2007

Modeling distributions of rare plants in the Great Basin, western North America

Leila M Shultz, Ph.D.

Western North American Naturalist 67(1), © 2007, pp. 26–38

MODELING DISTRIBUTIONS OF RARE PLANTS IN THE GREAT BASIN, WESTERN NORTH AMERICA

M. Aitken[1,2], D.W. Roberts[1,3], and L.M. Shultz[1,4]

ABSTRACT.—In this 2-phase study, we developed field-validated site and landscape-level predictive models for identifying potential rare and endemic plant habitat in the Great Basin of western North America. Four species were chosen to include a range of environmental variability and plant communities. Herbarium records of known occurrences were used to identify initial sample sites. The geographic coordinates, environmental attributes, and vegetation data collected at each site were used to develop 2 predictive models for each species: a field key and a probability-of-occurrence or predictor map. The field key was developed using only field data collected at the sites on environmental attributes and associated species. Predictive maps were developed with a geographic information system (GIS) containing slope, elevation, aspect, soils, and geologic data. Classification-tree (CT) software was used to generate dichotomous field keys and maps of occurrence probabilities. Predictions from both models were then field-validated during the 2nd phase of the study, and final models were developed through an iterative process, in which data collected during the field validation were incorporated into subsequent predictive models. Cross-validated models were >96% accurate and generally predicted presence with >60% accuracy. These models identified potential habitat by combining elevation, slope, aspect, rock type, and geologic process into habitat models for each species.

Key words: predictive model, Great Basin, endemic species, rare plant species, classification tree, habitat prediction.

ACRONYMS.—CEC—cation exchange capacity, CT—classification tree, DEM—digital elevation model, GAM—generalized additive model, GIS—geographic information system, GLM—generalized linear model, GPS—geographic positioning system, ROC—receiver operating curves, STATSGO—State Soil Geographic Database, USGS—United States Geological Survey, UTM—Universal Transverse Mercator.

Nomenclature and distributions: Welsh et al. 2003, Shultz et al. 2005.

Across landscapes the availability of suitable habitat limits the distribution of plant species. Vegetation models establish a relationship between species distributions and environmental variables (Roberts and Cooper 1989, Franklin 1995, Vayssières et al. 2000). Predictive vegetation modeling differs from vegetation modeling in both its use and production of spatial data (Miller and Franklin 2002). This form of modeling often employs GIS capture, storage, manipulation, analysis, and display of these spatial data.

The predictive power of vegetation models depends on the strength of correlation between the species occurrence and identified habitat variables (MacDougall and Loo 2002). A variety of statistical methods have been used to relate plant species distribution to spatial distribution of environmental predictor variables (Table 1); however, no single method is clearly superior. Some of these techniques are reviewed in Franklin (1995), Guisan et al. (1999), Guisan and Zimmermann (2000), Austin (2002), Ejrnæs et al. (2002), Elith and Burgman (2002), Elith et al. (2002), Guisan et al. (2002), Miller and Franklin (2002), Rushton et al. (2004), and Guisan and Thuiller (2005). Most studies concentrated on broad-scale predictions (e.g., about communities, alliances, or on widely distributed, abundant individual species; Table 1).

Particularly for rare species, small geographic ranges and/or narrow habitat specificity lead to acute habitat limitation that is often strongly correlated with specific environmental variables. Therefore, if the biotic and physical parameters of plant distributions can be quantified, models should be able to predict species distributions within a landscape. The

[1]Department of Forest, Range, and Wildlife Sciences, Utah State University, Logan, UT 84322-5230.
[2]Umatilla National Forest, 3502 Hwy. 30, La Grande, OR 97850.
[3]Department of Ecology, Montana State University, Bozeman, MT 59717-3460.
[4]Corresponding author. E-mail: shultz@cc.usu.edu

TABLE 1. A partial list and summary of relevant literature. ANN—artificial neural network, CART—classification and regression trees, CCA—canonical correspondence analysis, CT—classification trees, DA—discriminant analysis, EFNA—ecological niche factor analysis, GAM—generalized additive models, GARP—genetic algorithm for rule-set prediction, GIS—geographic information system, GLM—generalized linear model, GLMM—generalized linear mixed model, LiR—linear regression, LoR—logistic regression, MARS—multivariate adaptive regression splines, NMS—nonmetric multidimensional scaling, NN—neural network, PCA—principal components analysis, ZIP—zero inflated negative binomial regression.

Author	Year	Analysis method	Spatially explicit	Modeling unit
Bio et al.	1998	GAM, GAM, multiple LoR	no	Species
Bio et al.	2002	GAM, GLM	yes	Species
Boetsch et al.	2003	Mahalanobis distance	yes	Rare species
Brown	1994	GAM, GLM	yes	Vegetation types
Cairns	2001	ANN, CT, GLM	yes	Species
Cawsey et al.	2002	GAM	yes	Species
Cherrill et al.	1995	GIS matrices	yes	Rare species
Collingham et al.	2000	Stepwise LoR	yes	Rare and nonnative species
Dirnböck, et al.	2002	GAM	no	Local vegetation
Elith and Burgman	2002	GLM, GAM, GARP	no	Rare species
Engler et al.	2004	ENFA, GLM	yes	Rare species
Fertig and Reiners	2002	Logistic regression, CART	yes	Species
Franklin	1998	CT, GAM, GLM	yes	Species
Franklin	2002	CT	yes	Species
Gioia and Pigott	2000	GLM with SPMODEL software	yes	Species
Guisan et al.	1998	GLM	yes	Species
Guisan et al.	1999	CCA, GLM	yes	Species
Guisan et al.	2006	GAM	yes	Species
Harrison et al.	2000	Multiple LiR	yes	Endemic species richness
Hill and Keddy	1992	Stepwise multiple regression	no	Rare species
Hirzel and Guisan	2002	GLM	no	Virtual species
Hooten et al.	2003	Hierarchical Bayesian models	yes	Species
Imm et al.	2001	$\chi 2$	yes	Rare species
Jelaska et al.	2003	CT, DA, multiple LoR	yes	Species and communities
Luoto	2000	Multivariate LiR	yes	Rare species richness
Luoto et al.	2002	GLM	yes	Rare species richness
Miller and Franklin	2002	CT, GLM	yes	Vegetation alliances
Moisen and Frescino	2002	ANN, CT, GAM, MARS	no	Forest characteristics
Moore et al.	1991	CT	yes	Forest communities
Nilsson et al.	1988	Mann-Whitney	no	Rare species
Pearce and Ferrier	2001	GAM, GLM, ZIP	yes	Species abundance
Robertson et al.	2003	Multiple LoR, PCA	yes	Species
Sperduto and Congalton	1996	Equal and weighted $\chi 2$	yes	Rare species
Stahle and Chaney	1994	Identified co-occurring habitat variables	yes	Species
Thuiller et al.	2003	CT, GAM, GLM	yes	Species
van Horssen et al.	1999	Multiple stepwise logistic Gaussian regression and krieging	yes	Species
Vayssières et al.	2000	CT, GLM, LoR	no	Species
White and Miller	1988	Multiple regression	no	Community
Wiser et al.	1998	GAM, LoR	no	Rare species
Wu and Smiens	2000	Nonstatistical modeling	yes	Rare species
Zaniewski et al.	2002	ENFA, GLM	yes	Species
Zimmerman and Kienast	1999	LiR	yes	Community

potential role of GIS in predicting sensitive plant habitat was recognized at least 19 years ago (e.g., Myatt 1987). However, employing distribution models with GIS for predictive mapping of rare plant habitat and distribution remains insufficiently explored (Franklin 1995, Luoto et al. 2002).

