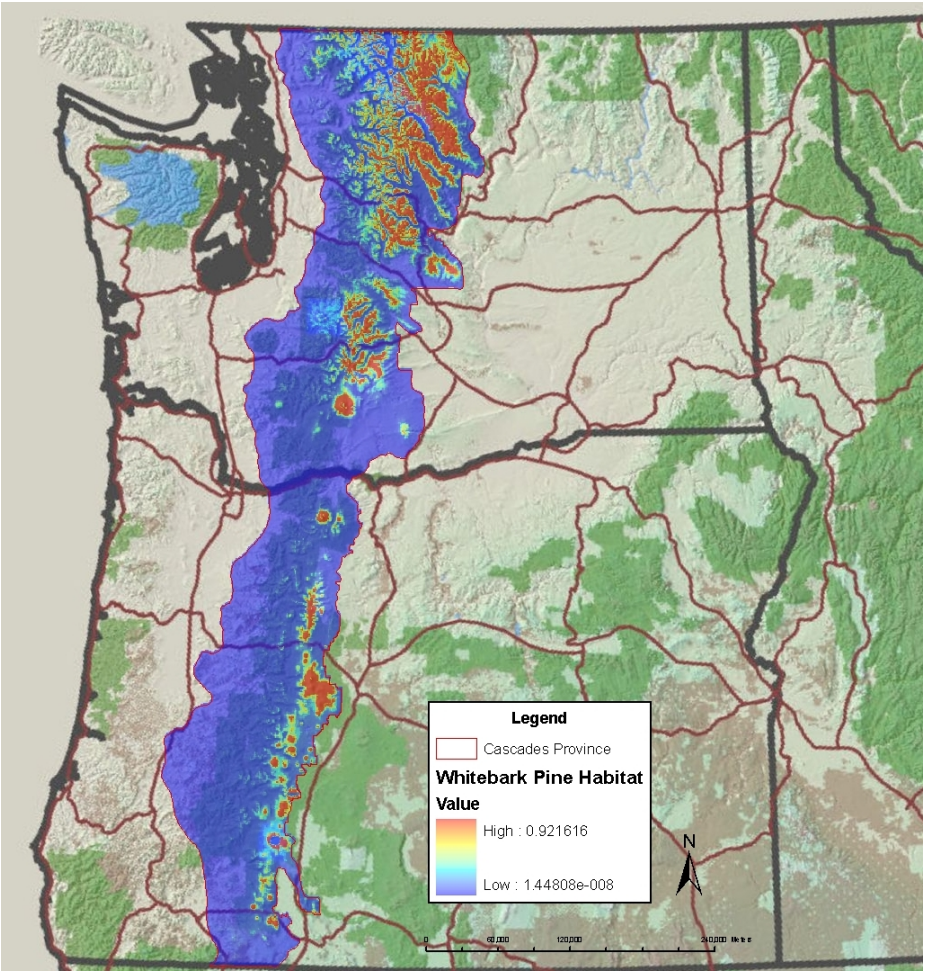
Use of Species Distribution Models in Conservation Biology

Report to the Interagency Special Status and Sensitive Species Program
USDA Forest Service, Region 6 and
USDI Oregon/Washington Bureau of Land Management

27 February 2013

Richard Helliwell and Josh Chapman
Umpqua National Forest, Roseburg, Oregon



Cite as: Helliwell, R and J. Chapman. 2013. Use of Species Distribution Models in Conservation Biology. Report on file with: Interagency Special Status/Sensitive Species Program. U.S. Department of Interior, Bureau of Land Management, Oregon/Washington and U.S. Department of Agriculture, Forest Service, Region 6. Portland, Oregon. 22p.

Table of Contents

Introduction …….…. 2
The Species Niche Concept ……..... 2
Types of Species Distribution Models ……. 4
Maxent Model Inputs ……... 6
Species Presence Records ……..…6
Environmental Variables ……...… 7
Running the Model ……. 8
Maxent Model Outputs ……. 9
Continuous Prediction Map …… 9
Model Evaluation ……. 10
Determination of a Threshold for Classification of Habitat …… 12
Management Applications of Species Distribution Models …… 13
Conservation Planning and Reserve Design …… 13
Model Extrapolation – Climate Change and Invasive Specie Range Expansion … 15
Models to Guide Surveys …… 16
Project Effects Analysis …… 17
Concluding Remarks …… 18
Literature Cited ……19

Introduction

As of this writing, the US Forest Service and Bureau of Land Management (BLM) in the Pacific Northwest were responsible for management of 821 sensitive species along with tracking another 466 strategic species and 315 survey and manage species. Although there is some overlap between the latter category with the first two, this represents well over one thousand fungi, lichens, bryophytes, vascular plants, mammals, birds, amphibians, reptiles, invertebrates and fish species of conservation concern. Responsible management of such an immense and diverse assemblage of organisms requires finding efficient means of prioritizing, inventorying and planning for these taxa. Species distribution models (SDM) can provide a valuable tool for identifying the range and distribution of species often revealing patterns that would not be otherwise evident from existing data. This information is in turn, valuable for guiding inventory, informing conservation strategies, designing preserves, and evaluating project effects.

The focus will be on Maxent, a SDM that can generate predicted habitat using only existing presence records (i.e. it does not require both presence and absence data). Maxent is a machine learning software package that uses a maximum entropy approach for species distribution modeling. Maxent uses species presence points and a set of environmental variables selected for their significance for the species whose habitat is to be modeled to predict environmental suitability for the species as a function of the given environmental variables (Phillips et al. 2006). In 1979, George E.P. Box declared that, "essentially, all models are wrong, but some are useful". The purpose of this paper is to introduce biologists and botanists to basic modeling theory, model creation, and suggest ways models may be used in practical ways to benefit conservation planning, species surveys, and project level effects analyses so that they may gain a better understanding of how some models can indeed be useful.

The Species Niche Concept

Species Distribution Models are based on the Hutchinsonian concept of the species niche. A basic appreciation of this concept is fundamental to understanding SDMs. In 1957 G. Evelyn Hutchinson expanded upon the concept of niche by describing it mathematically in three-dimensional space. While a predicted habitat map produced by a SDM is displayed in geographic space, the model that produced the map analyzed the environmental space of the target species. This environmental space reflects the various inputs that shape any species habitat requirements for food, shelter and reproduction. If an organism has only two requirements then its environmental space could be graphed on a simple figure with an x and y axis. However, organisms would typically have myriad requirements requiring a multi-dimensional graph or three-dimensional model. This space, or hypervolume, defined in three-dimensional space is what Hutchinson called the **fundamental niche** of a species (Figure 1). When this fundamental niche is overlapped with the available environment (what Hutchinson called the biotope) the area of intersection represents the area that is actually available for the species to occupy (Figure 1). This area is known as the **realized niche**. Because the fundamental niche would typically include an optimal part with suboptimal areas towards the boundaries, the realized niche may only be a subset of the overlap. This is actually a simplification of the real processes that shape habitat because it does not explicitly address

important biotic processes among them: competition, source/sink habitat, dispersal limitations, facilitation and symbiosis (Franklin 2009). All of these processes may result in a species occupying less than its full realized niche.

This species niche concept has remained a central tenet of ecology and is a fundamental principle to modeling of species distribution although the appropriate term for what modeled distributions represents remains in dispute (Austin 2002, Guisan and Thuiller 2005, Soberon and Townsend Peterson 2005, Franklin 2009, Sillero 2011). Some have suggested that because the known sites that are input into models are themselves products of both abiotic and biotic interactions, then models must, by extension, represent the realized niche (Guisan and Thuiller 2005). Others have suggested that since it is usually abiotic variables that drive the models, the output would logically represent the fundamental niche (Soberón and Townsend Peterson 2005). Biotic processes also often operate at a finer scale than the scale of the modeled output. For our purposes, the salient point is that any model must be understood within the context of the organism. Niche is a property of the species rather than of the environment.

