

Predicting Rare Plant Occurrences in Great Smoky Mountains National Park, USA

John R. Boetsch¹

National Park Service
Great Smoky Mountains National Park
Twin Creeks Natural Resources Center
1314 Cherokee Orchard Road
Gatlinburg, TN 37738 USA

Frank K. Van Manen²

Department of Forestry, Wildlife and Fisheries
University of Tennessee
Knoxville, TN 37901-1071 USA

Joseph D. Clark

Southern Appalachian Field Laboratory
Biological Resources Division
U.S. Geological Survey
274 Ellington Plant Sciences Building
University of Tennessee
Knoxville, TN 37901-1071 USA

¹ Corresponding author current address:
John R. Boetsch, Olympic National Park,
600 E. Park Avenue, Port Angeles,
Washington 98362-6757 USA;
John_Boetsch@nps.gov

² Current address: Southern Appalachian
Field Laboratory, Biological Resources
Division, U.S. Geological Survey, 274
Ellington Plant Sciences Building,
University of Tennessee, Knoxville, TN

ABSTRACT: We investigated the applicability of biometric habitat modeling to rare plant inventory and conservation by developing and field testing a geographically explicit model for *Cardamine clematitis* Shuttleworth ex A. Gray (mountain bittercress), an endemic plant of the southern Blue Ridge Mountains, USA. For each of 187 confirmed coordinates for *C. clematitis* in Great Smoky Mountains National Park, 13 habitat variables were measured with a geographic information system. These data were used to calculate Mahalanobis distances for each 30-m x 30-m pixel within the study area; small values of Mahalanobis distance represented site conditions similar to those of known locations of *C. clematitis*, whereas larger distance values represented dissimilar conditions. Following model development, we tested model performance by sampling 120 randomly distributed plots for *C. clematitis* presence. Logistic regression showed that Mahalanobis distance values were strongly related to *C. clematitis* occurrence ($P = 0.039$). Overall, 75% of all known occurrences of *C. clematitis* had associated Mahalanobis distance values below 17.7, and 95% of all occurrences were below 33.8; the median Mahalanobis distance value for the study area as a whole was 40.0. A habitat suitability cutoff value was defined which identified roughly 23,640 ha (19.5% of the study area) as suitable habitat. Although the model successfully predicted species absence in test plots with high Mahalanobis distance values, many sites with low values did not contain *C. clematitis*. Only 16.2% of test plots below the habitat suitability cutoff contained *C. clematitis*. The absence of *C. clematitis* from sites with low Mahalanobis distance values (low specificity) is not necessarily indicative of a poor model; metapopulation processes (e.g., recolonizations, local extinctions) have been shown to play a major role in presence or absence of many plant species. That may be partially the case with our model as evidenced by a relationship between *C. clematitis* presence and habitat patch size.

Predicción de Ocurrencia de Plantas Raras en el Parque Nacional Great Smoky Mountains, USA

RESUMEN: Investigamos la aplicabilidad del modelado biométrico del hábitat a los inventarios de plantas raras y su conservación mediante el desarrollo y la corroboración a campo de un modelo geográficamente explícito para *Cardamine clematitis* Shuttleworth ex A. Gray (mountain bittercress) una planta endémica del sur de Blue Ridge Mountains, USA. Para cada una de las 187 coordenadas en que confirmamos la presencia de *C. clematitis* en el parque nacional Great Smoky Mountains, se midieron 13 variables de hábitat con un sistema de información geográfica. Esos datos se usaron para calcular las distancias Mahalanobis para cada uno de los pixels de 30 m x 30 m dentro del área de estudio; valores pequeños de distancia Mahalanobis representaron sitios con condiciones similares a aquellos en que se presentaba *C. clematitis*, mientras que distancias grandes representaban condiciones diferentes. Una vez realizado el modelo, testamos su performance muestreando 120 plots distribuidos al azar en busca de *C. clematitis*. La regresión logística mostró que las distancias Mahalanobis estuvieron fuertemente relacionadas a la presencia de *C. clematitis* ($P = 0.039$). El 75% de todas las ocurrencias conocidas de *C. clematitis* tuvieron distancias asociadas menores de 17,7; y el 95% de todas las ocurrencias estuvieron por debajo de 33,8; la mediana de la distancia Mahalanobis para el área de estudio en general fue 40,0. Un valor de corte de conveniencia de hábitat fue definido identificando 23649 ha (19,5% del área de estudio) como área apropiada. Aunque el modelo predijo con éxito la ausencia de la especie en los plots con alta distancia Mahalanobis, muchos sitios con valores bajos no contenían *C. clematitis*. Sólo el 16,2% de los plots testeados bajo el punto de corte de conveniencia de hábitat contenían *C. clematitis*. La ausencia de *C. clematitis* en lugares con baja distancia Mahalanobis (baja especificidad) no es necesariamente indicador de un modelo pobre; los procesos de metapoblación (por ej. recolonización, extinciones locales) han demostrado que juegan un rol mayor en la presencia o ausencia de muchas especies de plantas. Este puede ser parcialmente el caso con nuestro modelo, como lo evidencia la relación entre la presencia de *C. clematitis* y el tamaño del parche de hábitat.

