methods (Domain, GARP, and Bioclim) for producing accurate results with sample sizes ranging from 5 to 25 occurrences. LR has markedly low accuracy with small data sets (Elith and Leathwick 2009; Stockwell and Peterson 2002; Wisz et al. 2008).

2.4.3 Data Type

The input data available will influence the modeling methods considered for a particular research effort. Some models require both data on where the species has been recorded as present as well as where the species has been recorded to be absent. These include GLMs, GAMs, and ANNs. When absence data is not available, a presence-only model will be required. These include BIOCLIM, DOMAIN, Maxent, and ENFA (Pearson 2007).

While multiple data types can be used as biological model inputs, raster grids of continuous data are most effective for environmental variables; discrete environmental data, like soil type or land cover, are not functional with many SDMs (Pearson 2007). Representing environmental data as a continuous surface ensures that every cell containing biological data will contain environmental data as well (Pearson 2007). The resulting cells containing both biological and environmental data are then used by the model to predict the missing biological data.

The need to discern the degrees of influence of predictor variables on model results will also influence model selection. Regression models, like geographically weighted regression (GWR), GAMs and GLMs, provide output that includes the relative influence of input variables (Guisan et al. 2002). However, “black box” models, like ANNs, do not provide such details of their functioning and so may not be suitable if relative variable influence is important (O’Sullivan and Unwin 2003).

The desired form of output data will also influence model selection. Output options might include a continuous prediction where probability values range from 0 to 1, or a binary prediction where, for example, 1 represents species presence and 0 represents absence. If a model produces continuous data, a threshold value can be used to apply binary classifications (Guisan et al. 2002).

2.4.4 Data Scale

To appropriately serve research goals, the spatial scale, including both the extent and the resolution, should be selected based on several influencing criteria: the phenomenon being modeled (Austin 2007; Elith and Leathwick 2009; Johnson and Gillingham 2005; Pearson 2007), the data, and the intended use of the results (Elith and Leathwick 2009). The extent is associated with the purpose of the analysis, for example, whether the study extent is global, or instead covers a smaller region. The resolution is more often than not a function of the data properties, spatial accuracy, and the precision of the biological data (Elith and Leathwick 2009). Correlative models are more appropriate for global or regional analysis while mechanistic approaches might be more appropriate for finer resolution output (Guisan and Zimmermann 2000; Helmuth et al. 2005; Richardson et al. 2004).

2.4.5 Data Relevance/Predictor Selection

Model data input requirements vary, determined by both the model being used and the ecology of the species under study. For the biological data, which represents the observed locations of a species, some models need presence-only data, while others require both presence and absence data.

Selecting the appropriate environmental variables which will serve as the predictors requires knowledge of the meaningful relationships between the biological and environmental data. Elith and Leathwick (2009) stressed that environmental variables must be ecologically relevant to the species under study in order to produce a well-fitted model. In addition to the goal of the study and data quality, model results are influenced by the biology of the organism and knowledge about the organism (Johnson and Gillingham 2005; Robertson et al. 2003). Mixed models might be utilized in certain circumstances. For example, species that have mobility (like birds and fish) require different modeling approaches than sessile ones because models with a

mobile component must include movement descriptors (Elith and Leathwick 2009). Temporal predictors are necessary for incorporating seasonal variation (Elith and Leathwick 2009). Elith and Leathwick (2009) conclude that prediction accuracy depends more on species characteristics than modeling technique differences.

There are many techniques to select predictors. The most appropriate methods will depend on the specifics of the study (Johnson and Gillingham 2005). For model results to predict species distributions accurately, the environmental influences on the species must be understood and represented in the data in a relevant and appropriate manner. To ensure relevance, the focus should be on environmental processes, both ecological and biophysical, that influence the species, data availability, and model purpose (Austin 2007). Error in predictor data can have a significant impact on results and for this reason Austin (2007) suggested using direct variables (for example sunlight) because they bring less potential for error than using indirect variables (for example slope and aspect) that have been derived from direct sources.

Methods for incorporating relevant biologic-environmental interactions include combining regression trees with other approaches, ridge regression, and stepwise selection. Methods to apply variable weighting of predictors include using regression, convex, or alpha hulls (Elith and Leathwick 2009). Envelope methods may be used to construct and apply predictors with equal weights.

2.4.6 Linear and Curvilinear Responses

Austin (2007) explained that while regression methods are commonly used to model species distributions, the curvilinear responses that are common in relationships between species and their environments are not well modeled in several regression methods. For example, structural equation modeling (SEM), described as strong for testing causality, and geographically weighted regression (GWR), which is good at examining spatial non-stationarity of ecological processes and taking into account spatial autocorrelation and factors that vary in space, are not adept at

addressing curvilinear relationships and interactions (Austin 2007). In an SDM comparative study by Lippitt et al. (2008), logistic regression parametric assumptions prevented the model from accurately characterizing the non-linear biological and environmental relationships. In SDM comparative studies by Austin (2007) and Wisz et al. (2008), Maxent is cited as being strong at predicting curvilinear species distributions.

2.5 Model Performance Measures

There are many different approaches for validating a model's performance. The AUC, Kappa coefficient, and correlation coefficients are commonly used to measure predictive performance (Elith and Leathwick 2009). The AUC is a single value that assesses model accuracy. For species distribution models, it measures the distinction between sites where species presence is likely against sites where species presence is unlikely. Swets (1988) classification or AUC can be used for interpretation and describes the possible values of AUC as follows: 0.09 to 1.00 is excellent, 0.80 to 0.90 is good, 0.70 to 0.80 is fair, 0.60 to 0.70 is poor, and < 0.60 is failing.

The Kappa coefficient shows the level of agreement between the observed and predicted data. A higher K value suggests better agreement between the observed and predicted values. It is possible for the Kappa coefficient to overestimate model performance (Robertson et al. 2003). AUC is positively related to the model's performance (Pearson 2007; Zuo et al. 2008). It is a common and desirable method for assessing model success (Austin 2002) and applicable to any species distribution model (Phillips and Dudik 2008).

Binary predictions can be validated with an error matrix. The results of the error matrix can then be statistically tested with, for example, the Kappa coefficient, binomial, or chi-squared tests. The choice of validation tests will depend on the data used and the purpose of the model. Some software packages include test statistic tools (e.g. ArcGIS and DIVA-GIS). Continuous predictions that are not converted to binary output can be assessed via threshold-independent assessment (Pearson 2007).

Spatial autocorrelation, the likelihood that nearby objects will be more similar to each other than those related at a further distance apart can cause a model to be overfit (O'Sullivan and Unwin 2003). In species distribution studies, spatial autocorrelation can result from correlated environmental variables, local and region sub-species populations that have variable ecologies, absolute density-independent dispersal, and potential population growth rates (Bahn et al. 2008). Moran's I and Geary's C are two methods available to measure spatial autocorrelation (O'Sullivan and Unwin 2003).

Chapter 3:

Data and Methods

3.1 Biological Data

The biological input data used in this research was a dataset of occurrences of *Hydromantes shastae* from the CNDDDB (Table 1). CNDDDB data have been used as a primary source for biological input data by other researchers modeling species distributions (Hernandez et al. 2006; Meentemeyer et al. 2004; Thorne et al. 2006; Williams et al. 2009). The data were acquired via the California Department of Fish and Game Biogeographic Data Branch online download portal in July 2010 (California Department of Fish and Game 2010). The original format of the downloaded data was an Environmental Systems Research Institute (Esri) polygon shapefile. The polygon dataset was chosen over the also available point dataset because, according to the California Department of Fish and Game Biogeographic Data Branch, the polygon dataset contains the accuracy necessary for spatial analysis, while the point dataset is provided for visualization purposes only (California Department of Fish and Game 2007). To prepare the data for the model, all of the 62 *H. shastae* occurrence records were extracted from the CNDDDB population occurrence polygon layer and clipped with the study area boundary using Esri ArcMap. Esri ArcToolbox was used to generate centroids from the polygons to produce an approximate point shapefile representing known *H. shastae* population occurrences. ArcMap was used to add and calculate longitude and latitude values in decimal degrees for each of the 62 occurrence records. To meet the input requirements of Maxent, the database table associated with the species occurrence point shapefile was converted to a comma-separated values (.csv) file using Microsoft Excel. Only the species name, longitude, and latitude columns were retained. This .csv file was the only biological data file used in Maxent. No additional occurrence data was used for model training (Section 3.3).

