

A MAXENT-BASED MODEL FOR IDENTIFYING LOCAL-SCALE TREE SPECIES
RICHNESS PATCH BOUNDARIES IN THE LAKE TAHOE BASIN OF CALIFORNIA AND
NEVADA

by

James J. Pollock

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)

August, 2015

DEDICATION

I dedicate this work to my parents, Scott and Maureen. Your pride in me is my strength and motivation. This document could not exist without you.

ACKNOWLEDGMENTS

My warmest thanks to the remarkable people that guided and supported my journey to a Master's degree.

Writing a thesis is a sometimes lonely and always daunting task. My thesis advisor, Dr. Travis Longcore expertly administered the perfect mix of sage advice, encouragement and occasional kicks in the pants to keep me on point, on task and on pace. More importantly, he pushed me to delve deeply into the ecology behind my topic—the final product is many times better than I would have produced on my own.

The members of my thesis committee, Dr. Karen Kemp and Dr. Su Jin Lee provided mentorship long before my thesis project began. Their enthusiasm for all things geospatial is infectious and I could not ask for a finer pair of mentors. My thesis is quite a tome; their patient critiques are most appreciated.

Finally, to my family, Tami, Madison, Morgan and Murphy. Your dinner deliveries, patience and hugs contributed to my success more than you know.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGMENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF EQUATIONS	xi
LIST OF ABBREVIATIONS	xii
ABSTRACT	xiv
CHAPTER 1: INTRODUCTION	1
1.1 Research Questions	3
1.2 Motivation	5
1.3 Workflow	7
CHAPTER 2: BACKGROUND	9
2.1 Species Richness	9
2.2 Conceptual Model	12
2.3 Appropriateness of Maximum Entropy for Species Richness Modeling	16
2.4 Maximum Entropy Theory	19
2.5 Additional Relevant Maxent Studies	20
2.6 Species Richness Boundary Detection and Definition	22

CHAPTER 3: DATA AND METHODS	24
3.1 Data	24
3.1.1 Map Projections	24
3.1.2 Tahoe Basin Existing Vegetation Map (TBEVM) v.4.1	26
3.1.3 Tahoe Basin Digital Elevation Model (DEM)	32
3.1.4 Tahoe Basin SSURGO Soils Survey	34
3.1.5 Tahoe Area Weather Stations	37
3.1.6 TRPA Land Use and Completed USFS Prescriptions	40
3.2 Data Processing	42
3.2.1 Precipitation Surface Creation Using Empirical Bayesian Kriging	42
3.2.2 Data Consolidation	46
3.2.3 Tree Species Richness Attribute Calculation	47
3.2.4 Data Split by Management Intensity	48
3.3 Models and Analyses	48
3.3.1 Exploratory Ordinary Least Squares Regression	49
3.3.2 Species Richness-Based Maxent Model	51
3.3.3 Species Range-Based Maxent Model	56
3.3.4 Known Species Location Versus Predicted Species Range Map	57
3.3.5 Split Moving Window Dissimilarity Analysis (SMWDA)	58

CHAPTER 4: RESULTS	61
4.1 Evaluation of EBK-Derived Precipitation Surface	61
4.2 Selection of Environmental Variables	65
4.3 Maxent Models	70
4.3.1 Common Procedural Issues	71
4.3.2 Species Range-Based Model	72
4.3.3 Species Richness-Based Model	77
4.3.4 Validation of Species Richness Models by Comparison	85
4.4 Split Moving Window Dissimilarities	87
CHAPTER 5: DISCUSSION AND CONCLUSIONS	93
5.1 Research Questions	93
5.1.1 Can the Maxent maximum entropy modeling package make valid predictions of tree species richness patches?	93
5.1.2 Can the location and properties of border regions between Maxent predicted species richness ranges be derived from Maxent output and if so, are they valid?	95
5.1.3 Can the use of Maxent and a GIS produce a valid and broad scale representation of tree species richness?	97
5.2 Discussion	97
5.2.1 Applicability of Species Richness Models to Forest Management	98
5.2.2 Relationship to Recent Research	100
5.2 Suggestions for Future Work	103

5.2.1 Modeling Realistically Dense Data	103
5.2.2 Field Validation	103
5.2.3 Temporal Study	104
REFERENCES	105
APPENDIX 1: COMPLETE JACKKNIFE ANALYSIS FOR SPECIES RICHNESS	113

LIST OF TABLES

Table 1 TBEVM Species	29
Table 2 Species Richness Calculation	48
Table 3 Forest Management Categories	48
Table 4 Environmental Variables Selected from Exploratory Regression	66
Table 5 Statistical Diagnostics Following OLS	67

LIST OF FIGURES

Figure 1 Study Area-Lake Tahoe Basin California/Nevada	2
Figure 2 Conceptual Model of Environmental Factors Leading to Species Richness Structure in the Lake Tahoe Basin CA/NV	12
Figure 3 Common LTB Map Projections	25
Figure 4 Tahoe Basin Existing Vegetation Map v. 4.1 (TBEVM)—Tree Species Richness Data Overview and Exploration	27
Figure 5 Species Richness Data Exploration Using Geostatistical Analysis	31
Figure 6 LTB 5m LIDAR-Derived Digital Elevation Model (Resampled to 30m Resolution)	33
Figure 7 SSURGO Soils Data—Mean Moisture Capacity (cu cm)	35
Figure 8 LTB Weather Stations and Preliminary Precipitation Surface	39
Figure 9 TRPA Land Use and Prescribed Treatments	41
Figure 10 Pre-Kriging Data Analysis (Stationarity, Autocorrelation, Normality)	44
Figure 11 EBK Precipitation Surface (Inches of Precipitation-Annual Mean)	45
Figure 12 Data Consolidation Tool (ArcGIS Model Builder)	47
Figure 13 Maxent Pre-Processing using AcGIS and Spreadsheet Applications	52
Figure 14 Maxent Iteration	54
Figure 15 Post Processing—Species Richness Method (ArcGIS)	56
Figure 16 Post-Processing--Species Range Technique (ArcGIS)	57
Figure 17 EBK Diagnostics (Model Fit and Residuals)	62
Figure 18 Precipitation Surface Prediction Standard Error	64
Figure 19 OLS Residuals—Histograms and Scatter Plots	68

Figure 20 Geographic Distribution of OLS Residuals (Over and Under Prediction—Fits Within 2 Standard Deviations)	69
Figure 21 Maxent SDMs for Tree Species Lake Tahoe Basin, CA/NV	73
Figure 22 LTB Maxent SDMs (binary) with Species Presence Overlay	74
Figure 23 Maxent Species Richness Regions—Species Range Method, LTB, CA/NV	76
Figure 24 Maxent Tree Species Richness Distribution Models, LTB, CA/NV	78
Figure 25 Maxent Area Under Curve (AUC) Analyses for LTB Tree Species Richness Distribution Models	80
Figure 26 Jackknife Analysis for 3 Species	81
Figure 27 Tree Species Richness Distribution Model for Zero Species: Raw Maxent Output for Lake Tahoe Basin CA/NV	82
Figure 28 Maxent Derived Tree Species Richness Border Regions—Species Richness Method: Lake Tahoe Basin CA/NV	84
Figure 29 Model Comparison: Tree Species Richness Models—Species Range Method (top) vs. Species Richness Method (bottom)	86
Figure 30 SMWDA Sampling Location North Lake Tahoe, CA/NV	88
Figure 31 Split Moving Window Dissimilarity Analysis (6 Pt. Window)	90

LIST OF EQUATIONS

(1) Simpson Diversity Index	50
(2) Upper Window Average	60
(3) Lower Window Average	60
(4) Dissimilarity Index	60
(5) Mean Dissimilarity (For Monte Carlo)	60
(6) Standard Deviation of Dissimilarity (For Monte Carlo)	60
(7) Dissimilarity Z-Score	60

LIST OF ABBREVIATIONS

ASCII	American Standard Code for Information Interchange
AUC	Area Under Curve
BEF	Biodiversity Ecosystem Functioning
CSV	Comma Separated Values
DEM	Digital Elevation Model
EBK	Empirical Bayesian Kriging
GIS	Geographic Information System
LIDAR	Laser Identification Detection and Ranging
LTB	Lake Tahoe Basin
LTBMU	Lake Tahoe Basin Management Unit
MEM	Maximum Entropy Model
NDVI	Normalized Difference Vegetation Index
OLS	Ordinary Least Squares
ROC	Receiver Operating Curve
ROR	Relative Occurrence Rate
SDM	Species Distribution Model
SMSS	Shifting Mosaic-Steady State
SMWDA	Split Moving Window Dissimilarity Analysis
SSURGO	Soil Survey Geographic Database
TBEVM	Tahoe Basin Existing Vegetation Map
TRPA	Tahoe Regional Planning Agency
UC	University of California

UNR	University of Nevada, Reno
USDA	United States Department of Agriculture
USFS	United States Forest Service
UTM	Universal Transverse Mercator

ABSTRACT

The Lake Tahoe Basin, California/Nevada is the setting for evaluating a species richness modeling technique that is both accessible and provides an apparently unique approach to studying forest diversity patterns. Species richness, the total number of species of a focal group present in an ecological community without regard to individual taxa, is an important indicator of biodiversity. Despite its importance to researchers and natural resource managers, predicting species richness patterns in forested landscapes is difficult and therefore, not common. The computationally powerful yet highly accessible Maxent package, specifically designed for modeling species distributions, is used to predict homogenous patches of species richness by treating species richness values as individual “species.” Areas where ranges of homogenous species richness overlap are then isolated and displayed as “border regions” similar to ecotones. Nowhere in the ecological literature is Maxent used in this manner, nor are transitional zones between regions of species richness viewed as spatial entities. Therefore, this thesis investigates if Maxent can make valid predictions about species richness and if areas where species richness predictions overlap constitute transition zones. To validate the model, traditional species distribution models for each included tree species were created using Maxent, stacked and then summed to produce a comparable species richness surface. Similar patterns between the two models indicate that Maxent accurately predicts species richness from environmental factors. Border regions were validated as legitimate spatial entities using split moving window dissimilarity analysis—a technique used to identify ecotones. Results indicate that using Maxent for this application is very likely valid and species richness border regions represent a promising spatial entity for studying diversity patterns. This spatially explicit approach provides an

accessible method for studying species richness patterns at multiple scales. Further, a temporal series of these models provides a method for examining how diversity changes over time.

CHAPTER 1: INTRODUCTION

Species richness is defined as the number of species of a focal group present in an ecological community without regard to any specific taxa. As an indicator of diversity, it is a useful indicator of the health, structure and productivity of forest ecosystems (Adams 2009). For example, high species richness is associated with high productivity due to factors such as interspecific interaction or niche partitioning (Morin et al. 2011, Zhang et al. 2012). Conversely, extremely high species richness is not necessarily desirable. Morin et al. (2011) observed that productivity increases with richness, but productivity levels out at a consistent species richness value. The interaction of disturbance and richness-induced productivity vary the species carrying capacity of individual patches (Adams 2009). These phenomena and many other diversity factors are of great interest to scientists and forest managers.

This research provides an accessible geospatial method that predicts local-scale species richness patches using a maximum entropy analysis (via the Maxent species distribution modeling package) and then treats overlap between them as transition zones. By treating homogenous richness patches (or collection of patches) as ecological communities, properties of transition zones that separate them such as their position, shape and abruptness can be mapped and observed. The primary assumption is that in the real-world landscape, individual pixels may only harbor one species richness value. Pixels that simultaneously predict multiple species richness values combine to form an ecotone-like structure separating areas of more stably defined diversity. This provides more information than a simple mapping of predicted richness. Border regions delineate islands of diversity and depending on representation, may indicate the abruptness of the transition. The Lake Tahoe Basin (LTB) (Figure 1) provides an excellent

setting for presenting this approach due to its rich data availability and its status as a distressed ecosystem.

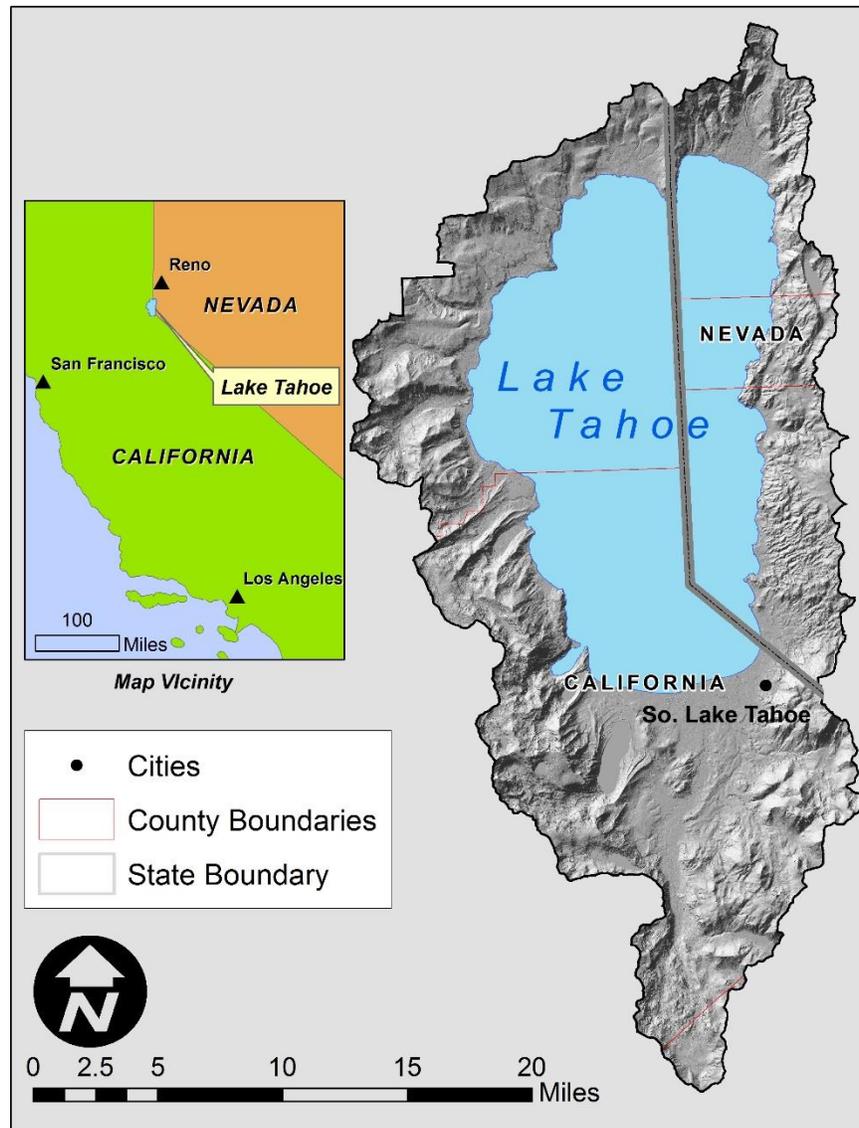


Figure 1 Study Area-Lake Tahoe Basin California/Nevada

The study area is defined by the jurisdiction of the Tahoe Regional Planning Agency (TRPA). The Basin straddles the States of California and Nevada and is situated in the Sierra Nevada Mountains approximately 200 mi. east of San Francisco, CA and 30 mi. south of Reno, NV.

Mid-nineteenth-century Anglo-American immigrants clear-cut ponderosa pines (*Pinus ponderosa*) in the LTB to supply timber to the nearby Comstock Lode. Later, tourism induced development and near complete fire suppression further altered its structure (Raumann and Cablk

2008). The LTB is now less resilient to natural disturbances so large tracts of trees are perishing from insect attack and disease; further, slash accumulating in the forest understory threatens devastating wildfire (Taylor 2014).

Currently, the US Forest Service is thinning and restructuring the Basin's forests into a pre-settlement structure as a first step toward a species configuration conducive to predicted future conditions. Patterns of diversity are indicators of both the forest's current state and the results of restoration efforts (U.S. Department of Agriculture 2013).

The Tahoe Basin is not alone. As of 2014, an estimated 2 billion ha of forest in the United States are degraded due to human activity. Four general approaches, revegetation, ecological restoration, forest landscape restoration and functional restoration, exist to restore degraded forests (Stanturf, 2014). In some manner, each of these methods affects species richness by (intentionally) inducing a disturbance. Given the influence of diversity on ecosystem health, any tool that clearly presents forest structure through the lens of species richness is an invaluable asset to those charged with researching and making forest management decisions.

1.1 Research Questions

A series of research questions form the core of this research. The questions build upon each other addressing: the viability of creating tree species richness predictions using Maxent, the validity of its output and finally, if that output may (or should) be further processed in a GIS.

1. Can maximum entropy habitat modeling produce a valid and broad scale representation of tree species richness?

A variety of ways are available to predict species richness. For example, at a regional scale a remote sensing algorithm expands the normalized difference vegetation index (NDVI) to predict

tree species richness (Waring et al. 2006). On more local scales, techniques that leverage species area models (SAM) or species accumulation curves are common (Ugland et al. 2003).

Unfortunately, these are often computationally difficult and they are rarely spatially explicit.

Ideally, tree species richness could be extrapolated from sampled data; however, this is notoriously difficult (Lam and Kleinn 2008). As a black-box maximum entropy modeling package, Maxent both handles the heavy mathematics and is capable of producing broad scale predictions based on sampled (presence only) data.

Maxent is designed to predict the ranges of individual species based on sparse observations and environmental data, not species richness. Therefore, the second research question addresses the validity of using Maxent in this way.

2. *Can the Maxent maximum entropy modeling package make valid predictions of tree species richness patches?*

This question of validity is somewhat subjective. The most important element of resolving this issue is defining criteria that relate to the model's purpose and context (Rykiel 2006). No perfect algorithm, however complex, exists for predicting species richness. Therefore, it is unreasonable to expect Maxent to predict the specific species richness value for any given surface pixel.

Rather, it is expected that, as a valid method Maxent will predict ranges of species richness that generally agree with actual observations and provide a useful indication of structure and change across the study area.

The study of borders and ecotones is an exciting and growing area of study in ecology, which leads to the next research question.

3. *Can the location and properties of border regions between Maxent predicted species richness ranges be derived from Maxent output and if so, are they valid?*

Maps of valid species richness border regions provide a cleaner and richer depiction of a forest's diversity structure. The question of validity is also relevant for this element of the model, however. Exactly what defines a border region or ecotone is a matter of debate in the literature. Given this model is based on predictions derived from probability, it is possible overlapping regions are simply the result of variation in the predictions. Techniques exist in the literature to identify legitimate ecotones, so if transitions between species richness ranges are meaningful, they should be detectable using those methods.

1.2 Motivation

Academic literature focused on species richness and diversity widely agrees that managing diversity is critical to effective forest conservation and restoration efforts. For instance, Xu et al. (2012) exclaim species richness information is critical for assessing “biodiversity, conservation, and extinction risk in the face of climate change and anthropogenic disturbance” (p. 1006). Morin et al. (2011) conducted an extensive study correlating tree species richness to forest productivity.

Many authors also concur that estimating species richness from sampled data is difficult. (Graham and Hijman 2006, Lam and Kleinn 2008, Xu et al. 2012). A geospatial technique that effectively predicts zones of consistent species richness and then identifies the location and properties of border regions between them addresses both of these issues.

A search of the literature produced no practical application of maximum entropy to create a species richness distribution surface. An investigation by Xu et al. (2012), however, addressed

the problem of species richness's sensitivity to scale (Crishom et al. 2013). The authors assessed whether twelve quantitative methods could upscale locally sampled species richness to a regional level. The authors opined that for heterogeneous landscapes, the maximum entropy method performed the best. The excellent performance of maximum entropy in a species richness application is encouraging given the popularity and relatively easy access to the maximum entropy algorithm provided by the Maxent modeling package (Phillips et al. 2006). The specific mathematics evaluated by Xu et al. differ somewhat from Phillips et al., but they are anchored by the same theory. Given this, an evaluation of Maxent output for this application is worthwhile.

The Lake Tahoe Basin provides an excellent opportunity to utilize and evaluate the maximum entropy method in a practical species richness investigation. An investigation of Northern boreal forests would primarily feature regional (γ) scales whereas a small conservation area would likely examine local (α) scale diversity. In contrast, the spatial extent of the Lake Tahoe Basin is such that diversity patterns are relevant at fine granularity as well as at wider scales (α - β) (Whittaker 1960). These factors add particular interest to investigating the performance and utility of a maximum entropy method given the findings of Xu et al. (2012).

While plotting raw Maxent predictions of species richness provides useful information regarding the dispersion of tree species diversity, the structure of Maxent output rasters allow for post-processing to display border regions. There are two benefits to extending the analysis this way. First is clarity of the representation. Maxent does not produce predictions of neat polygons with coincident borders. Rather, numerous pixels are identified as suitable for multiple species richness values. Displaying these areas as borders between areas of homogenous zones of species richness reduces clutter in the representation and better highlights patterns. Second, the

width, shape and abruptness of border regions provide important clues as to the underlying mechanisms that shape diversity patterns. Lastly, species richness border zones provide new, tangible spatial objects for scientists and managers to investigate (Kroger 2009).

1.3 Workflow

The workflow for this research followed a stepwise path from planning to interpretation.

Specifically, the research plan was to determine appropriate environmental factors, construct the models and then validate them.

1. Establish data needs (Maxent environmental variables)

The scientific literature and simple brainstorming identified a list of biotic, abiotic and anthropogenic factors that influence species diversity in the LTB.

2. Gather required datasets

Several spatial datasets provided input variables to the models constructed for this research:

- *Tahoe Basin Existing Vegetation Map (TBEVM)*
- *Tahoe Basin Digital Elevation Model*
- *Tahoe Basin Soils Survey (SSURGO)*
- *Tahoe Area Weather Stations*
- *TRPA Land Use*

3. Process datasets

Raw datasets were processed to consolidate datasets and derive attributes not explicitly provided by the dataset authors.

4. *Exploratory regression*

The ArcGIS Ordinary Least Squares (OLS) exploratory regression tool iteratively created regression models to select the most appropriate independent variables for use in the Maxent models.

5. *Species richness-based maximum entropy model (MEM)*

This model highlights ranges of tree species richness communities (e.g. 2 species, 3 species, etc.) and then intersects them to identify the position, width, shape and abruptness of overlapping predictions.

6. *Species based Maximum entropy model*

This validation model adds Maxent-derived species distribution models (SDM) for each relevant tree species—a stacked species distribution model. If both models in this study are valid, the resulting surfaces should be similar (e.g. Dubuis (2011)).

7. *Known species vs. predicted range map*

While Maxent is specifically designed to predict species ranges, this output must also be validated. If the Maxent predictions encompass point data of known species, the Maxent predictions may be considered reasonable.

8. *Split moving window dissimilarity analysis (SMWDA)*

SMWDA is a common method used in ecotone studies (e.g. Hennenberg 2005, Kroger 2009). This analysis validates that overlapped Maxent predictions of species richness represent transitions between areas of diversity.

Chapter 2 provides an overview of peer-reviewed literature that support the premises and methods used in this research.

CHAPTER 2: BACKGROUND

This chapter reviews literature related to this research. In some instances, evidence from reviewed literature is synthesized to provide the basis for assumptions made in this project (e.g., species richness conceptual model). The topics proceed from prevailing theory behind species richness and maximum entropy modeling to practical issues such as parameter setting in the Maxent package.

2.1 Species Richness

The thrust of this research is to provide scientists and managers a tool to view the distribution of tree species richness across a study area. The utility of such information derives from the nexus between species diversity and ecosystem function. A significant body of literature explores this connection, termed biodiversity ecosystem functioning (BEF; Morin, 2011), from a variety of angles ranging from climate change to management technique. In the context of this research, several of these studies are leveraged to develop a conceptual model of species richness pattern development.

An excellent starting point for exploring species richness is the deep analysis provided by Jonathan Adams in his book, *Species Richness* (2009). Adams defines species richness as the total number of species present without regard to density. His first chapter explores local scale species richness which in his words are “applicable to scales ranging from a few centimeters to several tens of kilometers” (p. 7). This differs from the classic *alpha*, *beta*, *gamma* diversity scales proposed by Whittaker (1960), but as Adams laments, the complexity of ecological systems precludes a unified theory. For the purposes of this investigation, Adam’s local scale seems most appropriate.

In sessile communities (e.g. trees), Adams lists three primary factors that affect the local species richness:

- Recovery from a single large disturbance
- Varying frequency and intensity of disturbances across an area
- Varying quality of substrate such as soils, slope, elevation, aspect

These criteria cover significant ground. Discrete disturbances might include a clear cutting event or perhaps a natural fire. Frequency of sustained disturbance can relate to climate patterns or perhaps the level and type of routine management activity. The last item encompasses any substrate variable such as slope, aspect or elevation.

Morin et al. (2011) provide a more pointed investigation that examines the relationship between diversity and productivity using a forest succession model. The authors document evidence that tree species richness promotes productivity in temperate forests via complementarity between species. Specifically, a forest community that combines many species with varying shade tolerances and abilities to grow in degraded environments enhances a forest's ability to recover from small-scale disturbance. The authors note that productivity is related to species richness levels. Adams (2009) discusses this concept as it relates to local-scale richness. He proposes that a balance between disturbance and productivity affects species-specific carrying capacity. A particularly relevant item gleaned from Morin et al. (2011) is their use of soil moisture capacity in their models. That variable is used as an environmental factor representing soil quality in this study.

DeClerck et al. (2006) provide insight into the specific effects of species richness on the Lake Tahoe Basin. The researchers studied species richness in the Desolation Wilderness area of the LTB with a focus on stand stability derived from species richness. Of particular interest are

their environmental variable selections of elevation, slope and aspect and community response variables of species percentage, basal area and canopy closure. The results of this study are interesting. The authors found that increasing species richness in the Desolation Wilderness correlates with increased resilience (the ability to *recover* from disturbance) but not resistance (*tolerance* of disturbance).

Martin-Quellar et al. produced two relevant studies (2011 and 2013) that address management effects on species richness. The caveat to the applicability of these studies is that they investigate *gamma* scale diversity in Central Spain. This limits their specific applicability to this study somewhat, but they are the only studies of this nature addressing this critical element. The 2011 study utilized geographically weighted regression to examine the effects of management on species richness. The particularly salient conclusion is that biodiversity models are more effective where management practices are incorporated and that management effects on species richness were stationary across their study area. Further relevant conclusions include observations that tree species richness closely correlates with basal area and annual precipitation.

Martin-Quellar et al. (2013) used Bayesian statistics to further investigate the biodiversity-management style connection focusing on the effects of silviculture. This is particularly applicable to the Tahoe Basin because active restoration efforts have been ongoing for some time. General observations by the authors indicate that moderate intensity disturbances (natural or induced) contribute to species richness. Also of interest to this study is the authors' conclusion that a hierarchy of variables, chiefly climate, affects the overall influence of management and other disturbance types on species richness.

2.2 Conceptual Model

Ideas and conclusions from the literature in section 2.1 were synthesized into a conceptual model for species richness structure graphically depicted in Figure 2. Specifically the graphic depicts a systems-based hypothesis regarding how the environmental elements used in the Maxent model combine to shape a forest structure made up of varying species richness patches.

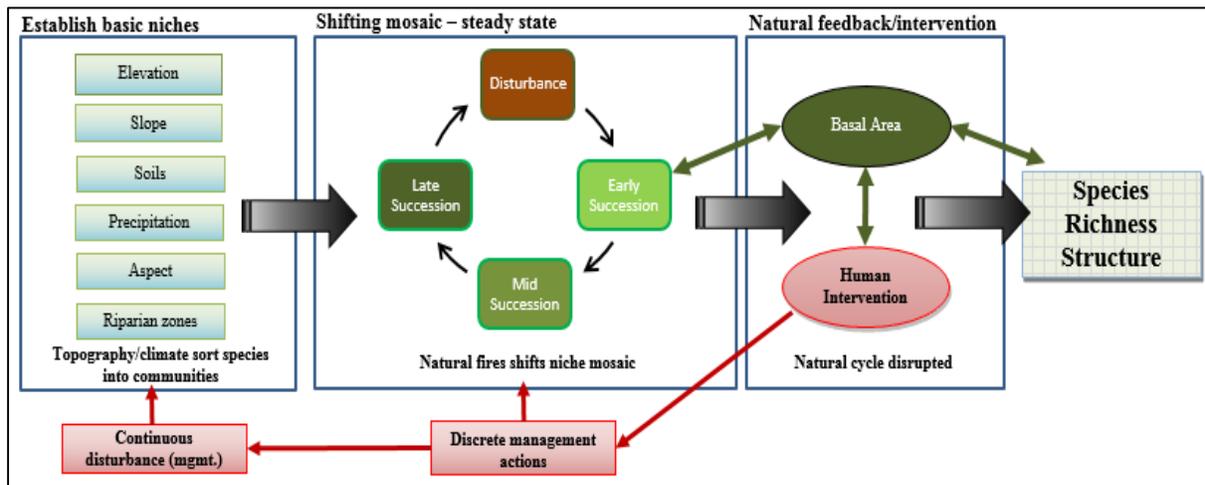


Figure 2 Conceptual Model of Environmental Factors Leading to Species Richness Structure in the Lake Tahoe Basin CA/NV

The conceptual model graphically depicts the biotic, abiotic and anthropogenic environmental factors captured by the species richness prediction models. Factors are arranged into a theoretical system based on evidence in related literature and Hubbell's Unified Neutral Theory of Diversity and Biogeography (Hubbell 2008). The goal is to produce a systems-based model whose output is tree species richness structure in the Lake Tahoe Basin.