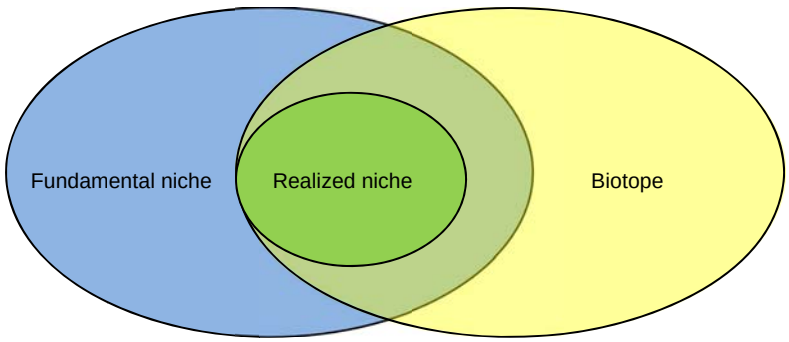


FIGURE. 1. Diagrammatic representation of the relationship between fundamental niche, realized niche and biotope (adapted from Figure 3.1.a of Franklin 2009).

Another implicit assumption of SDMs is that the species is in equilibrium within its environment. If either the species is occupying an unstable environment due to historically uncharacteristic disturbance or lack of disturbance, exotic species interactions, pollutants, or climate change this can affect the veracity of the model. Likewise if the species is exhibiting genetic drift or hybridization or simply if the taxon is inadequately circumscribed then this too may result in a less robust model. For these reasons, models that predict future distribution based on climate change or potential expansion of invasive species must be interpreted within the assumptions of the model. Such models are necessarily limited by incomplete knowledge of future biotic interactions and potential dispersal barriers. The following is a brief summary of types of SDM's and required information for creating SDM's. For a thorough review of modeling history, evolution and theory, see Elith and Leathwick 2009.

By way of a final note on terminology, it should be noted that there are different names in the literature for species distribution models. These include ecological niche model, habitat suitability models and others. For the purposes of this overview we will mostly use the terms

niche and habitat interchangeably with the acknowledgement that this imprecision is not desirable but it does reflect the entangled history of the two terms (Chase and Leibold 2001).

Types of Species Distribution Models

The extraordinary memory and speed of modern computers has allowed for an increasingly myriad number of modeling algorithms available for constructing SDMs, all of which have their strengths and weaknesses. The simplest means of identifying habitat involves a non-statistical filtering of the landscape through a Geographic Information System (GIS). Typically this would involve identifying, through empirical or intuitive means, one or more environmental variables and intersecting the spatial layers to create a map. This may produce a useful model if the target species exhibits strong fidelity to one or a few well-represented environmental variables in GIS and these relationships are well-understood. For example, if the target species occurs only above timberline in subalpine seeps and you are only concerned with the habitat under your jurisdiction, then you may not require a sophisticated model to meet your needs. These habitats are usually readily delineated and represent a small fraction of the landscape. However, if you have the apparent habitat but the species isn't known from your area, this simple model can't predict if past inventory has been inadequate or if habitat is actually unsuitable for some other reason. Because a simple filter exercise doesn't analyze how the environmental variables interact with each other or the target organism it offers only limited predictive or explanatory capacity.

Species distribution models generally conform to one of two types: statistical models and machine-learning models. Statistical models typically involve use of some form of regression to relate a dependent (or response) variable to an independent (or predictor) variable(s). There are number of statistical models that employ regressions including: Generalized Linear Models (GLM), Generalized Additive Models (GAM), Resource Selection Function (RSF) and Non-Parametric Multiplicative Regressions (NPMR). All of these use various link functions to fit the predictor function to the response function. Each can perform well depending upon the type of data available. One requirement of each of these models is that they all require that the input data be collected using a randomized sampling design that identifies both presence and absence locations for the target organism.

Having absence records can improve model performance, however absence data collection is expensive and can be difficult to confidently demonstrate, particularly for species that are not well understood. For example, fungi typically fruit inconsistently and only briefly so a sampling design based on sporocarp production would require multiple visits per season over several years to adequately determine absence. Distinguishing non-detection from absence is routinely an issue with sampling wildlife as well. In one mark-recapture study of small mammals, detection error estimates varied greatly among seven species with one species exhibiting a 23% rate of erroneous non-detection (Gu and Swihart 2004). Detection reliability typically increases with more intensive sampling but this can be cost-prohibitive.

Several models are available that require only presence data. Maxent and Ecological Niche Factor Analysis (ENFA) are two machine learning methods that compare environmental data at presence locations against the background data throughout the rest of the study area. Machine

learning methods are algorithms that inductively learn the mapping function or classification rules directly from the training data (Franklin 2009). Both statistical and machine learning models are correlative methods in that the algorithms can reveal environmental relationships that correlate with the areas where a species is known to occur but can't infer causal relationships. A third type of model that we will refer to only slightly is mechanistic models that attempt to link fine scale environmental data directly to physiological or functional traits of species, rather than simply presence (Kearney et al. 2010). This type of model can reveal causal relationships with the environment but is very data intensive and requires a greater understanding of the target species than is often the case.

Presence-only models allow existing specimen records and observation databases to be utilized often without any additional fieldwork. However, collection area bias can adversely influence the resulting model so the quality of these data requires review. For example, herbarium and museum records tend to be associated with roads where access is easy. Regions that are close to populated areas, particularly universities, also tend to be better represented than remote rural areas. Another factor that should be considered is the age of record. In many instances, a collection that was made from a forest or field a century ago is now an urban area. Also, presence records gathered during a different climatic cycle may obscure the role of climate in its current potential distribution. Sample bias leads to overfitting of the data, particularly for small data sets, in which the predicted habitat is too closely fitted to the presence points. Maxent utilizes a regularization feature to reduce overfitting and has been demonstrated to perform well with few species presence records.

Another consideration in model selection is the type of environmental data that are available. Some models can utilize only continuous data (a range of relative values) but not categorical data (discrete classes of data values). Most GIS data that are of interest to modelers are continuous. This includes variables such as climatic data, elevation and aspect related data. In some cases categorical data can be converted to binary values based on some predetermined threshold in order to be used as continuous data. Soil Series would be an example of categorical data. The values associated with each geographic unit stand on their own, that is to say the series name is not a value that can be quantifiably related to any other series name. On the other hand, a quantity associated with a soil series, such as percent rock, could be entered as continuous data.

Geospatially accurate, fine scale habitat variables are often very hard to come by and the modeler must look to surrogates for these variables. Vegetation variables are often not available across multiple landownerships, so often times modeled vegetation variables like those available from Landfire (www.landfire.gov), Gap Analyses (http://gapanalysis.usgs.gov/), LEMMA GNN (http://www.fsl.orst.edu/lemma/main.php?project=master&id=modelDownloads) or variables derived from Landsat (http://landsatlook.usgs.gov/) imagery (e.g., leaf cover, EVI (Enhanced Vegetation Index), NDVI (Normalized Difference Vegetation Index), brightness, greenness, wetnesss, reflectance, etc.) among others are used as surrogates to approximate finer scale habitat requirements. The modeler must have a good grasp on the ecology of the species of interest, or have access to species experts to review and suggest habitat variables that have geospatial data equivalents to use in creating SDMs.

Maxent, which is a contraction of the phrase maximum entropy, has been demonstrated to perform as well or better than other models (see Hernandez et al. 2006, Gibson et al. 2007, Townsend Peterson et al. 2007, Wisz et al. 2008, Elith and Graham 2009, Rupprecht et al. 2012). It requires only presence data, performs well with few presence points and can use both continuous and categorical data. It also is a relatively easy to use program that is available as a free download (http://www.cs.princeton.edu/~schapire/maxent/). For a more in-depth look at how Maxent works, and the statistics behind the model creation see: Phillips et al. 2006, Phillips 2008, Elith and Leathwick 2009 and Elith et al. 2011. Species Distribution models will be discussed in the context of Maxent throughout the remainder of this overview although the principles apply generally to other SDMs as well.

Maxent Model Inputs

Species Presence Records

Presence records inputs for model development are known as the **training data**. Because it is a density estimation method rather than a regression method, Maxent is able to perform well with very few presence points. For this reason having accurate, reasonably unbiased presence records can be more important than having large numbers of records, however more data are an advantage for cross-validation of the model. Maxent has been reported to perform well with as few as 5-10 presence points. Hernandez et al. (2006) found it to be little improved with more than approximately 50 known sites although Wisz et al. (2008) and Feeley and Silman (2011) found that predictions made from larger sample sizes generally outperformed those with smaller sample sizes. How many presence points are necessary will ultimately depend upon how representative the sample population is (i.e. how unbiased it is). The fidelity of the species to the environmental variables available in a GIS is also going to affect model predictions based on few training data. If presence data are based upon limited sampling from only a couple of locations with many observations, caution is needed in interpreting model results. Such a model can be useful for guiding additional inventory but it would likely be inadequate for other purposes.