Index terms: conservation planning, habitat modeling, landscape model, predicting plant species occurrence, rare species

INTRODUCTION

Information on the occurrence of rare and sensitive species is often incomplete and heavily influenced by where collectors have searched in the past. Although knowl-

edgeable field biologists can capably evaluate the suitability of a specific site for certain rare species, this approach can only provide qualitative guidelines for further surveys, is costly, and is impractical as a means of storing data. Alternatively, pre-

dictive habitat models may provide an effective and unbiased method of delineating areas where selected species are most likely to occur, or of identifying areas suitable for restoration or reintroduction. Equally important, large-scale habitat models can be used to evaluate the distribution and connectivity of habitat, which have important implications for metapopulation dynamics and, consequently, for the long-term survival of species.

Most previous attempts to predict individual species occurrences have employed logistic regression (Hosmer and Lemeshow 1989, ter Braak and Looman 1995), discriminant function analysis (Lachenbruch 1975), or classification tree (CART) analysis (Breiman et al. 1984, Moore et al. 1991). Those techniques require two sets of data to calibrate the model—the first set describes the conditions associated with species presence, and the second set corresponds to presumably unsuitable sites. An important assumption is that those two data classes are mutually exclusive and it is assumed that species absence indicates conditions that are unsuitable. These assumptions are easily violated, particularly with rare species, because presence is affected by the ability of the species to colonize new areas and because stochastic processes often lead to local extinction, both of which may be unrelated to habitat suitability. Moreover, absence data are frequently selected at random and it may be falsely assumed that such a random sample represents a suitable null model. Considering the large amount of conservation information that is stored as presence data (e.g., The Nature Conservancy's element occurrence data, used by Natural Heritage Programs across the United States), a modeling approach that requires only such data could be of great advantage in avoiding restrictive assumptions.

For these reasons, we selected the Mahalanobis distance statistic (Rao 1952) to model suitable habitat for rare plant species. Mahalanobis distance is a multivariate index of dissimilarity that can be calculated using data associated only with presence of the species, thereby avoiding the distributional assumptions of the other modeling techniques (Clark et al. 1993a).

We illustrate the application of this modeling approach with *Cardamine clematitis* Shuttleworth ex A. Gray (Brassicaceae), mountain bittercress, a vascular plant species endemic to the southern Blue Ridge Mountains of southwest Virginia, North Carolina, and Tennessee, USA. It is an evergreen, herbaceous perennial that flowers in May and fruits in May to June. In 1997, when modeling was initiated, *C. clematitis* was listed as a Federal Category 2 species (i.e., Species of Concern) because of its narrow range and rarity, and inventory work was already underway that underscored the immediate need for better information to guide field work. Most reports associate this species with forested stream banks and seepage areas above 1000 m elevation (Kearney 1892, White 1982, Wofford 1989, Weakley 1996), typically associated with cold montane mixed coniferous/deciduous forests, which in the southern Appalachian Mountains are themselves rare and geographically restricted. This species is considered to be distinct from *C. flagellifera* O. E. Schulz (Dudley 1974, Al-Shehbaz 1988), which is found on low slopes, streambanks and alluvial flats at lower elevations.