Table 1: Types and sources of biological and environmental data used in Maxent modeling of *Hydromantes shastae* distribution in Shasta County, CA

Type of Data	Source
Species Distribution	
Current occurrence data for state and federally listed, rare and sensitive species in the United States	California Natural Diversity Database: http://www.dfg.ca.gov/biogeodata/cnddb/ (subscription required)
Topography	
Elevation at 1/3 arc-second (approximately 10 m) resolution	National Elevation Dataset (NED): http://seamless.usgs.gov
Geology	
California state geologic units characterized by rock type	United States Geological Survey: http://tin.er.usgs.gov/geology/state/state.php?state=CA
Land Cover	
Dataset describing vegetation and land use in the United States	United States Geological Survey Biological Informatics Office National Biological Information Infrastructure (NBII) Gap Analysis Program (GAP) : http://lc.gapanalysisprogram.com/landcoverviewer/Map.aspx
Precipitation	
Annual average precipitation	United States Department of Agriculture (USDA) Natural Resources Conservation Service: http://datagateway.nrcs.usda.gov/GDGOrder.aspx
Soils	
United States soil types by region	USDA Natural Resources Conservation Service Soil Data Mart Soil Survey Geographic (SSURGO) Database: http://soildatamart.nrcs.usda.gov/

3.2 Environmental Variables

Environmental predictor variables were chosen based on the scientific literature describing *Hydromantes shastae* dependencies on the environment (Section 2). These variables were elevation, geology, land cover, precipitation, and soils (Table 1). Each predictor variable dataset required manipulation to prepare it for Maxent use. All of the variables were converted to Esri grid format and assigned values for 10 by 10 m grid cells using Esri's ArcGIS Desktop Spatial

Analyst Extension. Because attribute tables are not retained in grids, ArcMap was used prior to the grid conversions to add “value” fields containing the attribute data that would be used in Maxent to evaluate model fit and later to be used for joining grid layers to attribute data. ArcToolbox was used to convert the grids to the American Standard Code for Information Interchange (ASCII) format. ArcToolbox tool was then used to assign each variable the standard spatial reference chosen for this research: Universal Transverse Mercator, North American Datum 1983, Zone 10N, meters.

Elevation data used was acquired from the 1/3 arc-second National Elevation Dataset (NED), a product of the United States Geological Survey (USGS) (2010a). The 1/3 arc-second resolution approximates to the 10 m by 10 m cell size to which all the predictor variables were converted for Maxent input. The USGS (2010a) has reported that the NED dataset is the best coterminous United States elevation data available, with updates occurring approximately every two months. The data acquired for this research was obtained via download in July 2010. The data set was clipped to the study area boundary. Within the study area, the dataset contained elevation values ranging from 153 to 1711 m.

The geologic environmental predictor data used contained primary and secondary predominant rock type characterizations for the state of California. The data were acquired from the USGS online and was downloaded as a shapefile in May 2010 (United States Geological Survey 2005). The shapefile was clipped to the study area boundary. The field that contained the predominant rock types characterizing each area was generalized from 15 to three classifications: volcanic, limestone, and neither (Table 2). These classifications were based on the primary rock types described in the literature (Section 2) as predominant in *H. shastae* habitat. The generalization was performed because of Phillips (2005) direction that while Maxent can accept categorical data as input, the categories should be few in number.

Table 2: Descriptions of geology classes

Class code	Description
1	Volcanic
2	Limestone
3	Neither

Land cover data was acquired from the USGS Biological Informatics Office, National Biological Information Infrastructure (NBII) Gap Analysis Program (GAP) (United States Geological Survey 2010b). The downloaded shapefile contained three levels of detail. Level 1 contained nine classes generalized to the level of vegetative structure, as well as three undesignated records. Level 1 data was cleaned of unclassified records and reclassified with Spatial Analyst to the eight classes existing in the study area (Table 3). Level 2 contained 43 classes and incorporated elevation and climate data. Since elevation and climate data were included in the *H. shastae* model by other datasets, Level 2 was not utilized to avoid overfitting the model, which can degrade predictive performance (Hastie et al. 2001; Phillips and Dudik 2008). Level 3 contained 590 classifications, which are too many for this model, based on Phillips (2005).

Table 3: Descriptions of land cover classes

Class code	Description
1	Human land use
2	Aquatic
3	Sparse and barren systems
4	Forest and woodland systems
5	Shrubland, steppe, and savanna systems
6	Grassland systems
7	Recently disturbed or modified
8	Riparian and wetland systems

The precipitation data used consisted of annual average precipitation in inches as calculated from 1971 through 2000. This variable was obtained from the USDA NRCS Geo Spatial Data Gateway (Oregon State University and the Oregon Climate Service at Oregon State University 2010). The dataset, once clipped to the study area boundary, contained precipitation values ranging from 42 to 94 inches, or about 125 to 250 cm.

Spatial and tabular soil type data were acquired from the United States Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS) Soil Data Mart (United States Department of Agriculture 2008a-c). To cover the study area, three shapefiles were combined using ArcMap and clipped to the boundary. Editing was then performed to merge remnant coincident boundaries between areas from different source shapefiles with the same soil types. The field describing soil type contained values from two different map unit schemas; these were edited and made uniform by using map unit/soil type definitions present in the tabular data files that accompanied the spatial data.

3.3 Species Distribution Model

Figure 3 provides a conceptual summary of the modeling method. The Maxent model allowed for species distribution predictions to be made with one presence-only biological dataset. In addition to not requiring a species absence dataset, Maxent did not require two separate datasets for model training and testing purposes. Maxent separated a single species occurrence file into two subsets, one for training and one for testing. The sample size used included all of the 62 CNDDDB *H. shastae* population occurrence records. The user designated Maxent to divide the 62 biological presence records into 47 (75% of the total of 62) for training and 15 (25% of the total of 62) for testing.

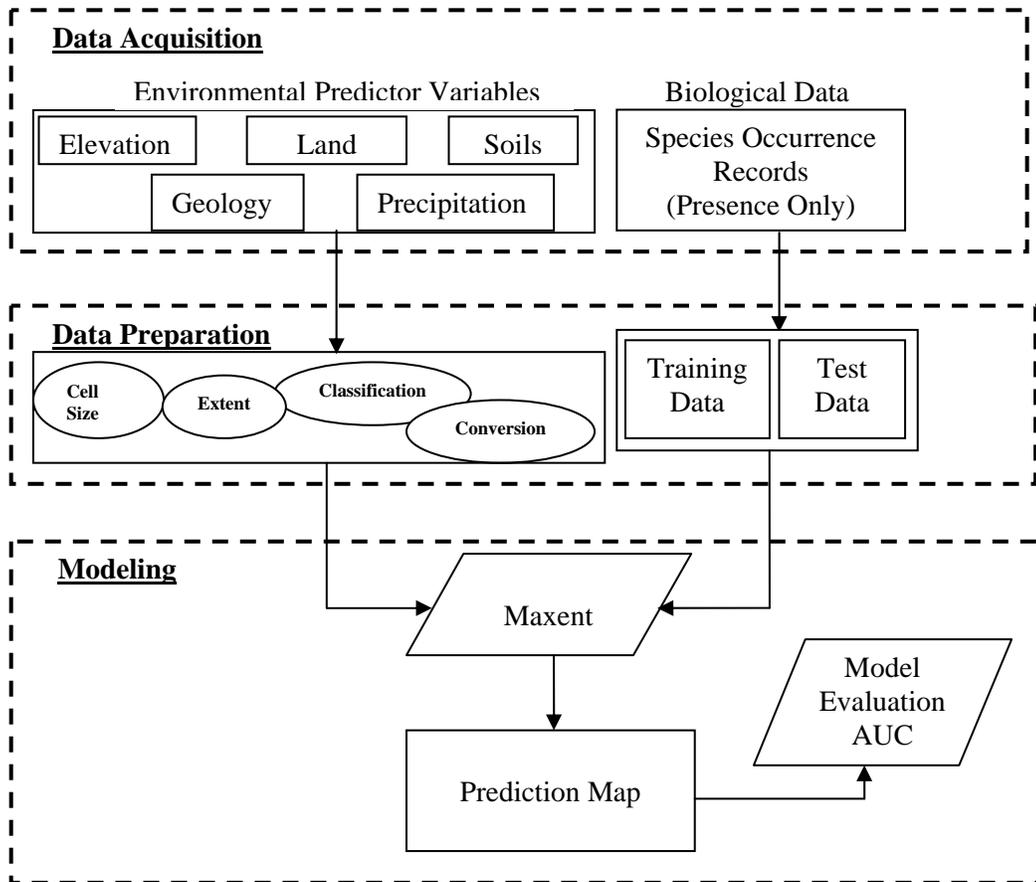


Figure 3: Conceptual summary of modeling method

The five environmental predictor variables (previously prepared for model input with conforming cell size, geographic extent, projection, and ASCII file type) were added to Maxent in a single file. Maxent can accept both continuous and categorical environmental variables (Phillips and Dudik 2008). Within the model settings interface, elevation and precipitation were identified as continuous and geology, land cover, and soils as categorical.