The conceptual model is rooted in Stephen Hubbell's Unified Neutral Theory of Biodiversity and Biogeography (Hubbell 2008). The theory treats trophically similar species (in this case, montane conifer and deciduous riparian trees) as ecologically identical (i.e. equally fit to their environment). Competition between species is discounted in favor of random births, deaths, speciation and dispersal as the mechanisms for creating biodiversity. This is a useful

assumption given the number of species present is the only important output of this model. The specific species contributing to the species richness values may vary and remain anonymous.

The inclusion of niche-determining factors (far left panel) is counterintuitive given Neutral Theory discounts niches as a contributing factor to biodiversity. Following the publication of Hubbell's theory in 2001, several papers suggested that Neutral Theory is inadequate in isolation, requires unrealistic auxiliary assumptions and spatially explicit versions are limited in their descriptive power (Etienne and Rosindell 2011). Gewen (2006) suggests that neutrality should be considered a backdrop or null hypothesis. Another paper opines that Neutral Theory by itself is too extreme. Stabilizing influences (e.g. niche-forming factors) play roles in biodiversity to degrees that vary by specific community (Adler 2007).

This model accepts that known biotic and abiotic factors, identified in the literature and prioritized using regression, influence tree species richness in the LTB; however, the niches are not defined classically. Hubbell's Neutral Theory is an extension of island biogeography. As such, niches formed by the factors in the far left panel most accurately represent distinct communities that support a finite number of tree species drawn from the pool of species present in the LTB. Clearly, these "island" communities are contiguous; therefore, allopatric speciation is unlikely to be a significant driver of biodiversity in the LTB. Random birth, death and especially seed dispersal (e.g., Barringer et al. 2012), however, are very plausible mechanisms.

Mixing ecological drift (neutral factors) and competitive factors into a single model is not unprecedented. For example, Orrock and Watling (2010) created a biodiversity model that tests their theory that community size is the mediating nexus between neutral and competitive factors.

The middle panel of the conceptual model incorporates shifting mosaic-steady state (SMSS) as proposed by Borrmann and Likens (1994) to represent natural disturbance and

resulting secondary succession in the ecosystem. Adams (2009) stresses the heavy role of acute and chronic disturbances in defining the species richness of local-scale plant communities. In fact, Adams illustrates that an intermediate level of disturbance (versus no disturbance or catastrophic disturbance) generally results in optimum species richness. SMSS is somewhat controversial (as is Neutral Theory) because it represents a climax condition. Perry (2002) presents an exhaustive discussion of this debate and the shift in thinking brought about by the emergence of spatial ecology. The intent here is not to make a claim as to the existence of SMSS; rather, its inclusion in the model provides a convenient way to represent disturbance and succession. SMSS is not entirely theoretical, however. A recent study documented evidence of SMSS approximately 100 miles north of Lake Tahoe in the Lassen Volcanic National Park, California (Wang and Finley 2011). These factors are commonly accepted across the ecological literature even if their specific ramifications are debated (Perry 2002).

SMSS represents niches (species richness “islands”) that are inherently stable. Neutral Theory presumes succession by random colonization of disturbed landscape patches versus a static climax state. This model presumes, however, climax states (or at least intermediate seres) exist for periods of time even if they are short. During those periods, community sub-patches are subject to localized disturbances such as natural fire, avalanches, etc. Disturbed patches move through successional stages (e.g. meadow/brush → sparse/sun-adapted forest → dense mature forest). The conclusions of DeClerck et al. (2006) regarding resistance and resilience tend to support this hypothesis in the LTB. The random nature of these disturbances ensures a shifting micro-landscape, but overall niche stability.

An SMSS disturbance model meshes well with Neutral Theory if we accept an individualistic view of succession (Gleason 1926). The individualistic view assumes that

patches left barren by disturbance will be randomly colonized by a species well-adapted to local conditions. The specific species that colonizes a barren patch is irrelevant; the aggregated result of multiple random disturbances and succession within a community will be a steady mean value of species richness. This process in isolation would logically result in a perpetual climax state (e.g. SMSS). Both the individualistic view of succession and Neutral Theory accept that environmental variables that define niches degrade over time resulting in ecological drift (and perhaps a change in maximum species richness). For instance, climate change may alter mean annual precipitation or, significantly in the LTB, management practices may mechanically remove certain species.

The final (far right panel) incorporates non-natural influences into the model. Regression analysis revealed that basal area is a major contributor to species richness in the LTB. Basal area is defined as the total forest floor coverage of living tree stems (expressed as area/area). A review of relevant literature reveals that its impact on species richness is well documented. Basal area is a common indicator of primary productivity, which is a proxy for forest health and structure as affected by biodiversity (Sagar and Singh 2006, Adams 2009, Christholm 2013). Basal area also reflects time since disturbance, land use history, stand heterogeneity, and ongoing succession; that is, it is an indicator of modification by the local environment (Lohbeck et al. 2012).

In sum, basal area is a community response variable. Its function in the model is to capture unusual (non-parameterized) events such as disease or insect attack that act to reduce basal area via mortality and, importantly, the presence or absence of human management activity such as prescribed thinning, burning or cutting for convenience (e.g. to facilitate development). Clearly, every removed tree reduces total basal area and constitutes a disturbance.

Finally, red boxes along the bottom of the model separate human management into discrete events and continuous disturbance. A single anthropogenic disturbance (a prescribed burn for instance) is likely absorbed into the ecosystem similarly to a natural event such as a wildfire or avalanche. Ongoing policy, such as maintaining a ski resort or an urban landscape fundamentally alters the underlying niche discriminators that produce species richness patches.

Species richness patterns are often predicted by stacking species distribution models tailored to each species present in the ecosystem (SDM, Dubuis et al. 2011, Cord et al. 2014). An individual species SDM is the output of an ecological system. This model illustrates that species richness distributions are also the output of an ecological system. In other words, a spatially explicit species richness model may be extracted from relevant biodiversity factors (Dubuis 2011). A major goal of this research is to show that both approaches (utilizing maximum entropy) will arrive at functionally equivalent solutions.

2.3 Appropriateness of Maximum Entropy for Species Richness Modeling

The maximum entropy algorithm for constructing area-based models is used in this study. The typical use of this algorithm is for range estimation for single species from presence-only data (Phillips et al. 2006). No practical use of this method for species richness modeling was found in the literature; however, several journal articles lend credibility to the method.

Graham and Hijmans (2006) compared the efficacy of three techniques for mapping species ranges and species richness. Specifically, the authors evaluated maps drawn by experts, maps produced by species distribution modeling, and hybrids of the two methods. The authors explicitly describe their use of Maxent for species range maps including variables, settings, and evaluation. As for species richness mapping, the authors state they evaluated a map created by a

“distribution model” (p. 583); however, they are unclear as to the specific procedure they used to create it. Given their use of Maxent for their species range map, it is likely they used Maxent for this application as well, although it is unlikely they treated species richness values as individual species. The relevant conclusions of their study are that distribution model-derived richness maps tended to overestimate species richness versus hybrid maps but performed well.

Xu et al. (2012) evaluated twelve procedures for estimating species richness at a regional scale from local scale samples. Six of these are non-parametric incidence-based methods, and six are area-based methods, including maximum entropy. The authors use the term, MaxEnt; however, they are referring to the maximum entropy method rather than the Maxent application itself. Further, Xu et al. are not attempting to derive species richness from environmental factors, they extrapolate species richness from area. The mathematics and end goal of Xu et al. are different from the present study; however, the underlying maximum entropy approach, estimating an unknown distribution based on sampled data, is similar. Importantly, the authors successfully applied maximum entropy to a species richness problem and concluded that the maximum entropy procedure was the best performing algorithm in their test set. Additionally, they made the same overall assessment as Graham and Hijmans (2006). That is, the algorithm overestimates species richness versus known plots by as much as 40%. Xu et al. does not provide conclusive evidence that maximum entropy is appropriate for this application, but their similar results to Graham and Hijmans and success in a species richness application add credibility to the procedure proposed by this research.

Another important study is Dubuis et al. (2011). The researchers directly compared species richness predictions derived from a macroenvironmental model (inferring species richness from relevant environmental predictors) and stacked SDMs. The authors did not create

spatially explicit models; rather, they used statistical techniques. The study concluded that the two methods were functionally identical; however, each had unique strengths and weaknesses. Interestingly, the authors suggest that combining the two techniques could provide the most accurate predictions. Aside from being non-spatially explicit, Dubuis et al.'s study is remarkably similar to this research. Their methods lend credibility to the procedures used in this thesis and their conclusions support validation by comparing stacked SDM and species richness-based models.

A less specifically relevant, but interesting, study is Kolström and Lumatjärvi (1999). The authors developed a decision support model designed to predict the effect of forest management decisions on species richness. The model is more complex than the maximum entropy-based model proposed by this research. Specifically it is a simulation that incorporates high granularity variables such as decaying wood, stem diameter, and distribution by species. The most relevant component of this extensive modeling effort is an opinion offered in its conclusion, "Our system does not give absolute values for number of species in a stand, but the results could be interpreted as a relative change in the number of species due to the change of habitats. . ." (p. 55). This reflects the specific intent of this research and implies that a less than optimal algorithm can deliver useful information.

None of this literature provides absolute assurance that Maxent is appropriate for this application. Taken together, however, they imply that validation efforts are worthwhile and likely to succeed.

2.4 Maximum Entropy Theory

Although Maxent is a black box-style modeling environment, a rudimentary understanding of its underpinnings are essential to optimizing its use. Three sources were utilized to gain this familiarity.

Elith et al. (2011) translates the machine-learning based mathematics of the Maxent algorithm into statistical language more familiar to ecologists. The paper explores many aspects of the Maxent environment, but the central topic is an explanation of maximum entropy theory itself. Specifically, it relates that the model minimizes the relative entropy (a measure of the distance between distributions) between two probability densities (one from the presence data and one from the landscape).

Merow et al. (2013) is an overview of the settings in Maxent and how they ought to be applied. The authors begin with an overview of the underlying mechanics of the Maxent algorithm from a statistical perspective similar to Elith et al. (2011). This information is used to support recommendations regarding model settings. The authors encourage Maxent-setting decisions based on biological and study related factors rather than convenience or a lack of understanding.

These articles are quite technical despite their intent to simplify maximum entropy. Maxent's web-based documentation is very accessible and was used to temper the difficult reading of the formal articles. The following brief summary is a simple language overview of Maxent based on these sources.

Maximum entropy is described by the statistical distribution that best fits the geographic spread of samples and the background. Maxent calculates the maximum entropy input data presence data as they relate to the "background." The background is the sum of environmental

variables the user provides as explanatory variables. For those accustomed to the ArcGIS environment, these raster datasets are counterintuitively known as “features.”

The process is similar to an exploratory regression; however, the maximum entropy method is nonparametric, meaning it makes few data assumptions and is capable of creating highly complex curves from samples drawn from multiple distributions. The benefit of this is that Maxent can accept mixtures of feature data types such as continuous, categorical, interval that would be difficult to process with ordinary regression techniques. Maxent internally converts categorical data into interval data.

The resulting (raw) output is a relative occurrence rate (ROR) for each raster cell. This is simply the likelihood that a cell is occupied by the same species as the sample set. This is not an explicit probability that a cell is suitable habitat. That would require knowledge of the population size. Rather, it is simply the likelihood that a cell is suitable habitat relative to the set of other cells.

2.5 Additional Relevant Maxent Studies

Prates-Clark et al. (2008) used Maxent with remotely sensed presence data and climate environmental data to investigate actual versus potential ranges of three tree species in the Amazon Basin. Specifically, the researchers were interested in the efficacy of using remotely sensed data in maximum entropy-type models. They compared statistical and predictive output for climate data alone, remotely-sensed alone, and combined models. They found that remotely sensed data is consistently superior as a stand-alone input (likely due to reduced sampling bias), but it is limited because it cannot capture all processes that contribute to forest species distribution. The study is highly relevant to the present study given Maxent is typically used to

analyze field-collected data. The TBEVM data used in this research is derived from IKONOS satellite imagery and shares similar limitations as the remotely sensed data in this article.

TBEVM data is limited to detecting species observable in the canopy. Satellite imagery cannot detect vegetation within the forest understory.

Smith et al. (2012) investigated the spatial extent and recruitment factors of acacia trees in Southern Africa. While the subject matter of this investigation is not particularly relevant, its scale is. The study area of Kruger National Park in South Africa is about 2.5 million ha—roughly equivalent to the Lake Tahoe Basin. Most examples of Maxent use in the literature involve much broader spatial extents than this study or the Lake Tahoe area (e.g. State or continental-level). This study provides validation that Maxent output can be effective at a more community-level scale.

York et al. (2011) modeled habitat for invasive tamarisk (a woody plant) and the Southwestern willow flycatcher. Tamarisk replaces nesting habitat for the flycatcher. Biologic control efforts introduced a beetle that is reducing tamarisk. The researchers used overlap of Maxent predictions of highly suitable tamarisk and flycatcher habitat to predict most-suitable areas for the biologic controls. This study is specifically post-processed Maxent output (e.g. overlay) to achieve a research goal. Further, this study implies that the overlap of Maxent ranges is meaningful, rather than artifacts of the prediction algorithm.

Maxent ranges are based in probability, so they inevitably contain some error. The potential for error is further compounded by the noisy nature of ecological data. Stacking predictions could compound their error to the point that the resulting surface contains more error than meaningful data. As such, this study rigorously evaluated stacked surfaces to ensure they are meaningful.

2.6 Species Richness Boundary Detection and Definition

An important objective of this research is to establish that overlapping Maxent species richness ranges mark legitimate, meaningful boundary regions and not simple variation in the prediction algorithm. Ecological boundaries and ecotones are a high interest topic in the ecological literature (Cornelius and Reynolds 1991, Hennenberg et al. 2005, Kark and van Rensberg 2006, Kroger et al. 2009). Although agreement is widespread that these zones are interesting and significant, there is little consensus on how to define them. Kark and van Rensberg (2006) surveyed recent trends on the subject and working from a variety of definitions in the literature, they crafted a working definition of ecotone: An ecotone is an “area of steep transition between communities, ecosystems, or biotic regions” often emphasizing the abruptness of the change (p. 32). Whether or not boundaries between areas of similar diversity are properly termed ecotones, boundaries or some other phrase, is less important than if it is useful to identify them. By the Kark and van Rensberg definition, ecotones are properly defined if they delineate homogenous regions forged by natural processes. Species richness and biodiversity easily meet that criterion.

Whether Maxent output can effectively identify border regions is another matter. Kroger et al. (2009) investigated boundary detection for riparian zones in the Kruger National Park, South Africa. They emphasize that border detection is a matter of identifying changes in underlying processes. Maxent predicts probability of occurrence based on variability in the environmental data that represent these underlying processes. In discussing methods of identifying ecological borders, Cornelius and Reynolds (1991) point to theory that these borders act similarly to semi-permeable membranes where elements of adjacent homogenous units percolate between each other or into an intermediate zone. It is not a great leap to suggest overlapping homogenous Maxent predictions could represent this process.

Cornelius and Reynolds (1991) further suggest that ecotones and borders can be identified by seeking discontinuities in transect data with a family of analyses called “split moving window.” Kroger et al. (2009) and Hennenberg (2005) utilized split moving window dissimilarity analysis (SMWDA) in their research involving border and ecotone identification. SMWDA is relatively simple statistical procedure, but it is limited to identifying discontinuities along two-dimensional transects. The Maxent method is less quantitatively rigorous, but it allows for creating mappable representation of border regions across a large area. In the context of this research, the established SMWDA method is used to validate the Maxent method. If Maxent range overlaps represent border regions, then SMWDA discontinuities should coincide with them.

This chapter summarized the role of species richness in ecosystems, provided theoretical framework for predicting species richness from environmental variables. Additionally, the use of Maxent to create species richness predictions was explored using examples from relevant literature. Chapter 3 is an exhaustive description of the data and techniques used to undertake this research.

CHAPTER 3: DATA AND METHODS

This chapter provides a detailed description of the datasets and methods used in this research. The first section provides an evaluation of the properties and appropriateness for use of each dataset followed by a description of how they were processed for use. The second section is an accounting of the rationale for, and construction of, each model and procedure used in this project.

3.1 Data

No spatial analysis is credible without a full evaluation of its input datasets. Further, appropriately projecting data for any spatial analysis is a critical step. The geographic position of the Lake Tahoe Basin makes choosing the best projection uniquely challenging, so this section begins with a brief discussion of that issue. The remainder of the section is a thorough evaluation of each dataset. The Tahoe Basin Existing Vegetation Map (TBEVM) is the primary source of species richness data, so a more thorough exploration of that dataset is included.

3.1.1 Map Projections

All datasets in this project are projected to **NAD 83 UTM zone 10**. Selecting an appropriate map projection for analysis in the Lake Tahoe area can be difficult. The Basin straddles the boundaries for multiple UTM and State Plane projection systems (Figure 2). These are the projection systems that best preserve area and distance at scales that usefully represent geographic regions of this size (e.g. ~1:300,000). Lake Tahoe's geography places it at a disadvantage in this regard because no system perfectly suits Basin-wide studies. The Tahoe Basin is distant from the central meridian of any usable standard projection and sits at high

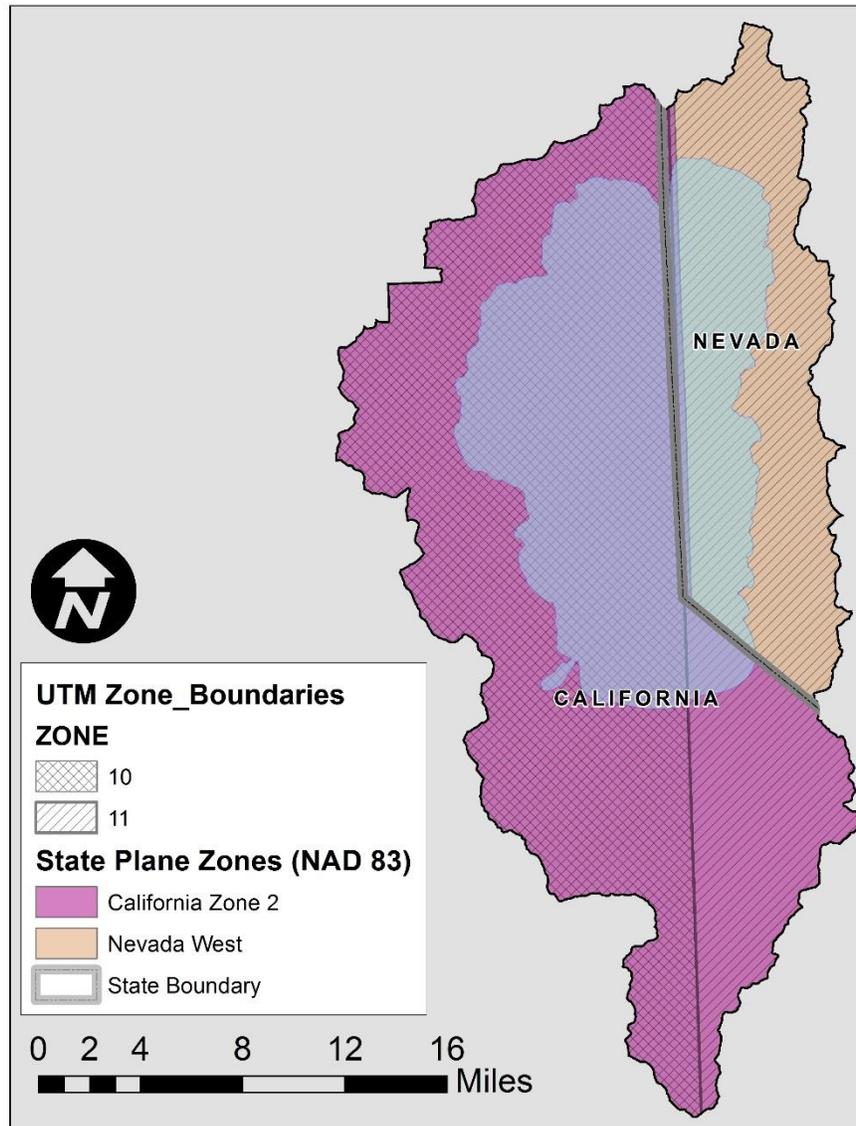


Figure 3 Common LTB Map Projections

The most common map projections for the Lake Tahoe Basin are provided to illustrate a unique problem facing spatial analysts in the LTB. The Basin straddles the boundaries of four map projection schemes, none of which encompass the entire Basin. For Basin-wide studies, State-Plane systems are unusable, leaving the extreme edge portions of UTM zones 10-11 as the most usable projections. Zone 10 covers the most area of the Basin and is the de-facto standard for LTB studies. In every case, some distortion is inevitable and must be considered when evaluating the results of spatial analyses in this region.

elevation, so some distortion is unavoidable. Custom projections are an option for fine accuracy work, but they make sharing datasets difficult. State Plane systems are not valid outside State boundaries, so researchers must select a UTM zone or a broader projection scheme such as Albers. While UTM zones 10 and 11 nearly bisect the Basin, zone 10 covers a slightly larger

area. NAD 83 zone 10 is not ideal, but it has become the (mostly) de facto standard for GIS work by governing agencies in the Lake Tahoe area.

3.1.2 Tahoe Basin Existing Vegetation Map (TBEVM) v.4.1

TBEVM is the product of a UC Davis Center for Spatial Technology and Remote Sensing project that used IKONOS satellite imagery to map vegetation in the Lake Tahoe Basin (Figure 4). The dataset provides high granularity vegetation information that exceeds the detail of any other publically available dataset.

- *Representation:* The dataset is a vector polygon model, so the data are technically discrete; however, the extremely fine resolution of the polygons make it comparable to many raster datasets (e.g. Landsat). The effect is to create a nearly continuous model when viewed at moderate scales.
- *Size:* The dataset contains 20,693 discrete polygon objects.
- *Scale:* The extent of the dataset covers the Lake Tahoe Basin that totals 1330 sq. km. Polygons in the dataset vary from 16m² to 100,000 m². The underlying raster data have a 1-m resolution.
- *Aggregation:* Polygons were aggregated from raster data on the basis of physiognomic vegetation classes. The polygons represent best-fit vector overlays of vegetation classes defined by agencies such as the US Forest Service (USFS) and the Federal Geographic Data Committee. The resulting data reflect distinct micro-regions of homogeneous vegetative communities.

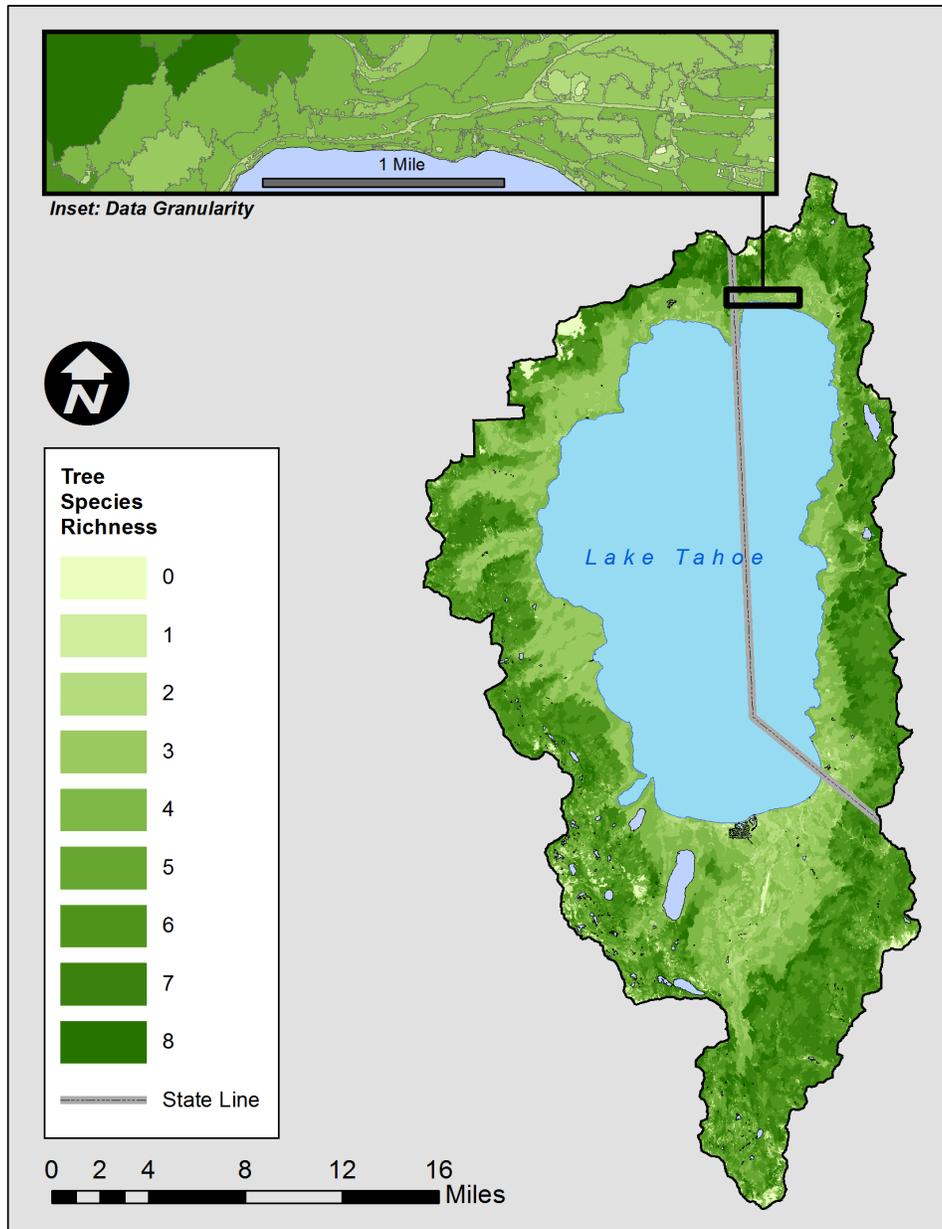


Figure 4 Tahoe Basin Existing Vegetation Map v. 4.1 (TBEVM)—Tree Species Richness Data Overview and Exploration

The Tahoe Basin Existing Vegetation Map is a product of the UC Davis Center for Spatial Technology and Remote Sensing that utilized 2007 IKONOS satellite imagery to classify existing vegetation in the LTB. Tree species richness data for this study was derived from this dataset. The main figure displays the resulting species richness surface and the inset illustrates the fine granularity of these data.

- *Attributes:* The dataset contains over fifty attributes derived from IKONOS classification and analysis. Data of particular interest are percent coverage figures for nine individual tree species.

Note that species richness is represented in Figure 4 to facilitate an initial overview of the presence-only data used in this research. Species richness is not a native attribute in TBEVM. The method used to create this attribute is described in section 3.2.3.

Over 40 additional attributes are included in the dataset. Some variables such as “basal area” were directly assessed and utilized as Maxent feature data. Others, such as “stems per hectare” provided the basis for derivative attributes.

- *Fuzzy/Crisp:* These data are matched to subjective ecological categories that crisply partition a continuously varying forest landscape. The polygons effectively model forest structure in aggregate; however, their individual boundaries are inherently fuzzy.
- *Error:* Error in these data is high as in any forest dataset. Remotely sensed data is error prone due to misclassification and atmospheric anomalies. Beyond inherent remote sensing error, this is an aging dataset. These data were obsolete the moment they were collected and that error multiplies with age. The authors’ metadata includes accuracy estimates that vary by attribute from 10 to 80%. The large size of the dataset works to balance error to a reliable 50% range. Age degrades that accuracy. The data do not represent the forest today, but it is an excellent representative snapshot.
- *Fitness for use:* TBEVM contains a large percentage of the data used in this research, so it was given a particularly close evaluation. The age of the data—circa 2007—is concerning given field validation of derived datasets is desired, but TBEVM is the only publically available dataset that can deliver species-level resolution. To maximize the probability of

success of this research, an extremely fine grained dataset was specifically sought. TBEVM meets this criterion (Figure 3 inset). Ecological experts at UC Davis assembled these data, so the data can be trusted as an accurate snapshot of Tahoe Basin forest structure as of the date it was assembled.