Efforts to predictively model rare plant habitats and distributions have been few, and the degree of success highly variable (Sperduto and Congalton 1996, Wiser et al. 1998, Imm et al. 2001, Edwards et al. 2005). Equally varied are the definitions of rare plants included in these studies. A species is generally regarded as rare if it has low abundance and/or a small range (Gaston 1994). More specifically, distributions of rare species fall into 3 categories (Rabinowitz 1981). First, a rare species may be known from only a few individuals occurring over a large area. Second, a rare species may occur as small populations, but with populations widely scattered in geographic ranges. Finally, populations may be large, but the number of populations may be very small. Many "rare species" in the literature are widespread and infrequent, an attribute that complicates modeling. However, Boestch et al. (2003) suggest that rare species well suited for modeling are relatively common in their respective habitats and respond to "large-scale" gradients.

The objective of this research was to develop predictive vegetation models of rare plant habitat for 4 species and test their utility. We used a GIS database to develop and evaluate models at 2 scales: a local (field-site) scale and a landscape scale.

METHODS

Study Area

The study area was located in the eastern Great Basin of west central Utah. The landscape consisted of north–south trending basin and range topography dominated by limestone and dolomite. Soils were predominantly deep and well-drained aridisols, entisols, and mollisols (Wilson et al. 1975). Annual precipitation ranged from 100 mm in the valleys to 450 mm in the mountains, and was generally lowest from summer through midautumn (Holmgren 1972). Average annual temperatures in the area ranged from 7ºC to 13ºC (45ºF to 55ºF). The growing season ranged from 40 to 200 days (Ashcroft et al. 1992) and was inversely related to elevation. The extreme environment of the study area provided an effective test of predictive models because physical parameters provided the primary constraints to species' distributions.

Species

We selected 4 endemic perennial species representing a range of habitat types, community associations, and elevation. *Sphaeralcea caespitosa* occurs mainly on limestone and dolomite outcrops of Cambrian through Devonian formations (Franklin 1996). *Penstemon concinnus* inhabits calcareous and igneous gravels. Both species occurred in lowland desert shrub communities (1510–2300 m; Albee et al. 1988). *Jamesia tetrapetala* and *Primula domensis* occur in crevices of limestone cliffs in the montane zone (2000–2750 m; Welsh et al. 1993). *Sphaeralcea caespitosa, P. concinnus,* and *J. tetrapetala* are considered endangered throughout their range, and *Pr. domensis,* is considered critically endangered throughout its range (Atwood et al. 1991); however, none are federally listed. Distributions of rare plant species with small geographic range and/or narrow habitat specificity generally correlate with physical factors (Shultz 1993, Gaston 1994); thus, we expected these species would lend themselves well to modeling, despite their rarity.

Modeling Overview

We used an iterative process of field visits and predictive vegetation modeling over 2 field seasons. Data collected in the 1st year were used to develop models that were tested in the 2nd field season. In addition to validation, the 1st year models served to effectively stratify the sampling effort in the 2nd field season. In order to maximize the number of samples available for modeling, validation data collected in the 2nd season were subsequently pooled with the 1st year's data to develop the final models. All models in both years were 10-fold cross-validated for error estimation, and samples used in model development were never used in assessing model error rates. Two types of models were developed for each of 4 rare species: a site-specific predictive field key and a spatially explicit map of potential habitat. The predictive field keys used data collected in the field and were designed for

subsequent use in the field. The spatially explicit models predicted occurrence probabilities as a map based on GIS-defined site properties and served to prioritize field survey sites. Our approach of modeling at 2 different scales was similar to approaches of Wiser et al. (1998) and Wu and Smeins (2000), and allows for different environmental influences operating at different scales.

Field Methods and Data Collection

Initial baseline information for plant locations came from herbarium voucher specimens. Even though herbarium records may be biased by nonsystematic sampling (i.e., based on access, site "appeal," and other preferences; MacDougall et al. 1998), vouchers often provide the best data available and have been used in other studies (MacDougall et al. 1998, Wiser et al. 1998, Gioia and Pigott 2000, Hijmans and Spooner 2001). We used only vouchers with locality descriptions precise to within 0.5 km^2 (quarter section). The 4 species selected were conspicuous enough, especially when flowering, that detection probability was very high.

We sampled 171 sites in 1996 and 296 sites in 1997. Sampling strategies during the 1st field season were voucher driven. While this approach potentially introduced spatial autocorrelation in the data, random or stratified sampling strategies could not produce sufficient numbers of presence sites for model development. During the 2nd season, field sampling was stratified by model predictions based on the 1st year's data and included equal proportions of samples of all prediction probability groups (absent 0–0.24, unlikely 0.25–0.49, likely present 0.5–0.74, very likely present 0.75–1.0).

Sites were defined as areas >100 m^2 of uniform slope, aspect, geology, and vegetation composition. Consequently, perimeters, areas, and shapes of sites varied. All sites were field mapped on U.S. Geological Survey (USGS) 7.5′ topographic maps and geographic coordinates recorded on GPS units. GPS coordinates were differentially corrected to 2-m accuracy. Each site was also assigned a geographic identity (i.e., Tule Valley, House Range) following USGS place names as they appeared on 7.5′ quadrangle maps.

Site data included elevation, percent slope, slope orientation, slope position, topographic position, and a checklist of all vascular plant species present. Slope position was described and coded as hilltop, upper slope, midslope, lower slope, or flat. Topographic landform classifications followed a 9-unit land surface model (Dalyrymple et al. 1968) with the addition of 2 categories describing erosion areas. Our field categorization was based on apparent evidence of landscape-scale processes.

Because soil texture influences vegetation patterns (Parker 1991, Knight 1993, McAulliffe 1994), we collected samples from the upper 10 cm of soil. Analysis included hand-texturing, wet and dry color identification, calorimetric pH measurements, and testing for calcium carbonate. Not all sites had soil present.