Our approach to habitat modeling depended on the assumption that the study species is associated with optimal and measurable habitat conditions across the study region (Knick and Rotenberry 1998), and that GIS variables at 30-m resolution are adequate surrogates to measure such conditions. Our objective was to determine whether such remotely sensed habitat characteristics corresponding to known locations of *C. clematitis* at GSMNP could be used to predict the occurrence of unknown populations. Following that, our goal was to use independent field data to test the hypothesis that Mahalanobis distance values are associated with the probability of occurrence of the species.

METHODS

Study Area

Most known occurrences of *C. clematitis* are within Great Smoky Mountains National Park (GSMNP), a 206,119-ha reserve along the boundary of Tennessee

and North Carolina. GSMNP is characterized by mountainous terrain with steep slopes, narrow valleys and coves, and fast-flowing streams (King and Stupka 1950). Elevations at GSMNP range from 270 to 2027 m. Thornthwaite (1948) classified this area as a temperate rain forest with an average annual precipitation of about 230 cm at the highest elevations. Approximately 20% of GSMNP was never logged, and an additional 30% has been only diffusely disturbed since European settlement (Pyle 1985, 1988). The study area was confined to that portion of GSMNP above 914 m, which encompassed all known occurrences of *C. clematitis* in the park, and totaled 121,531 ha in size. By confining the study area to higher elevations, we avoided testing the model in areas where the absence of the species was a near certainty, thus decreasing the probability of type I errors (false rejection of the null hypothesis).

Compilation of Location Data for Model Development

Information on the locations of *C. clematitis* populations was gathered from the North Carolina and Tennessee Natural Heritage Programs, and from records maintained at GSMNP. We surveyed 18 grouped occurrences of *C. clematitis* (defined as populations by the Heritage Programs) in 1997 to verify species presence and to maximize the accuracy of the position coordinates. Comprehensive surveying and mapping of these and newly encountered populations resulted in 187 unique coordinate positions for *C. clematitis*. Eighty-five of the input positions were recorded with an unassisted military Y-code signal from the U.S. Department of Defense's global positioning system (GPS) (PLGR+96, Rockwell International, Cedar Rapids, Iowa). The GPS unit yielded an average displacement error of 12 m with 50% of errors < 10 m and 95% < 20 m. Where satellite reception was hindered by vegetation, terrain, or ionospheric interference, positions were hand-mapped by relating altimeter and compass readings to the local features depicted on U.S. Geological Survey 7.5-minute quadrangle maps. Universal Transverse Mercator (UTM) coordinates were derived from hand-mapped positions with an estimated error of 50 m.

Collection of Model Calibration Data from the GIS Database

We used a database of 13 GIS habitat variables to characterize habitat conditions of the 187 sample locations (Table 1). All GIS coverages were continuous variables based on 30-m x 30-m pixels, with the exception of vegetation type (MacKenzie 1993), which was a categorical variable based on 90-m x 90-m pixels. This categorical variable was transformed into binary design (also “dummy”) variables prior to modeling. Because > 95% of the positions occurred in four of six forest types (i.e., spruce-fir, northern hardwood, cove hardwood, and mesic oak), a separate design variable was created for each of those four types.

Mahalanobis Distance Modeling

Mahalanobis distance is a multivariate statistic that characterizes dissimilarity (Rao 1952). This statistic was calculated for each pixel in the GIS coverage of the study area by combining the information from 16 habitat coverages (i.e., the 12 continuous topographic variables and 4 design variables for vegetation type) with the following formula:

$$\text{Mahalanobis distance} = (\underline{x} - \hat{\underline{\mu}})' S^{-1} (\underline{x} - \hat{\underline{\mu}}),$$

where $\underline{x}$ is a vector of habitat characteristics in the GIS grid, $\hat{\underline{\mu}}$ is the mean vector of habitat characteristics of the original sampling points, and S^{-1} is the inverse of the variance-covariance matrix calculated from the sampling points. The Mahalanobis distance statistic represents the standard squared distance between a set of sample variates, $\underline{x}$, and “ideal” habitat characterized by the presence data and represented by $\hat{\underline{\mu}}$. Small values of Mahalanobis distance represent habitat conditions similar to those of known locations of *C. clematidis*, whereas larger distance values represent dissimilar conditions. A variety of habitat combinations can result in identical distance values. Mahalanobis distance is dimensionless because it is a function of standardized variables, despite the different measurement scales among the original variables (Clark et al. 1993a). An additional advantage of this technique is that

collinearity among predictor variables is compensated for by the inclusion of the variance-covariance matrix in the calculations.