Maxent matched the species occurrence data to the environmental variables using regularization parameters and applied constraints (Table 4); these kept the data from matching too closely, called overfitting, which could degrade predictive performance (Hastie et al. 2001; Phillips and Dudik 2008). Maxent fit the regularization coefficient empirically based on the sample size, a method that was deemed quite effective by Phillips et al. (2006). According to

Maxent’s automatic regularization guidelines, the sample size of 62 will be calculated using a coefficient interpolated between 0.25 and 0.05, with 0.25 being the coefficient for 30 samples and 0.05 the coefficient for 100+ samples (Anderson et al. 2003). The regularization default settings were examined and accepted by the user, in part due to Phillips and Dudik (2008) which concluded Maxent defaults to be appropriate and produce results almost equivalent to models that incorporate true absence data. For each grid cell in the study area, the model used the regularization parameters to calculate values for the five predictor variables. It then used the training data to identify and rank the pixels, with higher values assigned to pixels most similar to those coincident with the training data. In order to produce binary results that could be interpreted as “suitable” and “unsuitable” habitat, a minimum training presence threshold was set, the same threshold rule used, with effective results, by Phillips et al. (2006).

Table 4: Regularization parameters and applied constraints

Number of Training Samples	47
Number of Test Samples	15
Random test percentage	25
Regularization multiplier	1
Max number of background points	10000
Replicates	1
Maximum iterations	500
Output format	Logistic
Convergence threshold	0.00001
Apply threshold rule	Minimum training presence

3.4 Model Performance Measures

The predictive performance of Maxent for modeling *H. shastae* occurrences was measured using AUC. The AUC method has been shown to be an accurate measure of the Maxent model’s overall performance (Elith and Leathwick 2009; Hernandez et al. 2006; Johnson and Gillingham 2005; Pearson 2007; Phillips and Dudik, 2008; Zuo et al. 2008). With a presence only model, AUC represents site ranking quality, the probability that a presence prediction is ranked above a

non-presence site. Rankings above 0.5 are considered better than random. Elith (2002) considered rankings of 0.75 as achieving potential usefulness. Swets (1988) AUC classification was used to assess the results of this study (Table 5). In order to assess the significance of each environmental predictor variable to the model, Maxent was set to perform a jackknife test. When run, the model sequentially eliminated environmental variables in order to calculate the gain contribution of each. It also ran each variable individually, thus identifying the one which contributed the highest gain.

Table 5: AUC classification

0.09 to 1.00	excellent
0.80 to 0.90	good
0.70 to 0.80	fair
0.60 to 0.70	poor
<0.6	fail

Chapter 4:

Results

4.1 Species Distribution Map

The Maxent model produced a species distribution map, AUC values representing model performance and response curves and a percentage table showing how each environmental predictor variable contributed to the Maxent prediction. The *H. shastae* species distribution map produced by the model illustrates where the species is predicted to occur (Figure 4).

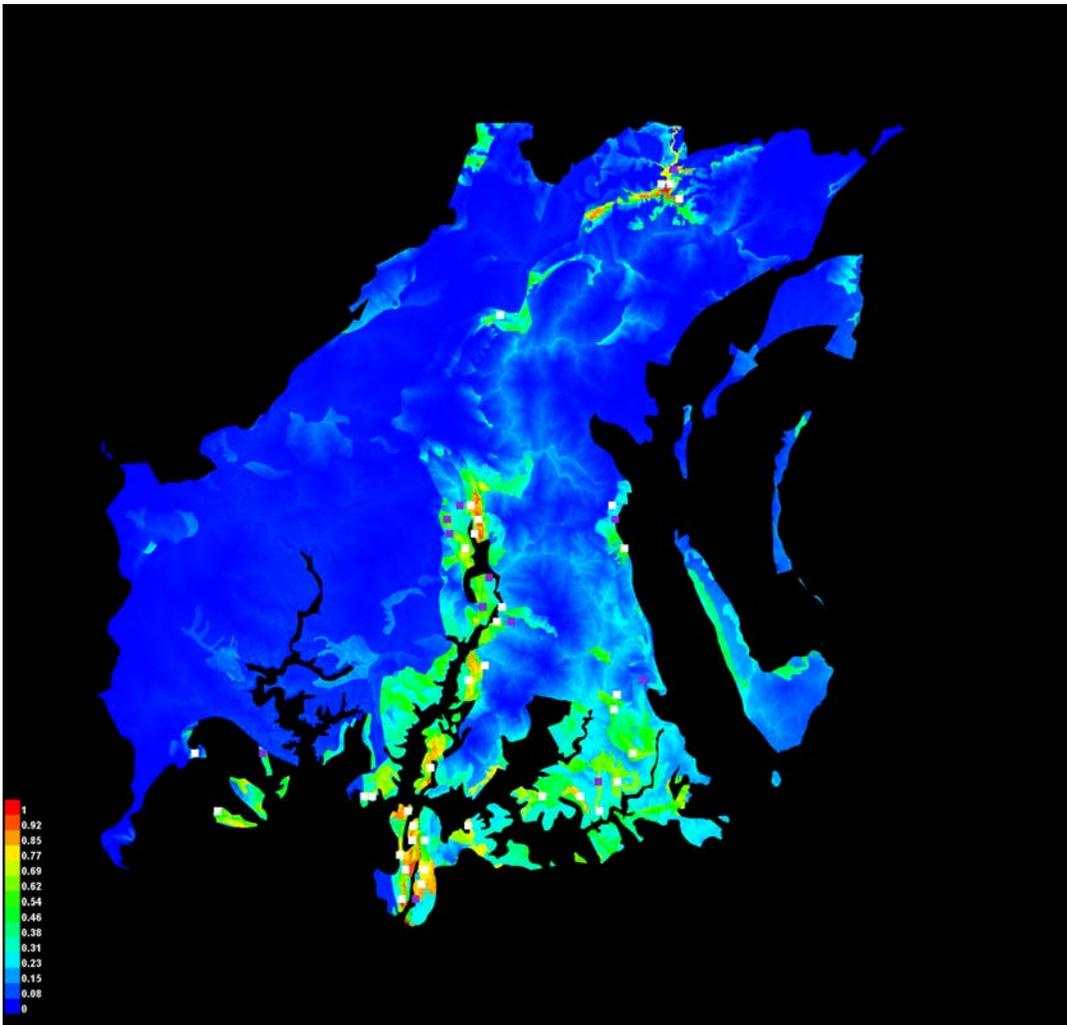


Figure 4: Maxent representation of predicted *Hydromantes shastae* distribution. Red and orange colors represent more accurate predictions and blues represent lower probability of occupancy. Dots represent training (white) and testing (purple) data

Figure 5 shows the same data with the inclusion of the study area boundary and a different color palette. Figure 6 shows an illustration of the model's prediction of suitable and unsuitable habitat. All of the test points occupy areas that the model predicted as suitable habitat.