While the data may be considered accurate as of their collection, they have inherent limitations. First, they are limited to the resolving power of the author’s classification algorithm. There are nine tree species classified (Table 1).

Table 1 TBEVM Species

Species Code	Scientific Name	Common Name
ABCO	<i>Abies concolor</i>	white fir
ABMA	<i>Abies magnifica</i>	red fir
JUOC	<i>Juniperus occidentalis</i>	western juniper
PIAL	<i>Pinus albucaulis</i>	whitebark pine
PICO	<i>Pinus Contorta</i>	lodgepole pine
PIJE	<i>Pinus jeffreyi</i>	Jeffrey pine
PIMO	<i>Pinus monticola</i>	western white pine
POTR	<i>Populus tremuloides</i>	quaking aspen
TSME	<i>Tsuga mertensiana</i>	mountain hemlock

Two prominent species, Ponderosa pine and Douglas fir (*Pseudotsuga menziesii*) are excluded from the dataset as these species are indistinguishable from Jeffrey pine (*Pinus jeffreyi*) using the author’s remote sensing technique. Second, remote sensors cannot “see” into the understory level of a forest. The data are limited to a canopy view of the forest, which may disguise pertinent elements of species richness such as early succession recruitment of shade tolerant species. With these limitations understood, the dataset is well suited for use in this project.

Geostatistical analysis was utilized to conduct an exploration of the TBEVM tree species richness data (Figure 5). The top panel is a histogram of species richness. The data appear to be

normally distributed except for a spike the lowest bin. This is probably attributable to urban areas disturbing the natural distribution system.

The second panel is a typical snippet of an entropy Voronoi analysis for the dataset. The data appear to be *non-stationary*.

The last panel is a trend analysis that graphically depicts an increasing richness values moving away from the lake (and increasing in elevation).

Notably absent is a semivariogram. The downside of a high granularity dataset covering a large area is that it may exceed software capabilities or computer memory limitations. Memory limitations precluded the construction of a semivariogram for these data.

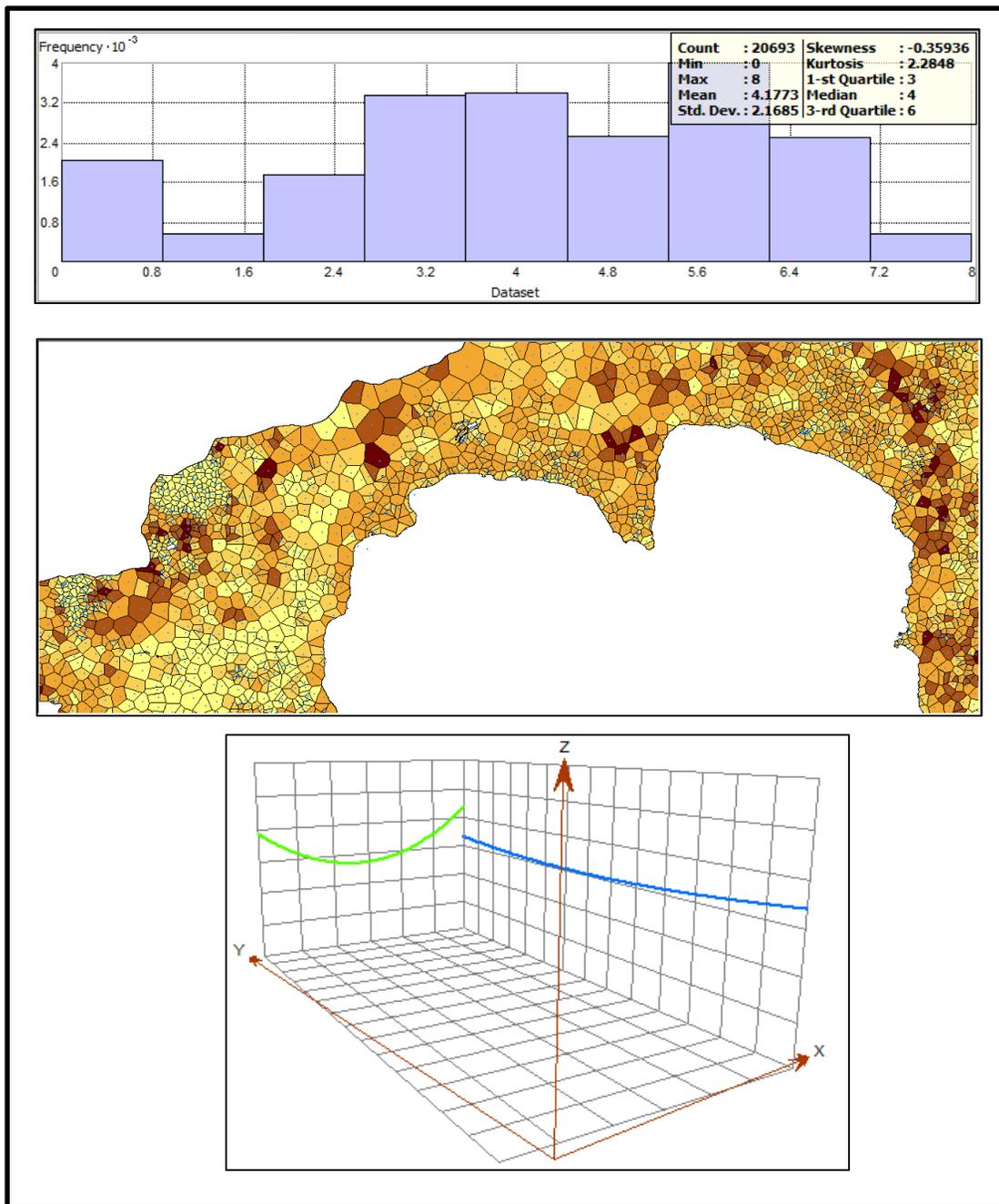


Figure 5 Species Richness Data Exploration Using Geostatistical Analysis

The ArcGIS geostatistical toolbar provides further exploration of the species richness data derived from the TBEVM dataset and displayed in Figure 4. The top panel is a histogram of the species richness data. The data approximate a normal distribution except for a spike in the lowest bin, possibly attributable to anthropogenic disturbance in urban areas. The middle panel is an entropy Voronoi diagram section used for assessing the stationarity of the data. The data appear non-stationary. The bottom panel is a trend diagram that suggests an increasing species richness trend moving away from the center of the dataset (Lake Tahoe) in the x (E-W) and y (N-S) directions.

3.1.3 Tahoe Basin Digital Elevation Model (DEM)

DEMs are a raster model of elevation data. This LIDAR-derived Digital Elevation Model (Figure 6) was assembled by the San Diego Supercomputing Center utilizing public domain point cloud data provided by TRPA.

- *Representation:* The native dataset is a 5x5 meter raster model whose data indicate the mean elevations of first returns (e.g. treetops) and ground returns (for bare ground) from a LIDAR sensor. LIDAR point densities for first returns average 11.82 points per square meter, while ground returns average 2.26 points per square meter. To save computer resources, the working data for elevation were resampled to a 30x30 meter resolution. Slope and aspect derivatives were computed from the 5x5 meter data as these are comparatively small files.
- *Size:* Lake Tahoe Basin totals 513 sq. mi. in area. At 5 m² resolution, this equates to approximately 266 million pixels.
- *Scale:* The extent of the dataset covers the Tahoe Basin with 5 sq. meter cells
- *Aggregation:* The data is continuous; therefore, it is not technically aggregated. The size of the dataset required it to be resampled into a coarser resolution—a form of aggregation.
- *Attributes:* These data are expressed in meters (ratio data). Although not specifically an attribute, the dataset is hydrologically enforced meaning drainages are forced to conform to known stream features.
- *Fuzzy/Crisp:* Viewed as ‘point’ data, raster elevation data are crisp; however, the point data are interpolated into a surface. Any interpolation is an educated guess and is, therefore, always fuzzy.

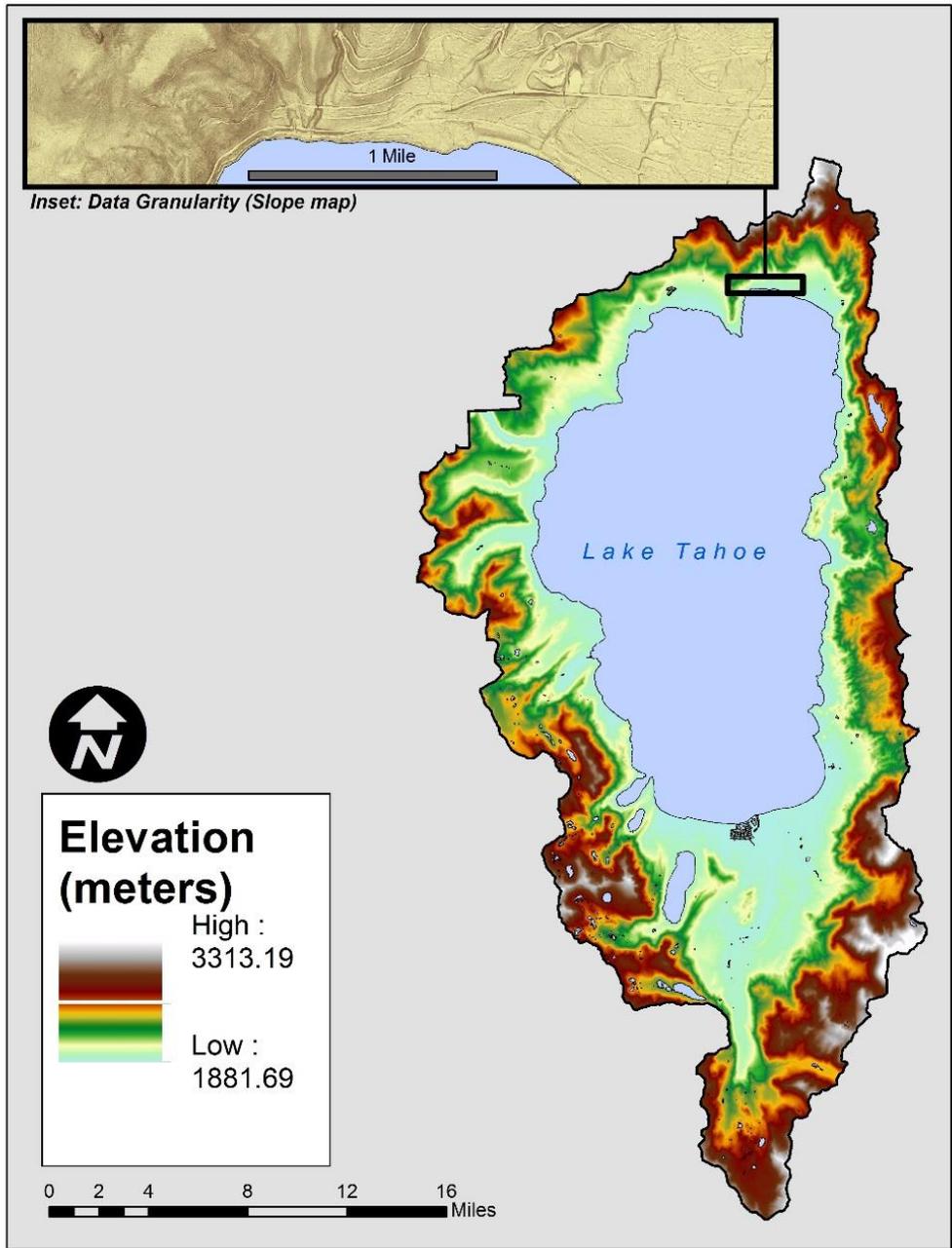


Figure 6 LTB 5m LIDAR-Derived Digital Elevation Model (Resampled to 30m Resolution)

This LIDAR-derived Digital Elevation Model was assembled by the San Diego Supercomputing Center utilizing point cloud data provided by the Tahoe Regional Planning Agency. The DEM is displayed at its native 5m resolution; however, for analysis, the model was resampled to 30m to reduce its impact on computer resources. The inset is a slope derivative of the 5m elevation surface provided to illustrate the granularity of these data. File sizes for slope and aspect derivatives at 5m resolution are much smaller than the original elevation surface. Those surfaces were analyzed at their native resolutions.

- *Error:* Vertical error in these data is quantified by root mean square error, which is estimated at .036 meters for this dataset. While this is an impressive number, LIDAR data have inherent error. The data reflect the first returns from the LIDAR sensor; this may not be the ground. The inset panel of Figure 6 illustrates this problem where trees and manmade objects such as streets and buildings are visible in the image.

Error in this dataset is larger than normal due to the forested terrain of the Tahoe Basin. Many of the LIDAR returns are likely to be treetops, which can be up to 50 feet tall in this area. This error is unlikely to present a significant problem for this application. The data are already aggregated and elevation error of ± 100 m is acceptable given the scale of the study area.

- *Fitness for use:* The primary fitness issue with this dataset is its unwieldy size. ArcGIS does not process extremely large datasets very well, and where it does, processing is lengthy and places large resource demands on typical personal computers. This is for the raw data—analysis functions can increase resource demands to the point they may not be possible. These data (and their derivatives) were resampled into a coarser resolution to decrease required computer resources.

3.1.4 Tahoe Basin SSURGO Soils Survey

The Soil Survey Geographic Database (SSURGO) is the output from an ongoing nationwide survey of soils (Figure 7).

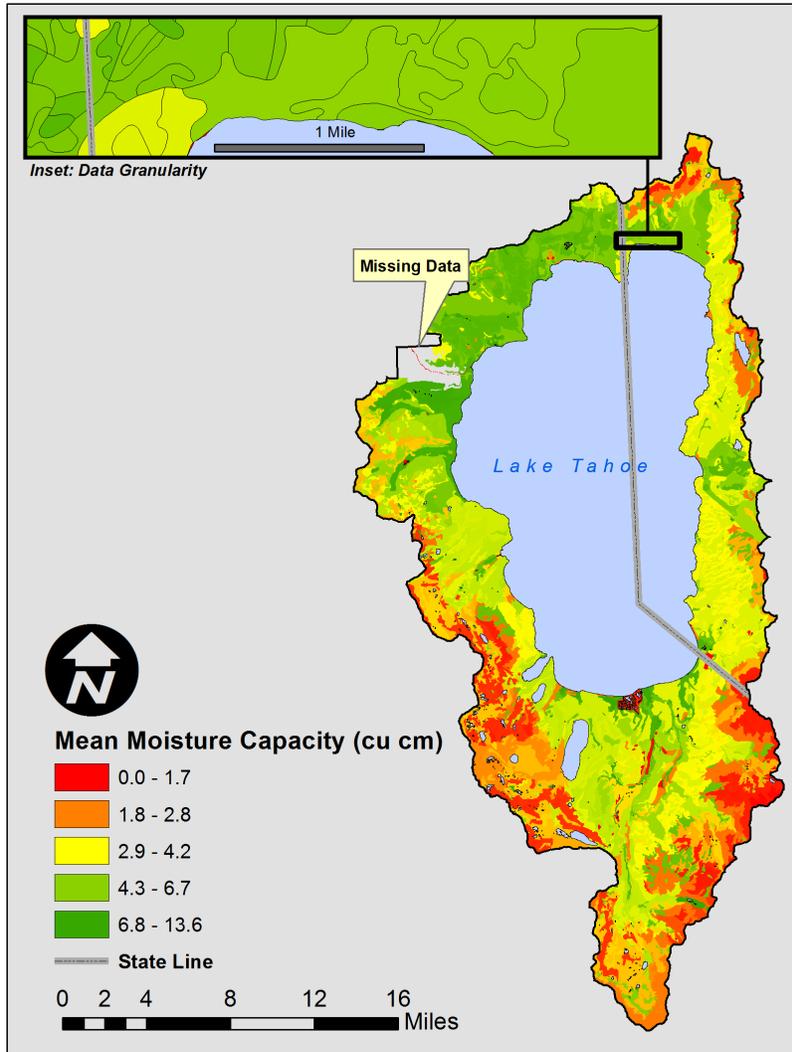


Figure 7 SSURGO Soils Data—Mean Moisture Capacity (cu cm)

The Soil Survey Geographic Database (SSURGO) is an extensive survey of nationwide soil conditions produced by the US Department of Agriculture’s Natural Resources Conservation service. The attribute of interest for this research is mean moisture capacity, and a surface symbolizing it is presented in the figure. Moisture capacity serves as a proxy for overall soil quality under the assumption that high moisture capacity corresponds to deep soils with upper horizons that contain ample organic material. The inset illustrates the typical granularity of these data, although polygon size tends to vary across the study area. The database is incomplete in one area of the Basin. That area is noted in the figure and, for consistency, all datasets were trimmed to exclude that area.

The data are collected by walking teams that physically survey soils across the country. Their data are compiled, and soil scientists establish polygons (map units) based on areas with similar

characteristics. The database is extensive—the metadata document outlining its attributes is approximately 250 pages long.

- *Size:* The dataset contains 3,978 discrete polygon objects
- *Scale:* The extent of the dataset covers the Lake Tahoe Basin that totals 1330 sq. km. Polygons in the dataset vary from 1698m² to 497947729m². The general size of these polygons is illustrated by the inset of Figure 7.
- *Aggregation:* These field data are aggregated into polygons. The intent is to keep the map units uniform, but averaging is inevitable with such a high number of attributes.
- *Attributes:* The dataset contains hundreds of attributes. In most instances, these are nominal data that correspond (sometimes literally) to the field notes of data collectors. This analysis requires ratio or interval data that can represent soil quality in a regression equation. Each polygon is assigned the weighted average of field-recorded moisture capacity readings (.01 cu cm precision) within the polygon. Moisture capacity correlates to organic content and nutrients and is used to represent soil quality for analyses. Note that polygons in Figure 7 are symbolized by their actual values; however, the legend is truncated to five classifications for brevity.
- *Fuzzy/Crisp:* These data are undoubtedly fuzzy based on many of the factors previously discussed. Soils typically merge gradually from one form into another over geographic space. A notable exception may occur around geologic features such as fault or fracture zones. These areas can form boundaries of distinctly different geologic formations.
- *Error:* Error in these data must be regarded as high. Scientists interpret notes from field workers and establish subjective areal boundaries from afar. The sheer scope of the field work required to produce this dataset ensures some level of error.

- *Fitness for use:* This dataset is produced by the US Department of Agriculture and is the national standard for soils data. Still, this is likely the most troublesome dataset involved in this analysis. One relatively minor problem is that there is a data gap in the northwest section of the study area (highlighted in Figure 7).

3.1.5 Tahoe Area Weather Stations

A shapefile compiled by the University of Nevada, Reno (UNR) Desert Research Institute Climate Center containing the geographic positions of weather stations around the LTB that maintain web-accessible temperature and precipitation data was used in this study (Figure 8). Location data from the shapefile were joined to a spreadsheet containing mean annual temperature and precipitation data from each of the stations. These data were imported into a geodatabase and used to interpolate microclimate surfaces for the LTB.

- *Representation:* The weather station dataset is a vector point dataset, so the data are discrete. The interpolated surface is a continuous raster derived from the point precipitation attributes
- *Size:* The dataset contains 29 discrete points representing mean annual temperature and precipitation.
- *Scale:* The extent of the dataset has a minimum bounding rectangle of 1750 sq. miles. Distances between points range from 1 to 58 miles and average 22.4 miles.
- *Aggregation:* The point data are not aggregated by area; the climate data were all recorded at their respective coordinates.
- *Attributes:* Each point contains two pertinent attributes. Mean annual temperature (time span varies—see error information) is interval data measured in degrees Fahrenheit and mean annual precipitation is ratio data measured in inches.

- *Fuzzy/Crisp* The point data are spatially crisp—their exact location is known. The climate attributes are fuzzy in that they are means that represent a temporal aggregation. Their values are representative, but they do not indicate a specific point in time.
Error: The attribute error may be significant. First, the historical data do not cover identical periods of time. Second, the data are collected and reported by differing entities and methods. Some of the data are volunteered by private individuals, so it is unknown if the instruments used are professional quality or if the data are accurately reported. Finally, some stations report summary statistics while others provide a report of daily numbers that may vary in completeness.
- *Fitness for use:* The purpose of this dataset is to create interpolated surfaces that represent microclimates across the Basin. Stations outside the study area were purposely included to mitigate edge effects. Unfortunately, areas to the south and west lack enough accessible data caches to produce complete coverage of the study area perimeter. The surface in Figure 8 was created for exploratory purposes using the inverse distance weighted technique. The resulting surface reflects the expected precipitation pattern with higher precipitation amounts on the western edge of the LTB and smaller amounts near the Carson Range along the eastern edge of the Basin.

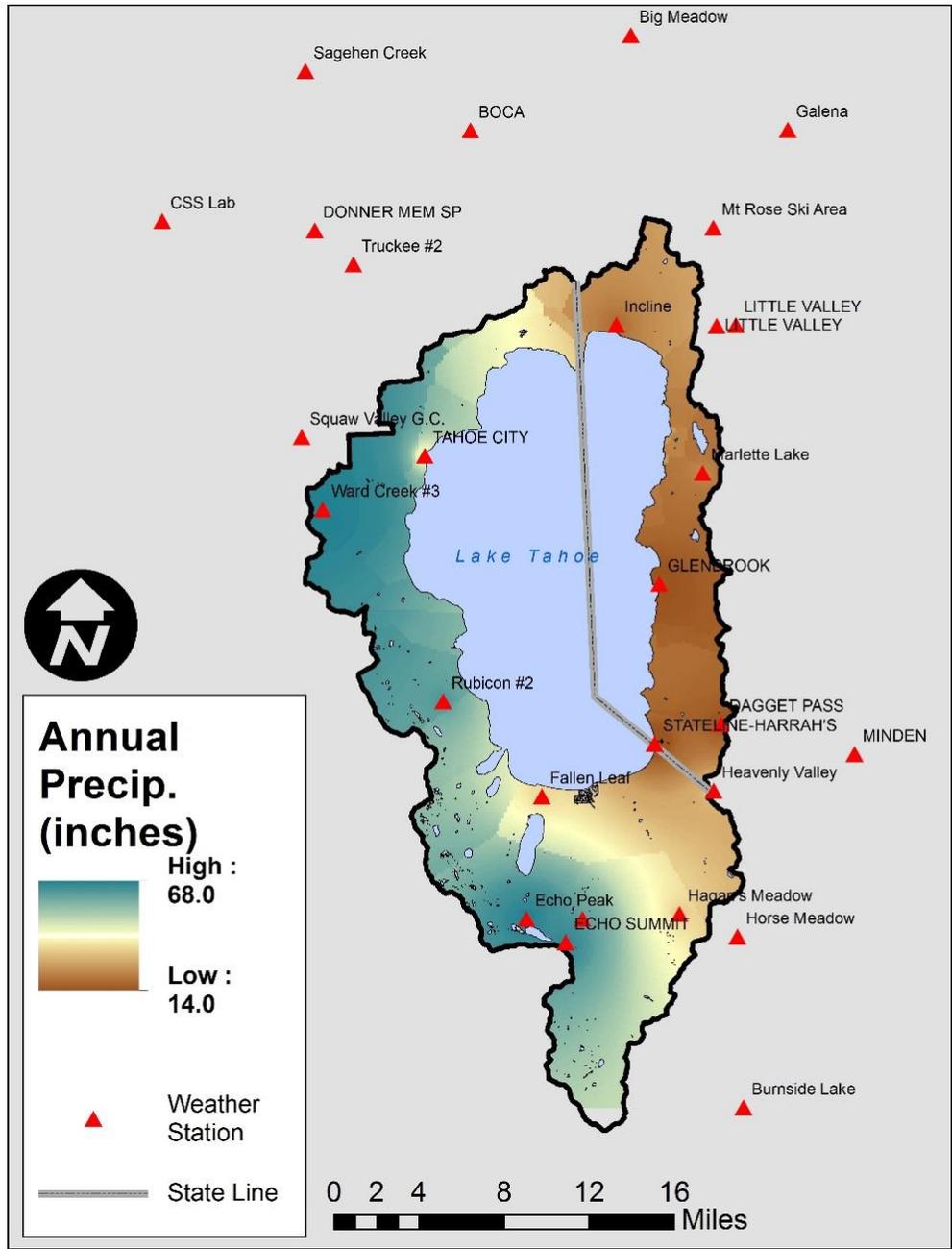


Figure 8 LTB Weather Stations and Preliminary Precipitation Surface

This figure illustrates the location of weather stations whose data were used to compute precipitation and microclimate surfaces for the Lake Tahoe Basin. Where possible, weather stations located both inside the study area and around its periphery were used to interpolate the surfaces in order to mitigate edge effects. The colored surface in the figure is a preliminary surface constructed using the inverse distance weighted interpolation technique to test the sufficiency of the collected weather station data. The surface generally mirrors the known climate pattern of the LTB and therefore, the data was deemed adequate for performing a more complex kriging procedure. Note that the interpolation did not reach the far southern tip of the Basin. This is acceptable given no land use data is available for this area and all datasets were clipped to exclude this part of the study area.

3.1.6 TRPA Land Use and Completed USFS Prescriptions

This dataset delineates land use zones as defined by the Tahoe Regional Planning Agency (TRPA) as well as areas that received prescribed forest burns or thinning (Figure 9). These data summarize the varying levels of management intensity around the Basin

- *Representation:* The dataset is a vector polygon model; the data are discrete.
- *Size:* The dataset contains 308 discrete polygon objects; however, these data were be dissolved to combine similar classes such as “backcountry” and “conservation”.
- *Scale:* The extent of the dataset covers most of the Lake Tahoe Basin with the exception of a 15.5 sq. km. portion of Alpine County for a total of 1313 sq. km. Polygons in the dataset vary from 6000m² to 90 km² (wilderness area)
- *Aggregation:* The polygons represent boundaries assigned by a government agency. They do not aggregate data.
- *Attributes:* The key attributes of this dataset are nominal and name the intended land use within each area. The prescriptive treatment polygons contain nominal attributes that indicate if the treatment was a thin or a burn.
- *Fuzzy/Crisp:* These data are truly crisp. They do not represent a natural phenomenon that is dynamic or loosely defined. The polygon boundaries are exact and will not change unless a government authority alters them.
- *Error:* There should be little error in these data because the authority that created them digitized them. The caveat to this is that the dataset must be properly digitized. Several polygon slivers were noted while exploring this dataset. A geodatabase topology identified and corrected those errors.