The precision of the training data should generally be that of the pixel or grain size of the environmental variables. If the model is trained only to variables at a 1 km^2 scale then the precision of the known sites is less of a concern then if 30 m^2 environmental data are being employed. There should not be more than one training point per pixel (Figure 2). Where the training point is placed within each pixel is irrelevant. The presence point for a terrestrial species may be located in the middle of a lake if that lake is entirely contained within the pixel that the species truly occurs in. That this is true highlights the importance that the scale of the species to be modeled must be assessed against the scale of available environmental data in GIS. A narrow endemic whose distribution is predicated by microsite landscape features cannot be adequately modeled with coarse scale global data.

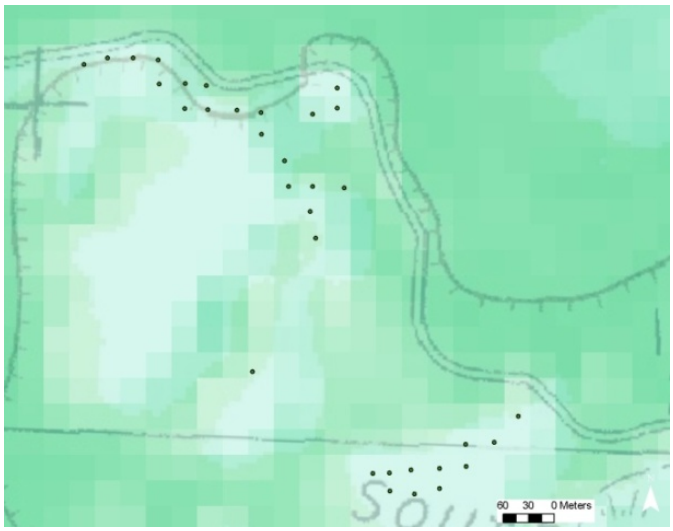


FIGURE. 2. An example of a data set in which redundant species locations have been removed so that there is only one location (black circles) per 30 m^2 pixel.

Environmental Variables

The model works within a Geographic Information System (GIS) environment so ultimately the precision of the model output is dependent upon the accuracy and precision of the spatial data that is available in GIS. Environmental variables that are chosen based on an assumed relationship with the organism will be more useful than simply throwing every available variable into the model to see what sticks. Environmental variables are generally selected based upon an individual's knowledge and scientific literature review of the habitat requirements of the species of interest, along with discussions with taxonomic experts for the species. The most commonly employed variables include: climate (e.g., WorldClim, DayMet), topography (DEMs, aspect, slope) and remote sensing derived vegetation variables (e.g., GNN or Landfire data). Some other variables such as waterbodies, streams, soils and geology are often useful but may not be consistently mapped across the intended study area. For example, mapped soils on private lands are typically mapped at a finer level of classification than on National Forest lands. Stream layer detail, particularly involving Class 4 streams, is often highly variable between watersheds and land owners. Detailed geology mapping is typically done by USGS quad and may not be available across the entire extent of the sampling region.

Vegetation data are typically derived from Landsat data in combination with topographic and climatic data. Gradient Nearest Neighbor (GNN) data available through the LEMMA website (http://www.fsl.orst.edu/lemma/main.php?project=master&id=modelDownloads) includes useful data on forest composition, structure and cover for Oregon and Washington. All variables must be converted from vector data (e.g. shapefile, feature class) to raster data. Digital elevation model (DEM) data may be utilized directly as a measure of elevation or transformed to represent units of aspect, slope, solar radiation or landform.

Spatial data are typically available at a range of scales. DEMs are available at a scale as fine as 10 m^2 while climate data are typically represented at 1 km^2. Variables of different grid sizes can be used together but they would all be masked to the scale of the finest variable. All variables used in a habitat model must be masked to the same extent and projection to ensure that each

variable's pixels extent and placement are identical. Highly correlated environmental variables should be removed otherwise they can result in misleading interpretation of the variable contribution to the model. When variables are closely correlated it allows for multiple, equally reasonable, pathways to generate the same model. It is common for many climatic variables, in particular, to be highly correlated with each other as well as to DEM variables. This evaluation is readily accomplished using ArcMap's Band Collection Statistics tool in the Spatial Analyst extension.

Finally, the accuracy of each of the environmental variables used needs to be considered in assessing the model. This is especially important in the case of vegetation variables, which are most often created from modeled relationships between orthophotography and fixed vegetation plots (like LEMMA GNN and Landfire data). Most data sources will provide the methodology in which the data were derived and in some cases accuracy assessments of specific products. For example, GNN evaluates each of their species distribution products against plot data using a Kappa statistic and their vegetation condition products with histograms. The accuracy of some of the individual plant species products, in particular, is quite low so this needs to be considered when using and interpreting these data.

Running the Model

Once the species presence records and environmental data have been properly formatted to a study area that is appropriate for species of interest, these data are input into Maxent's user friendly input screen (Figure 3). Maxent sequentially evaluates the values of each of the environmental variables for each pixel with a presence location against those pixels that don't have a presence noted (Figure 4). These non-presences are treated as background data (rather than pseudo-absences as in other methods) so the result of the analysis is a relative probability of occurrence for each pixel on a scale of 0 to 1. In actuality, the highest probability of occurrence is typically less than one therefore the values might read something like 0.00 to 0.89. By checking the "create response curves" box, the output will display graphically how each individual environmental variable affects the prediction (Figure 5). There are actually two sets of response curves. One set displays how the variables performed individually, the second displays how they performed in conjunction with the other variables.

Maxent Model Outputs

Continuous Prediction Map

The immediate output of Maxent is an .html file which includes a .png map of the predicted habitat rated on a continuous scale between 0.0 and 1.0 (Figure 6). The file also displays the contributions of each individual environmental variable to the model. In addition there will be a graph displaying the average omission and predicted area and a second graph illustrating sensitivity vs. specificity. If selected when the model is run, response curves of each variable will be displayed, both as they perform individually and in combination with the other variables. Maxent can also compute and display how each variable contributed to the model, both individually and in conjunction with the other variables. Post-model evaluation is facilitated by the inclusion of maps in ASCII

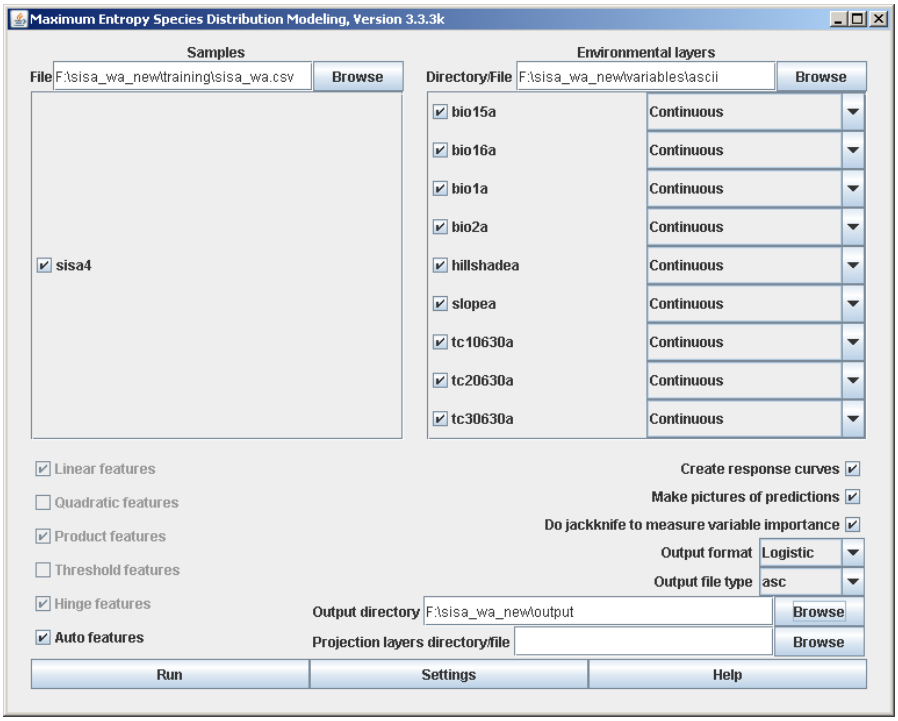


FIGURE. 3. An example Maxent data entry screen. The file with the species location or training data is entered at the left. Only a single species is being analyzed in this example. The environmental variables are on the right. Note that all the variables in this example are continuous variables.