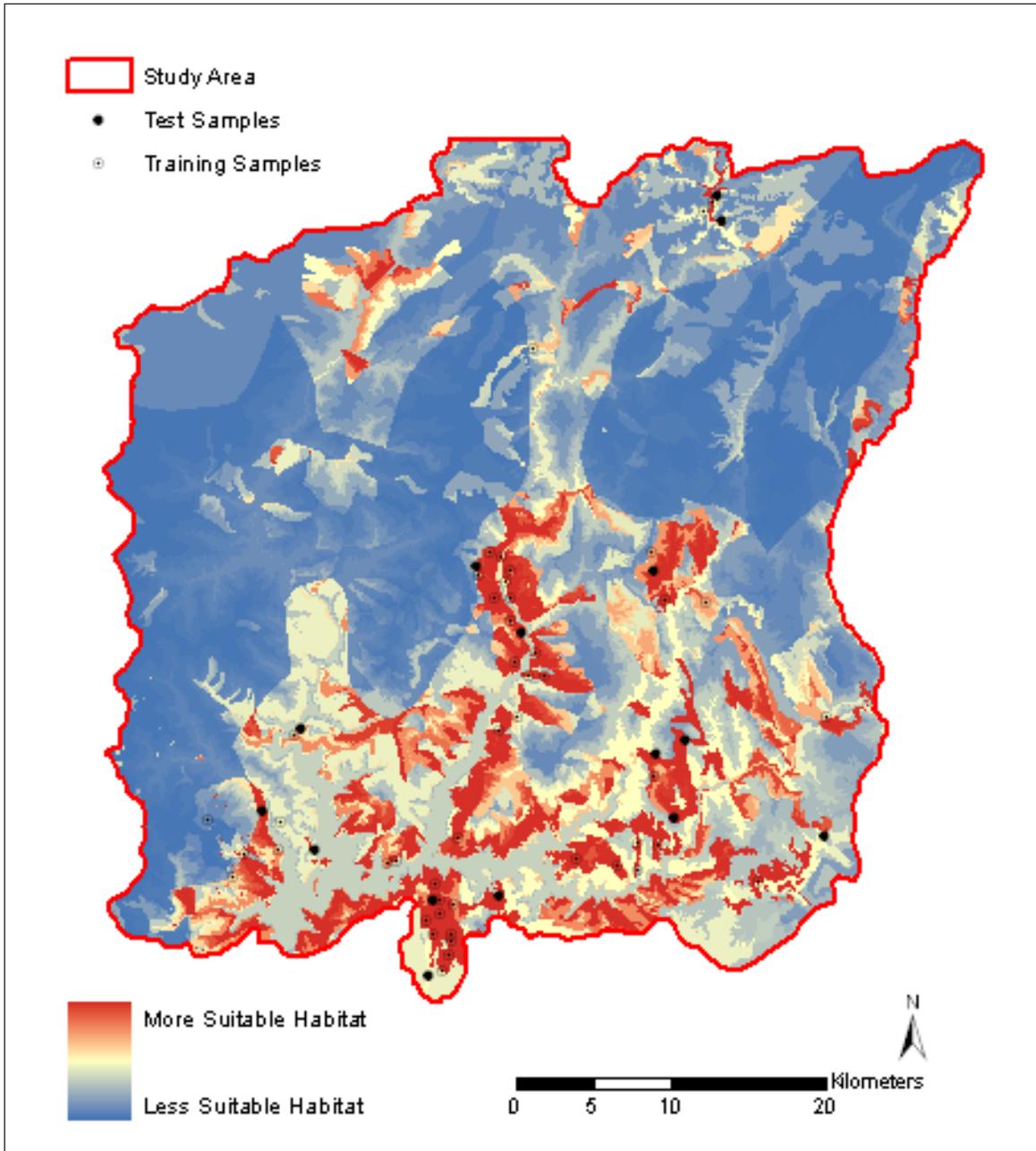


Figure 5: Maxent representation of predicted *Hydromantes shastae* distribution with study area boundary and an alternative palette for the habitat suitability values

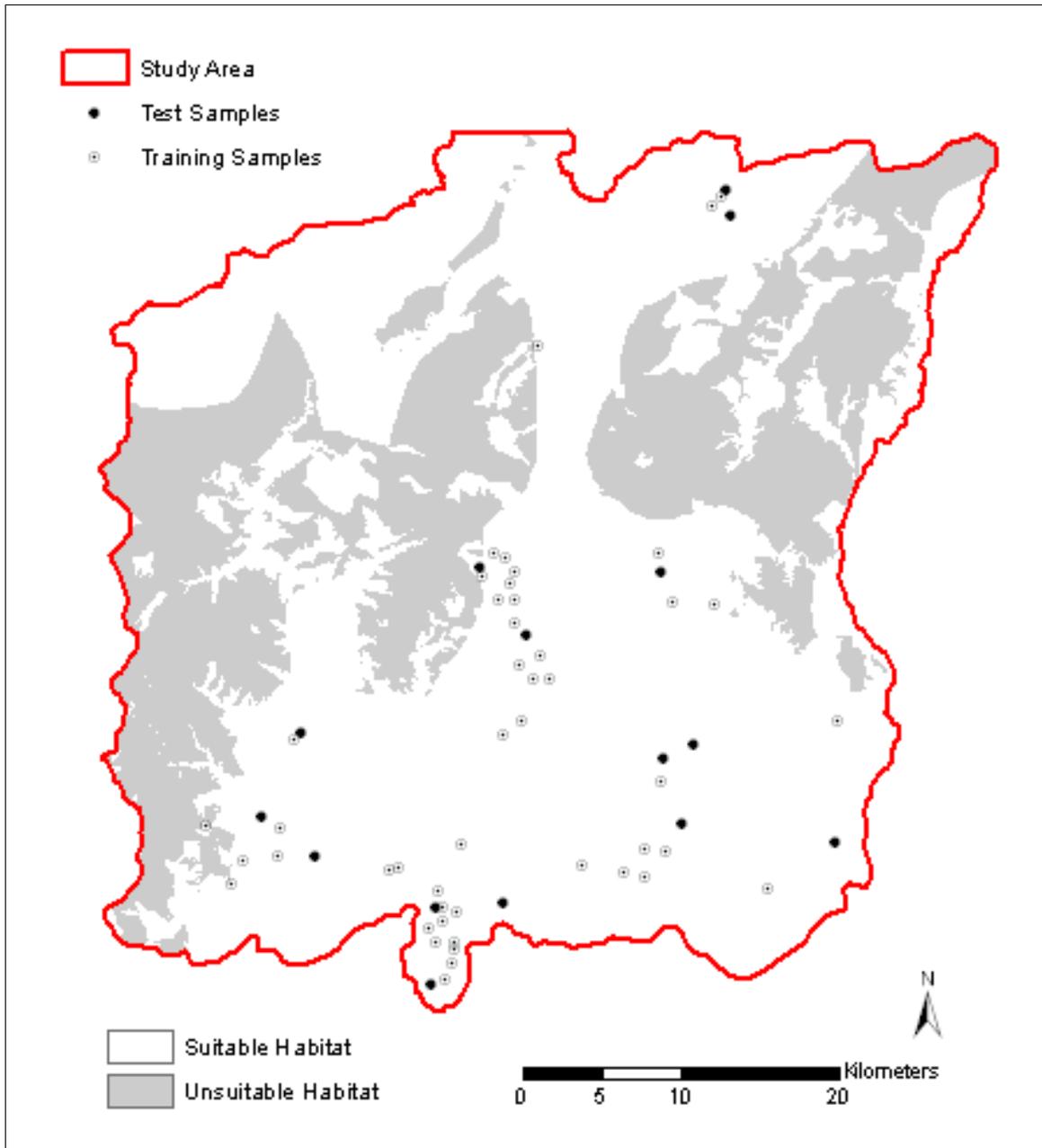


Figure 6: Binary model results from threshold, representing suitable and unsuitable habitat. All of the test samples fall within suitable habitat areas

4.2. Model Performance

Figure 7 shows the plot of sensitivity against specificity at different threshold values. This method identifies the optimum threshold where the sum of specificity and sensitivity are maximized. Liu et al. (2005) regarded AUC as one of the best methods for threshold

determination. The AUC for the test data was 0.879. According to Swets (1988) AUC classification, this ranks the model as “good” (Table 5).