We calculated $\hat{\underline{\mu}}$ and S^{-1} with SAS (Versions 6.0-8.0) based on the habitat characteristics of the 187 sample locations. We then used this information and the above formula to calculate Mahalanobis distance in Arc/Info® GRID (ESRI, Redlands, California) for each 30-m x 30-m pixel in the study area (Clark et al. 1993b).

Independent Field Testing

Because of the rarity of *C. clematidis*, a completely randomized sampling design would have resulted in an excessive number of test plots where the species was predicted to be absent. Therefore, we developed a stratified sampling design to concentrate our effort in areas with lower Mahalanobis distance values (presumably supportive sites), while still testing the full range of values in the study area. Although the chi-square distribution can be used to rescale the Mahalanobis distance values to probability classes (Clark et al. 1993a), we did not use this approach for stratification as the assumption of multivariate normality could not be met. Because the sampling intensity of different classes of Mahalanobis values is a function of their availability within the study area, we instead stratified the Mahalanobis distance classes according to the number of pixels within the study area, using the following geometric equation:

$$100 \text{ (percent total pixels)} = \frac{n + 2n + 4n + 8n + 16n + 32n}{n + 2n + 4n + 8n + 16n + 32n} \times 100$$

This equation doubles the amount of area in each successive class and, as such, $n = 1.59$. Thus, class 1 contained the lowest 1.59% of the pixels in the study area (i.e., those with the lowest Mahalanobis distance values), class 2 represented 3.17% of the study area (the next lowest values), class 3 represented the next 6.35%, etc. The ranges of Mahalanobis distance values for classes 1–6 were as follows: 3.8 (minimum distance)–10.5, 10.5–13.4, 13.4–17.3, 17.3–24.0, 24.0–39.7, and

39.7–1164.9 (maximum distance), respectively. We sampled 20 test plots in each of the six classes, thereby concentrating the per area effort to those sites with smaller Mahalanobis distance values. We used Arc/Info® to generate the stratified random sample positions.

We restricted the location of study plots to areas 100–500 m from trails or roads to facilitate sampling. The 100-m minimum distance was designed to reduce the influence of incidental effects associated with trails and roads (e.g., understory disturbance related to road/trail maintenance, or encroachment by weedy species into gaps created by this disturbance). Although this approach restricted testing to a smaller area, we assumed the sample was unbiased because the frequency distribution of the number of pixels in each of the six stratification classes in the test area did not differ from what would be expected if the entire study area had been sampled (asymptotic Kolmogorov-Smirnov statistic = 0.577, $P = 0.893$). The resulting test area was 42,236 ha in size, which represented 34.8% of the study area and 20.5% of GSMNP.

Field testing occurred during autumn 1997 and spring 1998. Ninety percent of the test plots ($n = 108$) were located with the GPS receiver (mean error = 11.7 m, median = 9.5 m), whereas the remaining 10% were located with map, compass, and altimeter. Once the sample position was established, a square area representing a 30-m x 30-m pixel centered on this point was surveyed, and the presence or absence of *C. clematidis* was recorded.

Data Analysis

We used logistic regression (Hosmer and Lemeshow 1989) to determine the relationship between the presence and absence of *C. clematidis* in the test plots and the distance values of the corresponding pixels. We tested for the assumption of linearity in the logit by use of the Box-Tidwell transformation (Guerro and Johnson 1982) and tested the fit of the model by use of the Hosmer-Lemeshow goodness-of-fit test (Hosmer and Lemeshow 1989).

Table 1. Variables used to calculate Mahalanobis distance values for *Cardamine clematidis* in Great Smoky Mountains National Park, USA, 1997–1998.