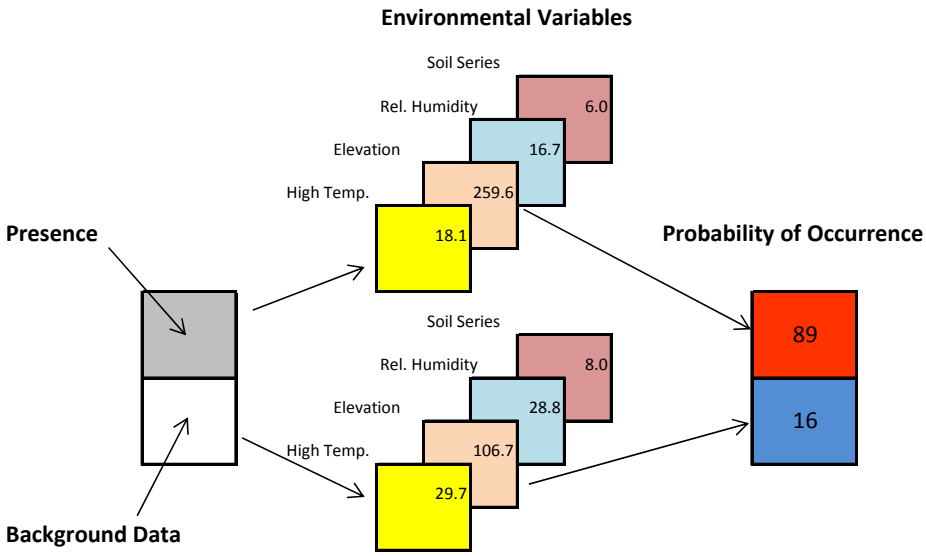


FIGURE. 4. A diagrammatic representation of how species presence data and environmental data interact to generate a mapped probability of occurrence on a pixel by pixel basis throughout a geographic area.

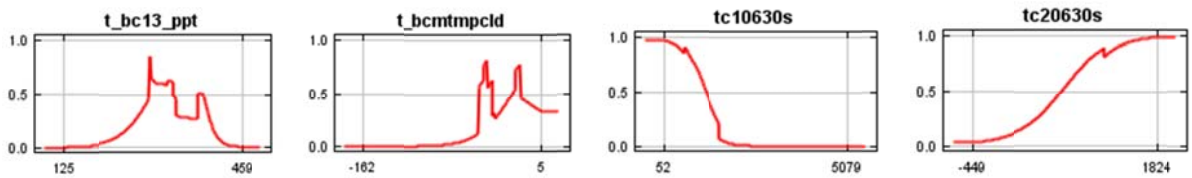


FIGURE. 5. Example response curves displaying how individual environmental variables affect the model.

format, lambda files that contain the computed values of the constants and excel tables with the predicted area, training and test omission data. The ASCII file can be converted into a raster file for inclusion back into GIS for further analysis.

Model Evaluation

Typically, if there are enough known sites then a proportion will be withheld from the model to be used for testing. The initially withheld data (testing data) would be used to rerun the model several times to see if a different set of known sites consistently results in the same predicted habitat as the initial data set (training data). With fewer known sites, sometimes the entire data set will be used to train the model and a subset of these data would be used to test the model. These multiple models can then be statistically compared to see how well it performed.

One of the most powerful tools for evaluating a model is the area under the receiver operating characteristic (ROC) curve or **AUC** (Figure 7). A value over 0.9 is very good; less than 0.5 means the model performed worse than a random. However, this is a better indicator of model strength for some models than for others. In particular, it is sensitive to the size and extent of the study area and the distribution of the species habitat within the study area. For example, let's say you are interested in identifying habitat for a species on a series of islands. You define the extent of your study area as a 100 mile radius around the group of islands and run your model. The model determines that all land is habitat while the ocean is not. You have now created a model that is 100% **accurate** (the ocean is indeed poor habitat) but completely useless because of the lack of model **precision** (we already knew this was a terrestrial species). Because the area of land within the study area is small relative to the ocean the model will score an excellent AUC value. Compared to open ocean, any dry land looks like good habitat to the model. This is how presence-only models work; they compare the environment at the known sites to the environment everywhere else within the extent of the study area. To correct for this we have to redefine our extent to include only the area where the organism we're interest in might reasonably be expected to occur. Ideally we would eliminate the ocean altogether from the model.

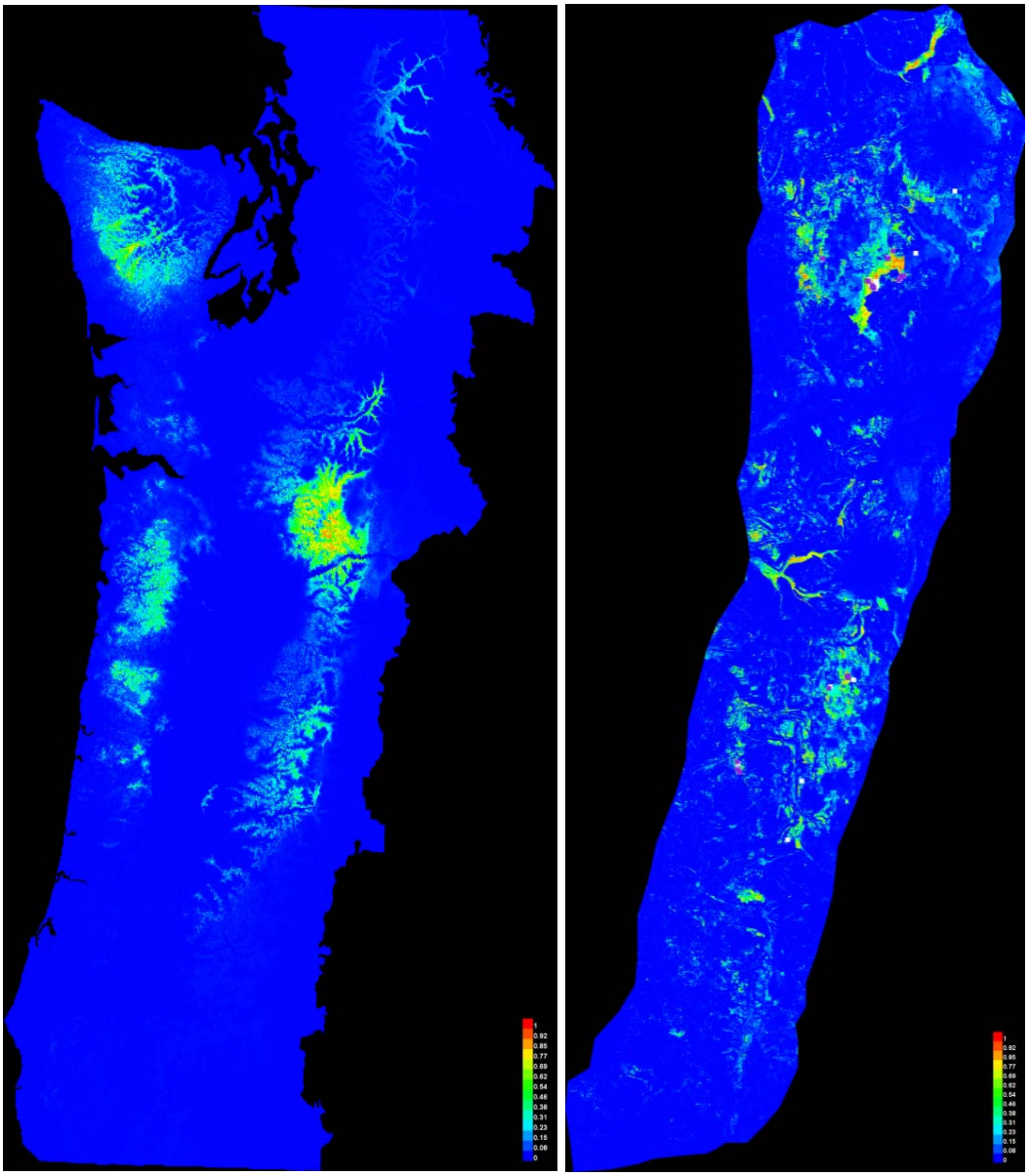


FIGURE. 6**.** Two examples of continuous prediction output maps generated by Maxent. Blue is non-habitat while increasingly warmer colors represent the best habitat. These maps are simply png. files but the ascii file can be converted into a raster for further analysis in GIS.