Statistical Methods and Predictive Modeling

We chose classification trees (CT) to develop the predictive models because their utility has been well documented (Roberts and Cooper 1989, Franklin 1998, De'ath and Fabricius 2000, Vayssières et al. 2000, Fertig and Reiners 2002). CTs make no assumptions about data distribution, often require only a few variables to achieve an accurate classification (Dobbertin and Bigging 1998), and are easy to interpret and incorporate into other models (LeMay et al. 1994, Dobbertin and Bigging 1998). CTs also have the advantage of using an explanatory variable more than once (Jelaska et al. 2003), and they allow for nested dependencies. For a more detailed discussion of the tree classification process, see Clark and Pregibon (1992). We developed the predictive models with a tree classification program in S-Plus (MathSoft 1998). All models were 10-fold cross-validated (Fielding and Bell 1997). The ultimate predictive model took the form of a dichotomous key that could be incorporated into a computer program for extrapolation over large areas (Roberts and Cooper 1989).

Field keys and the spatially explicit (GIS) models were derived from different sets of variables that reflected the data available at such a scale. Multiple models were developed for each species using various subsets of the variables to find the best model and to assess the utility of different variables. Elevation, aspect, and slope were common to both field

and GIS models. Aspect, a circular variable, was converted into aspect value, calculated as 0.5[cos(*Aspect*) + 1] (Roberts and Cooper 1989). We fitted separate models in which slope orientation was defined as aspect or aspect value. Data unique to either group of models are described in following sections.

Model Evaluations

Error matrices, comparing observed presence and absence to predicted presence and absence, were developed for each model to facilitate model evaluations and comparisons. Model predictions with probabilities ≥0.5 were treated as "presence"; probabilities <0.5 were treated as "absence." For rare species, a model that always predicted absence would have very high accuracy, but would not be useful. Therefore, we evaluated each model from 3 perspectives: accuracy, utility, and bias. Accuracy was the ratio of correct predictions to all predictions. For the purposes of this study, errors of commission (false positive) were considered less problematic than errors of omission (false negative; Franklin 1998). Errors of commission could be more easily field-corrected than errors of omission. Accordingly, models were also considered in terms of their ability to correctly predict presences. Sensitivity to presence, or "utility," was calculated as the ratio of correctly predicted presences to the sum of omissions, commissions, and correctly predicted presences. This yielded a value between 0 (a useless model) and 1 (a model with no errors of omission or commission). Utility is related to the statistic's "sensitivity" and "specificity," often used to analyze confusion matrices (Fielding and Bell 1997, Vayssières et al. 2000). Utility included errors of commission and omission in a single, more stringent statistic. The final term, bias, describes the direction of the model's errors with respect to predicted and actual presences, and was calculated as the difference between total predicted presences and total actual presence divided by total actual presence. Negative values indicated omission tendencies and positive values indicate commission tendencies. In summary, the ideal model had high accuracy, high utility, and low bias.

Field Key Model

A field key of potential species occurrence was derived from field measurements. In addition to environmental characteristics, a checklist of all vascular plant species present at the site was completed. Selection of possible indicator species was based on high conditional probabilities of occurrence (high relative frequency of occurrence of the rare species given that the associated species was present).

Spatially Explicit (GIS) Model

We developed models from a GIS database using only predictors that could be mapped across the entire study area. Data came from a database developed for a larger study by Sharik et al. (2000). Elevation, slope, and aspect values were derived from a digital elevation model (DEM) with 100-m^2 resolution. State Soil Geographic Database (STATSGO) data were mapped at a scale of 1:250,000 and included clay content, cation exchange capacity (CEC), litter layer depth, organic material content, permeability, pH, depth to bedrock, salinity, and depth to water table. In addition, we digitized Hintze's (1980) 1:500,000 geologic map because, across a broad elevation range, distribution of rare plant taxa can correlate strongly with specific geologic formations (Heil et al. 1993). The digital geologic map included formation name, geologic period of the formation (age), rock type (e.g., limestone), formation process, and the presence of carbonates. No land-use data existed in any organized or digital form. Although we knew that climatic factors also influence species distributions (Reichenbacher and Zamundio 1993), such data would be problematic given the very low density of climate stations in the study area. Because these data would have been derived from terrain-sensitive interpolations of very limited, "punctually-distributed" data, they would have had very low precision (Guisan et al. 1999). Therefore, we excluded such data from the modeling efforts.

In order to associate each sample point with all environmental attribute values, we intersected geographic UTM (Universal Transverse Mercator) northing and easting coordinates for each site with the database to produce a dataset used to develop the GIS-based models. Model predictions developed from the GIS data were integrated with raster-based geoprocessing software (ARCGRID; ESRI 1997) to produce a new raster map in which each cell reflected the prediction of the model.

Predictions developed from the 1st field season's data were used to guide field sampling during the 2nd field season.

In the final analysis we compared the utility and bias of all models with $\hat{K}$ (Congalton et al. 1993, Næsset 1996). This statistic estimates *K*, the level of confusion in the error matrix. *K* analysis is a discrete multivariate technique used in accuracy assessment and is appropriate for discrete, binomially or multinomially distributed data. Values for $\hat{K}$ range from $-\infty$ to 1; values between 0.4 and 0.8 represent moderate to substantial agreement (Næsset 1996). The lower the value, the greater the number of omission and/or commission errors in the matrix.

Many GLM- and GAM-based models assess model performance based on receiver operating curves (ROC), which plot *sensitivity* (true positive predictions) over 1 – *specificity* (false positive predictions or errors of commission) over a range of values of threshold probability. Because tree classifiers produce only a single confusion matrix, and not a range of values over a range of probabilities, their performance evaluation reduces to single point in ROC space and is difficult to evaluate or justify (Vayssières et al. 2000). Manel et al. (2001) found that Kappa provided a robust evaluator of model performance when compared to ROC.

RESULTS

Field Data

Sphaeralcea caespitosa was observed at 31 sites in the Halfway Hills and Tunnel Springs areas. We observed numerous individuals at each site. Field sites ranged in elevation from 1661 m to 1911 m. This species was observed in sites with slopes ranging from flat to 35%. There was no relationship with aspect. Soil textures included clay, clay loam, sandy clay loam, sandy loam, and silty clay. Soils were strongly to very strongly effervescent, and pH ranged from 7.6 to 8.6.

Penstemon concinnus was observed at 43 sites in the Mountain Home Range, the Burbank Hills, and the Tunnel Springs Mountains. Numerous individuals were found at each site. Elevation ranged from 1773 m to 2356 m, and the species was never observed on slopes greater than 35%. Soil textures at presence sites included clay, clay loam, sandy clay loam, and sandy loam. These soils had pH values from 7.5 to 8.5 and were moderately effervescent.

Jamesia tetrapetala was observed at 13 sites located in the House and Snake Ranges. Occurrences typically consisted of solitary individuals. Elevations at these sites ranged from 2176 m to 2499 m. This species was only found on north-facing slopes, typically in rock crevices near vertical cliffs or in talus.