Name	Description ^a	Range ^b	Source
Aspect	Aspect transformed using: $1 + \cos(45 - \text{aspect})$	0.0–2.0	Beers et al. (1966)
Elevation	Digital elevation model (m)	914–2027	U.S. Geological Survey
Landform Index	Index of meso-scale topographic exposure	-4.0–93.6	McNab (1993)
Planiform Curvature	Slope curvature in horizontal plane	-8.8–8.8	calculated from elevation with the CURVATURE command (Grid)
Profile Curvature	Slope curvature in vertical plane	-10.9–8.6	calculated from elevation with the CURVATURE command (Grid)
Relative Slope Position	Relative slope position (%)	0–100	Wilds (1996)
Slope	Slope steepness (degrees)	0.0–64.5	calculated from elevation with the CURVATURE command (Grid)
Solar Incidence	Index of exposure to sunlight; approximated for the solar equinox; considers both nearby and distant shadowing factors	1–227	calculated from elevation with the HILLSHADE command (Grid)
Shannon-Weiner Index	Index of topographic complexity	17–35	Miller (1986)
Topographic Convergence Index	Simulates the flow accumulation of water; $TCI = 10 \times \ln(A / \tan B)$, where A is drained surface area and B is drained surface slope	17–182	Beven and Kirkby (1979), Wolock and McCabe (1995), Halpin (1995)
Topographic Relative Moisture Index	Index of moisture considering the effects of slope position, aspect and elevation	0–85	Parker (1982)
Terrain Shape Index	Index of micro-scale topographic exposure	-169–71	McNab (1989)
Vegetation	Vegetation type (9 forest types, health bald, grassy bald, grape thicket, or treeless)	Categorical	MacKenzie (1993)

^a All variables were continuous with spatial resolutions of 30 m, except for vegetation, which was a categorical variable with a spatial resolution of 90 m.

^b Range values are based on the study area, which was defined as the area in Great Smoky Mountains National Park with elevations > 914 m.

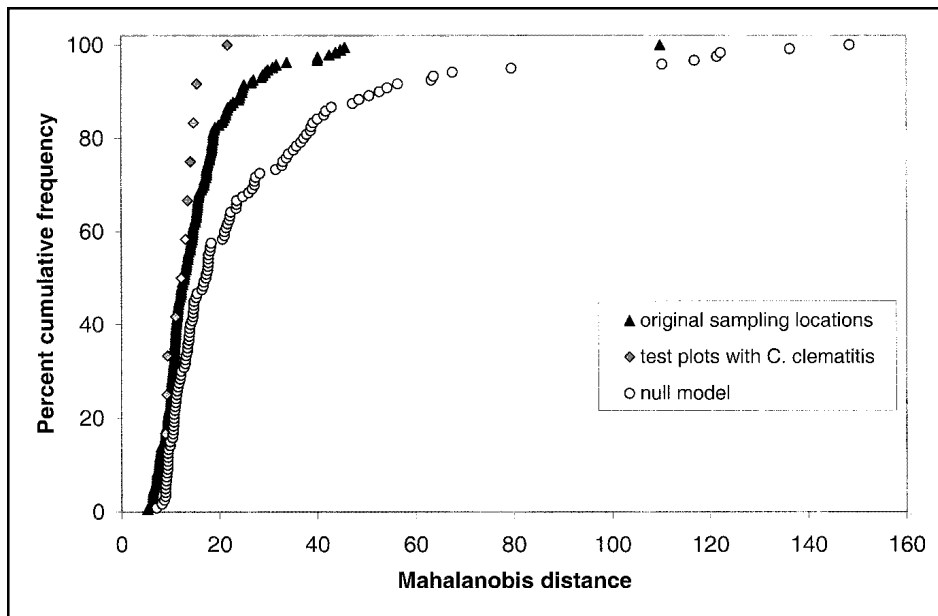


Figure 1. Cumulative frequency curves of *Cardamine clematidis* locations in Great Smoky Mountains National Park, USA, and Mahalanobis distance values. Original sampling locations are those used to build the model ($n = 187$). The null model curve represents a random distribution of *C. clematidis* across the range of Mahalanobis distance values sampled; those test plots where *C. clematidis* were found ($n = 12$) are depicted as well.

To evaluate the influence of habitat patch area on the occurrence of *C. clematidis* (e.g., patch size may influence extinction or recolonization of local populations), we calculated the total area of contiguous habitat patches at each test plot. These habitat patches were delineated by determining the value for Mahalanobis distance that maximized the sensitivity of the model (i.e., the lowest value that encompassed all occurrences in the independent test plots); all contiguous pixels at or below this value were included in the individual patch area calculations.

RESULTS

Mahalanobis distance values for the study area ranged from 3.8 to 1,164.9 (mean = 54.0, median = 40.0, $SD = 38.8$). Values for the original sampling locations (i.e., model calibration data) of *C. clematidis* ranged from 5.3 to 109.8 (mean = 15.8, median = 12.8, $SD = 12.3$). Ninety-five percent of the original sampling locations had distance values < 30 (Figure 1).