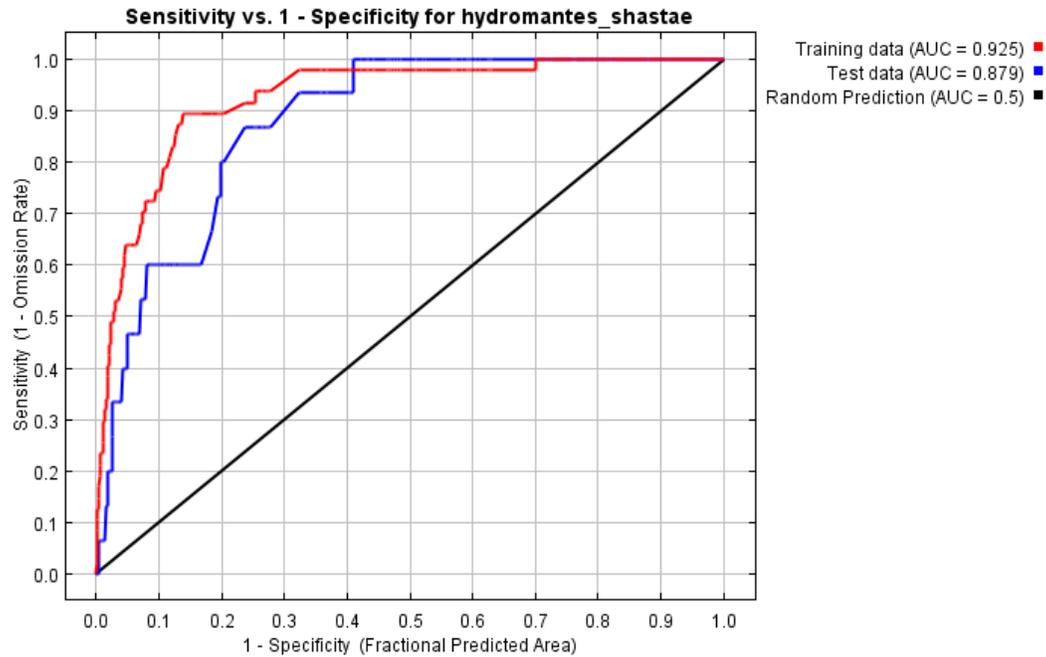


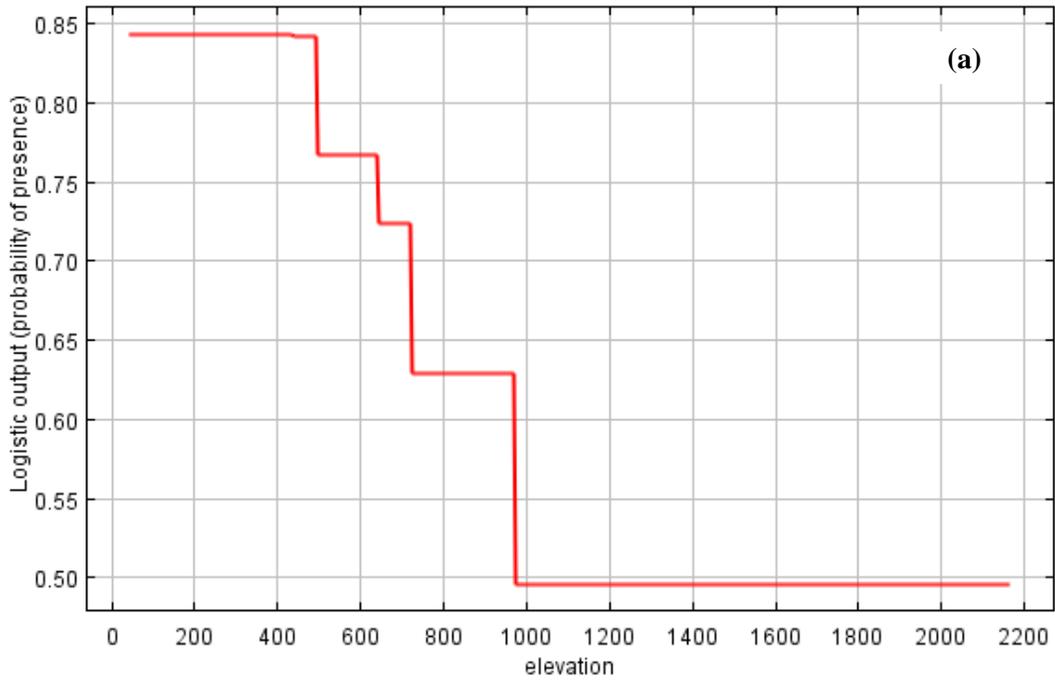
Figure 7: AUC curve for the model indicating model performance

4.3 Response Curves

Maxent’s response curves showed how the model’s predictions changed as the environmental variables varied (Figure 8 a-e). The probability of species presence declined as elevation values increased, with predictions of presence no better than random over about 304 m (100 ft).

Geology types showed a relationship with species presence for volcanic and limestone types (Table 2). Aquatic and grassland land cover types (Table 3) showed a stronger than random correlation with species presence. The precipitation response curve showed a better than random relationship with species presence for annual average precipitation values less than about 182 cm (72 in). Some soil types were strongly correlated with species presence with a majority well below random.

Response of *Hydromantes shastae* to elevation



Response of *Hydromantes shastae* to geology

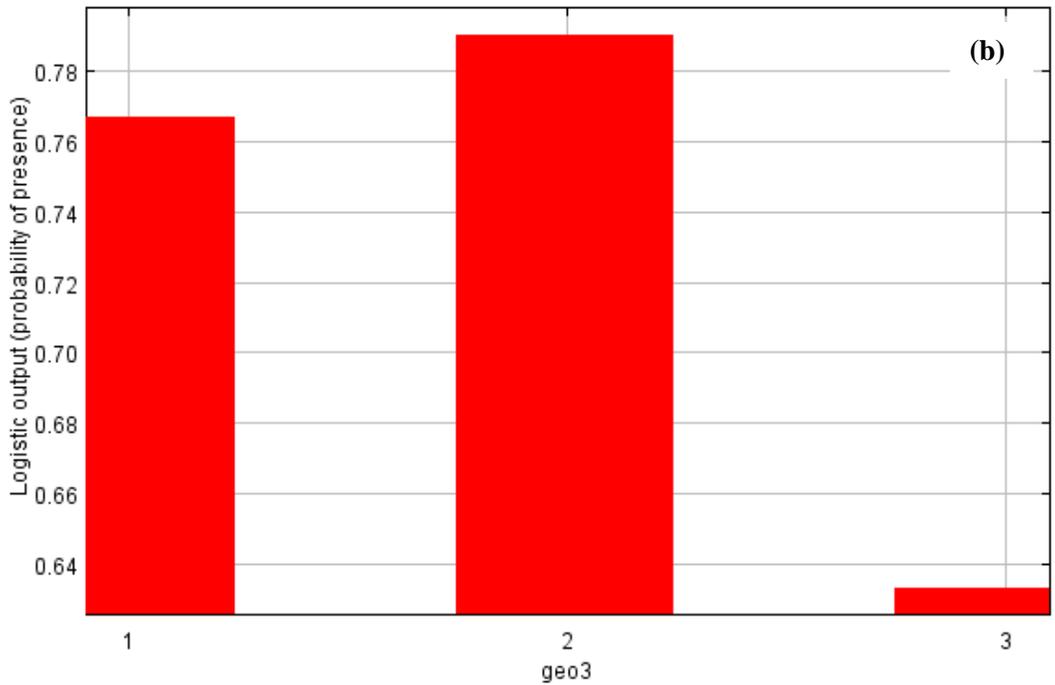
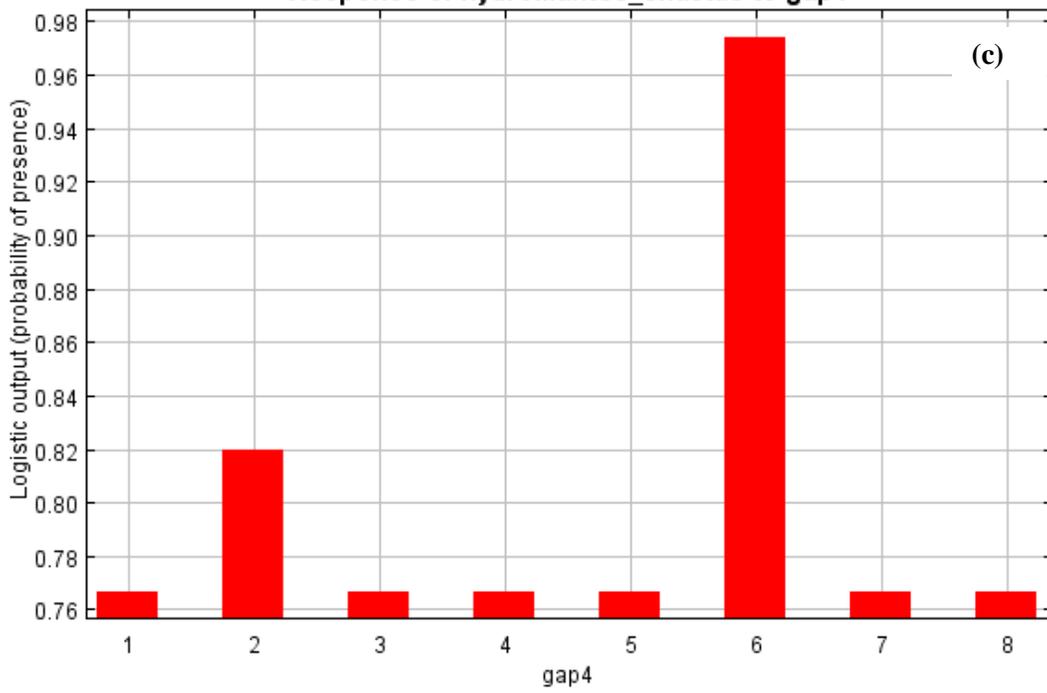