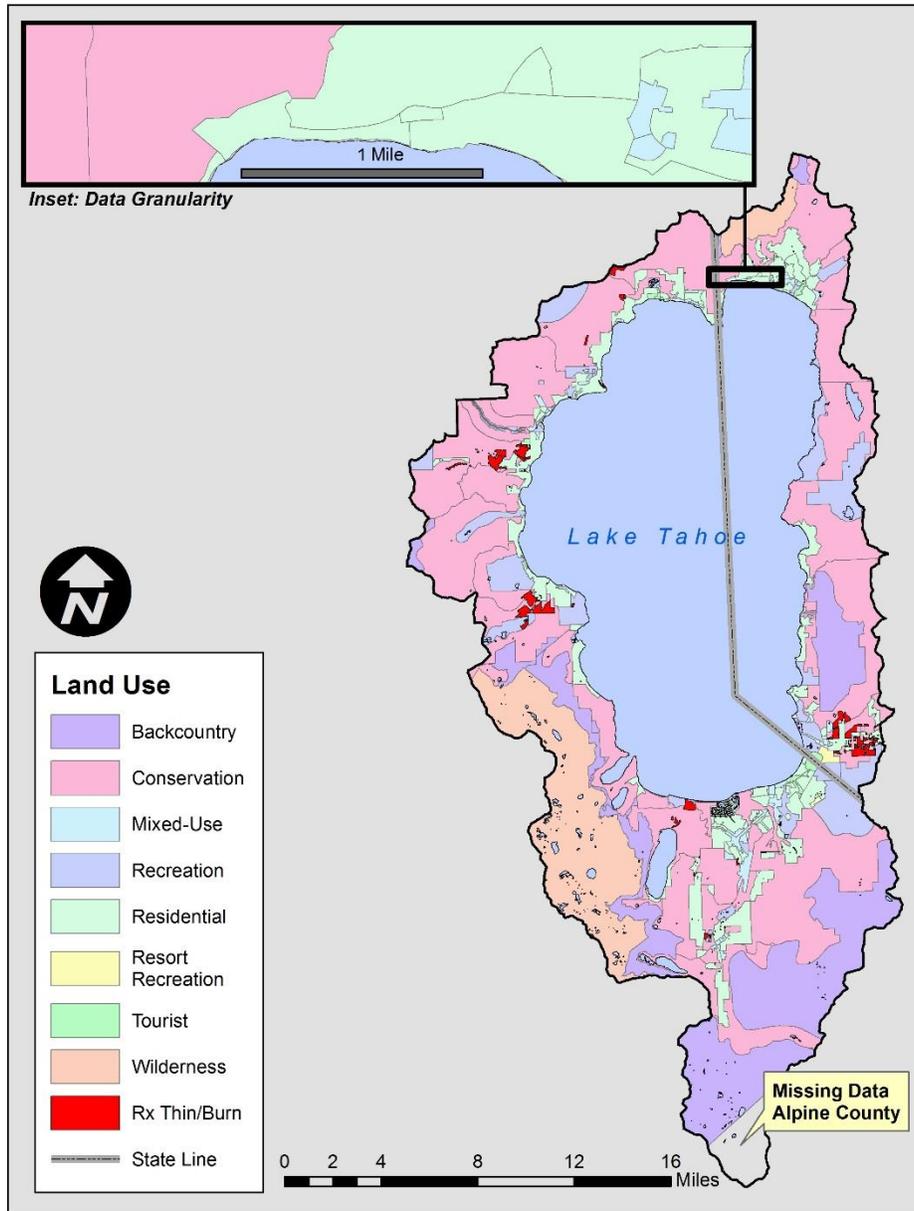


Figure 9 TRPA Land Use and Prescribed Treatments

This figure presents land use designations as assigned by the Tahoe Regional Planning Agency. These data were consolidated to ease data processing and represent discrete levels of management impact in the species richness models. Also depicted are specific instances of prescribed thinning or burning as of the date of data collection. Ultimately, these data were not included in the species richness models as regression analysis did not indicate they were a leading factor in predicting species richness. Similar to previous figures, an inset is provided to indicate typical granularity for these data. Lastly, no land use data for the far southern Alpine County section of the study area was available. This area was excluded from the analyses.

- *Fitness for use:* These data provide the spatial extent of forest management technique.

These techniques are reliant on land use, however, many land uses may relate to a given forest management policy. The data were further aggregated into three zones of management style, urban, heavy (ordinary forest management) and light (wilderness or conservation zones).

3.2 Data Processing

The raw data described in section 3.1 were further processed for use in the various analyses associated with this research. ArcGIS tools were used to manipulate the data. Minor processes such as creating slope or aspect surfaces are not described. The two major processing tasks were to create microclimate surfaces and consolidate the data into a single point dataset for exploratory regression.

3.2.1 Precipitation Surface Creation Using Empirical Bayesian Kriging (EBK)

A fine-grained precipitation surface was required for exploratory regression and as a Maxent environmental input layer. Empirical Bayesian Kriging (EBK) was used to interpolate this surface from the UNR Tahoe area weather station data. Kriging is a powerful interpolation technique, but it requires several data assumptions to create an accurate surface. Specifically, the interpolated data must be non-clustered, stationary, normally distributed and autocorrelated.

Several diagnostics were utilized to evaluate the weather station data for kriging suitability (Figure 10). In the top panel, the Voronoi map indicates the data are *mostly* stationary. The middle panel is a semivariogram that indicates the data are autocorrelated with a range of about 15 km. This is expected because weather data are typically autocorrelated. In the

bottom panel, the data appear normally distributed with several outliers in the wettest part of the spectrum. These discrepancies may be attributed to measurement error (i.e. amateur collection or instrument calibration) coupled with a small sample size.

EBK accounts for uncertainty by simulating many semivariograms. Further, moderate non-stationarity is tolerated by creating local models within subsets of the data.

The final surface depicts precipitation patterns very well (Figure 11). A visual inspection indicates that the surface reflects known precipitation patterns in the Tahoe Basin with the western Sierra Nevada range causing rain shadow and decreased precipitation in the eastern Basin. An in-depth, quantitative evaluation of the EBK results is presented in Chapter 4.

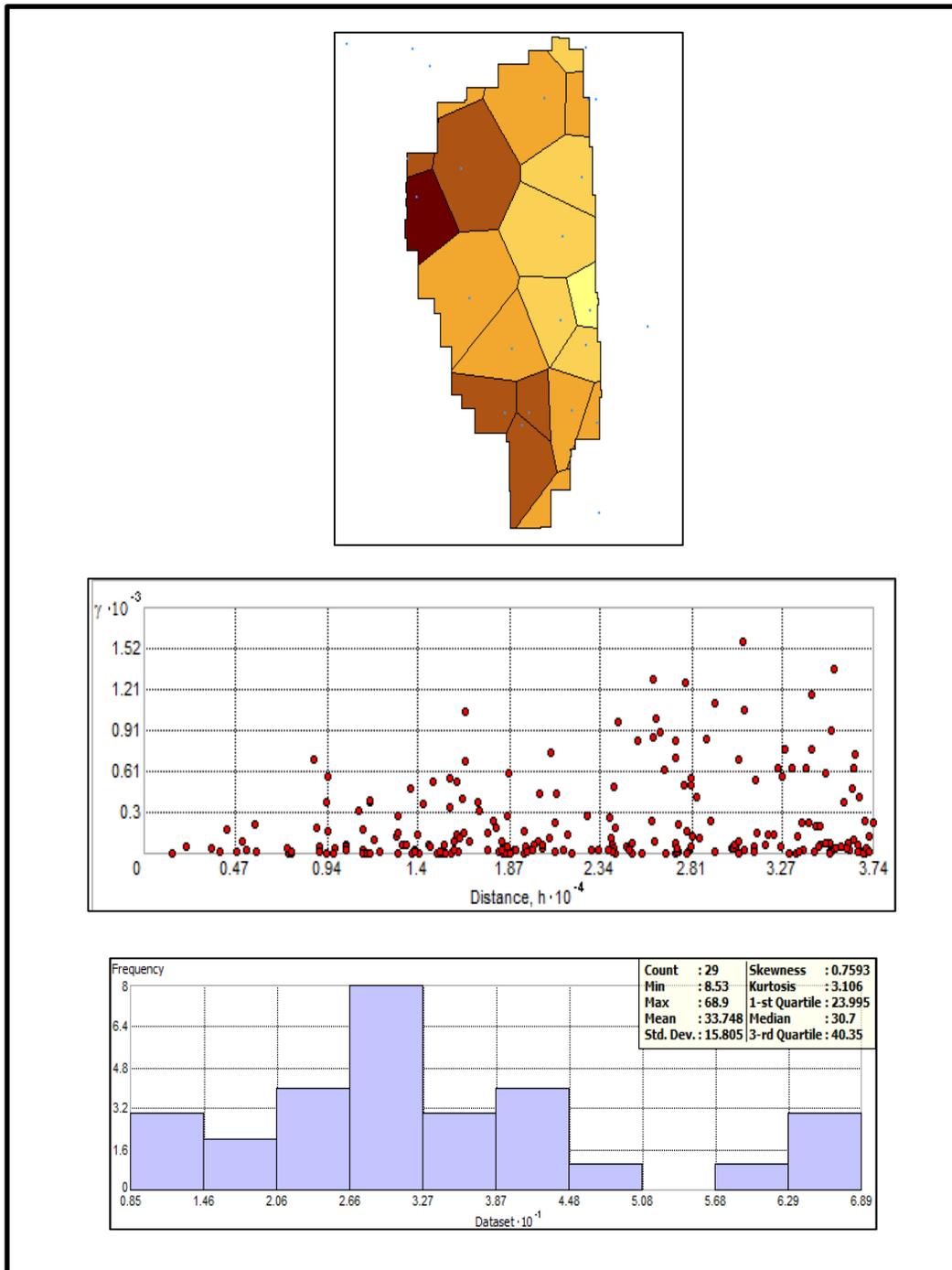


Figure 10 Pre-Kriging Data Analysis (Stationarity, Autocorrelation, Normality)

This figure provides an analysis of mean annual precipitation data gathered from web-accessible weather stations to measure their compliance with the kriging assumptions of stationarity, autocorrelation and normality. In all three cases, the data likely meet the assumptions; however, the figures are not convincing or conclusive given the small sample size of the precipitation data. Given these diagnostics Empirical Bayesian Kriging (EBK) was selected to create the precipitation surface. EBK is considered less reliable than traditional kriging; however, it is more robust to deviation from data assumptions.

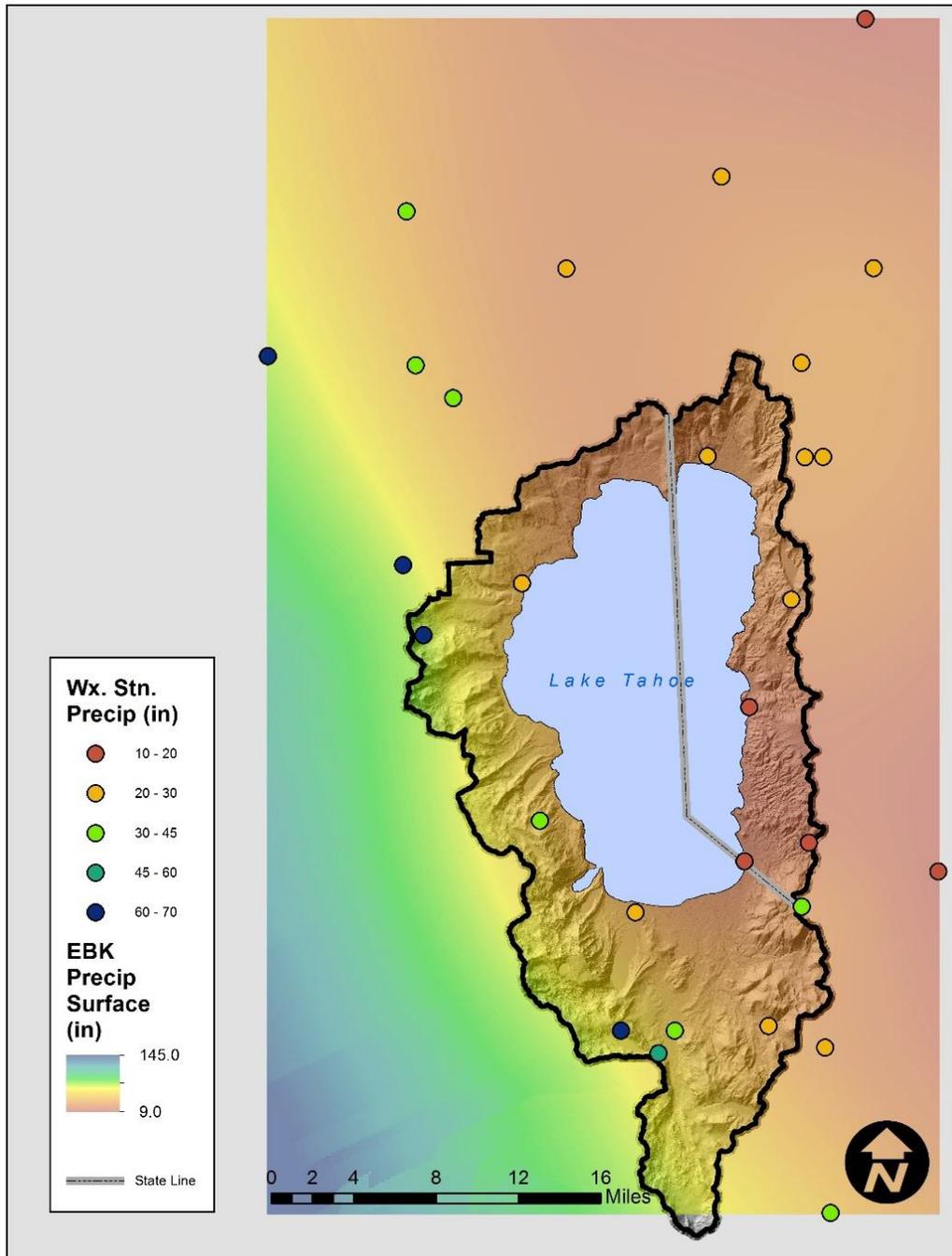


Figure 11 EBK Precipitation Surface (Inches of Precipitation-Annual Mean)

The final EBK precipitation surface provides an excellent representation of precipitation patterns in the Tahoe Basin. The Sierra Nevada crest (the Basin's western border) is highly effective at stalling northwesterly Pacific storms creating a rain shadow effect across the Basin. The Carson Range forming the eastern edge of the Basin prevent remaining moisture from reaching the Great Basin valleys to the east. The surface and point data accurately reflect these effects.

3.2.2 Data Consolidation

TEBVM was converted into a point dataset by calculating the centroid of each polygon and assigning the associated attributes to it. Interestingly and contrary to species-area theory, these species richness data do not strongly correlate with area (Arrhenius 1921). An ordinary least squares regression of species richness to area yielded an adjusted R^2 of only 0.03. Log-Log transforming the data only marginally improved this value to an R^2 of 0.13. This is not entirely surprising given the construction of species-area curves requires specific sampling techniques such as nested transects (Scheiner 2003). These data are intentionally biased; the TBEVM dataset authors delineated areas by encapsulating homogenous vegetative communities into polygons. Conversely, sampling techniques designed to construct species-area curves expect to capture an increasing number of homogenous communities with each widening transect. (Scheiner 2003).

TEBVM was converted into a point dataset by calculating the centroid of each polygon and assigning the associated attributes to it. This feature class along with all of the required raster datasets were placed into a geodatabase along with a copy of the TBEVM centroids (without attributes). The ArcGIS model builder tool (Figure 12) extracts raster values that underlie the non-attribute TBEVM “mirror” points and records them as attributes in an intermediate point dataset. The intermediate points were then spatially joined to the TBEVM centroid dataset. The output of this tool is a feature class that contains all the TBEVM attributes plus attributes extracted from the raster datasets. The consolidated dataset was then scrubbed for null values. These records, representing data gaps in the source datasets were removed to prevent sampling bias in the various analyses. The final dataset contains approximately 17,000 points.

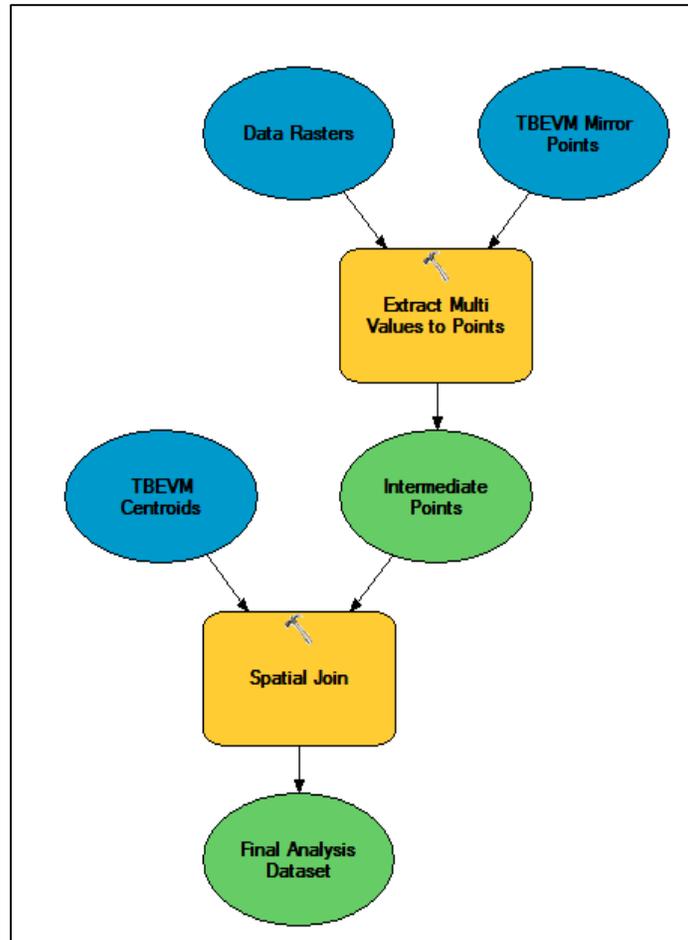


Figure 12 Data Consolidation Tool (ArcGIS Model Builder)

3.2.3 Tree Species Richness Attribute Calculation

TBEVM does not contain a species richness attribute. Several rows from the TBEVM attribute table (Table 2) illustrate how this value was calculated. Columns in the left portion of the table represent percent coverage of each tree species in the TBEVM polygons. To calculate tree species richness, counts of any species coverage other than zero for each row were tabulated and recorded in the far right column (Tree_specrich).

Table 2 Species Richness Calculation

<u>ABCO</u>	<u>ABMA</u>	<u>JUOC</u>	<u>PIAL</u>	<u>PICO</u>	<u>PIJE</u>	<u>PIMO</u>	<u>POTR</u>	<u>TSME</u>	<u>Tree specrich</u>
0	0.00543	0	0.2	0.3	0.13	0.12	0	0.09	6
0	0.04	0	0.23	0.13	0.08	0.09	0	0.00455	6
0	0	0	0.1	0	0	0	0	0.01	2

3.2.4 Data Split by Management Intensity

Initial attempts at exploratory regression using the consolidated data set resulted in highly skewed residuals. Martin-Queller (2011) reported stationary relationships between species richness and management effects in their GWR analysis. Given this, the consolidated dataset was split using the ArcGIS split tool according to the categories (Table 3) to account for data variation by management category.

Table 3 Forest Management Categories

<u>Land Mgmt. Category</u>	<u>Land Use Categories</u>
Urban	-Residential -Tourist -Resort -Mixed use
Heavy Mgmt.	-Backcountry -Recreation
Light Mgmt.	-Wilderness -Conservation

3.3 Models and Analyses

This section covers the various models and analyses produced to support this research. They serve three purposes. First is to derive the most appropriate explanatory variables for species

richness. Second is to produce the species richness border region model itself. Lastly, several models validate the border region model.

To be clear, the ultimate objective is two final models. The species richness-derived model (species richness derived from macroenvironmental factors and displayed as border regions) and the species richness model (traditional Maxent SDMs, stacked and summed to display species richness). These models validate each other and also provide an opportunity to evaluate the relative benefits of each approach.

3.3.1 Exploratory Ordinary Least Squares Regression

The ArcGIS exploratory regression tool was used to identify the best specified explanatory variables for tree species richness in the Lake Tahoe Basin. This technique was chosen as it provides an efficient method for sorting the numerous potential independent variables identified in the literature and compiled as attributes in the consolidated point feature class. The primary limitation of this method is that the ArcGIS tool only considers linear distributions. Non-linear regressions, such as logistic commonly used to fit ecological data are not supported.

A further complication of using a regression technique in this manner is that the desired independent variable, tree species richness, are count data with a very small range. Initial regression attempts using these were fragmented and unsatisfactory.

As a proxy for species richness, the continuous Simpson Diversity Index was calculated for tree species according to Equation 1 below where n_i is the number of individuals of each species and N is the total number of individuals of all species. The summation is across all species present (Simpson 1949).

$$\sum \left(\frac{n_i}{N}\right)^2 \quad (1)$$

TBEVM does not include data for number of individuals; however, estimates were calculated from the available data. Total area for each polygon was converted to hectares and the resulting value was reduced to equal the percent of each polygon that contains tree cover. Total individuals were then estimated by multiplying stem density per ha by hectares of tree cover. Coverage for each species is represented by percent of total area. A similar process was conducted for each species present to estimate the number of individuals per species. Clearly, the true stem density for each species may vary from the polygon means reported by TBEVM. Consequently, the resulting individual data certainly contains error. Short of conducting an extensive field survey, however, these data represent the best available estimates.

As mentioned in the previous section, categorical management effects were accounted for by splitting the dataset and performing separate regressions for each management category.

The exploratory regression tool iteratively performed regressions on combinations of all the attributes selected from the consolidated dataset. The tool automatically checks for issues such as multicollinearity and non-normally distributed residuals and then outputs a report indicating the best performing and best-specified independent variable combinations according to its statistical diagnostics. Residual plots for the best performing solution were checked for randomness and normality around the regression model. The independent variables for this regression model were further evaluated by checking for their use in related literature.

3.3.2 Species Richness-Based Maxent Model

This model is the primary model evaluated by this research. The typical use of the Maxent software is to input point occurrences of specific species to create a species distribution model (SDM). For this application, point observations of species richness counts were substituted for species data to produce predictions of species richness clusters. The goal of the model is to plot the location and dimensions of border regions separating homogenous patches of species richness. As mentioned by Kark and van Rensberg (2006), the abruptness of change that occurs within border zones is also a significant property.

To use Maxent, data that will be input into the application must be pre-processed to meet strict requirements (Figure 13). Raster environmental layers must be precisely coincident and cover exactly the same extent. Once coregistered, the raster data must be exported into an ASCII grid format. Presence-only (species) data are matched with X-Y coordinate data that coincides with the environmental raster data. These data must be exported into a comma separated value (csv) format. These tasks were accomplished using the Excel spreadsheet program.

As a black box-type modeling application, the heavy number crunching is done behind the scenes by Maxent's internal maximum entropy algorithm. Despite this, a full suite of user selectable parameters that affect Maxent's output is available. Some of those option choices and their rationale are listed below. Specific details regarding the effects of each parameter are as described by Merow (2013):

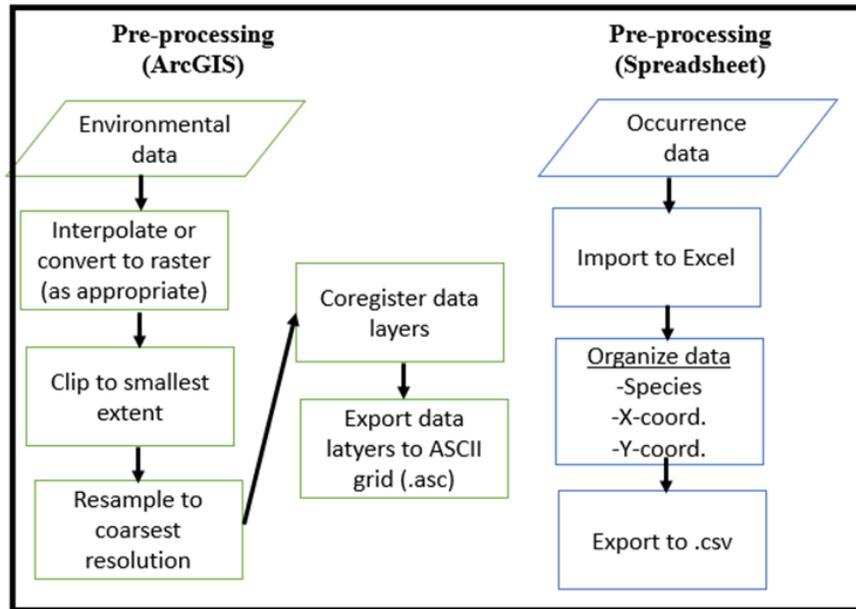


Figure 13 Maxent Pre-Processing using ArcGIS and Spreadsheet Applications

- *Output format*—The cumulative output option was selected given it is designed to produce crisp range boundaries and avoids using parameters not derived from the input data
- *Feature selection*—This selection allows for specifying the regression style for each input parameter. The specific distribution of each parameter is unknown, so the auto option was selected which allows Maxent to calculate the appropriate feature type for each input variable. Note that this relates to the “feature” discussion in Chapter 2. Feature type in this context relates to the appropriate distribution type of the input environmental data (e.g. linear, product, logistic, etc.), not the spatial data model involved. All environmental data are in a raster format.
- *Jackknife analysis*—Jackknife output was used to verify that each environmental variable is statistically significant within the model.

- *Threshold*—The threshold setting determines how Maxent evaluates the receiver operating curve (ROC) to determine if a cell is within or outside the calculated range. There are a number of possible settings; however, in this case an automated sensitivity analysis was desired. The equal specificity vs. sensitivity option ensures the output ROC value is no more sensitive than it is specific. In other words, some accuracy of each cell's prediction was sacrificed in order to create a stable (less sensitive to variable changes) model. Further, choosing a threshold option produces a binary raster (in addition to the range probability surface) where 1 is range and 0 is not range based the selected threshold option. This option is beneficial for removing bias as the decision of what ROC constitutes range is determined by the data as opposed to the user (e.g. arbitrarily selecting a 0.8 ROC cutoff).
- *Regularization*—This input allows the user to adjust the magnitude of Maxent's complexity penalty. The literature varies on how this option should be approached. Some advocate a default setting given the Maxent algorithm is highly effective. Others use an iterative approach to maximize the output kappa statistic similar to the method of using peak autocorrelation (e.g. Moran's I) to select an appropriate distance band in other spatial analyses. In this case this option was varied over several model runs to overcome Maxent's tendency to overfit presence data (Figure 14). An overfit model will display "halos" around presence data, while excessive regularization will be "flat" with little relief. The chosen models were the best subjective compromise between these extremes.

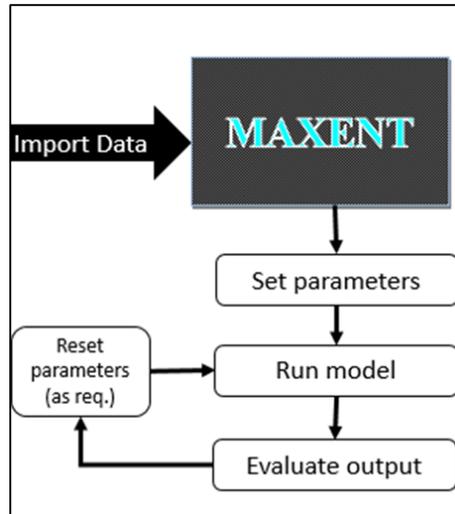


Figure 14 Maxent Iteration

- Test point percentage*—To construct model fit diagnostics, Maxent extracts a portion of the input point occurrences to validate the model. The user is charged with determining the ratio of test to model training input. There are ~16,000 input occurrence points in the TBEVM dataset which benefits the model given the Lake Tahoe Basin landscape is highly varied. The bootstrap sampling method was selected as it is a remove and replace technique that requires extra processing time because test points used for diagnostics are subsequently replaced and used to further train the model. Graham and Hijmans (2009) applied 25% of their presence data for test purposes. This study increased that figure to 30% test vs. training to ensure effective diagnostics, particularly bootstrap output, and to prevent any training degradation as a result of this high ratio.
- Bias layer*—A bias layer is an allowed input to account for sampling bias such as inaccessible terrain and proximity to roads. This layer was excluded from the analysis given the presence data are derived from IKONOS remote sensing imagery. Remotely sensed data provide the advantage of uniform coverage and avoid the inherent field-collection biases this input is designed to correct (Prates-Clark et al. 2008).

Validation of the Maxent output was accomplished via an evaluation against local knowledge and Maxent's built-in Area under curve (AUC) diagnostic. AUC measures the model's performance versus random by plotting test data, actual ROC and a random prediction. A value of 0.5 means that the model is predicting ranges as well as a random distribution. Something less than 0.5 indicates that the model is performing *worse* than random. Ideally, the actual ROC curve should peak in the upper left corner of the plot. A value of 0.7 is considered adequate for most applications (Graham and Hijmans, 2006). An evaluation against local knowledge also verified that richness predictions coincided with expected output. For instance, sub-alpine regions ordinarily harbor only two of the represented species. High elevation areas were checked for consistency with that knowledge.

To create a representation of border regions, Maxent species richness predictions were post-processed in ArcGIS (Figure 15). Choosing a threshold option in Maxent produces binary "range" or "not-range" rasters for each species richness value (8 total). Pixel values of "1" representing predicted range were reclassified to their respective species richness values. Next, the ArcGIS raster to polygon tool was used to create vector polygons that cover the extents of each species richness range. These polygons were then intersected to create a vector polygon layer that represents only those portions of the study area where species ranges overlap. This vector layer was used as a mask to extract values from each reclassified raster layer. This procedure trims the reclassified range rasters so that only pixels that Maxent classified as range for two or more species richness values are represented (i.e. border regions).