Mahalanobis distance values for the test locations ranged from 7.1 to 148.5. Twelve of the 120 test plots contained populations

of *C. clematidis* (Table 2), all with distance values below 21.7 (Figure 1). Overall, 75% of all known *C. clematidis* occurrences in GSMNP corresponded to distance values below 17.7, and 95% corresponded to values below 33.8.

The logistic regression analysis of the test data indicated that Mahalanobis distance values were related to occurrence of *C. clematidis* (parameter estimate = -0.132, $P = 0.039$; Table 3). That is, lower distance values were associated with presence of the species in the field test plots and higher values were associated with absence of the species. For any 10-point decline in Mahalanobis distance, the likelihood of occurrence increased by a factor of 3.8 (adjusted odds ratio; 95% confidence interval = 1.1–13.2).

A cutoff value of 21.7 was assigned to define suitable habitat; this value corre-

Table 2. Presence/absence of *Cardamine clematidis* in 120 independent test plots, Great Smoky Mountains National Park, USA, 1997–1998.

Stratification Class	Min Mahalanobis Distance	Max Mahalanobis Distance	Area (ha)	<i>C. clematidis</i> Present	<i>C. clematidis</i> Absent
1	3.8	10.5	1921	4	16
2	10.5	13.4	3843	3	17
3	13.4	17.3	7686	4	16
4	17.3	24.0	15,372	1	19
5	24.0	39.7	30,745	0	20
6	39.7	1,164.9	61,490	0	20
Total			121,357	12	108

Table 3. Logistic regression results to determine the relationship between Mahalanobis distance values and presence or absence of *Cardamine clematidis* in 120 test plots, Great Smoky Mountains National Park, USA.

Variable	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate
Intercept	-0.028	0.894	0.001	0.975	—
Mahalanobis Dist.	-0.132	0.064	4.244	0.039	1.98
Hosmer-Lemeshow goodness-of-fit statistic = 6.962, $df = 8$, $P = 0.541$					
Maximum rescaled $R^2 = 0.178$					