Figure 8: Maxent response curves for the five predictor variables: (a) elevation, (b) geology, (c) land cover, (d) precipitation, and (e) soils

Figure 8: Continued

Response of *Hydromantes shastae* to land cover



Response of *Hydromantes shastae* to precipitation

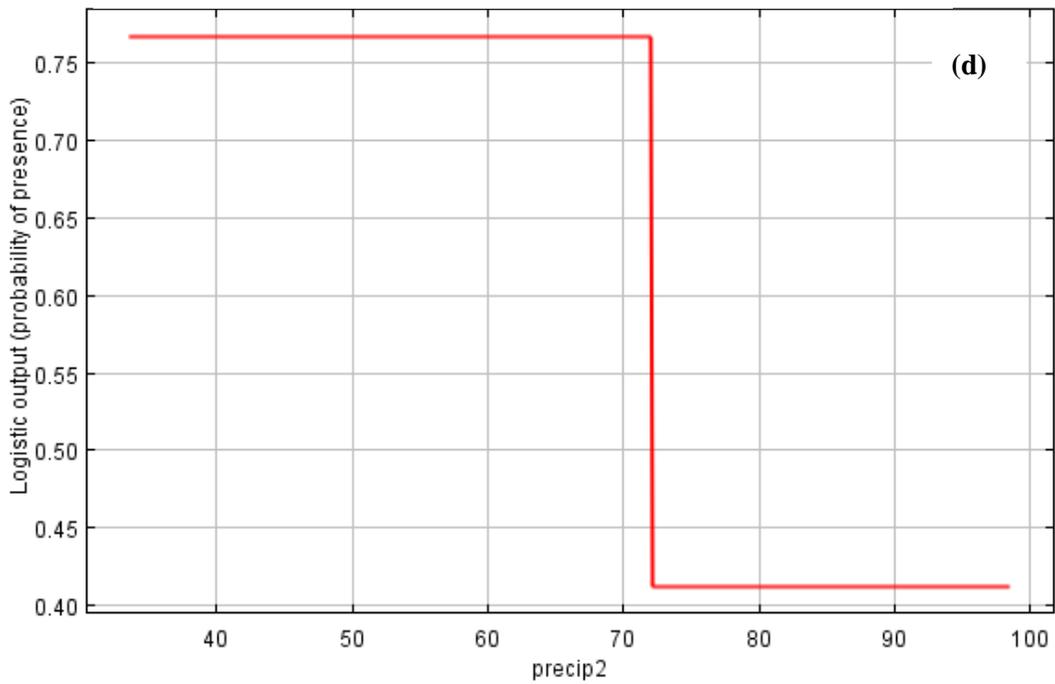
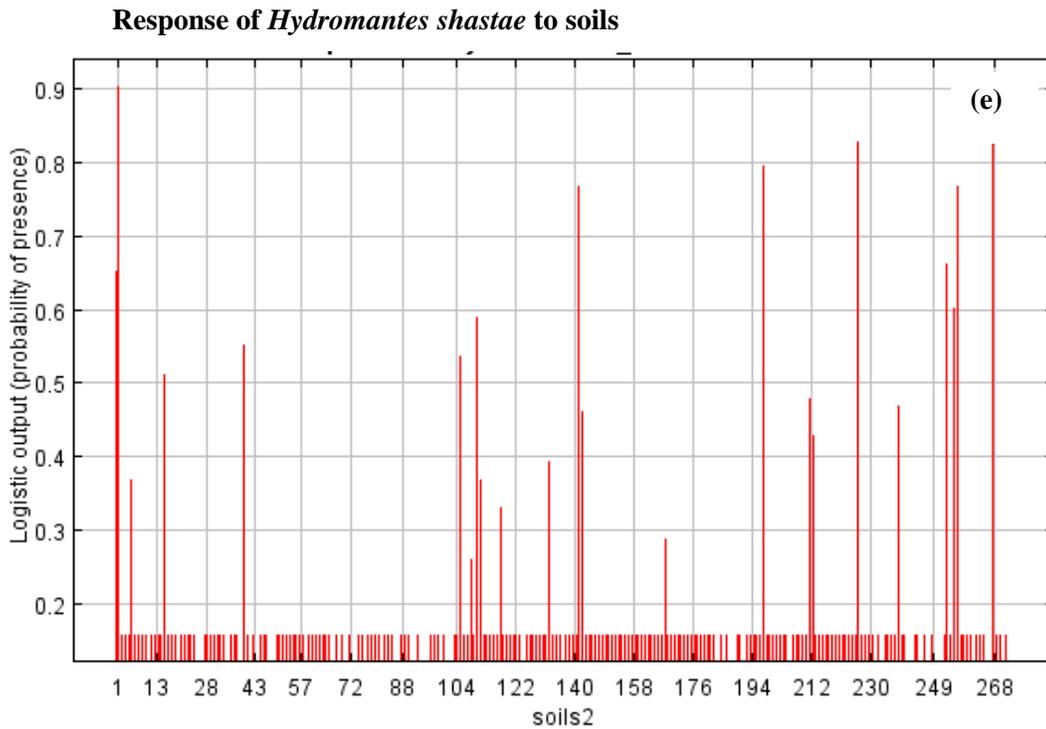


Figure 8: Continued



4.4 Predictor Variable Importance

The jackknife test showed that soil type was the most important variable in determining model prediction, in the training, test, and AUC evaluations (Figure 9 a-c). Soil type increased the gain more than any other variable when added to the jackknife test and decreased the gain the most when omitted. An estimate of the contribution of each predictor variable to the model showed that after soils, at 55.4%, elevation (19.3%) and precipitation (19%) contributed most to the model (Table 6). Geology (3.5%) and land cover (2.8%) contributed relatively little.

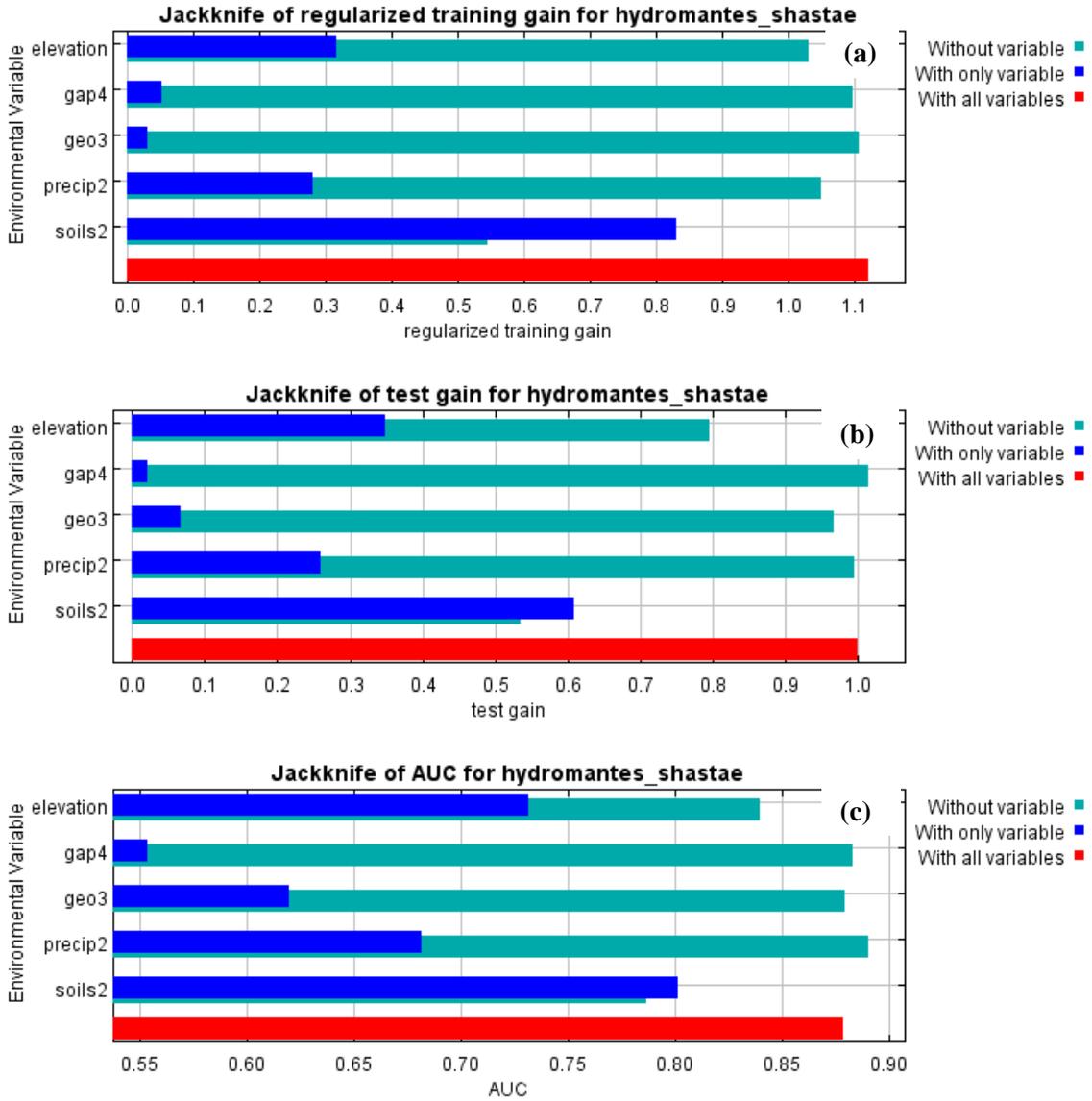


Figure 9: Jackknife results for (a) training gain, (b) test gain, and (c) AUC

Table 6: Variable contribution

Variable	Contribution (%)
soils	55.4
elevation	19.3
precipitation	19
geology	3.5
land cover	2.8