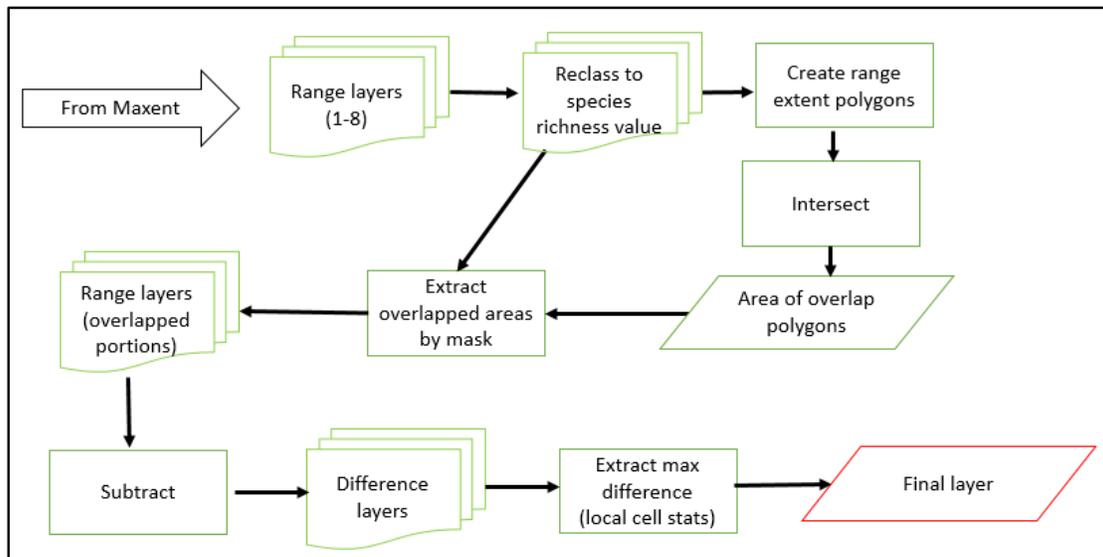


Figure 15 Post Processing—Species Richness Method (ArcGIS)

A benefit of using species richness as a Maxent species is that species richness categories are actually meaningful values. This property makes the abruptness of a border calculable by subtraction. For example, a border separating 6 species from a community of 2 species is much more abrupt than a possibly imperceptible shift from 3 to 4 species.

The abruptness value was extracted by stacking and subtracting the trimmed and reclassified range rasters. This was done iteratively to create a series of difference layers (e.g. 1–8 species, 2–8 species, 3–8 species. . .). The maximum difference value for each pixel was then extracted and displayed. The resulting layer is a depiction of transition zones between homogenous islands of species richness classified by abruptness.

3.3.3 Species Range-Based Maxent Model

The stated purpose of Maxent is to predict the ranges of specific species. Given this, an additional model deriving species richness data from single species range predictions was constructed—a stacked SDM (Figure 16). If predictions are significantly similar between the two models, the validity of predicting species richness with Maxent is enhanced. Note that this

method does not specifically address border regions as does the species-richness based model. This is beneficial as it serves to illustrate how patterns differ between a full representation of species richness versus border regions only.

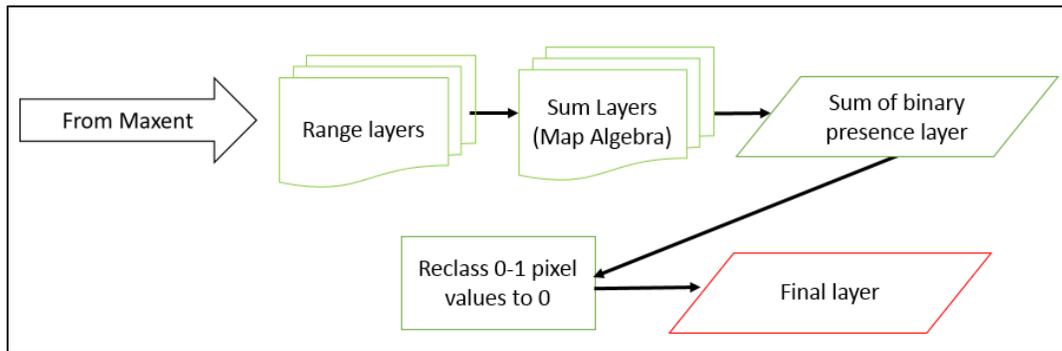


Figure 16 Post-Processing--Species Range Technique (ArcGIS)

Pre-processing and Maxent options are identical to the species richness technique. Post-processing differs in that species richness is calculated by summing number of species predicted to occupy each pixel. Pixels valued one in the resulting surface are reclassified to zero to reduce clutter on the surface and deemphasize areas with little or no diversity.

3.3.4 Known Species Location Versus Predicted Species Range Map

While Maxent is widely regarded as a valid tool for creating SDMs, the output of any model must still be validated; therefore, a known species location overlay was used to validate the Maxent output for each individual species (i.e. the individual layers stacked and summed to create a species richness model). For each species, TBEVM points where that species exists were overlaid onto its SDM. Ideally, these points will fall within the extent of Maxent range predictions.

If the Maxent species ranges coincide with TBEVM dominant species, the credibility of the Maxent prediction is enhanced.

3.3.5 Split Moving Window Dissimilarity Analysis (SMWDA)

The final validation model to support this research is an SMWDA analysis of the species richness boundary zones identified in the Lake Tahoe Basin using the modeling technique. The specific purpose of this analysis is to verify that the plotted boundary zones meaningfully represent a transition between species richness patches rather than mere variance in the predictive power of Maxent.

Hennenberg (2005) discusses several quantitative techniques for detecting borders and ecotones. Most of these, however, are multivariate. SMWDA as presented by Cornelius and Reynolds (1991) was selected for two reasons. First it is valid for univariate (and multivariate) analyses. Univariate analysis is desired for this application because it will isolate discontinuities in species richness only, eliminating ambiguity in the results. Second, while it is a moderately complex iterative procedure, its mathematics consist of common statistical concepts no more complex than summations, means and z-scores. Importantly, it does not require specialized computer applications or programming. It is computable with a carefully constructed spreadsheet.

The method described by Cornelius and Reynolds places a window along a series of ordered data (e.g. a transect). The window is split into two halves and averages for the variable(s) are calculated for each half. An index of dissimilarity is calculated to quantify the dissimilarity between the window sections. The window is shifted one plot and the procedure is repeated for the entire series. The expected mean and standard deviation are determined via a Monte Carlo procedure (1000 random iterations in this case). Z-scores for the dissimilarity index within the data are calculated and plotted. Significant z-score spikes indicate the presence of discontinuities in the data (interpreted as the presence of a border or ecotone). The scale at

which discontinuities occur can be detected by varying the size of the window. The specific mathematics are described by equations 2 through 7 below.

For this application, a series of TBEVM centroids that pierce or pass through plotted transition zones were extracted for analysis. If the plotted border regions represent true transition zones, discontinuities in the plotted z-scores (peaks) should correspond with the model-predicted border regions.

The primary issue with this technique is that the TBEVM plots are not evenly distributed as in field collected transects using evenly spaced quadrats along a compass bearing. This only invalidates the technique of varying window size to estimate scale given the number of plots along the data series will not correlate to a consistent Euclidean distance. There will likely also be some distortion in the shape of curves formed by a series of varying dissimilarity scores. This is inconvenient, but not fatal as the primary objective is to compare peaks with their associated window midpoint.

Chapter 4 presents the results of these analyses.

Upper Window Average
$$\bar{W}_{Ak+0.5i} = \frac{\sum_{j=k-Q/2}^k X_{ij}}{Q/2} \quad (1)$$

Lower Window Average
$$\bar{W}_{Bk+0.5,i} = \frac{\sum_{j=k+1}^{k+Q/2} X_{ij}}{Q/2} \quad (2)$$

Dissimilarity Index
$$DS_{k+0.5} = \left[\sum_{j=l}^v (\bar{W}_{Ak+0.5,i} - \bar{W}_{Bk+0.5,i})^2 \right]^{1/2} \quad (3)$$

Mean Dissimilarity
(for Monte Carlo)
$$\overline{DS}_{k+0.5} = \frac{\sum_{l=1}^M DR_{k+0.5,l}}{M} \quad (4)$$

Std. Deviation of Dissimilarity
(for Monte Carlo)
$$SD_{k+0.5} = \frac{\left[\sum_{l=1}^M (DR_{Ak+0.5,i} - \overline{DS}_{Bk+0.5,i})^2 \right]^{1/2}}{M - 1} \quad (5)$$

Dissimilarity Z-Score
(Univariate Only)
$$DZ_{k+0.5} = \frac{DS_{k+0.5}}{SD_{k+0.5}} \quad (6)$$

Where (for univariate analysis):

j = the data value

i = the position in the sequence

Q = window width

Q/2=total positions per half (even number)

k= sequential position (Q/2, Q/2+1, Q/2 +2...)

W_a=series half 1

k+0.5=window midpoint

W_b=series half 2

N= series length

v=number of measured variables (=1)

X_{ij}=i at point j

N-Q=total windows of width Q

DR_{k+0.5,i}=array of dissimilarities

M=Monte Carlo iterations (=1000)

CHAPTER 4: RESULTS

This chapter consolidates the results from each of the analyses described in Chapter 3.

Traditionally, scientific papers present results without commentary. Due to the highly graphic nature of spatial analyses, this chapter departs from that paradigm for the convenience of the reader. In cases where reference to the details of a figure aids the discussion, a brief interpretation follows the description of each figure. Broad-scale discussion and conclusions are reserved for Chapter 5. This strategy is not intended to discourage alternative interpretations or dissent; rather, the specific purpose is to provide easy reference to figures as they are discussed.

Results are not necessarily presented in the order they were produced. The subsections of this chapter are arranged to proceed logically toward the conclusions presented in Chapter 5.

4.1 Evaluation of EBK-Derived Precipitation Surface

The EBK-derived precipitation surface is an important input to both the OLS procedure and Maxent models in this thesis. As such, the resulting figure was carefully evaluated for performance and accuracy.

Parameters for the algorithm were set to optimize semivariogram fit and performance. Specifically, the thin plate semivariogram model with four sectors rotated to capture regions of significant variation was selected. No data transformation was chosen as none improved the model's fit to the data.

These parameters were adjusted dynamically in response to diagnostics within the ArcGIS EBK application. The final diagnostic output is displayed below (Figure 17). The top panel is the semivariogram used by the interpolator. Ideally, the blue semivariance averages should fit within the purple confidence envelope near the model function. The model performs relatively

well at short distances and significantly worse as distance increases. This is acceptable given weather stations were selected to provide many short distance data pairs near the study area.

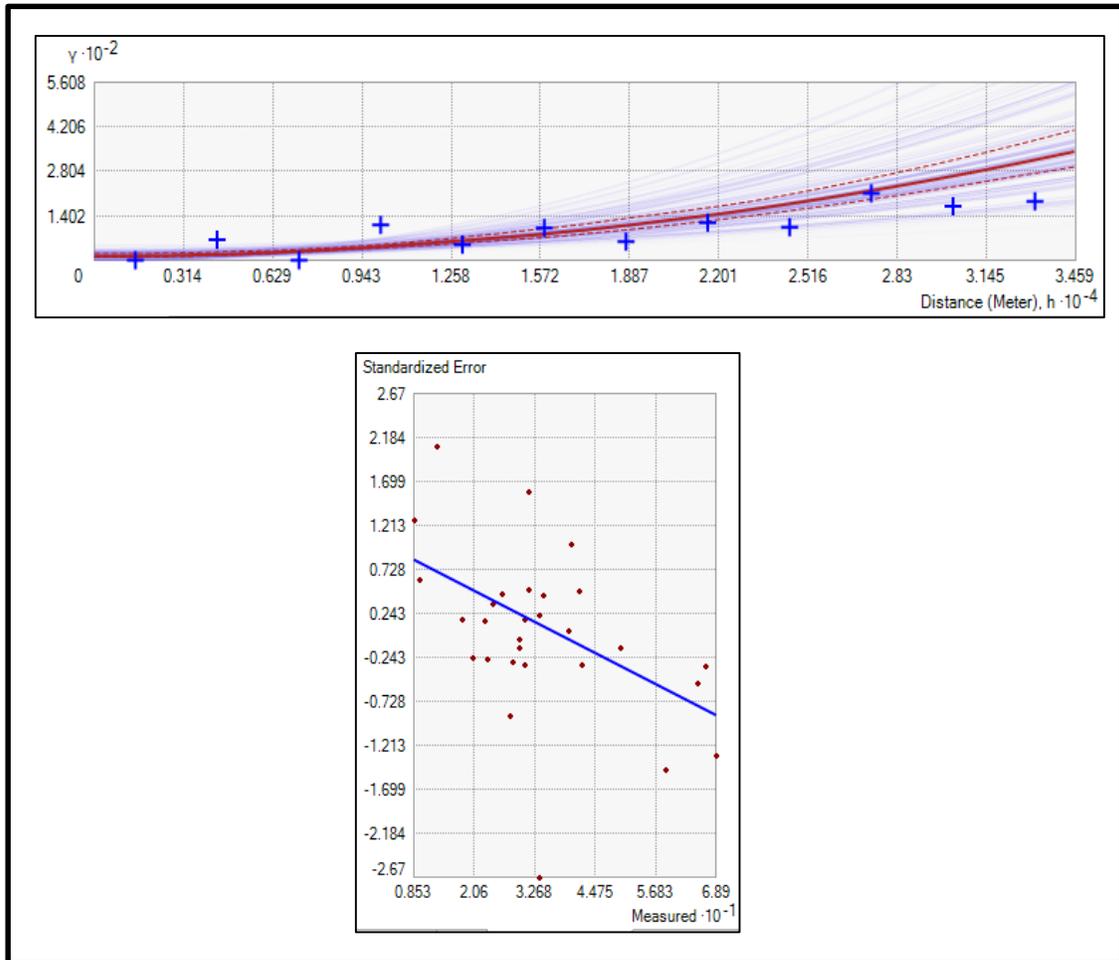


Figure 17 EBK Diagnostics (Model Fit and Residuals)

These diagnostics evaluate the output of the kriging operation for mean annual precipitation. The top panel is the semivariogram produced by the kriging algorithm. Ideally, the blue crosses should fall within the dashed confidence interval. The results indicate an imperfect fit particularly at long distances; however, error at long distances is irrelevant given data outside the study area was included to combat edge effects. The bottom panel represents error versus fits. The results exhibit the desired shotgun-style pattern, and generally follows the model although there are several outliers in the dataset.

The bottom panel is the standardized error plot for the model. The residuals appear widely scattered and follow the model indicating reasonable performance.

The quality of the kriging interpolation varies across space, and thus, prediction quality may be quantified using standard error. For the LTB precipitation surface, spatial distribution of this error was displayed as a validation surface (Figure 18). The lowest standard error values are centered on the study area due to favorable weather station distribution. Notably there are areas of higher error along the western study area fringe and in the extreme southern portion of the Basin. There are few web-accessible weather stations surrounding those (largely wilderness) areas, creating a data gap. Further, the surface does not reach the southern tip of the Basin. This is acceptable as this area (Alpine County) also lacks land use data. For data consistency, the far southern tip of the LTB was excluded from each of the models and analyses in this project.

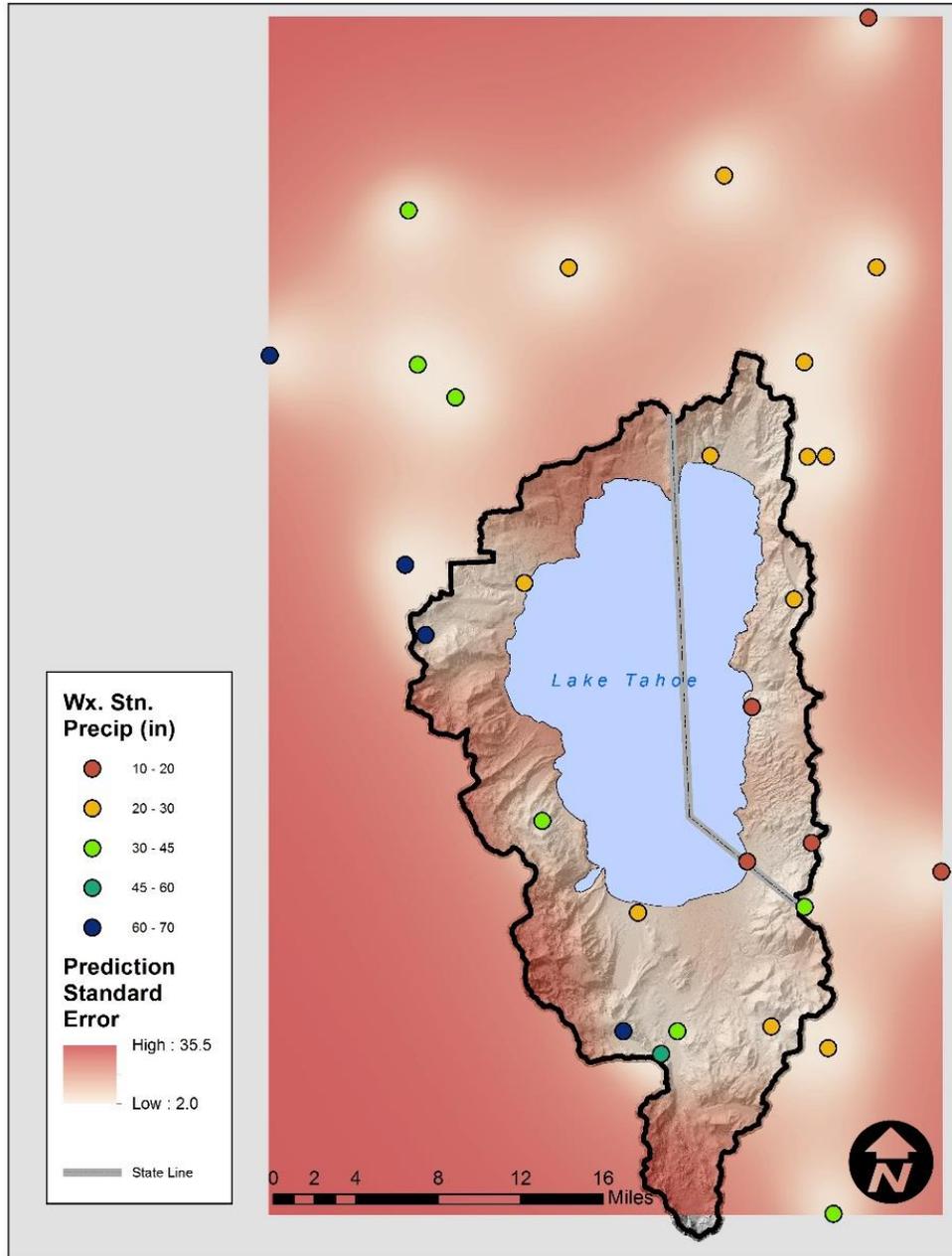


Figure 18 Precipitation Surface Prediction Standard Error

Prediction Standard Error provides a spatial estimation of the quality of a kriging prediction. In this figure, the majority of the study area is predicted to have low error although there are areas of concern. Specifically portions of the northwestern and western edges of the study area are predicted to have substantial error relative to the rest of the Basin. This is unavoidable due to a lack of accessible climate data in these areas. The far southern tip of the LTB also has considerable error, although that portion of the Basin is largely excluded from the analyses. Weather station data are included to illustrate the rate of confidence decay from known data points. Tight radii around data points suggest an overfit prediction; however, given the varied terrain of the LTB, tight predictions are preferred versus predictions for a flat, homogenous landscape.

4.2 Selection of Environmental Variables

Ordinary least squares regression analysis was performed to identify a global model that best describes species richness around the Basin and choose the most appropriate environmental variables for inclusion in the Maxent models. An initial exploration of 14 potential explanatory variables (identified in the literature) was undertaken using the ArcGIS Exploratory Regression tool. The Simpson Diversity Index was used as a continuous proxy for species richness and the dependent variable for the regression analysis.

Several high R^2 model combinations were returned, but no model was properly specified due to skewed (Jarque-Bera statistic $p < .05$) or autocorrelated (Morans I statistic $p < .05$) residuals.

Variables with coefficients that scored as significant for less than 95% of the model combinations were discarded. Value distribution histograms for the remaining variables were examined and most were not normally distributed. The majority of the significant variables mirror the variables used by Martin-Queller (2011) and discussed in DeClerck et al. (2006) and Adams (2009). These factors (Table 4) were transformed to better approximate a normal distribution and meet the assumptions of OLS regression.

Table 4 Environmental Variables Selected from Exploratory Regression

Variable	Transformation (for OLS)	Discussed in literature?	Eco Significance
Basal Area	Log	Y (Adams, DeClerck)	The area of ground physically occupied by trees. An important measure of productivity and also human disturbance (cutting/fires/thinning = less BA)
Soils (Moisture Capacity)	Log	Y (Adams)	Source of nutrients/competitive resource
Slope	Log	N	Niche delineator
Elevation	Log	Y Adams, Martin Queller)	Niche delineator-defines biome/community separation
Precipitation	Mean Difference	Y (DeClerck)	Moisture/Snowfall define niches and is competitive resource
Aspect	None	Y (Adams, Martin Queller)	Defines available photosynthetic energy (sunlight)
Stream Environment	Log	N	Local riparian communities are structured significantly different than other communities

Although the selected variables were highly significant, the misspecification of the model due to significant Jarque-Bera and Moran's I statistics was concerning. The model was troubleshot by including factors (individuals data) used to compute the dependent variable (Simpson Diversity Index) with the array of selected variables. The assumption was incorporating factors used to compute the dependent variable along with extremely significant environmental factors provided the best opportunity to produce a good fitting model. This was not the case statistically. Again, the best specified models had high R^2 values (~0.71), but each failed for skewed and autocorrelated residuals. Given these results, it is unlikely that a perfectly specified global model (using statistical diagnostics) is possible with these data (due to non-linearity or excessive outliers).

Due to these results, the data were split according to management intensity (as described in Chapter 3) and reanalyzed using the ArcGIS OLS tool. The statistical results indicated improved significance; yet, remained misspecified for residual normality and autocorrelation (Table 5). A manual examination of the OLS residuals was accomplished to evaluate if the model was usable despite its statistical misspecification.

Table 5 Statistical Diagnostics Following OLS

Management Area	Adj. R²	AICc	Koenker BP	Jarque-Bera	Morans I
Urban	.741	22104	.000	.000	.000
Light Management	.711	52709	.000	.000	.000
Heavy Management	.681	25829	.000	.000	.000

Ecological data are inherently noisy and typically contain significant numbers of outliers, so R² values are rarely higher than ~0.75-0.80. This is reflected in the literature. Bocard et al. (1992) attempted to partition variation in ecological data across several ecosystems ranging in scale from bacteria to boreal forests. Despite sophisticated procedures, they found 61% of their forestry data variation was unexplained due to the natural complexity of biotic and abiotic explanatory variables. Given that, these models are performing reasonably. AICc for light management is high compared to the other models, however. More concerning are the significant JarqueBera and Moran's I figures. Histograms of residuals (Figure 19) indicate that the residuals are closer to normally distributed than the statistics would imply.

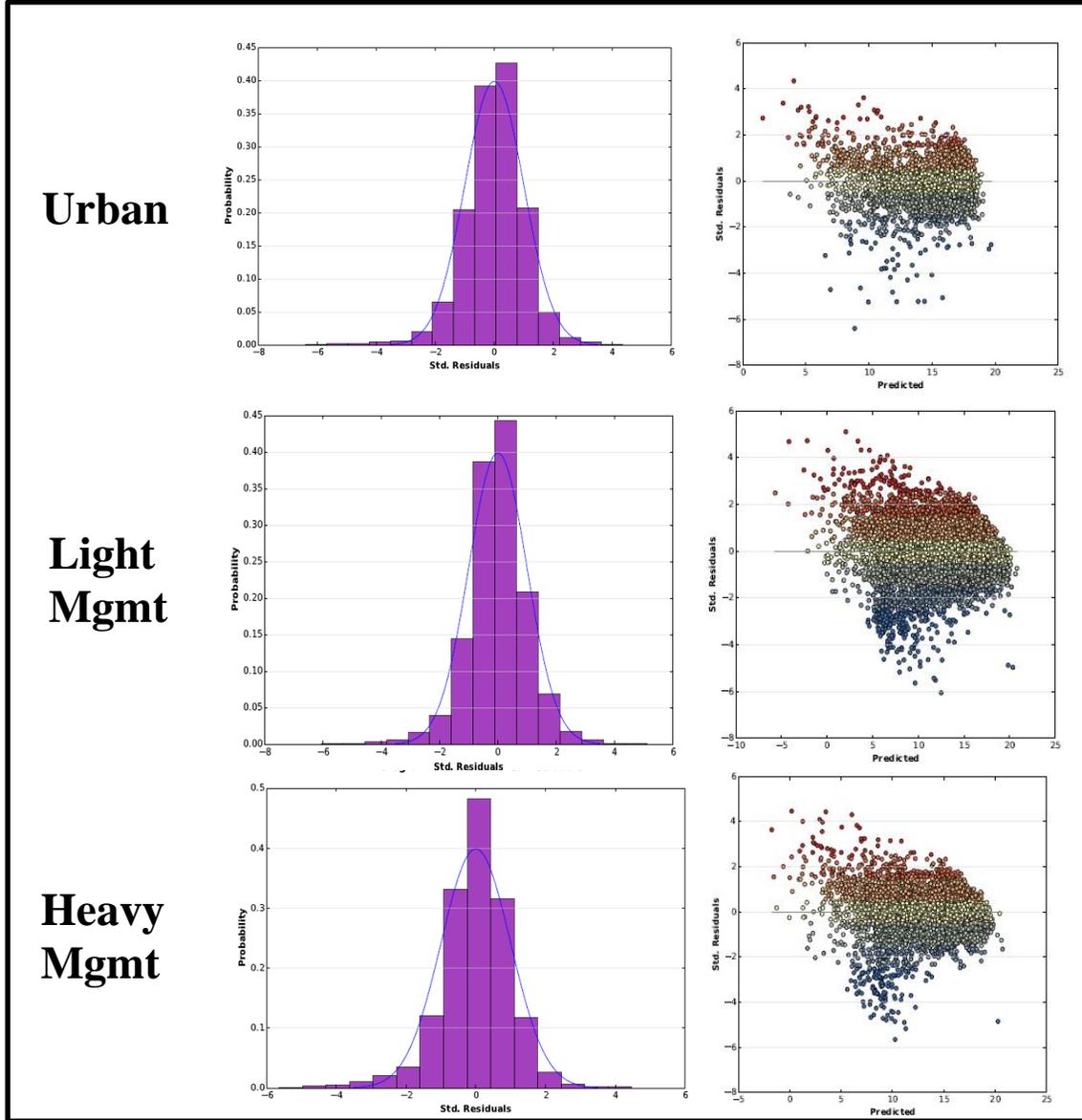


Figure 19 OLS Residuals—Histograms and Scatter Plots

Histograms and scatter plots graphically depict the goodness of fit of Simpson Diversity Index data to the ordinary least squares regression models. The histograms indicate the distribution of residuals around the model. Ideally, the residuals should be normally distributed around the model. That is, the distance of most data points from the model's prediction should be small and the volume of residuals should decrease as distance increases. The scatter plots depict the distance in standard deviations each residual is from the prediction. For these ecological data, known data within two standard deviations of the predicted value is considered a good fit.

The distributions are somewhat leptokurtic, with more residuals gathered around the mean than expected. Leptokurtic distributions such as LaPlace and Student's T tend to have fat tails and this is reflected in the histograms (Zar, 2010). All three scatter plots have residuals

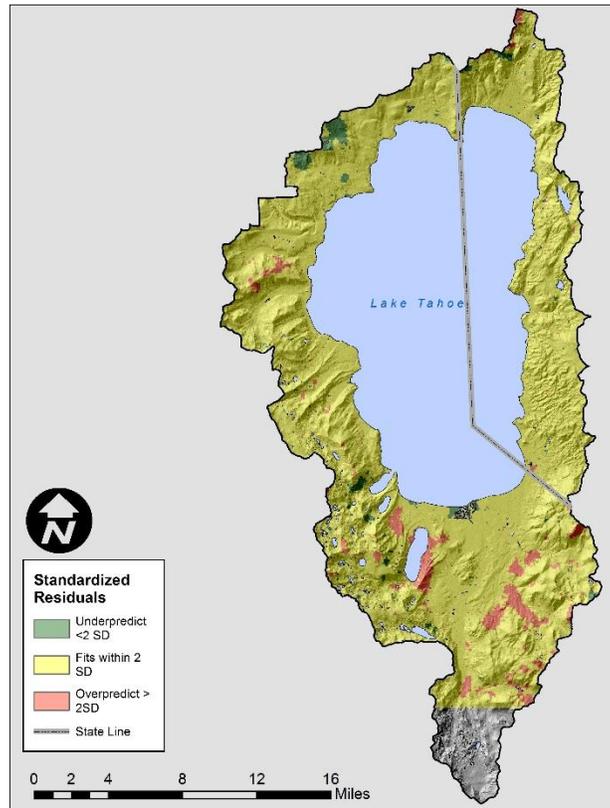


Figure 20 Geographic Distribution of OLS Residuals (Over and Under Prediction—Fits Within 2 Standard Deviations)

This study accepts residuals within two standard deviations of the model prediction as a good fit. Not every prediction fits that criterion; therefore, it is useful to examine how poor fits are spatially distributed. It appears from this figure that most of the LTB acceptably fits the diversity regression model. There are clusters of ill-fitting data points that warrant further examination; however, for the purposes of this study, the diversity model encompasses an acceptable percentage of the Basin.

clustered within two standard deviations of the mean. In all likelihood this represents the natural variance of the system. Several significant clusters sit outside the 2 standard deviation plane along with a large number of outliers. Residuals from the models were recombined into a figure that depicts the location of data points that fall more than two standard deviations from the model (Figure 20).