Primula domensis was observed on 11 sites. These sites typically had 10–20 plants. All were in the House Range above 2440 m, typically on 33%–58%, north- to southeast-facing slopes and cliffs. This species was generally observed growing in loamy soils and duff-covered colluvium. Soils, when present, were very effervescent and had pH values between 7.5 and 8.7.

Field-based Models

Field models used primarily combinations of site topographic characteristics and associated species. *Sphaeralcea caespitosa* was predicted to be present in the Halfway Hills at elevations below 1919 m on slopes greater than 9.5%. Probabilities were highest between 1815 m and 1919 m (Fig. 1, Table 2). The best field model for *Penstemon concinnus* predicted presence in several mountain ranges at elevations between 1912 m and 2254 m. In this elevation range, *P. concinnus* was predicted on convex creep slopes, colluvial foot slopes, and alluvial toe slopes over most of its range. The best field model for *J. tetrapetala* predicted presence on fall faces and channel beds at elevations above 2190 m and in association with *Petrophytum caespitosum.* For *Pr. domensis,* the best field model predicted presence at elevations greater than 2613 m. The model predicted presence at all aspects, but with greater probability on north-facing slopes. Predictions for all species were consistent with our field observations, as well as published literature (Atwood et al. 1991, Welsh et al. 1993, Franklin 1996). Accuracy, utility, and bias measures for these models are presented in Table 2. Among the field models, elevation proved to be the most consistently useful variable. Slope position, orientation, and landform were variously used in the hierarchical ranking of critical habitat for some species but not for others. Combining associated species and the USGS

TABLE 2. Summary of field-based model evaluation statistics. "Observations" refers to the total number of presences observed.

Model	Observations	Accuracy	Utility	Bias	$\hat{K}$
Sphaeralcea caespitosa	31	97	0.611	–0.068	0.743
Penstemon concinnus	43	96	0.563	–0.256	0.696
Jamesia tetrapetala	13	98	0.611	0.303	0.725
Primula domensis	11	99	0.769	0.091	0.866

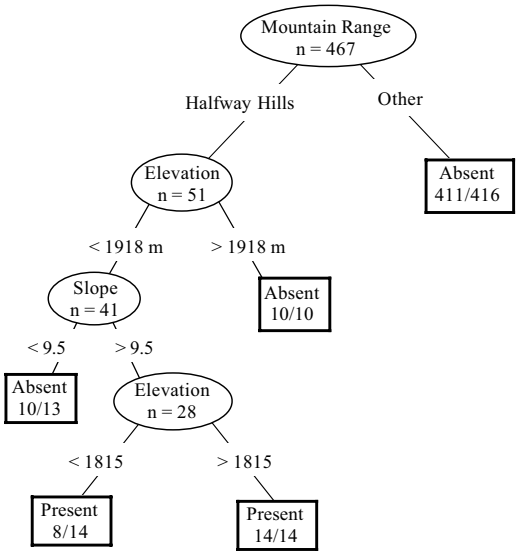

Fig. 1. Pruned classification tree model for *Sphaeralcea caespitosa* based on field data. Ellipses indicate internal nodes; boxes indicate terminal nodes. Ratios at the terminal node are the proportion correctly classified.

place-name variable with site characteristics further improved utility and minimized bias for all species except *S. caespitosa.* Associated species variables were tested in every model for which the data were made available. Overall, however, the addition of associated species had unexpectedly negative results on model performance. The tree classifier selected species we observed as fairly common and widespread, but not the "indicator" or "associated" species we noticed in the field.

In general, soil data collected during the field season never contributed to any accurate, useful, unbiased model. This may have resulted more from the type of data collected, than from any biophysical factor. Our soil analyses produced a limited, fairly homogeneous data set.

GIS-based Models

The most accurate GIS-based model for *S. caespitosa* predicted the species on a restricted set of geologic formations with low CEC soils (Fig. 2). Highest probabilities were obtained on limestone-shale and dolomite, with slightly lower probabilities on limestone or alluvium. *Sphaeralcea caespitosa* exhibited a complex relationship with elevation, depending on geology, but generally occurred above 1625 m. The best GIS-based model predicted *P. concinnus* on a restricted set of geologic formations on sites with shallow slopes at elevations above 1752 m and aspects greater than 147°. For *J. tetrapetala,* the GIS model predicted presence on slopes >25%. The best GIS model for *Pr. domensis,* predicted presence on dolomite at elevations greater than 2538 m. Accuracy, utility, and bias measures for these models are presented in Table 3.

The tree classifier consistently selected geologic formation, rock type, and percent slope for predictions from the GIS data. Model utility increased when slope orientation was calculated by azimuth rather than as a categorical variable. Like the field models, initial GIS models were not geographically constrained. Following the 2nd field season, a variable that described distance to known populations was tested in the model. This variable overly restricted predictions and was not pursued further during the study.

Models that described orientation as azimuth had utility values greater than or equal to those that described orientation as aspect value. Underprediction occurred in 3 of 8 azimuth models and 4 of 8 aspect value models. These differences probably resulted from the mathematical transformation rather than from any ecological phenomenon. When orientation data were described as a value between 0° and 360°, the data were treated linearly with the result that the difference between 10° and 350° appears greater than the difference between