Chapter 5:

Discussion and Conclusions

5.1 Model Strengths

Stockwell and Peterson (2002) identified that machine learning methods, like Maxent, have been shown to perform at 90% of maximum accuracy rate with only 50 sample points, and more accurately than other modeling methods with as few as 5 to 25 sample points (Hernandez et al. 2006). For this research, the model was set to divide the total available occurrence points into 75% for training the model and 25% testing points. Thus, the Maxent model might be expected to perform with at least an accuracy rate of 90% of the maximum with 15 sample points. In fact, the habitat suitability map placed all test samples in areas deemed “suitable” by the model. The model produced this result consistently for each of five iterations. This suggests that the model’s evaluation of “suitable” and “unsuitable” habitat, created using the training samples, was accurate, at least for the 15 test samples employed. The AUC value of 0.879 suggests that the model fit is good, far closer to a perfect fit than a random one (Swets 1988). The soils predictor variable, identified as contributing most to the model, contained a large proportion of limestone soils. This agrees with the literature describing the predominance of limestone soils in *Hydromantes shastae* habitat (Section 2).

5.2 Model Challenges

Although the geology variable classifications of limestone and volcanic rock types showed strong correlations with species presence in the response curve, the variable’s percentage of contribution to the model was evaluated at only 3.5%. This might be because while the occurrence data used as model input were collected during the seasons of fall, winter, and spring, when the species is active on the surface, the description of the species geologic range in the literature may include additionally its inactive occurrence locations, which are subterranean fissures and caves in

limestone and volcanic geologic regions. Also, while the elevation variable ranked at 19.3% importance to the model, its response curve showed species presence no better than random over 304 m (100 ft), which is in stark contrast to the literature reports of species occurrence ranging from 330m (1100 ft) to 773m (2550 ft) (Section 2). In addition, while land cover was found to be the least important variable to the model, the Maxent response curve showed a high potential for species presence in the grassland systems land cover type, a correlation not supported by literature accounts of the species' habitat. In terms of precipitation, Maxent's evaluation of annual average precipitation importance to the model (19%) may reflect the amphibious species moisture requirements as described in the literature (Section 2).

There are certainly sources of error and bias contained in all species distribution models (Section 2). In this model, there are specific factors that must be considered when interpreting *H. shastae* habitat suitability predictions. Some come from the modeling method itself. For example, the arbitrary data partitioning method used may have influenced model results (Fielding and Bell 1997). Splitting one set of occurrence data into one training set and one testing set gave different results than would have collecting two sets of independent data (Chatfield 1995). Test and training data contained the same spatial autocorrelation, contributing the same error in the results of both. Also, the arbitrarily held out testing sample may not have produced the 'best' model (Fielding and Bell 1997). The selection of predictor variables may not have produced the 'best' model and without real absence data, Maxent estimated absences utilizing the background matrix, which might have skewed the results, particularly with a small sample size of 15, if a pseudo-absence was assigned in error, either to a point where there was an occurrence, or where there was suitable habitat.

In addition to the modeling method, factors of species ecology and the data used must be considered during interpretation and exploration into possible explanations for this model's non-conformity to the literature. The CNDDDB dataset used may not have well informed the model of

actual species occurrence. The dataset was influenced by the species' patchy distribution, its rarity throughout the landscape, its seasonal surface occurrence, and its location (when active) on steep and rocky (often impassable) terrain. Literature explains that this species has a dynamic range in response to climatic and other changes (Parmesan 2006). Because CNDDDB data is based on biological surveys conducted primarily in areas identified based on historic species assumptions and CNDDDB data sources are from project-driven discoveries that occur in focused geographic areas, areas not surveyed may hold significant predictive information that was left out of this model. Key environmental parameters may be missing and those used may be misclassified. In addition, the model may have erroneously calculated an area as suitable that really is unsuitable because of geographic barriers that limit dispersal, or competition from other species (Pearson 2007).

5.3 Model Viability for *Hydromantes shastae*

Chatfield (1995) suggested that models can be accurate and useful regardless of their statistical validity. The value of this model should, at least in part, be determined by how the model will be applied. Minimally, the model identified areas for more focused study and survey. This Maxent model produced an accurate occurrence prediction. With additional model validation, like field validation, the model could support the delineation of suitable habitat. The model indicated the soils variable as the most important variable for making *H. shastae* habitat suitability predictions using Maxent. This agrees with expert understanding of the species that describes limestone soils and rocks (geology) as ecologically important for the species. While the model demonstrated its suitability to predict *Hydromantes shastae* occurrences utilizing presence-only data, its evaluation of environmental predictor variable importance is less satisfactory. Species impact studies and conservation efforts could use this model to inform their research and as a basis for further investigation.

5.4 Future Uses of Research

To extend the usefulness of this research, multiple model runs could be done to produce variations in predictions (see Amaral 2007), allowing for more conclusive results. In addition, Maxent and/or other modeling methods could be used to apply this model to making new population discoveries in new areas and under altered environmental conditions. Alternatively, other species distribution modeling methods could be used with these same datasets to generate superior results (Pearson et al. 2006; Phillips and Dudik 2008; Randin et al. 2006) or to extend the accuracy and value of the research reported herein. The results of this research project do confirm the viability of modeling rare and endangered species and the potential for using these kinds of analyses to support biological conservation and restoration.

References

- Amaral S, Costa C B, Renno, C D 2007. Normalized Difference Vegetation Index (NDVI) improving species distribution models: an example with the neotropical genus *Coccoloba* (Rubiaceae). *Simpósio Brasileiro de Sensoriamento Remoto*, 13.
- Anderson R P, Lew D, and Peterson A T 2003 Evaluating predictive models of species' distributions: Criteria for selecting optimal models. *Ecological Modelling* 162: 211-232.
- Austin M P 2002 Spatial prediction of species distribution: An interface between ecological theory and statistical modeling. *Ecological Modelling* 157: 101-118.
- Austin M P 2007 Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling* 200: 1-19.
- Bahn V, Krohn W B, and O'Connor R J 2008 Dispersal leads to spatial autocorrelation in species distributions: A simulation model. *Ecological Modelling* 213: 285-292.
- Beaumont L J and Hughes L 2002 Potential changes in the distributions of latitudinally restricted Australian butterfly species in response to climate change. *Global Change Biology* 8: 954-971.
- Bronson L, Thuiller W, Araujo M B, and Hirzel A H 2004 Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27: 437-448.
- California Department of Fish and Game 2007 *Proper use of the CNDDDB Point GIS Layer*. Sacramento, Biogeographic Data Branch. Retrieved from http://www.dfg.ca.gov/biogeodata/cnddb/pdfs/CNDDDB_AboutGISPointData.pdf
- California Department of Fish and Game 2009 *Special Animals*. Sacramento, Biogeographic Data Branch.
- California Department of Fish and Game 2010 *California Natural Diversity Database* [Data file]. Retrieved from <http://www.dfg.ca.gov/biogeodata/cnddb/>
- Chatfield C 1995 Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society, Series A* 158: 419-466.
- Elith, J 2002 Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In Ferson S and Burgman M (eds) *Quantitative Methods for Conservation Biology*. Berlin: Springer: 39-58.
- Elith J and Leathwick J R 2009 Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40: 677-697.
- Engler R, Guisan A, and Rechsteiner L 2004 An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology* 41: 263-274.