Red and green areas highlight the geographical locations where the model over and under predicts, respectively. The grey area in the southern portion of the figure represents an area of incomplete data. Data from that region of the study area was not included in the model. Clearly, the model is not perfectly specified. If a natural two standard deviation variance is accepted, however, the model adequately describes diversity across the majority of the LTB.

Rykiel (1996) investigated the validation of ecological models. In part, he asserts that models can be validated outside traditional statistics. One of his key points is that models can be validated for pragmatic purposes given that they are credible. In this case, heavy credibility for these models lies in the fact that the explanatory variables have been used for identical purposes in the peer-reviewed literature. Given this and the high model fit rate illustrated in Figure 20, these models were deemed credible enough to be used as the environmental factors in the maximum entropy analysis in this research.

4.3 Maxent Models

Maxent models designed to predict species richness are the primary elements under study by this research. Two models were produced: the primary species richness-based model and an individual species range-based model, produced primarily for validation purposes. The models were created using the parameters and procedures described in Chapter 3. This section begins with a brief discussion of issues common to both procedures and continues to the specific results and diagnostics for each model. Finally, the two models are compared for evaluation purposes.

4.3.1 Common Procedural Issues

With appropriate presence data (TBEVM tree species and species richness) and environmental layers in hand, the modeling process moved to the Maxent processing stage as described in Chapter 3.

In every species richness case, several runs were necessary for two primary reasons. First, the Maxent models are stochastic, so each run is likely to result in slightly different results. Second, as mentioned in the methods chapter, Maxent has a literature-noted propensity for overfitting models. The fix for that issue is to examine individual model runs and adjust the regularization output to apply a complexity penalty that “loosens” the model output.

For each species richness case, Maxent was run three times at the default regularization to evaluate consistency in the model. Each run involved up to 500 training iterations (the default maximum). The specific number of iterations for any run is an internal Maxent function. In every case, the resulting predictions surfaces were similar, but the initial runs (at default regularization) appeared overfit. This is a subjective measurement; however, the jackknife diagnostic output was used to apply a measure of quantification to this issue. Specifically, the jackknife bar graph for many initial runs indicated that *every* environmental input had an equally high influence on the model output. Further, the graphical range output in overfit runs appeared to hug the input presence data and exclude legitimate probable range in favor of known presence data. The regularization was iteratively increased until full, stable (i.e. consistent results over three runs) coverage was achieved.

4.3.2 Species Range-Based Model

The species range-based model was created as a method of validation for the species richness-based model; it is presented first because it proved to be the less effective model (Figure 21).

An initial check of the model, based on local knowledge, indicates that the predicted species occurrence probabilities are reasonable. For instance, white bark pine (*Pinus albicaulis*) is a high elevation species native to the sub-alpine biome. Zones of predicted white bark pine range indeed occur in patches along mountain crests.

Known points of species presence (from TBEVM) were then overlaid onto the Maxent “thresholded” (binary) range model (Figure 22). When a threshold parameter is selected in Maxent, the package uses the selected technique to make a best guess of which probabilities constitute a species’ range and which do not. This may be a user-selected probability, but in this case, the “equal sensitivity and specificity” function was selected to eliminate bias.

Ideally, known, real world species locations should directly correspond with their predicted ranges. The results were mixed. For white bark pine, the results were excellent. Other SDMs, such as for red fir (*Abies magnifica*) and white fir (*Abies concolor*) performed much less consistently.

These inaccuracies should probably be expected. The environmental variables used to predict each species range are the same factors that were used to predict ranges of species richness. In other words, environmental factors were selected via research and analysis to predict general diversity, not each individual species. Defining a tailored suite of environmental variables for each species would extend far beyond the scope of this research. Individual species across the spectrum of flora and fauna tend to establish ranges based on specific niches. In some

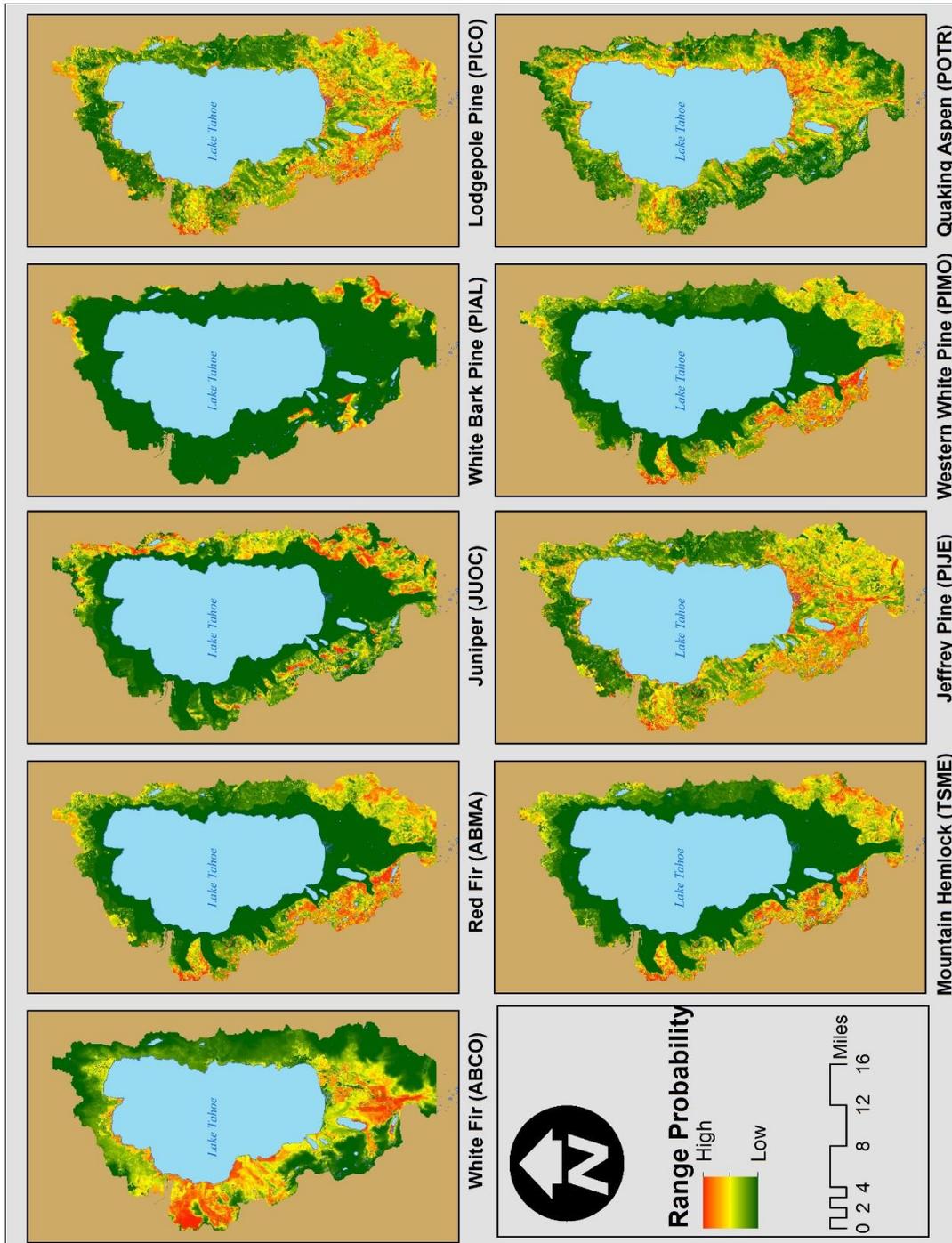


Figure 21 Maxent SDMs for Tree Species Lake Tahoe Basin, CA/NV

Maxent Species Distribution Models (SDM) for tree species in the Lake Tahoe Basin are illustrated. For consistency and convenience, each SDM utilized the same general diversity environmental variables used in the species richness models. Red areas of the surface indicate a high likelihood of habitat for each species; whereas, green areas indicate a low habitat probability.

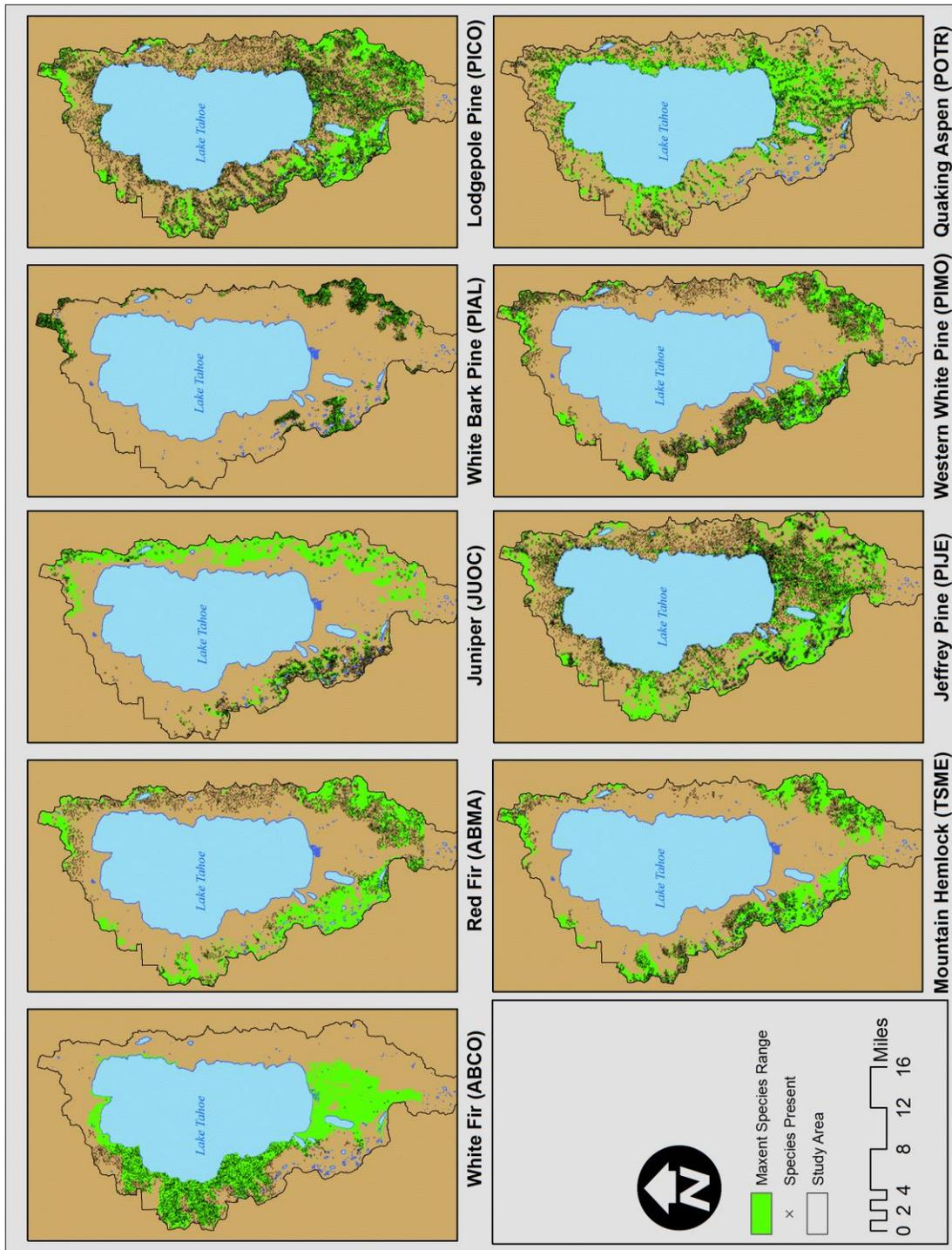


Figure 22 LTB Maxent SDMs (binary) with Species Presence Overlay

Green areas in this figure are Maxent’s best estimates of each species’ range. Unlike Figure 21, this is a binary “range” or “not range” representation as opposed to a probability continuum.. Small x symbols represent locations of known species presence. The best performing SDMs coincide with these presence data.

cases these align well with diversity factors such as elevation or soil quality. In other cases specific factors such as canopy cover (i.e., shade tolerance) be involved.

On the surface, the species range technique for predicting species richness appears more straightforward in a technical sense; however, a detailed evaluation of the pertinent environmental factors affecting each niche would be required for fine accuracy work.

While there are accuracy questions with some of the species range predictions, the general patterns are sufficiently correct to provide a reasonable validation device for the species richness prediction method. Most notably, in nearly every case, a species presence data point occurs where a species range is predicted; although in some cases there are very few (e.g. white fir).

A final stacked SDM surface (Figure 23) consolidates each SDM into a validation surface. Note that species richness border regions cannot be isolated using the stacked SDM technique, although in many cases they are visible. The stacked SDM model is a true species richness surface. Its validation function is to compare its species richness patterns with those produced by the species richness-based method and displayed as border regions (they should be similar if both techniques are valid).

Notable features in the stacked SDM model include the density of richness ranges in the wilderness areas (set off by insets). Although it will not be explored by this research, these dense regions could represent areas of shifting mosaic-steady state. The bottom inset highlights an area of urban-wildland interface. The primary feature of interest is a line of red-orange patches of high species richness along the interface.

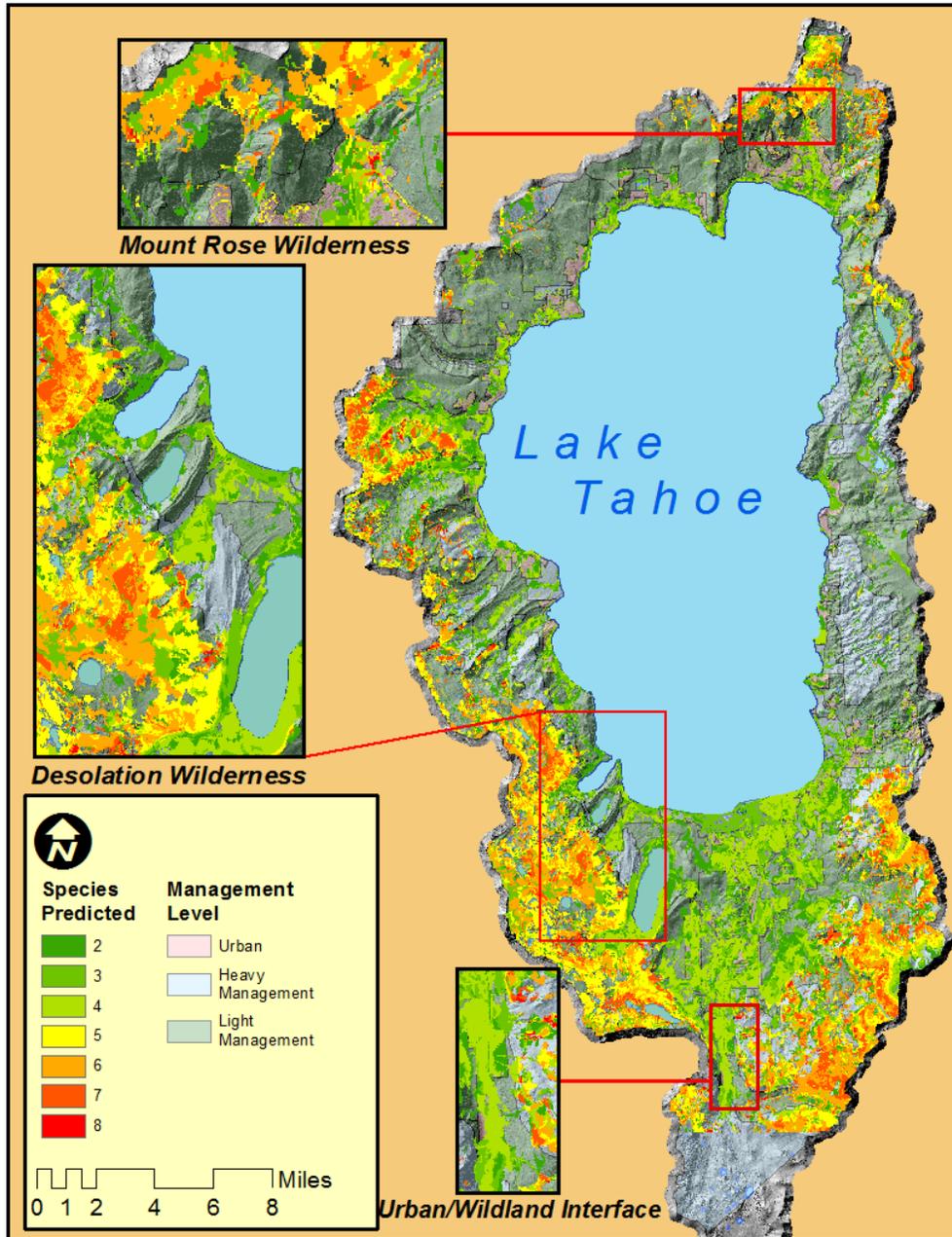


Figure 23 Maxent Species Richness Regions—Species Range Method Lake Tahoe Basin CA/NV

This figure is the result of summing the total number of species ranges that cover each pixel. The result is a species richness surface. Border regions appear as areas of extremely high diversity. They are areas that incorporate species from multiple homogenous zones of species richness. Areas where this is apparent are magnified by insets. Management level is also incorporated into the figure to illustrate portions of the study area where species richness follows management style.

This area of high diversity is likely an area of overlap between species ranges to the left (2–3 species) and the right (5–6 species)—a border region. This north-south string of abrupt border regions reflects a known zone of species richness change between adjacent woodland and urban zones (the airport in this case).

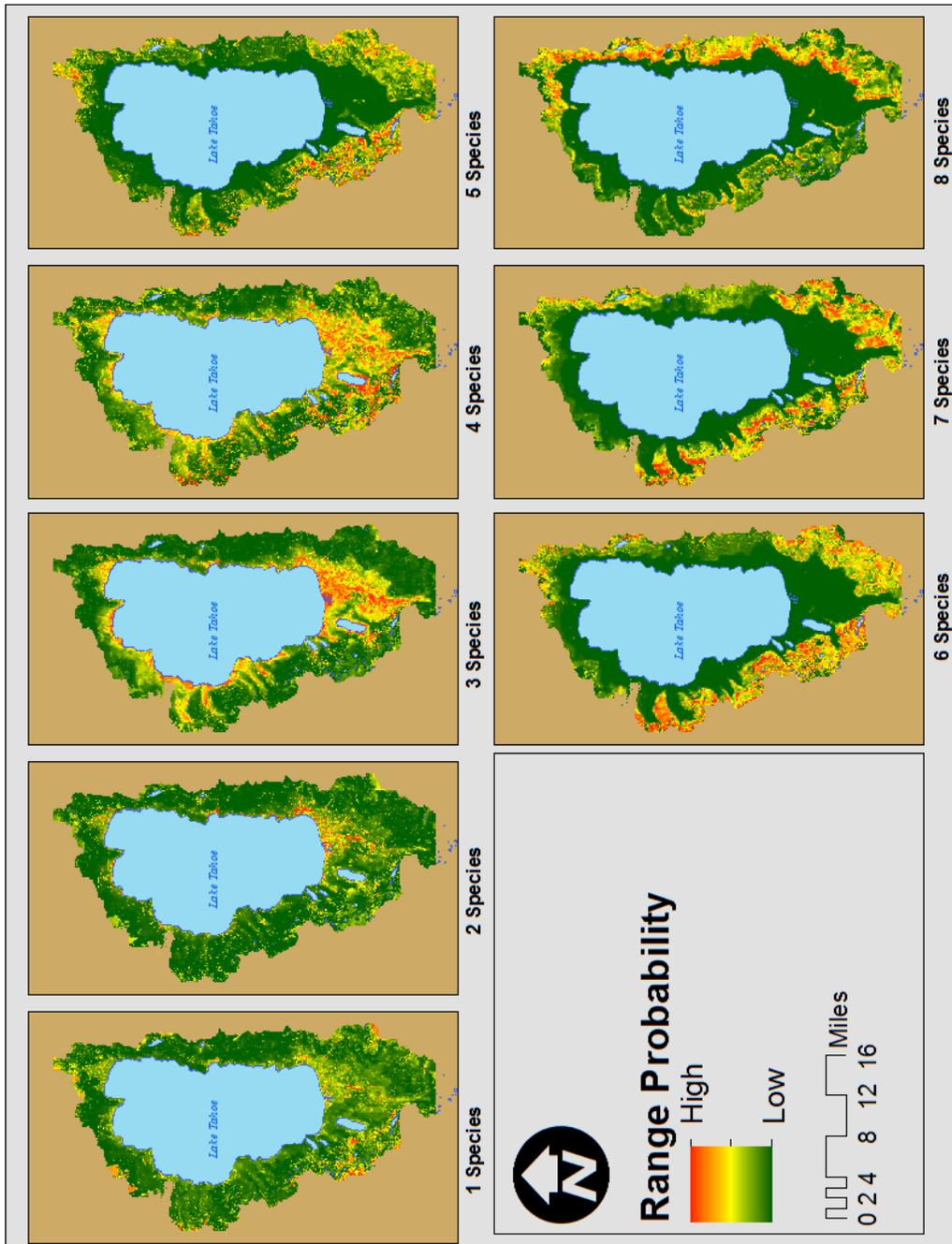
Finally, regions of management intensity are depicted on the map. In many areas, patterns of species richness align with these zones indicating a possible link between species richness and management style and/or intensity.

4.3.3 Species Richness-Based Model

The species richness-based model is the primary model being investigated for credibility and utility by this research. Similar to the model presented in section 4.2.2., the model was constructed using the stepwise procedure described in Chapter 3. A critical intermediate step in producing the species richness-based model is to use Maxent to predict species richness. Given the validity of these predictions represent one of the major research questions addressed by this thesis, emphasis is placed on statistical diagnostics bundled with the Maxent predictions.

Specifically, the results of Area Under the Curve (AUC) and jackknife analyses are presented to bolster the credibility of using Maxent to predict homogenous regions of tree species richness.

The range probability output for tree species richness values in the LTB (Figure 24) provides for an initial evaluation. As a first step, the output ranges were examined against local knowledge similar to the process used for the individual species ranges in Figures 21 and 22. The ranges make sense. For instance, the ‘1’ species range indicates high range probability at the Basin’s highest elevations. This makes good-sense given that the high elevation sub-alpine zone generally hosts a limited number of hearty, long-lived tree species such as white bark pine.



**Figure 24 Maxent Tree Species Richness Distribution Models
Lake Tahoe Basin CA/NV**

This figure presents the Maxent distribution models for tree species richness. Typically, Maxent output predicts the likelihood a pixel is habitat for an individual species by combining species presence data with relevant environmental factors. These models substitute species richness observations for species presence and incorporate factors that contribute to tree diversity in the LTB. The models do not predict any particular species. Rather, the models predict the likelihood a specific number of species will exist within in a particular pixel.

The '8' species zone incorporates an area of the Basin where there is evidence of succession from sun-loving Jeffery pine to fir species whose recruits better tolerate the dense, shady understory of forests that have been actively fire-suppressed for decades (Taylor et. al. 2014).

The primary statistical check of each SDM was an evaluation of Area Under the Curve (AUC) (Figure 25). The curves being referenced are the Receiver Operating Curve (ROC) and the associated curve of test values (randomly chosen data points from the input presence-only data). In each graph, the red curves represent ROC, the blue curves represent test data and the grey lines are a reference random prediction.

ROC is the probability that a cell contains a species richness value versus the rest of the background cells. Where the blue line closely matches the red line, the model is performing well as predictions are closely following the test data.

The axes of each chart represent specificity (x) versus sensitivity (y). Precisely how Maxent calculates ROC (and hence separates range from not range) is adjustable by the user. As a safeguard against a biased or invalid (due to excessive sensitivity), Maxent was set to equalize sensitivity versus specificity. The Maxent prediction algorithm adjusts or omits predictor variables and checks for excessive change in the final range predictions. Maxent adjusts the threshold until the output values are no more sensitive than they are specific

Each of the final eight species richness predictions met the 0.7 AUC benchmark and their test data closely mirrored the ROC plot. Creating eight comparable (i.e. use the same input variables) that are also statistically well-specified is challenging. That each of the models surpassed the established benchmark is a strong argument for their validity.

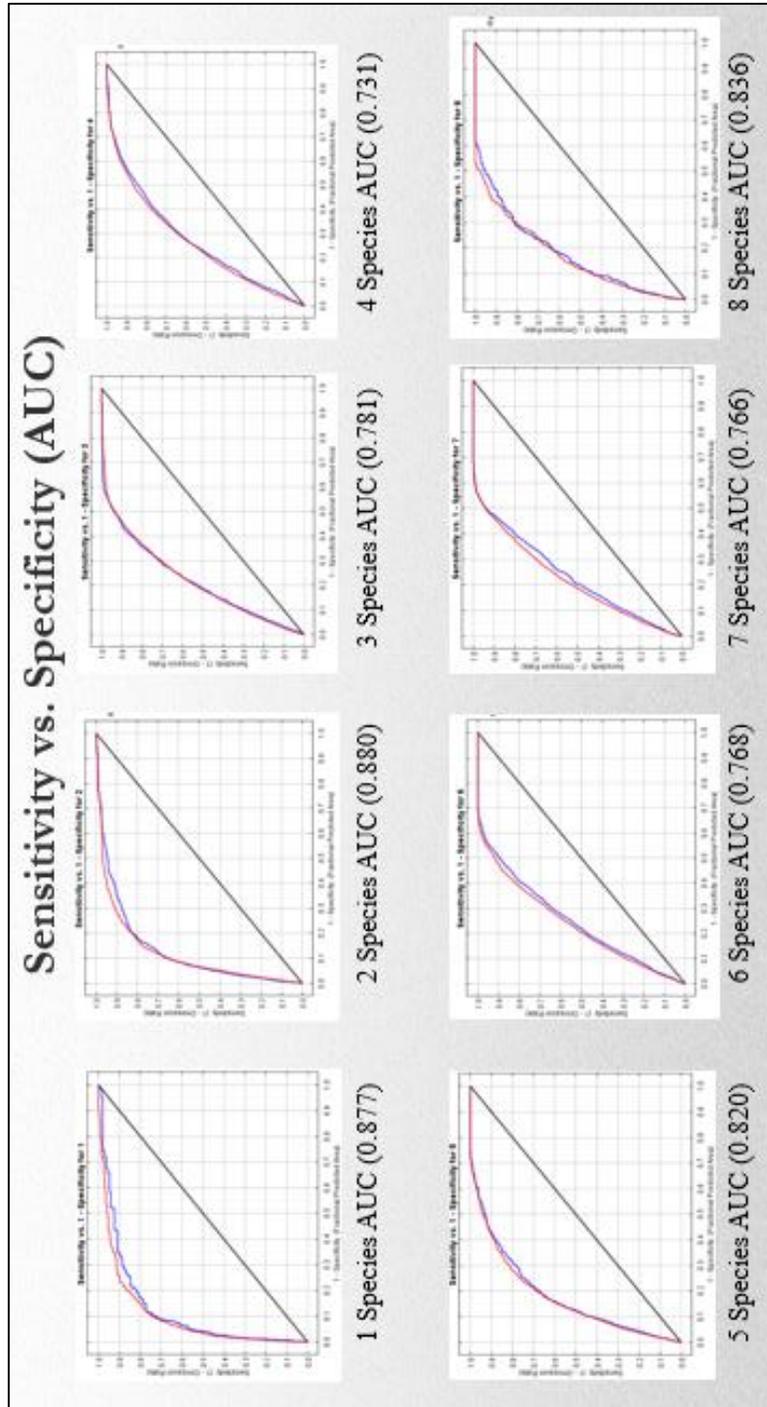


Figure 25 Maxent Area Under Curve (AUC) Analyses for LTB Tree Species Richness Distribution Models

The AUC charts plot specificity (x) vs. sensitivity (y). 0.5 AUC is considered a random prediction. 0.7 considered an adequate prediction. Ideal curves peak in upper left corner.