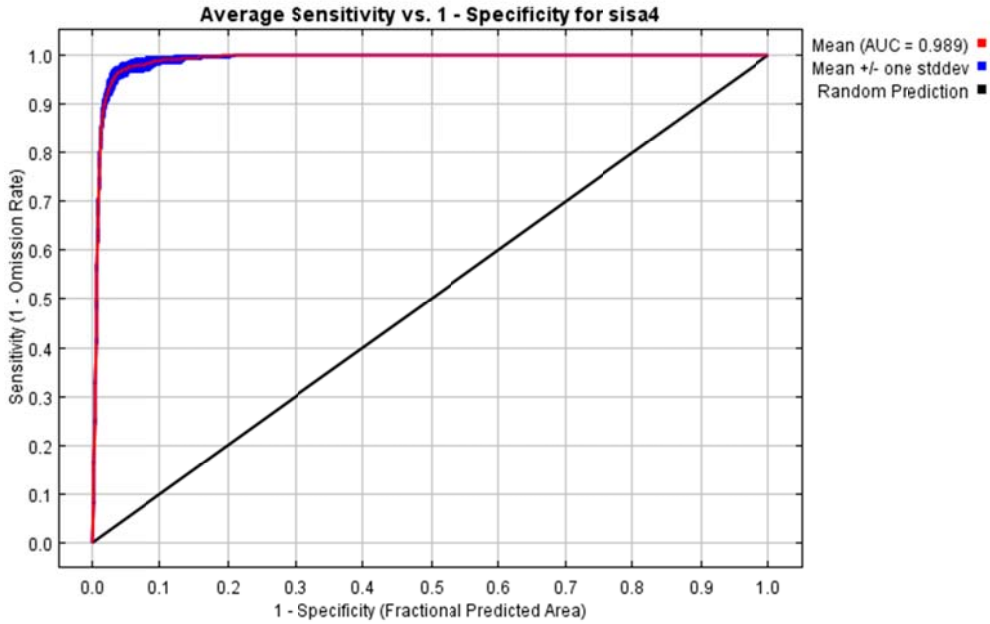


FIGURE 7. An example display of AUC. The black diagonal line represents what an output would look like if it performed only as well as would be predicted by random chance. The red line displays the actual output and blue is the spread of the data within one standard deviation. The AUC of 0.989 of this model is very high.

A more pertinent analogy here in Pacific Northwest would be an organism with limited vagility that occupies certain types of small meadow openings, rock outcrops or canopy gaps within a forested matrix. Modeling a species with such specific niche requirements is going to have a high AUC simply because of the limited and highly specific nature of its potential habitat within the landscape. It's not generally practical to confine the study area to such small openings but we can limit the extent of the study area to the range of the species. Because the range may not be known at the outset, this is often an iterative process to narrow the project extent. On the other hand, working with a wide-ranging bird or mammal with less specific habitat requirements is more likely to result in a model in which AUC is the most useful tool for assessing the veracity of your model.

Determination of a threshold for classification of habitat

The immediate spatial output of the Maxent model is displayed as continuous data. The best predicted habitat occurs at values closest to 1.0 while the poorest habitat is closest to 0.0. This raises the question of what value along this scale should we interpret as good habitat? The easiest way to address this is to evaluate at what values the model is capturing known sites or known absences. If the model is failing to adequately capture known sites within predicted habitat, this is known as an **error of omission**. Conversely, if the model is predicting too much suitable habitat then that would be an **error of commission**. For a presence-only model, such as Maxent, there are no empirical absence records so the model is evaluating predicted presence and absence between the training and test data. So for this situation, **sensitivity** refers to the proportion of presences that the model correctly predicted while **specificity** conversely identifies the absences that were correctly predicted. Identifying the habitat value where sensitivity and

specificity balance is one proven method of establishing a threshold for high-probability habitat. However, it is important to keep in mind that converting a continuous prediction map into high and low probability potential habitat is simply categorizing the data for convenience of interpretation; it doesn't change the value. Depending upon the objectives for the model, it may make more sense to emphasis maximum sensitivity, particularly if one is attempting to locate new species occurrences in the field with a limited budget.

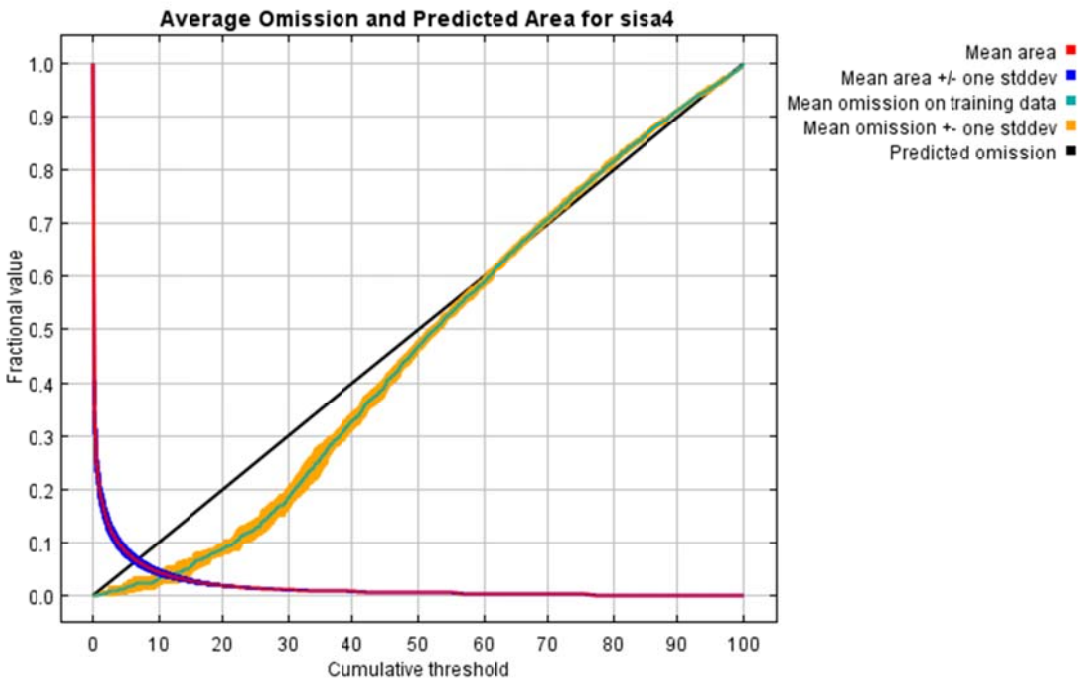


FIGURE 8. In this figure the black diagonal line represents the predicted omission so you want the omission on the training data (the green line) to closely fit this line. The yellow around the green line represents the spread around the mean.

Ultimately, for these models to be useful they need to be applied in the context of the organism that is being modeled and the environmental variables that contributed to the model. It is critical to appreciate that the pixels represent statistical approximations so they shouldn't be taken too literally. The pixels, after all, are square; certainly no one is suggesting that habitat occurs as square blocks.

Typically when designing an inventory based on a SDM, the predicted habitat should be interpreted to fit naturally occurring features that make sense for the organism. So that if a species habitat requirements are limited to rock outcrops or talus slopes, for instance, one would logically filter the predicted mapped habitat by these features. It should be mentioned however that we have had success with target species being found in predicted habitat that would not have been expected based on previously presumed habitat preferences.