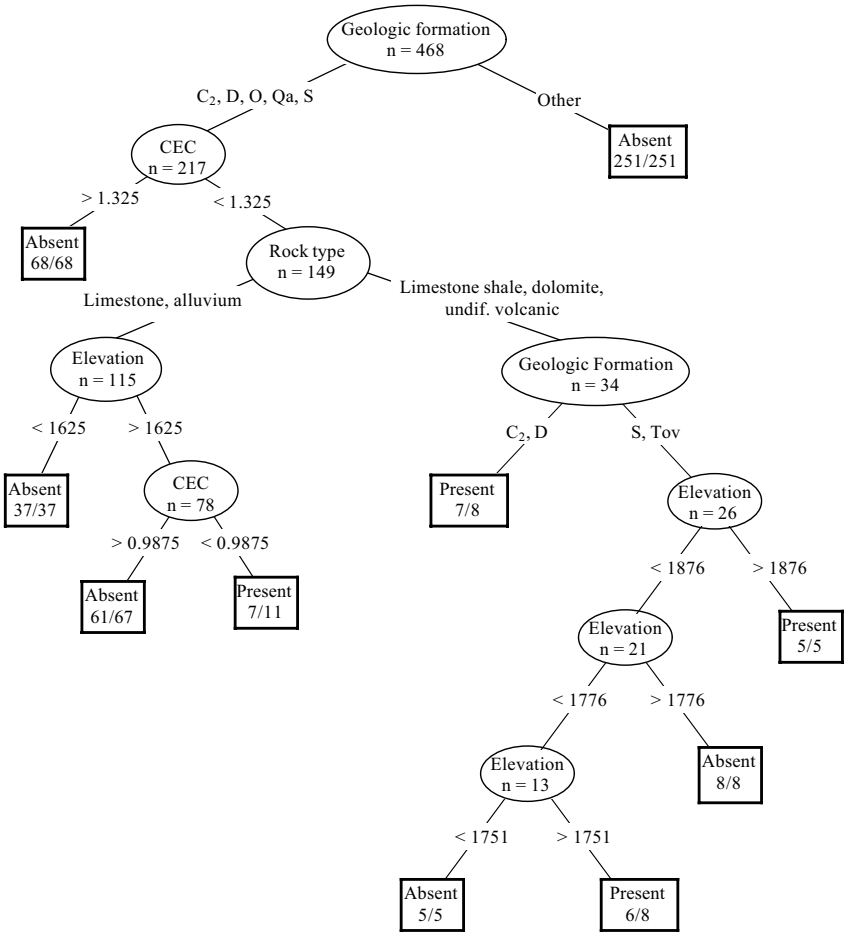

Fig. 2. Pruned classification tree model for *Sphaeralcea caespitosa* based on GIS data. Ellipses indicate internal nodes; boxes indicate terminal nodes. Ratios at the terminal node are the proportion correctly classified.

TABLE 3. Summary of GIS-based model evaluation statistics.

Model	Observations	Accuracy	Utility	Bias	$\hat{K}$
Sphaeralcea caespitosa	31	97	0.658	0.032	0.779
Penstemon concinnus	43	97	0.740	0.714	0.835
Jamesia tetrapetala	13	97	0.353	0.231	0.510
Primula domensis	11	99	0.770	0.300	0.866

10° and 300°. Transforming the data to aspect value eliminated this problem, but made it difficult to distinguish 90° from 270°.

Mapped Predictions

We converted the model algorithms to predictive mapping routines in the GIS and prepared maps of predicted presence by species, assigning each pixel the probability of presence of the species. To avoid excessive spatial extrapolation of the models, we determined the maximum distance between known presences, buffered these sites by that distance, and then applied the predictions to that total

TABLE 4. Summary of mapped predictions. Likelihood values are a percentage of the total area.

Model	Total area (ha)	Very likely present	Likely present	Unlikely	Absent
Sphaeralcea caespitosa	263,183	4.5	2.8	30.5	62.1
Penstemon concinnus	1,778,805	1.4	4.6	2.4	91.6
Jamesia tetrapetala	2,664,754	0	1.0	1.2	97.8
Primula domensis	4741	13	0	2.6	84.4

area. Consequently, the prediction areas for each species varied (Table 4). A map example is provided in Figure 3.

DISCUSSION

We successfully developed predictive models and useful maps of potential habitat for the narrow endemics in this study. Our selection included 2 species that occurred in small numbers of large populations (*Sphaeralcea caespitosa* and *Penstemon concinnus*), 1 species that had small populations in a large range (*Jamesia tetrapetala*), and 1 species that had small populations in a very small range (*Primula domensis*). While all models performed well, models for *S. caespitosa* and *P. concinnus* performed better than models for the other species. In all cases, maps produced during the 1st phase of this study greatly enhanced the efficiency of our field work. Mapped habitats represent extremes of environmental conditions ranging from desert alluvial fans to midelevation mountain slopes. Using cross-validated models, we predicted presence accurately even though presence data were often <10% of the total data set. Presence sites were predicted and field-validated, demonstrating the models' worth in conservation efforts.

When we did not find our target species at the locations mapped as "likely present," we found habitats that fit every other definition of critical habitat, including appropriate sets of associated species. The errors of commission are likely inherent in attempts to predict the distribution of species known to be rare, and possibly represent population limitations related to factors other than habitat. In many instances we suspect that actual occurrence in suitable habitat was limited by disruption of seed establishment by human disturbance, grazing, or biological constraints such as lack of pollinator and dispersal agents. Our finding that errors of commission are more common than errors of omission is likely not a critical problem, and likely makes the system especially powerful for identifying conservation areas.

The importance of landform in the field-based models suggests that an analogous variable may have benefitted GIS-based models. Future research might benefit from DEM data at finer resolution. MacDougall and Loo (2002) recommended GIS data at 1:20,000. Presence–absence information for associated species may also be helpful in future predictive modeling efforts (Edwards et al. 2005), an analysis that will be possible with new versions of the Atlas of Utah Plants (Ramsey and Shultz 2004, Shultz et al. 2005).

When applied to rare plant species, predictive vegetation modeling is a potentially powerful tool for both biologists and land managers. Rare species habitat modeling can be used to guide searches for unknown populations (thus reducing expensive field searches), to indicate site suitability for restoration and reintroduction efforts, to predict impacts of habitat degradation, and to provide a framework for further research on specific physiological requirements (Wiser et al. 1998). The predictive models proved useful in making field searches for new populations more efficient. This may be particularly important given limitations of time and funding for field studies. Our approach can be applied at site and landscape scales, and is best-suited for species with strong correlations to environmental variables. The techniques used in this approach are likely to become more readily available and more powerful with higher-resolution and more physiologically meaningful digital data

ACKNOWLEDGMENTS

This study was conducted as part of a multiresource inventory of the Hill Air Force Base Military Operation Area, Utah. Funding for the survey (including USDA Forest Service,