- Fielding A and Bell F 1997 A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24(1): 38-49.
- Fitzpatrick M C, Weltzin J F, Sanders N J, Dunn, R R 2007 The biogeography of prediction error: Why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography* 16: 24–33.
- Gorman J 1956 Reproduction in plethodontid salamanders of the genus *Hydromantes*. *Herpetologica* 12: 249-259.
- Gorman J and Camp C 1953 A new cave species of salamander of the genus *Hydromantes* from California, with notes on habits and habitats. *Copeia*: 39-43.
- Guisan A and Zimmermann N E 2000 Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.
- Guisan A, Edwards Jr T C, and Hastie T 2002 Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling* 157: 89-100.
- Hansen R and Papenfuss T 1994 Shasta salamander: Life on the edge. In Thelander (ed) *A Guide to California's Endangered Natural Resources Volume I: Wildlife*. Santa Cruz: Biosystems Books: 256-257.
- Hastie T, Tibshirani R, and Friedman J 2001 *The Elements of Statistical Learning*. Berlin: Springer .
- Helmuth B, Kingsolver J G, and Carrington E 2005 Biophysics, physiological ecology, and climate change: Does mechanism matter? *Annual Review of Physiology* 67: 177-201.
- Hernandez P, Graham C, Master L, and Albert D 2006 The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29: 773-785.
- Hirzel A H, Hausser J, Chessel D, and Perrin N 2002 Ecological-niche factor analysis: How to compute habitat-suitability map without absence data. *Ecology* 83: 2027-2036.
- Jaynes E T 1990 Probability theory as logic. Presented at the Ninth Annual Workshop on Maximum Entropy and Bayesian Methods, Dartmouth College, New Hampshire, August 1989. In the Proceedings Volume Fougère P F (ed) *Maximum-Entropy and Bayesian Method*. Dordrecht: Kluwer. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.8618>
- Jepson W and Hickman J 1993 *The Jepson Manual: Higher Plants of California*. Berkeley: University of California Press.
- Johnson C J and Gillingham M P 2005 An evaluation of mapped species distribution models used for conservation planning. *Environmental Conservation* 32(2): 177-128.

- Johnson C J and Gillingham M P 2008 Sensitivity of species-distribution models to error, bias, and model design: An application to resource selection functions for woodland caribou. *Ecological Modelling* 213: 143-151.
- Keen W H 1982 Habitat selection and interspecific competition in two species of plethodontid salamanders. *Ecology* 63: 94-102.
- Lewendal P 1995 Habitat-use by the Shasta salamander (*Hydromantes shastae*). California Department of Fish and Game. Retrieved from http://www.fs.fed.us/psw/programs/wildlife/emphasis_area_projects/shasta_sal.shtml
- Lindstrand L. 2000 Discovery of Shasta salamanders in atypical habitat. *California Fish and Game* 86: 259-261.
- Lippit C D, Rogan J, Toledano J, Sangermano F, Eastman J R, Mastro V, and Sayer A 2008 Incorporating anthropogenic variables into a species distribution model to map gypsy moth risk. *Ecological Modelling* 210: 339-350.
- Liu C R, Berry, P M, Dawson, T P, and Pearson, R G 2005 Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28: 385-393.
- Meentemeyer R, Rizzo D, Mark W, and Lotz E 2004 Mapping the risk of establishment and spread of sudden oak death in California. *Forest Ecology* 200(1-3): 195-214.
- O'Sullivan D and Unwin D J 2003 *Geographic Information Analysis*. Hoboken: John Wiley & Sons.
- Olson D and Lewendal P 1999 Survey Protocol for the Shasta Salamander (*Hydromantes shastae*). Version 3.0. USDA Forest Service, Pacific Northwest Research Station. Retrieved from <http://www.fs.fed.us/pnw/ar2000.pdf>
- Oregon State University and the Oregon Climate Service at Oregon State University 2000 *Average Annual Precipitation, 1971-2000* [Data file]. Retrieved from <http://datagateway.nrcs.usda.gov/GDGOrder.aspx>
- Papenfuss T and Brouha P 1979 The status of the Shasta salamander (*Hydromantes shastae*): Shasta-Trinity National Forest comprehensive species management plan and a species status report. Berkeley: United States Department of Agriculture Forest Service.
- Papenfuss T and Cross D 1980 The status of the Shasta salamander *Hydromantes shastae* in the Mountain Gate and Cedar Creek areas of Shasta County. Berkeley: United States Department of Agriculture Forest Service.
- Parmesan, C 2006 Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology Evolution and Systematics* 37: 637-669.
- Pearson R G 2007 Species' Distribution Modeling for Conservation Educators and Practitioners. Retrieved from <http://ncep.amnh.org>

Pearson R G, Dawson T P, Berry P M, and Harrison P A 2002 Species: A spatial evaluation of climate impact on the envelope of species. *Ecological Modelling* 154: 289-300.

Phillips S J 2005 *A Brief Tutorial on Maxent*. Retrieved from <http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc>

Phillips S J 2010 *Maxent Software: Release 3.3.3a*. Retrieved from <http://www.cs.princeton.edu/~schapire/maxent/>

Phillips S J and Dudik M 2008 Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography* 31: 161-175.

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

Randin C F, Dirnbock T, Dullinger S, Zimmerman N E, Zappa M, and Guisan A 2006 Are niche-based species distribution models transferable in space? *Journal of Biogeography* 33: 1689–1703

Richardson D M, Rouget M. and Rejmanek M 2004 Using natural experiments in the study of alien tree invasions: Opportunities and limitations. In Gordon M .S and Bartol S M (eds) *Experimental Approaches to Conservation Biology*. Berkeley: University of California Press: 180-201.

Robertson M P, Peter C I, Villet M H, and Ripley B S 2003 Comparing models for predicting species' potential distributions: A case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling* 164: 153-167.

Stebbins R 1985 *A Field Guide to Western Reptiles and Amphibians*. 3rd Edition. Boston: Houghton Mifflin Company.

Stockwell D R B and Peters D P 1999 The GARP modelling system: Problems and solutions to automated spatial prediction. *International Journal of Geographical Information Systems* 13, 143-158.

Stockwell D R B and Peterson A T 2002 Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148: 1-13.

Swets JA 1988 Measuring the accuracy of diagnostic systems. *Science* 240: 1285-1293.

Thorne J, Gao S, Hollander A, Kennedy J, McCoy M, Johnston R, and Quinn J 2006 Modeling potential species richness and urban buildout to identify mitigation sites along a California highway. *Transportation Research Part D: Transport and Environment* 11(4): 277-291.

United States Department of Agriculture 2008a *Intermountain Area, Parts of Lassen, Modoc, Shasta, and Siskiyou Counties, California* [Data file]. Retrieved from <http://soildatamart.nrcs.usda.gov/>

United States Department of Agriculture 2008b *Shasta County Area, California* [Data file]. Retrieved from <http://soildatamart.nrcs.usda.gov/>

United States Department of Agriculture 2008c *Shasta-Trinity National Forest Area, Parts of Humboldt, Siskiyou, Shasta, Tehama, and Trinity Counties, California* [Data file]. Retrieved from <http://soildatamart.nrcs.usda.gov/>

United States Department of the Interior, Bureau of Reclamation 1999 Appraisal Assessment of the Potential for Enlarging Shasta Dam Reservoir. Retrieved from http://www.usbr.gov/mp/slwri/docs/appraisal_assess/02_part1.pdf

United States Geological Survey 2005 *Geology of California* [Data file]. Retrieved from <http://tin.er.usgs.gov/geology/state/state.php?state=CA>

United States Geological Survey 2010a *National Elevation Dataset* [Data file]. Retrieved from <http://ned.usgs.gov>

United States Geological Survey 2010b *National Land Cover Gap Analysis Project* [Data file]. Retrieved from <http://lc.gapanalysisprogram.com/landcoverviewer/Map.aspx>

Wake D B and Papenfuss T J 2005 *Hydromantes shastae*. In Lannoo M J (ed) *Amphibian Declines: The Conservation Status of United States Species*. Berkeley: University of California Press: 784-785.

Williams J, Seo C, Thorne J, Nelson J, Erwin S, O'Brien J, and Schwartz M 2009 Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions* 15: 565-576.

Wisn M S, Hijmans R J, Li J, Peterson AT, Graham C H, Guisan A, and NCEAS Predicting Species Distributions Working Group 2008 Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14(5): 763-773.

Zuo W, Lao N, Geng Y, and Ma K 2008 GeoSVM: an efficient and effective tool to predict species' potential distributions. *Journal of Plant Ecology* 1(2): 143-144.