It must be noted that accurate Maxent predictions are highly dependent on the careful selection of environmental variables. While great care was taken to select the most relevant variables via literature research and regression analysis, the Maxent jackknife analysis provides an excellent final verification that each of the selected environmental elements indeed meaningfully impacts the SDM. A representative sample of the analysis (Figure 26) for three

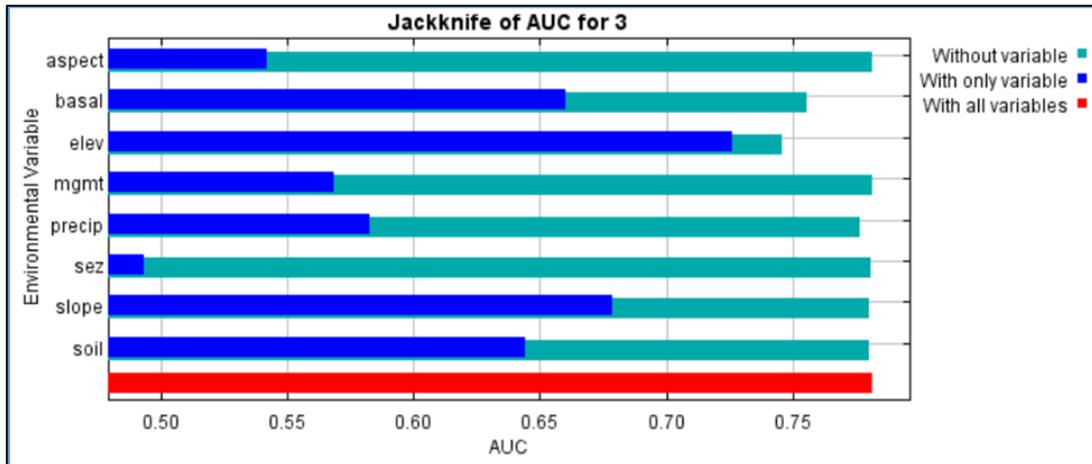


Figure 26 Jackknife Analysis for 3 Species

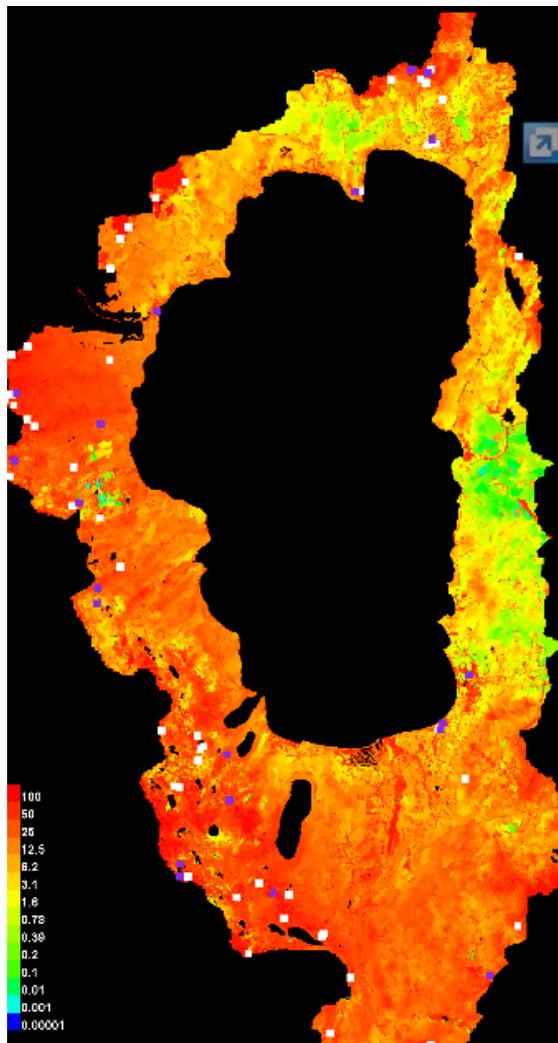
species (the full set appears in Appendix 1) is presented here. The dark blue lines indicate that each variable is significant (not left of the y-axis) and that their influence varies. The more right the blue line, the greater its influence on the model. While this graphic is highly representative, the specific contribution of each variable differed from model to model. Not surprisingly, elevation is most consistently significant across all the species richness SDMs, given elevation is the primary determinant of biome in the LTB. Basal area is also a consistently significant variable although its influence diminishes as species richness increases.

This is consistent with basal area’s function as a “wildcard.” It captures many of the unspecified noise variables in the system that produce a significant effect, but cannot be captured in a global variable due to their infrequency (e.g. disease, insect infestation or discrete

management events). Very diverse forests likely experience fewer of these random disturbances, hence basal area's generally inverse relationship with species richness.

The light blue bars quantify the effect of eliminating a variable from the model. For each variable, the further its bar reaches to the left, the heavier its omission affects the SDM. That is, leaving a variable out will significantly change the prediction.

An additional important result is the decision to omit an SDM for zero species from the modeling process. The SDM for zero species (Figure 27) suggests that, with few



**Figure 27 Tree Species Richness Distribution Model for Zero Species:
Raw Maxent Output for Lake Tahoe Basin CA/NV**

exceptions, suitable conditions for zero tree species can occur throughout the study area. This is a reasonable prediction. Rocky outcrops and treeless meadows are common across the LTB; however, the high uniformity of these data make them more apt to contribute noise than information to the models.

Beyond their specific utility, the validity of these zero species data is questionable due to the mechanics of Maxent itself. As a presence-only algorithm, is it reasonable for it to predict the 'presence' of nothing? This is an interesting question, but it certainly exceeds the scope of this investigation.

Conceivably, these data could produce a more complete model. Their potential for contributing noise along with their questionable validity, however, were deemed to outweigh their potential benefit, so they were omitted.

The final species richness border region output produced by this model provides a clean depiction of species richness patterns in the LTB (Figure 28). Note that border regions with the lowest abruptness value (1 species) were assigned no value as these likely represent the most uncertainty in the model, are least interesting and tend to clutter the figure.

The model's most notable features mirror the features highlighted in Figure 23. The bottom inset isolates an instance of wildland/urban interface near South Lake Tahoe, CA. This model identifies this known, highly abrupt zone of species richness change very clearly.

Portions of the Desolation and Mount Rose Wilderness areas are also inset. Wilderness ecosystems are left to regulate themselves barring extreme circumstances. The myriad of species richness zones could support the theoretical shifting mosaic/steady state pattern, while urban and more heavily managed areas appear to have a more uniform species richness pattern.

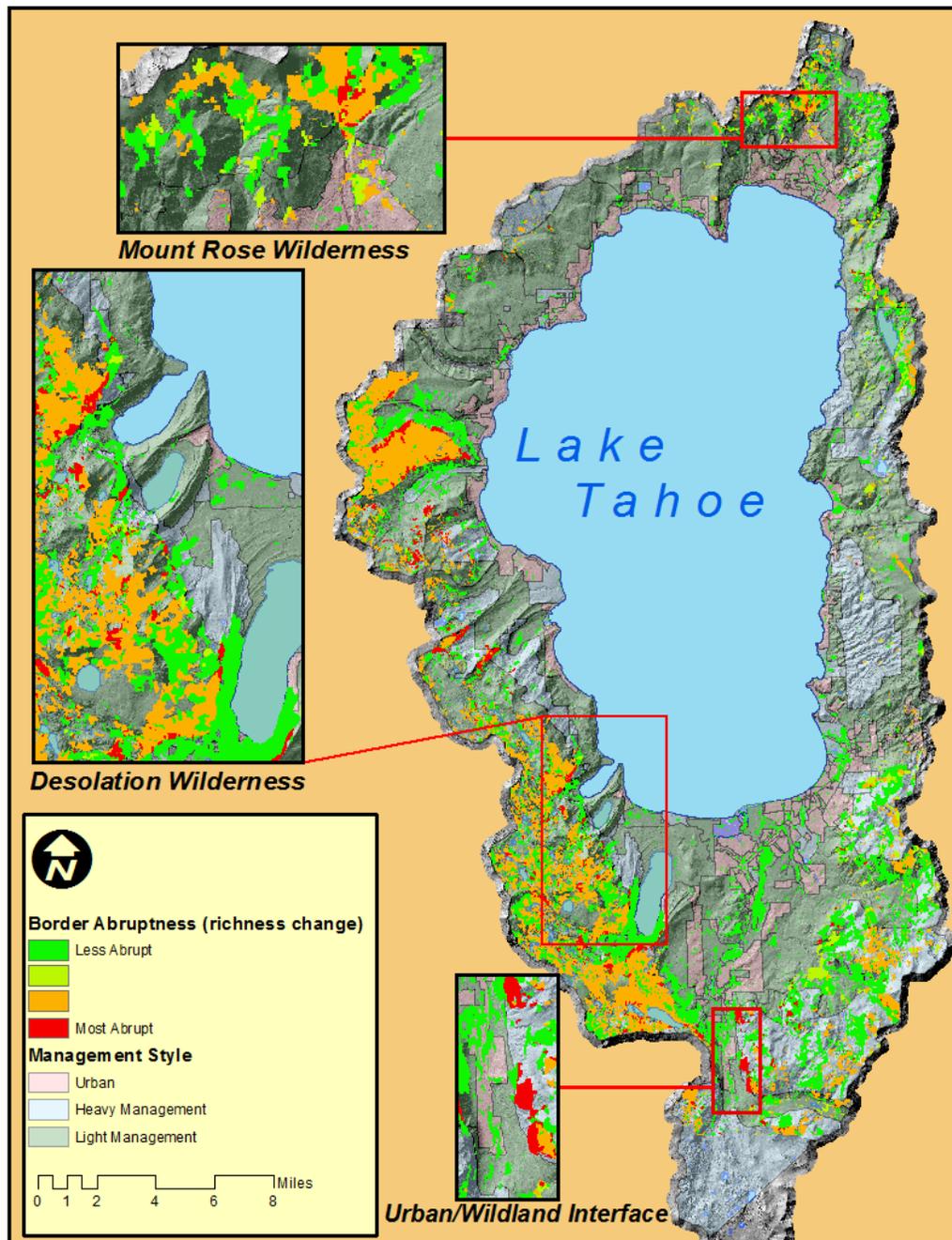


Figure 28 Maxent Derived Tree Species Richness Border Regions—Species Richness Method: Lake Tahoe Basin CA/NV

Final species richness surface proposed by this research. Border regions are classified by abruptness indicating the relative differences in species richness each border region represents. Insets highlight areas of the Basin where species richness patterns are dense. The Urban-wildland interface inset highlights an instance of species richness change known to be extremely abrupt (forested landscape abutting the clearcut airport area). Zones of management style are included to provide a visualization of management effects on diversity.

4.3.4 Validation of Species Richness Models by Comparison

For the purpose of comparing the two surfaces, the species richness-based and species range-based models are displayed side by side (Figure 29). As mentioned previously, the purpose of the species range-based (stacked SDM) model is to validate the species richness-based (border region) model. Specifically, the primary element that requires additional credibility is the use of Maxent to predict regions of species richness.

The underlying basis for this validation is that the validity of Maxent for the purpose of predicting individual species ranges is widely accepted in the literature. This is not the case for predicting species richness. If the two models, using identical environmental and presence data, and independent procedures designed to utilize Maxent species richness and species range data, respectively, arrive at significantly similar output, the credibility of using Maxent to predict species richness is greatly enhanced.

The earlier comparison of tree species SDMs versus known locations of respective species revealed some accuracy issues (likely due to choice of environmental variables); however, the general patterns were deemed acceptably accurate. Despite this limitation, the patterns displayed by the two models are extremely similar.

Note that the two models do not present identical information. The stacked SDM model presents homogenous zones of total species richness, while the species richness-derived model presents border regions classified by abruptness. The important observation is that the overall diversity patterns presented in the two models are extremely similar across the study area. This confirms that Maxent is predicting the same patterns using different approaches.

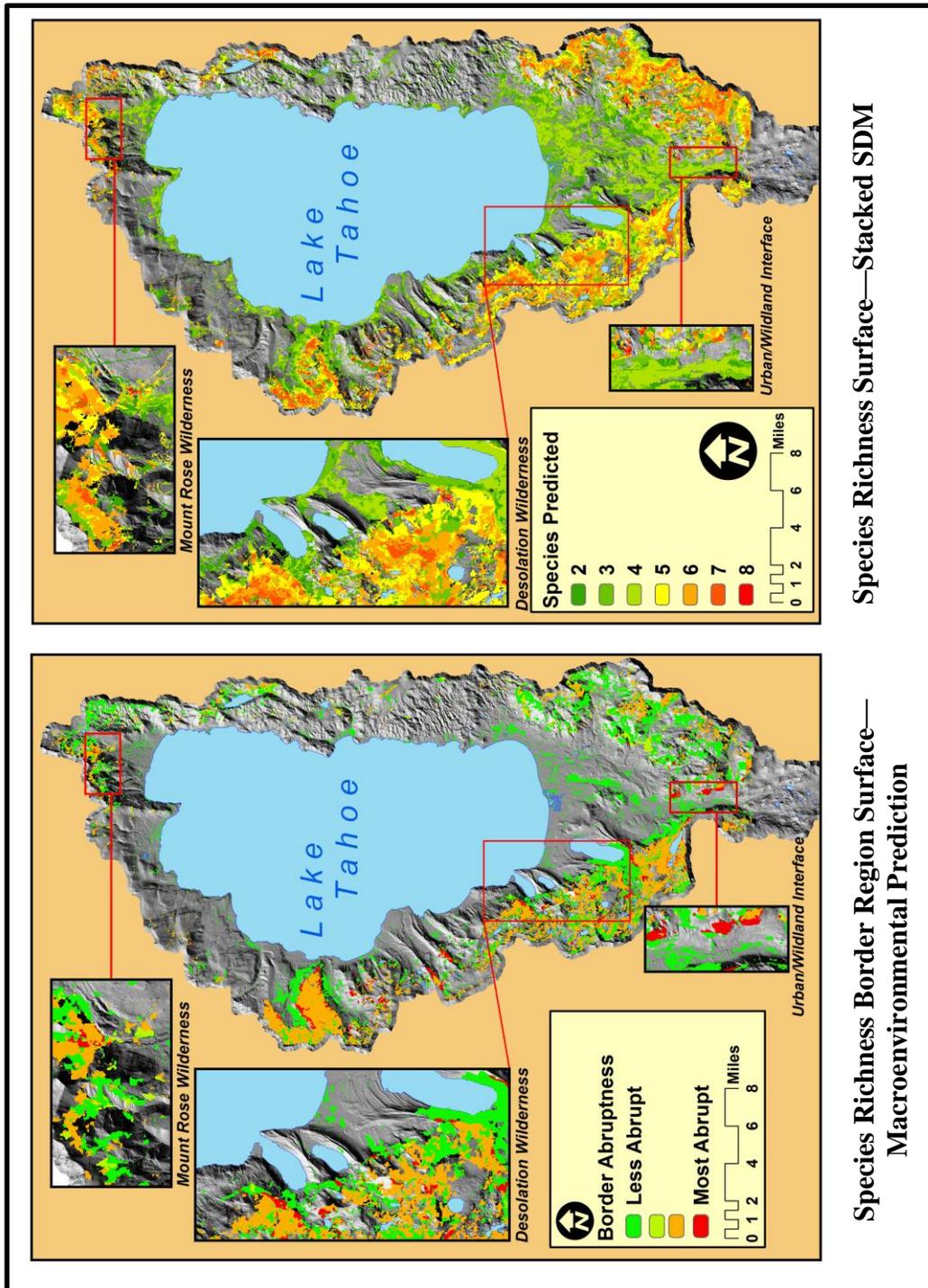


Figure 29 Model Comparison: Tree Species Richness Models—Species Range Method (top) vs. Species Richness Method (bottom)

This figure facilitates a visual comparison of the models produced by this research. The models support the validity of Maxent species richness models given they are functionally equivalent surfaces produced via different techniques. Both models effectively represent LTB tree species richness patterns although their specific representations differ.

The detailed insets provide more a fine-grained comparison of species richness patterns. The wildland-urban interface is quite visible in both models as dark red or orange patches. In the stacked SDM model, these patches relate to high species richness resulting from overlapping ranges. Note that these are generally bordered by yellow-orange areas of less richness. Conversely, the same area in the border region model contains isolated red patches representing very abrupt areas of transition—border regions. The relative isolation of these patches in an area of known richness change indicates that border regions were successfully isolated and displayed.

Similar patterns are evident within the wilderness area insets. Dense mosaics of species richness are evident in the within the stacked SDM model and the border region model outlines them well.

In sum, these models are functionally equivalent in that they represent the same diversity patterns. Their differences allow for studying them in different ways. In particular, the border region model provides an opportunity to investigate species richness in a new and interesting way.

4.4 Split Moving Window Dissimilarities

The species richness-based model depicts the border regions that divide homogenous plots of tree species richness. These are defined by portions of the Maxent predicted species richness ranges that overlap. These regions potentially define a class of spatial entity that is not described elsewhere in the literature. The purpose of SMWDA in the context of this research is to establish that these border regions are detectable entities and not simply variance in the predictive power of the Maxent algorithm.

Border and ecotone researchers consider SMWDA z-score peaks to be evidence of borders or ecotones. For this application, the principle idea is that if species richness border regions represent legitimate ecotone-like entities, they should be detectable using SMWDA.

The sampling area chosen for SMWDA analysis is in the northern portion of the LTB (Figure 30). This area was selected because it encompasses a large number of predicted border regions and as well as areas of significantly homogenous species richness (i.e. gray area). Further, the sampled zone spans a wilderness area, a developed recreation area and non-wilderness primitive area. Lastly, the sampling zone contains a large sample of TBEVM data (74 data points) for analysis.

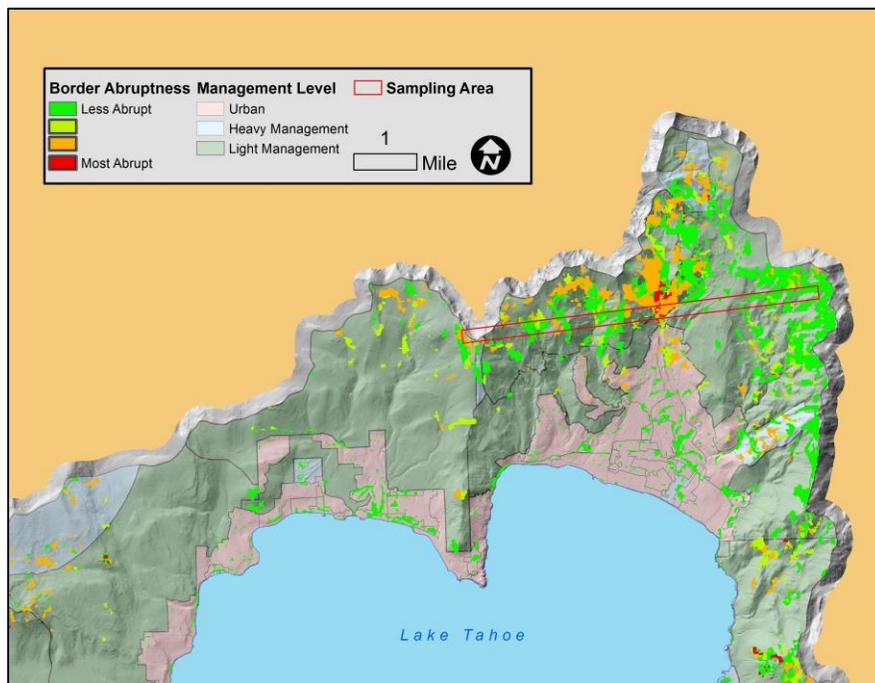


Figure 30 SMWDA Sampling Location North Lake Tahoe, CA/NV

The red rectangle represents the location of species richness data that were extracted from the TBEVM dataset for analysis using SMWDA. The location, size and shape of the sampling area were chosen to encompass a sufficient sample size of data points, approximate a linear dataset and capture a large number of border region predictions.

The results of the analysis (Figure 31) are moderately conclusive. As mentioned in a previous chapter, species richness is very sensitive to scale. Varying the window size used in SMWDA analysis fine tunes the scale of the analysis. Several iterations of SMWDA were conducted on these data with window sizes ranging from four to ten data points. As expected, no scale perfectly captured every border region; however, a six data point window reflected the expected pattern very well. For reference, a Monte Carlo simulation predicted that across the entire transect, the expected species richness change for any six point window is 2.1 species. The z-score for each point represents how many standard deviations from expected its six point window is. Note that for clarity (and reference to the related map), a z-score is plotted for each data point although this is not precisely accurate. Each window necessarily contains an even number of data points so that it may be split into even halves. The z-score relates to the midpoint of each window which is midway between the third and fourth points in this case. For example, the results graph in Figure 31 begins with data point number 3 representing the first position of the six point moving window. Strictly speaking, the related z-score refers to a geographic point between data point 3 and 4. The next window position (represented by data point 4) refers to a point between data points 4 and 5 and so on.

A precise geographic representation of each SMWDA z-score is unimportant given the data restrictions of this analysis. SMWDA performs best using field collected transect data where the data are evenly spaced. Practical limitations precluded field collection of samples for this analysis. The TBEVM species richness data used for this analysis is not specifically linear nor are the data evenly spaced. This skews the z-score spread and geographic relationship between the results graph and the dataset. These are relatively minor issues, but must be kept in mind during interpretation of the analysis.

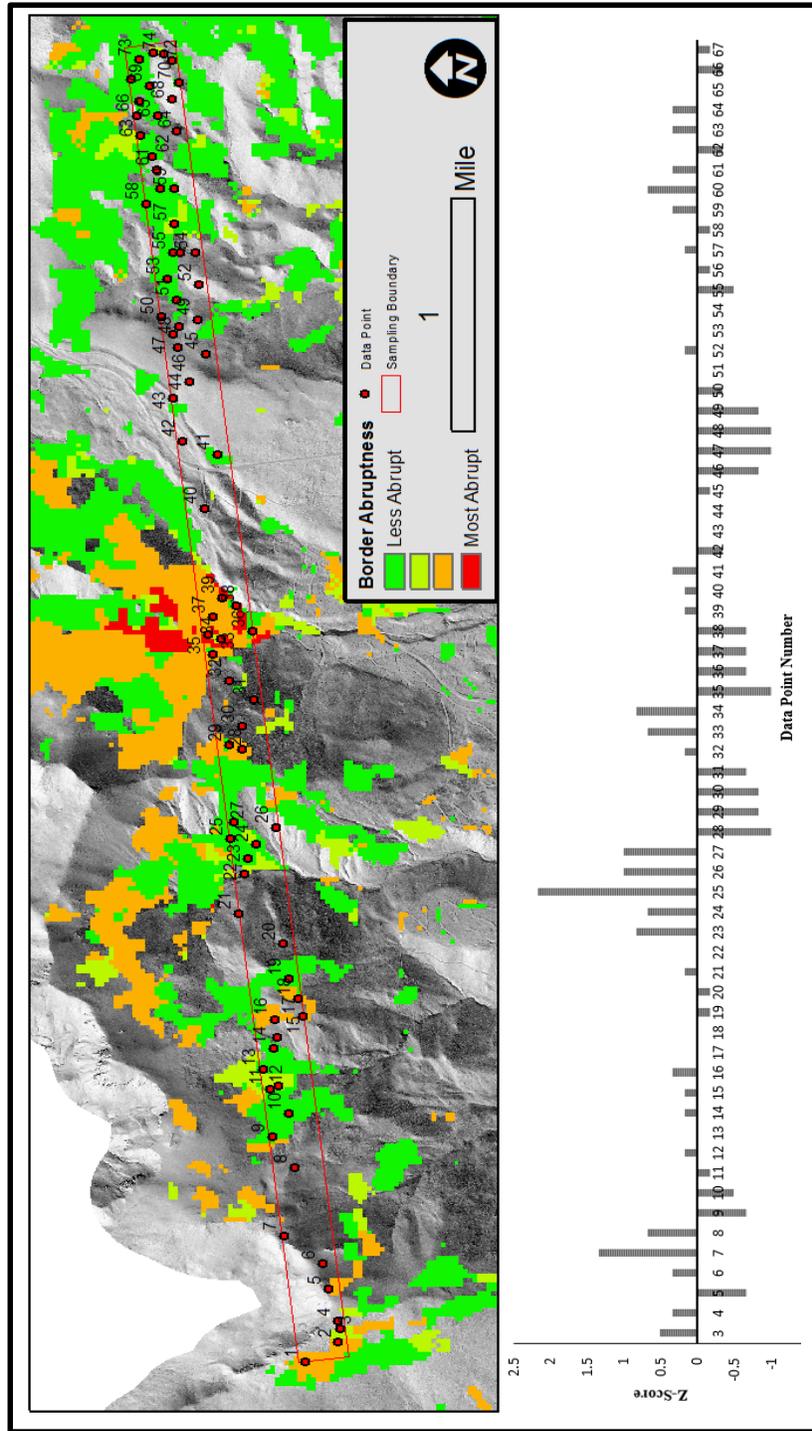


Figure 31 Split Moving Window Dissimilarity Analysis (6 Pt. Window)

SMWA uses a z-score to identify dissimilarity between halves of a moving window. In this case, the z-score represents the degree of difference in tree species richness three points to left and right of each subject data point. This plot highlight several important themes in the analysis. Data problems (far left), ideal response (~pt 23), and issues of scale (~pts. 35–41).

Moving along the results graph, the numbered z-score bars correspond with the numbered data points. The SMWDA results are derived strictly from the TBEVM data. The border region predictions are presented for reference only.

Beginning at the far left portion of the results graph, the results loosely follow the predicted pattern until data point 23. This portion of the results set is the least consistent with the Maxent predictions. This is likely due to some variance in the Maxent predictions. It must be noted that data points 4-6 sit atop a ridge that marks the Tahoe Basin boundary and the northern extent of the prediction surface. A small portion of the area north of the ridge was inadvertently included in the sampling region. This small data gap and perhaps some edge effect may explain some of this inconsistency. Additionally, the TBEVM data runs coincident with a predicted border region (points 9–19). The results in this area indicate window half dissimilarity near expected values as opposed to a sharp peak. It is likely that sharper peaks would be produced by a bisection of the predicted border area.

From point 23 through approximately point 50, the results detect border regions as would be expected. Sharp peaks correspond with predicted borders and negative z-scores correspond with highly homogenous areas.

The final section of the results set (50–67) is interesting. The area is a series of ridges in a managed, but primitive (non-wilderness) area. Slightly abrupt border regions are predicted along the ridges. This is ecologically logical given some rain shadow on the lee (eastern) side of the ridges would likely alter the predominant species community. The results set detects a series of alternating borders and homogenous areas that is highly consistent with the model prediction.

The results of the SMWDA analysis are not perfectly aligned with the model predictions nor is the analysis procedurally optimal. Given this, the existence of species richness border

regions cannot be conclusively validated; however, this analysis did consistently detect borders and homogenous zones that align with the model predictions. For the purposes of this thesis, these border regions will be considered valid, however extended research is necessary to further validate their existence.

Chapter 5, Discussion and Conclusions expands on the discussion presented with these results, and address each specific research question. Further, it provides some discussion regarding how this thesis relates to existing literature and finally provide some suggestions for further research.

CHAPTER 5: DISCUSSION AND CONCLUSIONS

A good deal of discussion is integrated into the results chapter of this thesis. As a result, this chapter focuses on providing summary answers to the proposed research questions, discussion of this research as it relates to the existing species richness literature and finally suggests avenues for further research.

5.1 Research Questions

The research questions posed by this thesis aimed to explore the utility and validity of a Maxent-based procedure to produce a species richness surface for tree species in the Lake Tahoe Basin, CA/NV. The method moves beyond creating a probability surface for each possible species richness value and eliminates areas of homogenous diversity from the representation. That is, pixels containing more than one range of tree species richness are presented as border regions between homogenous tracts of species richness. Thus, research questions address the validity of utilizing Maxent to extrapolate species richness from sampled data and environmental variables, whether or not border regions are meaningful entities and if the resulting surface is a useful tool for use in forestry research and management.