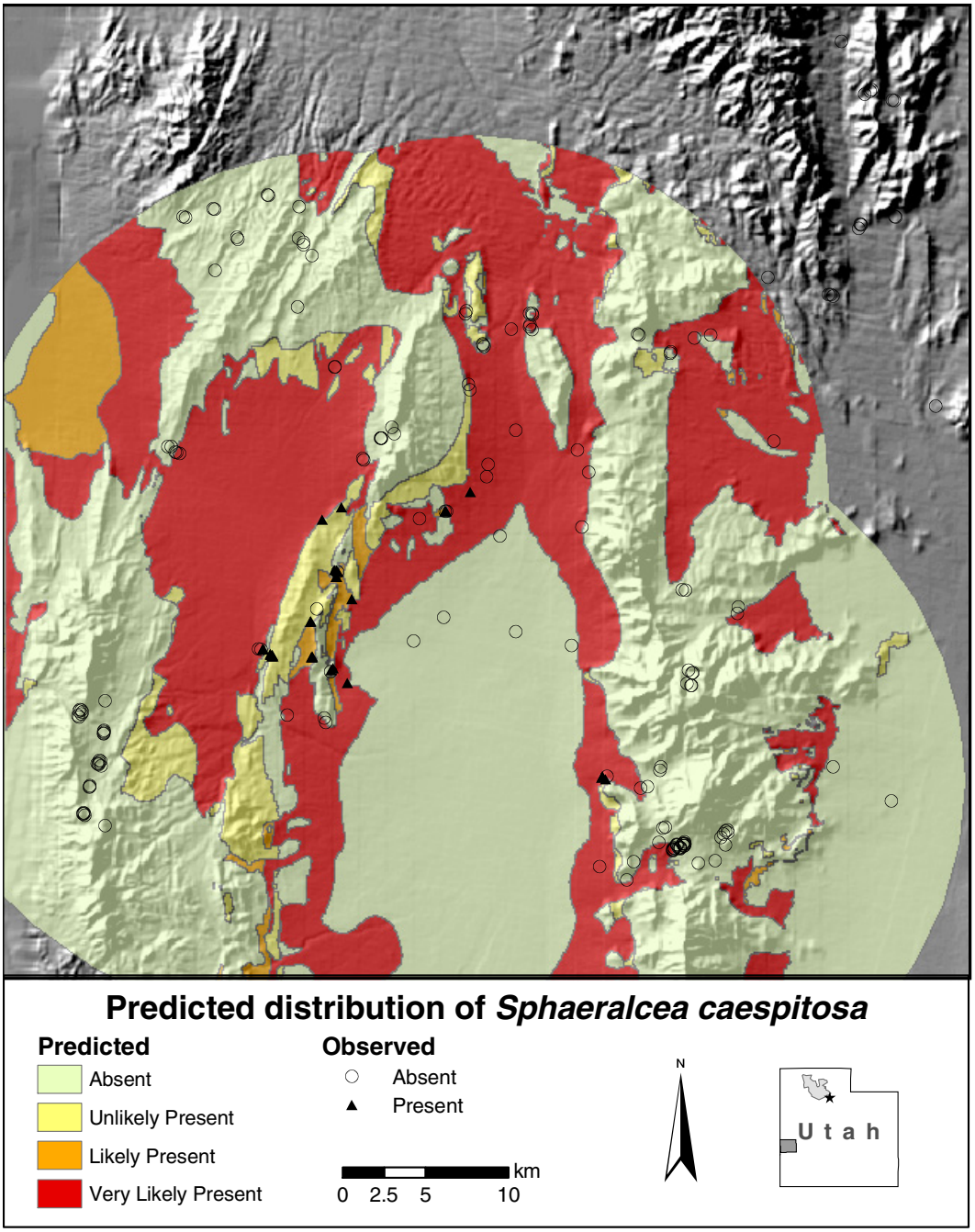

Fig. 3. Predicted distribution of *Sphaeralcea caespitosa*.

Bureau of Land Management, Department of Defense, and state lands) was provided from DOD/USDI Bureau of Land Management contract D910-A3-0210 with Utah State University. We are grateful to the curators of the Intermountain Herbarium, Utah State University; the Garrett Herbarium, University of Utah; and the S.L. Welsh Herbarium, Brigham Young University, for access to collections. We thank Robert Fitts and Trent Toler for help with field observations; M.A. (Ben) Franklin for generously sharing his mapped observations; Marcus Blood of the Hill Air Force Base Environmental Office for facilitating our work on Department of Defense properties; Dr. T.L. Sharik for his oversight of the project; and Jane Hayes, Andy Youngblood, Steven Smith, and 2 anonymous reviewers for their helpful suggestions in preparing this manuscript.

LITERATURE CITED

ALBEE, B.J., L.M. SHULTZ, AND S. GOODRICH. 1988. Atlas of the vascular plants of Utah. Utah Museum of Natural History, Salt Lake City.

ASHCROFT, G.L., D.T. JENSEN, AND J.L. BROWN. 1992. Utah climate. Utah Climate Center, Utah State University, Logan.

ATWOOD, D., J. HOLLAND, R. BOLANDER, B. FRANKLIN, D.E. HOUSE, L. ARMSTRONG, K. THORNE, AND L. ENGLAND. 1991. Utah endangered, threatened and sensitive plant field guide. USDA Forest Service, Intermountain Region, Ogden, UT.

AUSTIN, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling 157:101–118.

BIO, A.M.F., R. ALKEMADE, AND A. BARENDREGT. 1998. Determining alternative models for vegetation response analysis: a non-parametric approach. Journal of Vegetation Science 9:5–16.

BIO, A.M.F., P. DE BECKER, E.D. BIE, W. HUYBRECHTS, AND M. WASSEN. 2002. Prediction of plant species distribution in lowland river valleys in Belgium: modelling species response to site conditions. Biodiversity and Conservation 11:2189–2216.

BOESTCH, J.R, F.K. VAN MANEN, AND J.D. CLARK. 2003. Predicting rare plant occurrence in Great Smoky Mountains National Park, USA. Natural Areas Journal 23:229–237.

BROWN, D.G. 1994. Predicting vegetation types at treeline using topography and biophysical disturbance variables. Journal of Vegetation Science 5:641–656.

CAIRNS, D.M. 2001. A comparison of methods for predicting vegetation type. Plant Ecology 156:3–18.

CAWSEY, E.M., M.P. AUSTIN, AND B.L. BAKER. 2002. Regional vegetation mapping in Australia: a case study in the practical use of statistical modeling. Biodiversity and Conservation 11:2239–2274.

CHERRILL, A.J., C. MCCLEAN, P. WATSON, K. TUCKER, S.P. RUSHTON, AND R. SANDERSON. 1995. Predicting the distributions of plant species at the regional scale: a hierarchical matrix model. Landscape Ecology 10: 197–207.

COLLINGHAM, Y.C., R.A. WADSWORTH, B. HUNTLEY, AND P.E. HULME. 2000. Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. Journal of Applied Ecology 37:13–27.

CLARK, J.A., AND D. PREGIBON. 1992. Tree-based models. Pages 377–419 *in* J.M. Chambers and T.J. Hastie, editors, Statistical models in S. Wadsworth and Books, Pacific Grove, CA.

CONGALTON, R.G., K. GREEN, AND J. TEPLY. 1993. Mapping old-growth forests on national forest and park lands in the Pacific Northwest from remotely sensed data. Photogrammetric Engineering and Remote Sensing 59:529–535.

DALYRYMPLE, J.B., R.J. BLONG, AND A.J. CONACHER. 1968. A hypothetical nine-unit landsurface model. Zeitschrift für Geomorphologie 12:60–76.

DE'ATH, G., AND K.E. FABRICUS. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178–3192.

DIRNBÖCK, T., R.J. HOBBS, R.J. LAMBECK, AND P.A. CACCETTA. 2002. Vegetation distribution in relation to topographically driven processes in southwestern Australia. Applied Vegetation Science 5:147–158.

DOBBERTIN, M., AND G.S. BIGGING. 1998. Using the nonparametric classifier CART to model forest tree mortality. Forest Science 44:507–516.

EDWARDS, T.C., JR., D.R. CUTLER, N.E. ZIMMERMANN, L. GEISER, AND J. ALEBRIA. 2005. Model-based stratifications for enhancing the detection of rare ecological events. Ecology 86:1081–1090.