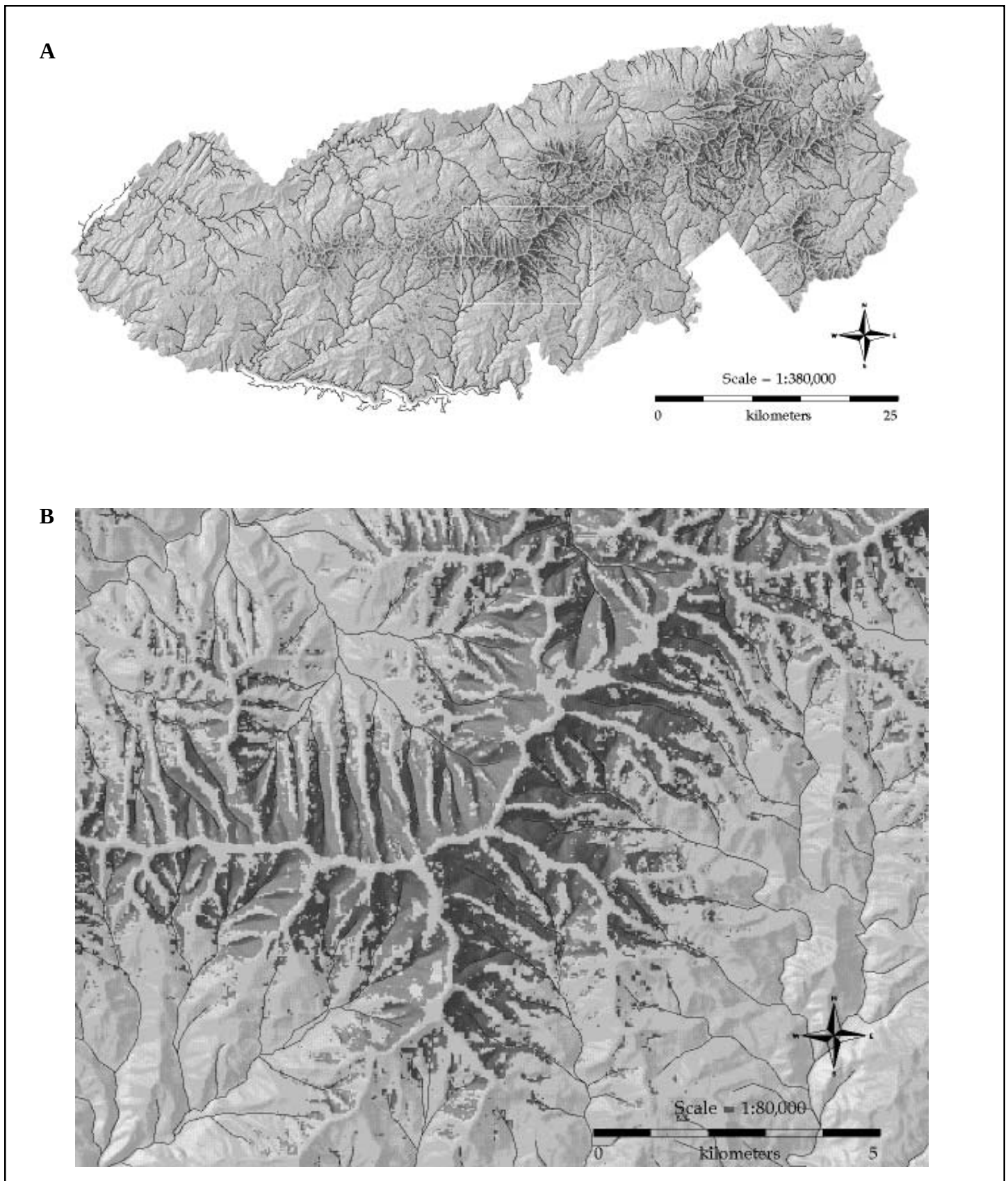


Figure 2a-b. Habitat for *Cardamine clematidis* (mountain bittercress) in Great Smoky Mountains National Park, North Carolina/Tennessee, USA. Habitat was defined as distance values below 21.7, and comprised 19.5% of the study area. Suitable habitat is shaded in green, with the remainder of the study area shaded in yellow. An inset for the Clingmans Dome area in Figure 2a is shown in close view in Figure 2b.

sponded to the lowest value that encompassed all occurrences of *C. clematidis* in the independent test plots. This cutoff identified approximately 23,640 ha (19.5%) of the study area as suitable habitat for the species (Figure 2). Of the 74 plots with Mahalanobis distance values below the cutoff, only 16.2% contained *C. clematidis*.

Suitable *C. clematidis* habitat, as we defined it, was dispersed across the study area in 19,863 distinct patches of varying size. When charted against Mahalanobis distance, the habitat patch size distribution for test plots below the cutoff value showed a possible threshold. *Cardamine clematidis* was absent from plots in patches below approximately 40 ha in size (Figure 3). Of the 62 observed absences below the suitable habitat cutoff, 19 (30.6%) had corresponding patch sizes below 40 ha.

DISCUSSION

From an initial sample of 187 *C. clematidis* positions, a multivariate statistic was generated that accurately predicted previous-

ly unknown occurrences. A comparison of the cumulative frequency curve of the test plots with the expected curve under the null model (i.e., species occurrences are randomly distributed across the range of Mahalanobis distance values within each of the six sampling strata) indicates that the Mahalanobis distance model effectively identified site conditions that contribute to species presence (Figure 1). Furthermore, *C. clematidis* was found in a narrow range of Mahalanobis distance values, beyond which habitat quality declined rapidly. This suggests relatively specific habitat requirements.

Where habitat conditions were unfavorable (large Mahalanobis distance), the model reliably predicted the absence of *C. clematidis*. However, the species frequently was not found where the model predicted favorable habitat conditions (low Mahalanobis distance). The rate of occurrence in the test plots suggests a habitat saturation rate under 20% for this species. We speculate that factors other than site quality may have contributed to that result. According to metapopulation theory (Han-

ski and Simberloff 1997, Hanski 1998), small and/or distant patches of suitable habitat have a reduced likelihood of successful colonization and persistence. Our finding that patch size is related to species occurrence is consistent with this theory. It is probable that *C. clematidis*, for which water and gravity are the primary agents of seed dispersal, could be severely limited by successful dispersal events, which suggests a metapopulation at nonequilibrium.