5.1.1 Can the Maxent maximum entropy modeling package make valid predictions of tree species richness patches?

A 2008 study examined the use of maximum entropy and state variables in macroecology (Harte et al. 2008). Harte et al. make three observations that are relevant to evaluating the success or failure of the species richness predictions in this thesis. First, they note that, as a procedure, the maximum entropy method has been applied to many areas of science and that it is a proven

method for inferring “most likely probability distributions” where knowledge-based constraints are incorporated into the procedure. (p. 2701–2). Second, the application of maximum entropy to a complex system requires only a numerically defined entity (that need not be narrowly defined), and a set of constraints on the probability distributions (p. 2702). Finally, where assumed constraints fail to provide an adequate prediction, then either, some constraints did not hold or additional constraints must be explored (p. 2709).

Observations by Harte et al. (2008) directly address the novelty of using Maxent to predict species richness. Specifically, they acknowledge the use of maximum entropy as widespread throughout the sciences and that studied entities need not be narrowly defined. Indeed, Harte et al. opine in their own example that distinguishing between “species” or “individuals” is unimportant—the critical criterion is that entities are numerically defined. Species richness meets this requirement.

This thesis further addresses the requirements of Harte et al. with a thorough investigation of constraints on species richness in the LTB. Candidate variables from the literature were analyzed quantitatively and then ranked by their impact on diversity. The specific contribution of each highly ranked diversity predictor on species richness was further explored by arranging them into a conceptual model grounded in ecological theory. These efforts directly address the knowledge-based constraints requirement espoused by Harte et al.

Finally, as Harte et al. (2008) indicate, the quality of a Maxent prediction weighs heavily on the inclusion of constraints that adequately predict the data. This thesis utilized Maxent to create SDMs for tree species in the LTB species pool and distribution models for species richness “entities” (one model for each possible value). Identical constraints (selected for their effect on diversity) were used for both procedures. Initial assessments based on local knowledge

indicated the surfaces generally reflect known species patterns in the LTB. Objective assessments (AUC, dominant species overlay), however, indicate the Maxent predictions of species richness entity distributions outperformed individual species SDMs. This is a positive and expected outcome for two reasons. First, given the predictor variables are tailored for diversity, not individual species, valid species richness distributions should outperform individual SDMs. Second, high performing species richness predictions, based on environmental factors separate from competitive factors affecting individual species, support the Neutral Theory notion that species richness arises from a landscape patch's ability to accept the random colonization of a specific number of tropically similar species. That is, the system that defines species richness patterns is distinct from the systems that define individual species distributions.

In sum, the distribution models created in this thesis strongly support that Maxent, given appropriate environmental variables, can *validly* predict ranges of species richness if certain ecological theories are accepted. Theories such as Hubbell's Neutral Theory (Hubbell 2008) and shifting mosaic-steady state (Borrmann and Likens 1994) are controversial (Perry 2002). The evidence, therefore, does not necessarily indicate that Maxent is the *best* tool for predicting species richness, just a valid one.

5.1.2 Can the location and properties of border regions between Maxent predicted species richness ranges be derived from Maxent output and if so, are they valid?

Two methods for arriving at a species richness distribution surface were investigated in this thesis. The primary method, stacking and subtracting Maxent species richness ranges, produces patches most correctly defined as "border regions." Stacking and then summing individual SDMs displays areas that border homogenous zones of species richness, but it does not isolate

them. This is evident in Figure 29 where the species range method model is arguably more cluttered; yet does not provide significantly more information. As to the capability of deriving border regions, the answer is yes; however, the species richness method is more effective at isolating them.

The more important component of this question pertains to validity. The concern is that overlapping ranges of predicted species richness might reflect simple variance in the prediction algorithm rather than legitimate ecological entities similar to an ecotone. Border region validation relies on an analytical technique called split moving window dissimilarity analysis (Cornelius and Reynolds 1991).

SMWDA is a technique designed to locate borders and ecotones by identifying discontinuities in ordered ecological data. The TBEVM data used in this thesis are not specifically ordered; however, a transect of these data that is approximately linear and covers several suspected border regions was extracted for analysis. The results (Figure 31) are mixed. The analysis identified several patches as discontinuities, strongly implying that they are legitimate borders. Conversely, several patches were either not detected or not detected consistent with the predicted abruptness. In these cases, plausible explanations such as edge effect or patches running parallel to the transect decrease the impact of these inconsistencies. Hennenberg (2005) reported similar issues with edge effect and detecting ecotones that are not perpendicular to the transect. Still, these inconsistencies as well as the analytical limitations of the TBEVM dataset cannot be completely discounted. Therefore, the best conclusion is that border regions are likely to be legitimate entities; however, not every suspected border region identified by this method is necessarily valid.

5.1.3 Can the use of Maxent and a GIS produce a valid and broad scale representation of tree species richness?

This is the penultimate question of this research. Broken into its three parts, the first element asks if Maxent and a GIS can produce a representation of tree species richness. The answer is a resounding yes. In fact, this research created two distinct representations using two different techniques that are functionally equivalent. Further, it was demonstrated that the Maxent surfaces are malleable enough to undergo substantial post-processing in a GIS, extending their utility beyond merely defining the range of a species or other entity.

The second element, validity, is less conclusive. The best answer is probably. As detailed by the previous two research questions, there is good evidence that Maxent produces a valid prediction; however, its validity requires the acceptance of several ecological principles that are heavily debated in the literature.

Lastly, the spatially explicit output and wide spatial scope of these models, if considered valid, represent a significant tool for addressing the persistent problem of scale in species richness applications. Patterns of diversity may be investigated as they vary from a few pixels near the resolution limit of the surfaces all the way out to their full extents.

5.2 Discussion

This section delves briefly into the literature to assess two issues. First, why compiling these or similar models would be of value to forest managers and researchers and second, how this research meshes with similar studies.

5.2.1 Applicability of Species Richness Models to Forest Management

Looking at the final species richness models produced during the course of this research (e.g. Figures 23, 28 and 29) it is evident that management policies play a heavy (though not exclusive) role in determining the final species richness structure of tree species in the LTB. For instance, species richness border regions often parallel the edges of management zones (e.g. along urban-wildland interface). In other cases, management zones (e.g. wilderness) encompass patterns of diversity that visibly differ from adjacent areas. Interestingly, in no case does a jackknife analysis indicate that management style is the most significant influence on species richness. Clearly, human managers cannot meaningfully affect factors such as elevation or aspect. Management style is, however, a forest manager's best opportunity to influence diversity in the interest of forest health.

The question then, given the multitude of factors and prescriptions available to forest managers, is why species richness matters and how models such as these are helpful.

Invasive species are a matter of great concern in the LTB (LTBMU and USFS 2007, LTBMU 2013, Taylor, 2014). A recent study indicates that high species richness relative to the available species pool correlates with higher invasibility of adventive species (Akatox and Akatova 2013). This is seemingly counterintuitive; however, it does correlate with the contemporary ecology of the Tahoe Basin. High species richness along Lake Tahoe's eastern shore, is accompanied by (or perhaps due to) ongoing succession of Jeffery Pine stands by shade tolerant fir species (Taylor 2014). Managing diversity-connected issues such as this is more complicated than it may seem. The issue of scale in diversity issues is again applicable. Issues such as succession are often rooted at regional or even geographic scales (e.g. climate change); however, land managers most often operate and make decisions at the human level. Models such

as these can highlight diversity patterns at multiple scales or perhaps reveal “hidden heterogeneity” that provides important clues toward addressing unwanted change (Bestelmeyer et al. 2003).

The model proposed by this research emphasizes the identification of ecotone-like structures. The purpose and value of identifying these structures in the context of forest management moves beyond decluttering the surface (although that is beneficial). The prime utility of identifying and validating these structures is that they represent points of change within the forest ecosystem. In this case, change in tree species diversity.

Changes, whether due to local urban development or global climate change, are points of heavy interest in distressed ecosystems such as the LTB. A seminal article on Sierran ecotones, Heath (1971), illustrates their use. The authors plotted the movement of an ecotone across a number of years. They noted that it traveled linearly and varied in size, shape and species composition without any sign of retrograde movement. The authors could not hypothesize a specific cause; however, they stressed that this was an important phenomenon and urged further research.

More recent research suggests the vertical movement of ecotones (i.e. upslope) are indicators of climate change. Specifically, thermally regulated ecotones such as at the alpine-woodland interface are indicated for this purpose (Kupfer and Cairns 1996). Bekker and Malanson (2008) present several hypotheses for lateral patterns in forested (specifically sub-alpine) landscapes. Winds, topography and seed dispersal are common themes; however one is of particular interest to ecotonal movement.

The authors describe a process of “wave mortality” where high tree mortality in one patch begets mortality in adjacent communities. This is an interesting insight for heavily

managed ecosystems such as the LTB. Common causes of patchy tree mortality in the Tahoe Basin are natural occurrences (e.g. insect attack, avalanche) or human activity (e.g. prescribed thinning or fires).

In essence, wave mortality and climate change returns the present discussion to one of the primary determinants of species richness as discussed in Adams (2008)—disturbance. If disturbance drives change in diversity, then models capable of pinpointing where and at what scale change is occurring within an ecosystem are valuable tools for land managers. These models are unlikely to pinpoint the underlying drivers of change, but a series of these models are more than capable of identifying points of change that may have otherwise gone undetected and warrant further management attention.

Finally, recall that these models are not mere surveys of known species richness. They are predictions drawn from sampled data. As such, advanced versions of these models could be capable of creating what-if scenarios. This serves two purposes. First, it overcomes the inherent downside of temporal studies. Forest landscapes evolve across decades, so identifying the effects of management efforts within the real-world landscape are inevitably delayed by many years. Second, the ability to test prescribed treatments in for their potential effects would be an extremely valuable resource.

5.2.2 Relationship to Recent Research

Chiarucci (2012) provides commentary on a piece of research cited throughout this thesis, Xu et al. (2011). Chiarucci laments that despite Xu et al.'s exhaustive labor evaluating the species richness prediction capabilities of a dozen area-based and non-parametric methods, no approach succeeded in hitting an accurate mark. That is, every procedure tested by Xu et al. either over or

under estimated true species richness. Chiarucci further cites many similar research efforts over the last two decades—including interpolation and extrapolation—and decries the fact that no real progress has been made. Chiarucci concedes that truly accurate species richness predictions are limited to areas that do not extend far beyond the area sampled.

Chiarucci’s commentary expresses “difficulty” in the sense that highly accurate estimators are elusive despite decades of research. This research addresses difficulty in the context of accessibility. It is fair so assume the accuracy range of the procedure reviewed by Chiarucci represent the state of the art. No quantitative assessment of accuracy was conducted as part of this research; however, empirical assessments were positive enough to reasonably assume the output of these models fall within the accuracy range of other documented procedures. A common thread among species richness prediction algorithms in the literature is that they are complex and difficult to perform without high-level mathematical knowledge and skills. The models in this thesis can be created with intermediate geospatial skills. As such, this research contributes a procedure that provides access to the state of the art for managers and researchers with modest mathematical backgrounds.

The accuracy of the models in this research is likely related to the dense presence-only data used with the Maxent application. TBEVM data meets the best accuracy criterion espoused by Chiarucci (2012) as it covers much of the Basin. Very little upscaling is required by the Maxent algorithm. Its primary function is to etch out the most probable boundaries of species richness homogeneity based on background factors and within the dense dataset. This is clearly a less rigorous task than extrapolating or interpolating between sparse (and possibly inaccurate) field samples.

Most species richness prediction algorithms depend on quadrat or transect data (e.g. Dupuis and Goulard 2011). Field collection of species richness data is tedious at best and is commonly fraught with significant biases. This leads to datasets that are insufficient, inconsistent and often inaccurate (Engemann et al. 2014).

The TBEVM data used in this research is remote sensing-based. Remotely sensed data is capable of providing much more study area coverage than field-collected samples. A good deal of research is centered on deriving species richness from remotely sensed data (Camathias et al. 2013, Cord et al. 2014, Fricker et al., 2015, abstract). Applying these dense, remotely sensed datasets to mathematical models has the potential for increasing accuracy in both area-based extrapolation and distribution models.

Remotely sensed datasets are not an accuracy cure-all, however. Classification algorithms limit the accuracy of satellite imagery. As noted in Chapter 3, TBEVM for example is incapable of discriminating Jeffrey Pine from Ponderosa Pine. In effect, one type of bias is being exchanged for another. The increased density of remotely sensed datasets, however, is likely a reasonable tradeoff for these biases.

Lastly, the work of Dubuis et al. is regularly cited in this thesis. Their work provided a non-spatially explicit framework for the conduct of this research. Dubuis et al. note that directly estimating species richness from environmental factors eliminates knowledge of specific species composition. They list this as a serious limitation. That is not necessarily the case. Human nature inevitably induces bias into decision-making. Anonymous data provide an opportunity to support or refute management paradigms without bias for particular species at least until field surveys are conducted. This is a good thing. Providing an opportunity to reassess or reaffirm management practices based on blind data may be the most significant outcome of this research.

5.2 Suggestions for Future Work

This final section provides a few suggestion for making this work more complete, as well as deriving additional benefit from the results of this research.

5.2.1 Modeling Realistically Dense Data

The extremely dense TBEVM dataset was purposely chosen for its expansive coverage of the LTB. The dataset is older; however, that limitation was countered by the high likelihood of creating an accurate surface. Even with emerging remote sensing techniques, the likelihood that users of this model will have access to extremely dense and carefully vetted data is unlikely. An investigation into how the model degrades with decreasing density is worthwhile.

5.2.2 Field Validation

Time restrictions combined with the winter season precluded field validation of the models in this research. A field investigation should determine which species richness borders are accurate, which are not and collect evidence that sheds light on variables that affect the predictions.

Border regions predicted by the model, particularly unexpected ones, should be investigated for a biological or ecological explanation. This research extension could provide valuable insight to forest managers interested in incorporating diversity indicators into their decision making processes.

5.2.3 Temporal Study

The study of borders and ecotones is most useful when done over time. An animated time series of models will provide a level of insight into the evolution of diversity around the LTB that a single snapshot cannot approach. Clearly, the most significant barrier to this research extension is the availability of adequate data. As mentioned in the preceding section, however, remote sensing of species richness is a heavy focus of recent research. Options from the latest literature should be explored.

Alternatively, the model could be expanded to include the capability for creating the what-if scenarios described in the previous section. Carefully altering existing data and then inputting them into the model could accomplish this. Whatever the method, a temporal expansion of this technique is necessary to extract its maximum benefit.

REFERENCES

- Adams, J. 2009. "Local-Scale Patterns in Species Richness." In *Species Richness*. ed. Adams, J. 1–46. Springer: Berlin Heidelberg.
- Adler, P.B., J. HilleRisLambers, and J. M. Levine. 2007. "A Niche for Neutrality." *Ecology Letters* 10, no. 2:95–104.
- Akatov, V. V., and T. V. Akatova. 2013. "Species Pool, Species Richness, Density Compensation Effect, and Invasibility of Plant Communities." *Russian Journal of Biological Invasions* 4, no.1:1–11
- Arrhenius, O. 1921. "Species and Area." *Journal of Ecology* 9: 95–99.
- Barbour, M., E. Kelley, P. Maloney, D. Rizzo, E. Royce, and J. Fites-Kaufmann. 2002. "Present and Past Old-Growth Forests of the Lake Tahoe Basin, Sierra Nevada." *U.S. Journal of Vegetation Science* 13:461–472.
- Bekker, M. F., and G. P. Malanson. 2008. "Linear Forest Patterns in Subalpine Environments." *Progress in Physical Geography* 32, no. 6 (December):635–53.
- Bestelmeyer, B. T., J. R. Miller, and J. A. Wiens. 2003. "Applying Species Diversity Theory to Land Management." *Ecological Applications* 13, no. 6 (December):1750–61.
- Borcard, D., P. Legendre, and P. Drapeau. 1992. "Partialling Out the Spatial Component of Ecological Variation." *Ecology* 73, no. 3 (June):1045–55.
- Bormann, F. H., and G. E. Likens. 1994. *Pattern and Process in a Forested Ecosystem: Disturbance, Development, and the Steady State Based on the Hubbard Brook Ecosystem Study*. Springer: Verlag.

- Camathias, L., A. Bergamini, M. Küchler, S. Stofer, and A. Baltensweiler. 2013. “High-Resolution Remote Sensing Data Improves Models of Species Richness.” *Applied Vegetation Science* 16, no. 4:539–51.
- Chiarucci, A. 2012. “Estimating Species Richness: Still a Long Way off!” *Journal of Vegetation Science* 23, no. 6:1003–5.
- Chisholm, R. A., Muller-Landau, H. C., Abdul Rahman, K., Bebber, D. P., Bin, Y., Bohlman, S. A., Bourg, N. A., Brinks, J., Bunyavejchewin, S., Butt, N., Cao, H., Cao, M., Cárdenas, D., Chang, L.-W., Chiang, J.-M., Chuyong, G., Condit, R., Dattaraja, H. S., Davies, S., Duque, A., Fletcher, C., Gunatilleke, N., Gunatilleke, S., Hao, Z., Harrison, R. D., Howe, R., Hsieh, C.-F., Hubbell, S. P., Itoh, A., Kenfack, D., Kiratiprayoon, S., Larson, A. J., Lian, J., Lin, D., Liu, H., Lutz, J. A., Ma, K., Malhi, Y., McMahon, S., McShea, W., Meegaskumbura, M., Mohd. Razman, S., Morecroft, M. D., Nytch, C. J., Oliveira, A., Parker, G. G., Pulla, S., Punchi-Manage, R., Romero-Saltos, H., Sang, W., Schurman, J., Su, S.-H., Sukumar, R., Sun, I.-F., Suresh, H. S., Tan, S., Thomas, D., Thomas, S., Thompson, J., Valencia, R., Wolf, A., Yap, S., Ye, W., Yuan, Z., Zimmerman, J. K. 2013. “Scale-Dependent Relationships Between Tree Species Richness and Ecosystem Function in Forests.” *Journal of Ecology*. 101:1214–1224.
- Cord, A. F., D. Klein, D. S. Gernandt, J. A. Pérez de la Rosa, and S. Dech. 2014. “Remote Sensing Data Can Improve Predictions of Species Richness by Stacked Species Distribution Models: A Case Study for Mexican Pines.” *Journal of Biogeography* 41, no.4:736–48.
- Cornelius, J. M. and J. F. Reynolds. 1991. “On Determining the Statistical Significance of Discontinuities within Ordered Ecological Data.” *Ecology* 72, no. 6:2057

- DeClerck, F. A. J., M. G. Barbour, and J. O. Sawyer. 2006. "Species Richness and Stand Stability in Conifer Forests of the Sierra Nevada." *Ecology* 87:2787–2799.
- Dubuis, A., J. Pottier, V. Rion, L. Pellissier, J. Theurillat, and A. Guisan. 2011. "Predicting Spatial Patterns of Plant Species Richness: A Comparison of Direct Macroecological and Species Stacking Modelling Approaches." *Diversity and Distributions* 17, no. 6:1122–31.
- Dupuis, J. A., and M. Goulard. 2011. "Estimating species richness from quadrat sampling data: A general approach." *Biometrics* 67, no. 4:1489–97.
- Elliot-Fisk, D. 1996. "Lake Tahoe Case Study." In *Sierra Nevada Ecosystem Project-- Final Report to Congress, Vol II: Assessment and Scientific Basis for Management Options*, ed. D. C. Erman, 217–268. Davis, CA: University of California, Centers for Water and Wildland Resources.
- Elith, J, S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. "A Statistical Explanation of MaxEnt for Ecologists." *Diversity and Distributions* 17, no. 1:43–57.
- Engemann, K., B. J. Enquist, B. Sandel, B. Boyle, M. Jorgensen, N. Morueta-Holme, R. K. Peet, C. Violle, J. Svenning. 2014. "Limited Sampling Hampers "Big Data" Estimation of Species Richness in a Tropical Biodiversity Hotspot." *Ecology and Evolution* 5, no. 3: 807–20.
- Etienne, R. S. and J. Rosindell. 2011. "The Spatial Limitations of Current Neutral Models of Biodiversity." *PLoS ONE* 6, no. 3:e14717.
- Fricker, G. A., J. A. Wolf, S. S. Saatchi, and T. W. Gillespie. 2015. "Predicting Spatial Variations of Tree Species Richness in Tropical Forests from High Resolution Remote Sensing." *Ecological Applications* (March): in press
- Gewin, V. 2006. "Beyond Neutrality--Ecology Finds its Niche." *PLoS Biology* 4 no. 8:e278.

- Gleason, H. A. 1926. "The Individualistic Concept of the Plant Association." *Bulletin of the Torrey Botanical Club* 53, no. 1 (January):7–26.
- Graham, C. H. and R. J Hijmans, 2006. "A Comparison of Methods for Mapping Species Ranges and Species Richness." *Global Ecology and Biogeography* 15:578–587.
- Harte, J., T. Zillio, E. Conlisk, and A. B. Smith. 2008. "Maximum Entropy and the State-Variable Approach to Macroecology." *Ecology* 89, no. 10 (October):2700–1.
- Heath, James P. 1971. "Changes in Thirty-One Years in a Sierra Nevada Ecotone." *Ecology* 52, no. 6 (November):1090–2.
- Hennenberg, K. J., D. Goetze, L. Kouamé, B. Orthmann, and S. Porembski. 2005. "Border and Ecotone Detection by Vegetation Composition Along Forest-Savanna Transects in Ivory Coast." *Journal of Vegetation Science* 16, no. 3:301–10.
- Hubbell, S. P. 2008. *Monographs in Population Biology, Volume 32: Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ, USA: Princeton University Press.
- Hymanson, Z. P., and M.W. Collopy. 2010. *An Integrated Science Plan for the Lake Tahoe Basin: Conceptual Framework of Research Strategies*, Pacific Southwest Research Station General Technical Report. Washington D.C.: U.S. Department of Agriculture, Report Number, PSW–GTR–226.
- Kark, S. and B. J. van Rensburg. 2006. "Ecotones: Marginal or Central Areas of Transition?" *Israel Journal of Ecology & Evolution* 52, no. 1:29–53
- Kolström, M, and J. Lumatjärvi. 1999. "Decision Support System for Studying Effect of Forest Management on Species Richness in Boreal Forests." *Ecological Modelling* 119, no. 1:43–55

- Kroger, R., L. M. Khomo, S. Levick, and K. H. Rogers. 2009. "Moving Window Analysis and Riparian Boundary Delineation on the Northern Plains of Kruger National Park, South Africa." *Acta Oecologica* 35:573–580.
- Kupfer, John A., and David M. Cairns. 1996. "The Suitability of Montane Ecotones as Indicators of Global Climatic Change." *Progress in Physical Geography* 20, no. 3 (September):253–72.
- Lake Tahoe Basin Management Unit, and United States Forest Service. 2007. Proposed Action for South Shore Fuel Reduction and Healthy Forest Restoration Project: A Healthy Forest Restoration Act Project. Available from http://www.fs.usda.gov/Internet/FSE_DOCUMENTS/fsm9_046293.pdf
- Lam, T. Y. and C. Kleinn. 2008. "Estimation of Tree Species Richness from Large Area Forest Inventory Data: Evaluation and Comparison of Jackknife Estimators." *Forest Ecology and Management* 255 no. 3:1002–1010.
- Lohbeck, M., L. Poorter, H. Paz, L. Pla, M. van Breugel, M. Martínez-Ramos, and F. Bongers. 2012. "Functional Diversity Changes During Tropical Forest Succession." *Perspectives in Plant Ecology, Evolution and Systematics* 14 no. 2:89–96.
- Martin-Queller, E., A. Gil-Tena, and S. Saura. 2011. "Species Richness of Woody Plants in the Landscapes of Central Spain: The Role of Management Disturbances, Environment and Non-Stationarity." *Journal of Vegetation Science* 22:238–250.
- Martín-Queller, E., Diez, J. M., Ibáñez, I., Saura, S. 2013. "Effects of Silviculture on Native Tree Species Richness: Interactions between Management, Landscape Context and Regional Climate." *Journal of Applied Ecology* 50:775–785.

- Merow, C, M. J. Smith, and J. A. Slander. 2013. "A Practical Guide to Maxent for Modeling Species Distributions: What it Does, and Why Inputs and Settings Matter." *Ecogeography* 36:1058–1069
- Orrock, J. L. and J.I. Watling. 2010. "Local Community Size Mediates Ecological Drift and Competition in Metacommunities." *Proceedings of the Royal Society of Biological Sciences* 277 no. 1691:2185–2191.
- Perry, L.W. 2002. "Landscapes, Space and Equilibrium: Shifting Viewpoints." *Progress in Physical Geography* 26:339–359.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. "Maximum Entropy Modeling of Species Geographic Distributions." *Ecological Modelling* 190 (3–4) (1/25):231–59.
- Prates-Clark, C. D., Sassan S. S., and D. Agosti. 2008. "Predicting Geographical Distribution Models of High-Value Timber Trees in the Amazon Basin using Remotely Sensed Data." *Ecological Modelling* 211, no. 3–4:309–23
- Raumann, C. G. and M. E. Cablk. 2008. "Change in the Forested and Developed Landscape of the Lake Tahoe Basin, California and Nevada, USA, 1940–2002." *Forest Ecology and Management* 255:3424–3439.
- Rykiel Jr., E. J. 1996. "Testing Ecological Models: The Meaning of Validation." *Ecological Modelling*. 90, no. 3:229–44.
- Sagar, R., and J. S. Singh. 2006. "Tree Density, Basal Area and Species Diversity in a Disturbed Dry Tropical Forest of Northern India: Implications for Conservation." *Environmental Conservation* 33, no. 3:256–62.
- Scheiner, S.M. 2003. "Six Types of Species-Area Curves." *Global Ecology and Biogeography* 12:441–447.

- Simpson, E. H. 1949. "Measurement of Diversity." *Nature* 163:688.
- Smith, A., B. Page, K. Duffy, and R. Slotow. 2012. "Using Maximum Entropy Modeling to Predict the Potential Distributions of Large Trees for Conservation Planning." *Ecosphere* 3:56.
- Sprugel, D. G. 1991. "Disturbance, Equilibrium, and Environmental Variability: What is Natural Vegetation in a Changing Environment?" *Biological Conservation* 58:1–18.
- Stanturf, J. A., B. J. Palik, M. I. Williams, R. K. Dumroese, and P. Madsen. 2014. "Forest Restoration Paradigms." *Journal of Sustainable Forestry* 33:S161–S194.
- Taylor, A. H., A. M. Vandervlugt, R. S. Maxwell, R. M. Beaty, C. Airey, and C. N. Skinner. 2014. "Changes in Forest Structure, Fuels and Potential Fire Behaviour since 1873 in the Lake Tahoe Basin, USA." *Applied Vegetation Science* 17:17–31.
- Ugland, K. I., J. S. Gray, and K. E. Ellingsen. 2003. "The Species-Accumulation Curve and Estimation of Species Richness." *Journal of Animal Ecology* 72, no. 5 (September):888–97
- U.S. Department of Agriculture Forest Service. Pacific Southwest Southwest Region. 2013. *Volume II Forest plan: Revised Land and Resource Management Plan, Lake Tahoe Basin Management Unit*. Report Number, R5–MB–254B. Available at http://www.fs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb5440880.pdf
- Wang, P., and J. C. Finley. 2011. "A Landscape of Shifting-Mosaic Steady State in Lassen Volcanic National Park, California." *Ecological Research* 26:191–9
- Waring, R. H., N. C. Coops, W. Fan, and J. M. Nightingale. 2006. "MODIS Enhanced Vegetation Index Predicts Tree Species Richness Across Forested Ecoregions in the Contiguous U.S.A." *Remote Sensing of Environment* 103:218–26.

- Whittaker, R. H. 1960. "Vegetation of the Siskiyou Mountains, Oregon and California." *Ecological Monographs* 30, no. 4 (Oct.):407
- Xu, H., Liu, S., Li, Y., Zang, R., He, F. 2012. "Assessing Non-Parametric and Area-Based Methods for Estimating Regional Species Richness." *Journal of Vegetation Science*, 23:1006–1012.
- York, P., P. Evangelista, S. Kumar, J. Graham, C. Flather, and T. Stohlgren. 2011. "A Habitat Overlap Analysis Derived from Maxent for Tamarisk and the South-Western Willow Flycatcher." *Frontiers of Earth Science* 5:120–129.
- Zar, J. 2010. *Biostatistical Analysis*. 5th ed. Upper Saddle River, NJ: Pearson.
- Zhang, Y., H. Y. Chen and P.B. Reich. 2012. "Forest Productivity Increases with Evenness, Species Richness and Trait Variation: a Global Meta-Analysis." *Journal of Ecology*. 100:742–749.

APPENDIX 1: COMPLETE JACKKNIFE ANALYSIS FOR SPECIES RICHNESS

This appendix contains the Maxent jackknife analysis charts against AUC for each species richness SDM.