EJRNÆS, R., E. AUDE, B. NYGAARD, AND B. MÜNIER. 2002. Prediction of habitat quality using ordination and neural networks. Ecological Applications 12:1180–1187.

ELITH, J., AND M. BURGMAN. 2002. Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. Pages 303–313 *in* J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall, and F.B. Samson, editors, Predicting species occurrences: issues of accuracy and scale. Island Press, Washington, DC.

ELITH, J., M.A. BURGMAN, AND H.M. REGAN. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. Ecological Modelling 157:313–329.

ENGLER, R., A. GUISAN, AND L. RECHSTEINER. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology 41: 263–274.

ESRI. 1997. ARC/GRID. Environmental Systems Research Institute, Inc., Redlands, CA.

FERTIG, W., AND W.A. REINERS. 2002. Predicting presence/absence of plant species for range mapping: a case study from Wyoming. Pages 483–489 *in* J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall, and F.B. Samson, editors, Predicting species occurrences: issues of accuracy and scale. Island Press, Washington, DC.

FIELDING, A.H., AND J.F. BELL. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38–49.

FRANKLIN, J. 1995. Predictive vegetation mapping: Geographic modeling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography 19:474–499.

_____. 1998. Predicting the distribution of species in southern California from climate and terrain-derived variables. Journal of Vegetation Science 9:733–748.

_____. 2002. Enhancing a regional vegetation map with predictive models of dominant plant species in chaparral. Applied Vegetation Science 5:135–146.

FRANKLIN, M.A. 1996. Field survey for *Sphaeralcea caespitosa* M.E. Jones in the Beaver River and Warm Springs Resource Areas, Beaver and Millard Counties, Utah. Final Report for 1994/1995 Challenge Cost Share Project. Prepared for Utah Department of Natural Resources and USDI Bureau of Land Management.

GASTON, K.J. 1994. Rarity. Population and Community Biological Series 13. Chapman and Hall, London.

GIOIA, P., AND J.P. PIGOTT. 2000. Biodiversity assessment: a case study in predicting richness from the potential distributions of plant species in the forests of south-western Australia. Journal of Biogeography 17:1065–1078.

GUISAN, A., O. BROENNIMANN, R. ENGLER, M. VUST, N.G. YOCCOZ, A. LEHMANN, AND N.E. ZIMMERMANN. 2006. Using niche-based models to improve sampling of rare species. Conservation Biology 20:501–511.

GUISAN, A., T.C. EDWARDS, JR., AND T. HASTIE. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157:89–100.

GUISAN A., J.-P. THEURILLAT, AND F. KIENAST. 1998. Predicting the potential distribution of plant species in an alpine environment. Journal of Vegetation Science 9:65–74.

GUISAN, A., AND W. THUILLER. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8:993–1009.

GUISAN, A., S.B. WEISS, AND A.D. WEISS. 1999. GLM versus CCA spatial modeling of plant species distribution. Plant Ecology 143:107–122.

GUISAN, A., AND N.E. ZIMMERMANN. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147–186.

HARRISON, S., J.H. VIERS, AND J.F. QUINN. 2000. Climatic and spatial patterns of diversity in the serpentine plants of California. Diversity and Distributions 6: 153–161.

HEIL, K., M. PORTER, R. FLEMMING, AND W. ROME. 1993. Rare plant diversity between Capital Reef National Park and Canyonlands, Arches National Park and Bridges National Monument of Southeastern Utah. Pages 78–102 *in* R. Sivinski and K. Lightfoot, editors, Southwestern rare and endangered plants: proceedings of the Southwestern Rare and Endangered Plant Conference. New Mexico Forestry and Resources Conservation Division, Santa Fe.

HIJMANS, R.J., AND D.M. SPOONER. 2001. Geographic distribution of wild potato species. American Journal of Botany 88:2101–2211.

HILL, N.M., AND P.A. KEDDY. 1992. Prediction of rarities from habitat variables: coastal plain plants on Nova Scotian lakeshores. Ecology 73:1852–1859.

HINTZE, L.F. 1980. Geologic map of Utah. (1:500,000). Utah Geological and Mineral Survey, Salt Lake City, UT.

HIRZEL, A., AND A. GUISAN. 2002. Which is the optimal sampling strategy for habitat suitability modeling. Ecological Modelling 157:331–341.

HOLMGREN, N. 1972. Plant geography of the Intermountain region. Pages 77–161 *in* A. Cronquist, A.H. Holmgren, N.H. Holmgren, and J.L. Reveal. Intermountain flora: vascular plants of the Intermountain West, USA. Volume 1. New York Botanical Garden, Bronx.

HOOTEN, M.B., D.R. LARSEN, AND C.K. WIKLE. 2003. Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. Landscape Ecology. 18:487–502.

IMM, D.W., H.E. SHEALY, JR., K.W. MCLEOD AND B. COLLINS. 2001. Rare plants of southeastern hardwood forests and the role of predictive modeling. Natural Areas Journal 21:36–49.

JELASKA, S.D., O. ANTONIĆ, T. NIKOLIĆ, V. HRŠAK, M. PLAZIBAT, AND J. KRIŽAN. 2003. Estimating plant species occurrence in MTB/64 quadrants as a function of DEM-based variables—a case study for Medvednica Nature Park, Croatia. Ecological Modelling 170:333–343.

KNIGHT, P.J. 1993. A survey and field analysis for *Poa-triplex pleiantha* in San Juan County, NM. Pages 38–52 *in* R. Sivinski and K. Lightfoot, editors, Southwestern rare and endangered plants: proceedings of the southwestern rare and endangered plant conference. New Mexico Forestry and Resources Conservation Division, Santa Fe.

LEMAY, V.M., D.E. TAIT, AND B.J. VAN DER KAMP. 1994. Classification of cedar, aspen and true fir trees as decayed versus sound. Canadian Journal of Forest Resources 24:2068–2077.

LUOTO, M. 2000. Modelling of rare plant species richness by landscape variables in an agricultural area in Finland. Plant Ecology 149:157–168.

LUOTO, M., T. TOIVONEN, AND R.K. HEIKKINEN. 2002. Prediction of total and rare plant species richness in agricultural landscapes from satellite images and topographic data. Landscape Ecology 17:195–217.

MACDOUGALL, A.S., AND J.A. LOO. 2002. Predicting occurrences of geographically restricted rare floral elements with qualitative habitat data. Environmental Reviews 10:167–190.

MACDOUGALL, A.S., J.A. LOO, S.R. CLAYDEN, J.G. GOLTZ, AND H.R. HINDS.1998. Defining conservation priorities for plant taxa in southeastern New Brunswick, Canada, using herbarium records. Biological Conservation 86:325–338.

MANEL, S., H.C. WILLIAMS, AND S.J. ORMEROD. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38:921–931.

MATHSOFT. 1998. S-Plus 5 for UNIX: guide to statistics. Data Analysis Products Division, MathSoft, Seattle, WA.