The scale of the environmental variables that contributed the most to the model should also be considered. If climatic variables at a 1 km^2 scale account for most of the contribution to the

model, then it is probably unrealistic to assume a precision of more than about a kilometer. Typically, 30 m^2 is the finest resolution that is practical so this would be about the highest degree of precision that could be expected. SDMs wouldn't generally be able to identify microsites within a meadow complex that a species occupies because the scale is too small for the environmental variables available. Returning to the concept of niche, appreciation of the model requires looking at the variables that contributed to the model and putting this in the context of what we know about the species. Some models, for example, will necessarily map more habitat than will be occupied at any given time because the species has dispersal limitations or competes poorly with another species whose niche overlaps.

Management Applications of Species Distribution Models

While SDMs provide fascinating conceptual insight into a species' range, distribution and habitat requirements, it is the practical applications that are of immediate concern to land managers. Model outputs can be used in a variety of ways for conservation planning, species survey design and layout, or for analyzing the effects of proposed projects on species habitat. In this way, habitat models can augment expert opinion to provide a quantitative means of filling gaps in the knowledge base for species. This is particularly the case where species data are extremely limited, incomplete or spatially biased.

Conservation Planning and Reserve Design

Probably the most common application of SDMs is as a quantifiable approach for the identification of areas of high conservation value for designation of nature preserves (Wilson et al. 2005, Elith and Leathwick 2009). Historically, designation of conservation preserves has often been opportunistic rather than strategic. For example, wilderness designation in the United States has used as its principle limiting criterion simply the absence of roads. While designated wilderness areas offer tremendous scenic, recreational and conservation benefits, they do not necessarily represent a strategic, prioritized network for conservation of biological diversity. In this case, conservation of biological diversity was never the principle objective of wilderness areas so it would be unfair to evaluate them using this standard. Here we consider the utility of SDMs where the objective is to preserve biological diversity within a given landscape, either as conservation of a single species or multiple species.

The US Forest Service and Bureau of Land Management (BLM) in the Pacific Northwest use Conservation Strategies for guiding preservation of sensitive/special status species and Management Recommendations for Survey and Manage Species. Each of these documents requires an assessment of the range, distribution and habitat of the species in question although this information is typically incomplete. It is usually impossible to assess or rank the conservation value of individual sites or estimate what percentage of the actual range, distribution and abundance the currently known sites represent. As a result, the precautionary principle is invariably applied to all currently known and subsequently discovered sites, just as would likely have occurred without a conservation strategy.

In many cases, a SDM can provide the basis for a quantifiable assessment for prioritizing the value of specific areas across the landscape so that conservation efforts may be focused in the

most promising areas. In situations where the known or suspected species distribution is significantly incomplete, an SDM can provide a more meaningful basis for evaluation of the distribution by land allocation. This would be a quantifiable assessment of acreage and habitat quality by land allocation rather than a simple count of sites per allocation, thus assuring that the analysis represents actual habitat rather than simply a likely biased or opportunistic collection of sites.

Because modeled habitat is spatially explicit it also allows for predicted core habitat to be assessed against current habitat quality for a quantitative evaluation of restoration activities. Hunter et al. (2012) used Maxent to identify potential wetland mitigation sites in New York State. They found that the model correctly classified wetlands 92% of the time while the expert opinion model performed at only a 62% success rate. Taecker (2007) found Maxent to outperform CART for identifying and classifying Piedmont Prairie remnants for potential restoration. Similarly, SDM have been used to identify areas for reintroduction of threatened and endangered species. Examples of this include modeling habitat for reintroduction of an endangered tree (Adhikari et al. 2011) and an endangered freshwater mussel (Wilson et al. 2010). In some cases, habitat models may help to identify areas that are being prevented from being recolonized by recovering populations because of natural or anthropogenic dispersal barriers (Hernandez-Santin et al. 2012). Analysis of connectivity issues often involve habitat modeling as the first phase of the process to identify habitat patches which may be followed by further analysis such as graph theory (Decout et al. 2010) or least-cost modeling (Poor et al. 2012) to examine permeability of the landscape between the patches.

Perhaps the most exciting recent development in the application of habitat models is their use in conjunction with phylogenetic analysis to inform delineation of cryptic or otherwise difficult to define species. By identifying niche differentiation between genetically unresolved lineages, habitat models may clarify the potential for gene exchange between parapatric and allopatric populations (Rissler and Apodaca 2007). A habitat model for the rare orchid, *Cypripedium fasciculatum* validated the lack of potential habitat between the highly disjunct eastern Washington and southwestern Oregon population centers for this species (Helliwell and Benz 2011). The model also suggested that there may be slightly different habitat preferences between the population centers that would be consistent with long genetic isolation and perhaps the initial phase of allopatric speciation.

Model Extrapolation – Climate Change and Invasive Species Range Expansion

An obvious application of SDMs is to extrapolate habitat under a projected changed climate or future potential expansion of the distribution of an introduced invasive species – or both. The results of published studies on climate change projections are sometimes alarming. For example, Loarie et al. (2008) used Maxent to model 591 of California's 2387 endemic plants under current and projected climates. They concluded that up to 66% of these endemic species will have range reductions of 80% or more. On a global scale, Thomas et al. (2004) projected that, by 2050, 15-37% of plant and animal taxa within the representative geographic regions that they sampled would be "committed to extinction". This paper was widely reported in the press. Some popular accounts extrapolated extinctions from this paper to be over a million species to up to a third of all species on the planet (Ladle 2004)!

Although modeling future distribution based on changed climatic conditions is one of the most common applications of SDMs, it is also possibly the most contentious application of SDMs (Jeschke and Strayer 2008, Franklin 2009). Much of the debate returns to the question of what is actually being modeled. Jeschke and Strayer (2008) identified three key assumptions of bioclimatic models that may not be met: that the biotic interactions are unimportant or don't change, that there is no genetic response to change and that there are no dispersal limitations. Recognize that these are all factors that describe the realized niche of a species. There is no way to fully account for all the various biotic interactions, complexities associated with dispersal and genome expression that cascade through systems in response to changes in the climate. If, on the other hand, there is reason to believe that the projected potential distribution is substantially constrained by current biotic interactions, the model becomes much more difficult to interpret. It is unlikely that the relationship between climate change, the target organism and the constraining biotic factors will occur as a linear function therefore the component relationships would have to be analyzed separately and then brought back together in a mechanistic model. Other criticisms of species response to climate change projections include the failure to adequately display uncertainty and overly simplistic interpretation of species-area relationships (Franklin 2009). Solutions to these deficiencies include working with local, fine-scale data rather than global data (Randin et al. 2009) or combining multiple types of models to identify points of congruence (Kearney et al. 2010).

Despite the identified shortcomings of climate change model impacts on species distribution and extinction risk, there is general agreement that such models provide an essential tool for assessing first order risks to taxa under future climate scenarios but, like all models, they do need to be understood within the context of their underlying assumptions (Heikkinen et al. 2006). Beyond the alarmist predictions of mass extinction, such models are beginning to be applied to potential management for species conservation. Carroll et al. (2010) assessed current and projected modeled habitat for the northern spotted owl (*Strix occidentalis caurina*) and 130 survey and manage species against the system of reserves established under the Northwest Forest Plan. Pauly et al. (2008) estimated the future distribution of evidently introduced populations of red-legged frogs (*Rana aurea*) in Alaska and British Columbia under global climate change and then evaluated management options to address the projected trend.

Predicting future expansion of invasive species is similar to predicting response to climate change in that both are predicated on a changed condition. In the case of invasive species the environmental variables, including climate, are held constant (although projecting both invasive species spread and effects of climate change is routinely attempted). Invasive species expansion models assume dispersal into its full potential distribution is principally limited by time since introduction into the new alien landscape so dispersal represents the changed condition. Since it is assumed that dispersal is limited, the SDM must necessarily predict the fundamental niche of the invading species to be useful or at least that something broader than the realized niche is being projected.