MCAULIFFE, J. 1994. Landscape evolution, soil formation, and ecological processes in Sonoran Desert bajadas. Ecological Monographs 642:111–148.

MILLER, J., AND J. FRANKLIN. 2002. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. Ecological Modelling 157:227–247.

MOISEN G.G., AND T.S. FRESCINO. 2002. Comparing five modelling techniques for predicting forest characteristics. Ecological Modelling 157:209–225.

MOORE, D.M., B.G. LEES, AND S.M. DAVEY. 1991. A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. Environmental Management 15:59–67.

MYATT, M.M. 1987. Predicting the habitat geography of sensitive plants and community types. Pages 51–60 *in* T.S. Elias, editor, Conservation and management of rare and endangered plants: proceedings from a conference of the California Native Plant Society. California Native Plant Society, Sacramento.

NÆSSET, E. 1996. Use of weighted Kappa coefficient in classification error assessment of thematic maps. International Journal of Geographic Information Systems 10:591–604.

NILSSON, C., G. GRELSSON, M. JOHANSSON, AND U. SPERENS. 1988. Can rarity and diversity be predicted in vegetation along river banks? Biological Conservation 44:201–212.

PARKER, K.C. 1991. Topography, substrate and vegetation patterns in the northern Sonoran Desert. Journal of Biogeogaphy 18:151–163.

PEARCE, J., AND S. FERRIER. 2001. The practical value of modeling relative abundance of species for regional conservation planning: a case study. Biological Conservation 98:33–43.

RABINOWITZ, D. 1981. Seven forms of rarity. Pages 205–217 *in* H. Synge, editor, The biological aspects of rare plant conservation. John Wiley & Sons, Ltd., New York.

RAMSEY, R.D., AND L.M. SHULTZ. 2004. Evaluating the geographic distribution of plants in Utah from the *Atlas of the Vascular Plants of Utah.* Western North American Naturalist 64:421–432.

REICHENBACHER, F.W., AND D.V. ZAMUNDIO. 1993. Use of a climatological model in an endangered plant survey of international scope. Pages 16–37 *in* R. Sivinski and K. Lightfoot, editors, Southwestern rare and endangered plants: proceedings of the Southwestern Rare and Endangered Plant Conference. New Mexico Forestry and Resources Conservation Division, Santa Fe.

ROBERTS, D.W., AND S.V. COOPER. 1989. Vegetation mapping and inventory: concepts and techniques. Pages 90–96 *in* D. Ferguson, P. Morgan, and F.D. Johnson, editors, Land classifications based on vegetation: applications for resource managers. General Technical Report INT-257, USDA Forest Service, Intermountain Research Station, Ogden, UT.

ROBERTSON, M.P., C.I. PETER, M.H. VILLET, AND B.S. RIPLEY. 2003. Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modeling techniques. Ecological Modelling 164:153–167.

RUSHTON, S.P., S.J. ORMEROD, AND G. KERBY. 2004. New paradigms for modeling species distribution? Journal of Applied Ecology 41:193–200.

SHARIK, T.L., M.R. CONOVER, E.W. EVANS, C.R., HAWKINS, R.D. RAMSEY, D.W. ROBERTS, L.M. SHULTZ, ET AL. 1999. Population, distribution, and habitat study for threatened, endangered, and sensitive species of plants and animals within the Hill Air Force Base (HAFB) restricted air space and associated public lands. College of Natural Resources, Utah State University–Bureau of Land Management Agreement No. D910-A3-0210. Logan, UT.

SHULTZ, L.M. 1993. Patterns of endemism. Pages 249–269 *in* R. Sivinski and K. Lightfoot, editors, Southwestern rare and endangered plants: proceedings of the Southwestern Rare and Endangered Plant Conference. New Mexico Forestry and Resources Conservation Division, Santa Fe.

SHULTZ, L.M., R.D. RAMSEY, AND W. LINDQUIST. 2005. Digital atlas of Utah plants. College of Natural Resources Website, Utah State University. Logan. Available from: http://www.gis.usu.edu/geography-department/utgeog/utvatias/ut-vascatlas.html

SPERDUTO, M.B., AND R.G. CONGALTON. 1996. Predicting rare orchid (small whorled pogonia) habitat using GIS. Photogrammetric Engineering and Remote Sensing 62:1269–1279.

STAHLE, D.W., AND P.L. CHANEY. 1994. A predictive model for the location of ancient forests. Natural Areas Journal 14:151–158.

THUILLER, W., M.B. ARAÚJO, AND S. LAVOREL. 2003. Generalized models vs. classification tree analysis: predicting spatial distributions of different species at different scales. Journal of Vegetation Science 14: 669–680.

VAN HORSSEN, P.W., P.P. SCHOT, AND A. BARENDREGT. 1999. A GIS-based plant prediction model for wetland ecosystems. Landscape Ecology 14:253–265.

VAYSSIÈRES, M.P., R.E. PLANT, AND B.H. ALLEN-DIAZ. 2000. Classification and regression trees: an alternative nonparametric approach for vegetation analysis. Journal of Vegetation Science 11:679–694.

WELSH, S.L., N.D. ATWOOD, S. GOODRICH, AND L.C. HIGGINS, EDITORS. 1993. A Utah flora. 2nd edition. Brigham Young University, Provo, UT.

_____. 2003. A Utah flora. 3rd edition. Brigham Young University, Provo, UT.

WHITE, P.S., AND R.I. MILLER. 1988. Topographic models of vascular plant richness in the southern Appalachian high peaks. Journal of Ecology 76:192–199.

WILSON, L., M.E. OLSEN, T.B. HUTCHING, A.R. SOUTHARD, AND A.J. ERICKSON. 1975. Soils of Utah. Agricultural Experiment Station, Bulletin 492. Utah State University, Logan.

WISER, S.K., R.K. PEET, AND P.S. WHITE. 1998. Prediction of rare-plant occurrence: a southern Appalachian example. Ecological Applications 8:909–920.

WU, B., AND F.E. SMIENS. 2000. Multiple-scale habitat modeling approach for rare plant conservation. Landscape and Urban Planning 51:11–28.

ZANIEWSKI, A.E., A. LEHMANN, AND J. MCC. OVERTON. 2002. Predicting species spatial distributions using presence-only data: a case of native New Zealand ferns. Ecological Modelling 157:261–280.

ZIMMERMANN, N.E., AND F. KIENAST. 1999. Predictive mapping of alpine grasslands in Switzerland: species versus community approach. Journal of Vegetation Science 10:469–482.

Received 9 September 2005
Accepted 5 June 2006

Western North American Naturalist

Volume 67 | Number 1 Article 4

3-27-2007

Modeling distributions of rare plants in the Great Basin, western North America

M. Aitken
Utah State University, Logan

D. W. Roberts
Utah State University, Logan

L. M. Shultz
Utah State University, Logan