Modeling invasive species expansion can be done one of three ways: presence data from the native range may be projected onto a novel geographic area; presence data from the invaded area may be projected into a larger geographic extent beyond the currently invaded area; or data from both the native and invaded range may be employed. Modeling invasive species spread

represents a particular challenge since it violates one of the assumptions of SDMs: that the organism is in equilibrium with the environment (Václavík and Meentemeyer 2009). For this reason it has been suggested that using data from the native range of the invasive species where the species has achieved distributional equilibrium should be an advantage (Jimenez-Valvérde et al. 2011). However, using only data only from the native range has been demonstrated to produce misleading models (Loo et al. 2007, Ibáñez et al. 2009). This is particularly true of regression models that effectively treat areas that have yet to be colonized as unsuitable habitat (Sutherst and Bourne 2009). Maxent, on the other hand, does not use pseudo-absences but rather treats non-presence pixels simply as background data, which avoids this particular problem. It has been suggested that predictions of both invasive species spread and species distribution shifts under climate change could be improved by combining correlative SDMs with mechanistic models (Kearny and Porter 2009, Kearny et al. 2010).

Despite identified shortcomings, models of invasive species spread continue to be useful as a tool to focus prevention and inventory measures. For example, the California Invasive Plant Council (2011) used Maxent to model the potential spread of 29 noxious weed species throughout California under both current and projected 2050 climate conditions with the intent of identifying the areas that would be most at risk in the near future. Habitat models can also be used to help understand the factors associated with the spread of invasive species. Kumar et al. (2009) determined that mean temperature during the warmest quarter was the most important factor influencing the distribution of the invasive diatom, *Didymosphenia geminata*. Although water chemistry would have been a more direct means of determining suitable habitat, these data are not widely available. It is important to understand that non-mechanistic models, such as Maxent, are describing only correlations rather than causal relationships. This is still useful information as long as the model is applied in the context of the organism rather than the other way around.

Models to guide surveys

One of the challenges in managing rare species is distinguishing truly rare taxa from simply underreported taxa. This is particularly the case with difficult to detect or inconspicuous species. (Molina and Marcot 2007). Habitat models can help guide inventory so that surveys are better targeted and more efficient than expert opinion surveys for many species. In some cases this can be accomplished with remarkably few known sites. Jackson and Robertson (2011) used Maxent successfully with only four locations of an endangered mole (*Neamblysomus julianae*) to discover two more know sites which allowed further refinement of their model. Buechling and Tobalske (2011) found that Random forest-based algorithms outperformed Maxent for some species but found both successfully modeled habitat for four rare plant species of NW Oregon resulting in the discovery of 22 new occurrences.

In some cases, SDM have guided inventory beyond the previously known range of species resulting in range extensions. Rebolo and Jones (2010) found Maxent outperformed ENFA in producing a habitat model for a rare bat, *Barbastella barbastellus*. Subsequent survey based on the modeled habitat resulted in 15 new occurrences being discovered and extending the range of the species c. 100 km to the south in Portugal. Here in the Pacific Northwest, a SDM generated using Maxent for the Forest Service sensitive species, *Sisyrinchium sarmentosum* resulted in discovery of four new sites on the Gifford Pinchot NF and two new sites on the Willamette NF

ca. 14 km to the south of any previously known occurrences (Helliwell, unpublished data 2012). Conversely, a habitat model for a rare moss, *Tetraphis geniculata*, indicated that it was unlikely to be found south of the McKenzie Highway in Oregon suggesting that it should be removed as a suspected species from the Forest Service and BLM sensitive species lists for units south of this area (Helliwell and Benz 2011).

Project Effects Analysis

Forest Service and BLM biologists are faced with evaluating the potential effects of proposed projects on a wide range of plant, animal and fungi species. For some species, a habitat model can provide for a quantitative assessment of the probability of species occurrence within the project. This would allow for concentrating surveys in the most appropriate areas, potentially saving time and money. Furthermore a habitat model allows a means of evaluating how important the loss or alteration of habitat is relative to the rest of the potential habitat in the vicinity. For example if the bulk of the best predicted habitat for a species thought to be associated with late-successional/old-growth stand condition lies outside the affected project boundaries in a late-successional reserve (LSR) or other reserved land allocations, then the loss of a small amount of low quality habitat may not have much effect on species viability. The concept of reserve identification can be utilized during project planning if the scale of the model is applicable to the planning area. Larger patches of higher quality habitat may be identified as areas where treatments should be skipped if treatments would reduce the suitability of the habitat, or areas of lower habitat quality could be proposed if treatments to improve habitat characteristics important to the species of interest are possible. The possible role of the potentially affected habitat on landscape functions such as connectivity or dispersal may also need to be evaluated but the SDM would provide a spatial means of assessing this as well.

Northern spotted owl habitat models depicting suitable nesting, roosting and foraging (NRF) habitat are often used in Section 7 Endangered Species Act consultations. Level one teams utilize NRF habitat models as an initial step for identifying habitat within a project's action area. After the team has verified the accuracy of the model on the ground it may be edited to refine the model to a suitable level of accuracy. The edited model then serves as the basis for the GIS analysis of impacts to NRF habitat for owl sites various scales which are used to quantify effects to habitat as a result of the project. Spotted owl habitat models have also been used to as the baseline for developing critical habitat proposals by the USFWS.

Concluding Remarks

In entitling this final section we pay homage to G. Evelyn Hutchinson whose principles of niche theory were introduced as "concluding remarks" at the Cold Harbour symposium on quantitative biology in 1957. The Hutchinsonian concept of niche has been central to the development of the numerous algorithms for mapping species distributions. This concept is equally central to understanding the resultant models and their application in a land management context. The brightly colored habitat maps that Maxent and other SDMs produce are only spatial representations based on the inputs and algorithms employed. They have no reality beyond the context of the organisms and environment that they deem to characterize. Whether the model represents real relationships or hollow presumption requires both the modeler and those

employing the model to understand its underlying assumptions as they apply to the organism in question.

One particular aspect that is worth noting is the question of scale. Our experience has been that there is a tendency for users to misapply habitat models too literally at the scale of the individual pixel without regard to the actual niche requirements of the species or the scale of the environmental variables that form the basis of the habitat model. For example, if the principle contributing factors to the model are global climatic variables at a 1 km^2 scale then that is, more or less, the scale at which the model should be applied. So a model that is projecting shifts in habitat suitably due to global climate change must be applied at a regional scale rather than at the scale of a timber stand even if the arrangement of pixels seems to distinguish adjacent stands. Finer grain data such as vegetation, slope, aspect or soil data may refine climatic data such that model outputs can distinguish habitat at a 100 m^2 or less scale but only if those data are both accurate at that scale and contributing significantly to the model. The scale at which the species utilizes its habitat also needs to be considered. In some situations, it may be possible only to accurately predict the geographic distribution of a species with narrow microsite requirements but not the specific microhabitat in which it occurs. If the intent of such a model is to target areas for inventory then it may be appropriate to filter the distribution map with local knowledge of the area.

Literature Cited

Adhikari, D., S.K. Barik and K. Upadhaya. 2012. Habitat distribution modeling for reintroduction of *Ilex khasiana* Purk., a critically endangered tree species of northeastern India. Ecological Engineering 40: 37-43.

Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. Ecological Modelling 157: 101-118.

Box, G.E.P. 1979. Robustness in the strategy of scientific model building. In: Robustness in statistics. Launer, R,L. and G.N. Wilkinson, eds. 202 pp.

Buechling, A. and C. Tobalske. 2011 Predictive Habitat Modeling of Rare Plant Species in Pacific Northwest Forests. Western Journal of Applied Forestry 26: 71-81.

California Invasive Plant Council. 2011. Prioritizing regional response to invasive plants in the Sierra Nevada. http://www.cal-ipc.org/ip/mapping/sierra/index.php. Berkeley, CA. 179 pp.

Carroll, C., J.R. Dunk and A. Moilanen. 2009. Optimizing resiliency of reserve networks to climate change: multispecies conservation planning in the Pacific Northwest, USA. Global Change Biology 16: 891-904.

Chase, J.M. and M.A. Leibold. 2001. Ecological niches, linking classical and contemporary approaches. The University of Chicago Press. 212 pp.

Decout, S., S. Manel, C. Miaud and S. Luque. 2010. Connectivity loss in human dominated landscape: operational tools for the identification of suitable habitat patches and corridors on amphibian's population. In: Forest landscapes and global change – new frontiers in management, conservation and restoration. Proceedings of the IUFRO landscape ecology working group international conference, September 21-27, 2010, Bragança, Portugal. J.C. Azedo, M. Feliciano, J. Castro and M.A. Pinto eds.

Elith, J. and C.H. Graham. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32: 66-77.

Elith, J. and J.R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution and Systematics. 40: 677-697.

Elith, J., S.J. Phillips, T. Hastie and M. Dudı´. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17: 43–57.

Feeley, K.J. and M.R. Silman. 2011. Keep collecting: accurate species distribution modelling requires more collections than previously thought. Diversity and Distributions 17: 1132–1140.

Franklin, J. 2009. Mapping species distributions, spatial inference and prediction. Cambridge University Press. 320 pp.

Gibson, L., B. Barrett, and A. Burbidge. 2007. Dealing with uncertain absences in habitat modelling: a case study of a rare ground-dwelling parrot. Diversity and Distributions 13: 704–713.

Gu, W. and R.K. Swihart. 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. Biological Conservation 116: 195-203.

Guisan, A. and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8: 993-1009.

Heikkinen, R.K., M. Luoto, M.B. Araújo, R. Virkkala, W. Thuiller and M.T. Sykes. 2006. Methods and uncertainties in bioclimatic envelope modeling under climate change. Progress in Physical Geography 30: 751-777.

Helliwell, R. and B. Benz. 2011. A species distribution model for *Cypripedium fasciculatum* and *Tetraphis geniculata*. Unpublished report to the Interagency Sensitive and Special Status Program. USDA Forest Service, Umpqua National Forest, Roseburg, Oregon. 15 pp.

Hernandez, P.A., C.H. Graham, L.L. Master and D.L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29: 773-785.

Hernandez-Satin, L., P.M. Harveson and L.A. Harveson. 2012. Suitable habitats for cougars (*Puma concolor*) in Texas and northern Mexico. The Southwestern Naturalist 57: 314-318.

Hunter, E.A., P.A. Raney, J.P. Gibbs and D.J. Leopold. 2012. Improving Wetland Mitigation Site Identification Through Community Distribution Modeling and a Patch-Based Ranking Scheme. Wetlands, 10 July 2012.

Hutchinson, G.E. 1957. Concluding remarks. Cold Spring Harbor Symposium on quantitative biology 22: 415-457.

Ibáñez, I, J.A. Silander, Jr., A.M. Wilson, N. LaFleur, N. Tanaka and I. Tsuyama. 2009. Multivariate forecast of potential distributions of invasive plant species. Ecological applications 19: 359-375,

Jackson, C.R. and M.P. Robertson. 2011. Predicting the potential distribution of an endangered cryptic subterranean mammal from few occurrence records. Journal for Nature Conservation 19: 87-94.

Jeschke, J.M. and D.L. Strayer. 2008. Usefulness of bioclimatic models for studying climate change and invasive species. Annuals of the New York Academy of Sciences 1134: 1-24.

Jiménez-Valverde, A., A.T. Peterson, J. Soberón, J.M. Overton, P. Aragón and J.M. Lobo. 2011. Biological Invasions 13: 2785-2797.

Kearney, M. and W. Porter. 2009. Mechanistic niche modeling: combining physiological and spatial data to predict species' ranges. Ecology Letters 12: 334-350.

Kearney M.R. B.A. Wintle and W.P. Porter. 2010. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. Conservation Letters 3: 203-213.

Kumar, S., S.A. Spaulding, T.J. Stohlgren, K.A. Hermann, T.S. Schmidt and L.L. Bahls. 2009. Potential habitat distribution for the freshwater diatom *Didymosphenia geminata* in the continental US. Frontiers in Ecology and the Environment 7: 415-420.

Ladle, R.J., P. Jepson, M.B. Araújo and R.J. Whittaker. 2004. Nature 428: 799.

Loarie, S.R., B.E. Carter, K. Hayhoe, S. McMahon, R. Moe, C.A. Knight and D.A. Ackerly. 2008. Climate change and the future of California's endemic flora. PLOS ONE 3: 1-10.

Loo, S.E., R Mac Nally and P.S. Lake. 2007. Forecasting New Zealand mudsnail invasion range: model comparisons using native and invaded ranges. Ecological Applications 17: 181-189.

Molina, R. and B. Marcot. 2007. Definitions and attributes of little-known species. In: Conservation of Rare or Little-Known Species, M.G. Raphael and R. Molina eds. Island Press. 93-124.

Pauly, G.B., S.R. Ron and L. Lerum. 2008. Molecular and ecological characterization of extralimital populations of red-legged frogs from western North America. Journal of Herpetology 42: 668-679.

Phillips, S.J. 2008. A brief tutorial on Maxent. www.cs.princeton.edu/~schapire/Maxent/tutorial/tutorial.doc

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

Phillips, S.J., M. Dudı´k and R.E. Shapire. 2004. A Maximum Entropy Approach to Species Distribution Modeling. Proceedings of the 21st International Conference on Machine Learning, Banff, Canada, 2004.

Phillips, S.J. and M. Dudı´k. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161–175.

Poor, E.E., C. Loucks, A. Jakes and D.L. Urban. 2012. Comparing habitat suitability and connectivity modeling methods for conserving pronghorn migrations. PLOS ONE: 71-12

Randin, C.F., R. Engler, S. Normand, M. Zappa, N.E. Zimmermann, P.B. Pearman, P. Vittoz, W. Thuiller and A. Guisan. 2009. Global Climate Change Biology 15: 1557-1569.

Rebolo, H. and G. Jones. 2010. Ground validation of presence-only modelling with rare species: a case study on barbastelles *Barbastella barbastellus* (Chiroptera: Vespertilionidae). Journal of Applied Ecology 47: 410-420.

Rissler, L.J. and J.J. Apodaca. 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). Systematic Biology 56: 924-942.

Rupprecht, F., J. Oldeland and M. Finckh. 2012. Modeling potential distribution of the threatened tree species *Juniperus oxycedrus*: how to evaluate the predictions of different modeling approaches? Journal of Vegetation Science 22: 647–659.

Sillero, N. 2011. What does ecological modeling model? A proposed classification of ecological niche models based on their underlying methods. Ecological Modelling 222: 1343-1346.

Soberón, J. and A. Townsend Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodiversity Informatics 2: 1-10.

Strittholt, J.R., D.A. Dellasala and H. Jiang. 2006. Status of mature and old-growth forest in the Pacific Northwest. Conservation Biology 20: 363-373.

Sutherst R.W. and A.S. Bourne. 2009. Modeling non-equilibrium distributions of invasive species: a tale of two modeling paradigms. Biological Invasions 11: 1231-1237.

Taecker, E.M. 2007. Identification and prioritization of lands for restoration of Piedmont Prairie in North Carolina. Master's Thesis. Duke University, North Carolina. 40 pp.

Thomas, C.D., A. Cameron, R.E. Green, M. Bakkenes, L.J. Beaumont, Y.C. Collingham, B.F.N. Erasmus, M. Ferreira de Siqueira, A Grainger, L. Hannah, L. Hughes, B. Huntley, A.S. van Jaarsveld, G.F. Midgley, L. Miles, M.A. Ortega-Huerta, A. Townsend Peterson, O.L. Phillips and S.E. Williams. 2004. Extinction risk from climate change. Nature 427: 145-148.

Townsend Peterson, A., M. Papes¸ and M. Eaton. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography 30: 550-560.

Václavík, T. and R.K. Meentemeyer. 2009. Invasive species distribution modeling (iSDM): are absence data dispersal constraints needed to predict distributions? Ecological Modelling 220: 3248-3258.

Wilson, C.D., D. Roberts and N. Reid. 2011. Applying species distribution modeling to identify areas of high conservation value for endangered species: a case study using *Margaritifera margaritifera* (L.). Biological Conservation 144: 821-829.

Wilson, K.A., M.I. Westphal, H.P. Possingham and J. Elith. 2005. Sensitivity Of Conservation Planning To Different Approaches To Using Predicted Species Distribution Data. Biological Conservation 22: 99-112.

Wisz, M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, A. Guisan and NCEAS Predicting Species Distributions Working Group. 2008. Effects of sample size on the performance of species distribution models. Diversity and Distributions 14:763–773.

Zhang, M.G., ZK Zhou, W.Y. Chen, J.W. Ferry Slik, C.H. Cannon and N. Raes. 2012. Using species distribution modeling to improve conservation and land use planning of Yunnan, China. Biological Conservation 153: 257–264.