In addition, the data resolution may be insufficient to fully capture the fine-scale habitat characteristics for this species. Our observations were that *C. clematidis* was found in areas with relatively well developed microtopography and near distinct breaks in topography, features that are difficult or impossible to detect at a 30-m pixel resolution. As fine-scale elevation data become available, models based on topographic variables may improve accordingly. Identifying important variables that are difficult to include at the landscape scale (e.g., vegetation structure, or the presence of evergreen *Rhododendron* L. species) may also improve model specificity.

We could have tested the Mahalanobis distance model by evaluating how well it fit the input data or by repeatedly withholding a portion of the input data for testing (i.e., jackknifing). Instead, we tested our model with independent data to confirm species presence in suitable areas and species absence in unsuitable areas. Without such an independent testing effort, the percentage of suitable habitat where the species was present could not have been calculated. Neither could absence from unsuitable sites be demonstrated. Moreover, independent field data allowed detection of potential sampling biases among the input data, which is particularly important in cases where the input data depend heavily on historical records of where species have been observed in the past.

Habitat models for plant species are relatively uncommon, particularly compared with wildlife habitat models, which have been an active area of research in recent decades (e.g., Verner et al. 1986, Corsi et

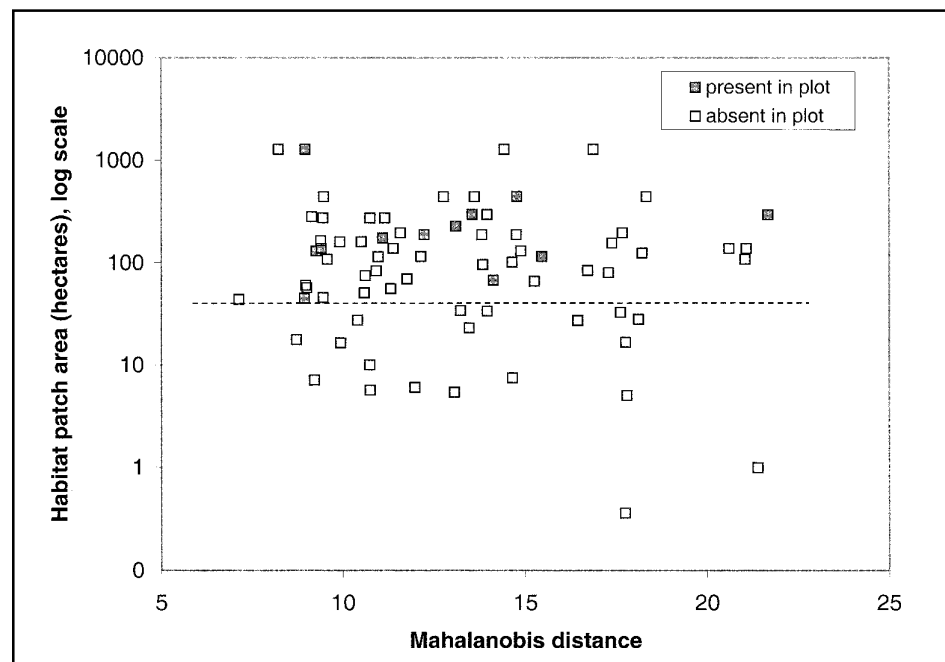


Figure 3. Scatterplot of habitat patch size against Mahalanobis distance for test plots below the habitat suitability cutoff value of 21.7. Of the 74 plots below this cutoff value, 12 contained *C. clematidis* (filled squares) and 62 did not (empty squares). A threshold patch size of approximately 40 ha (dashed line) can be discerned below which *C. clematidis* was not observed. This threshold explained 19 of the 62 cases in which *C. clematidis* was absent.

al. 1999, Scott et al. 2001). Terrestrial plants may be superior candidates for habitat modeling, however, because they are sessile and more directly affected by microclimate, topography, and other environmental factors for which remotely sensed data may provide adequate surrogates. Also, animals are more likely than plants to be encountered in suboptimal habitat, simply because animals are more conspicuous than plants as they disperse through a landscape of variable habitat quality. This argument favors the application of biometric modeling to plant species, given the important requirement that model calibration data represent optimal habitat conditions across the landscape (Knick and Rotenberry 1998). Finally, field testing of predicted locations is straightforward because plants are often easier to detect than animals.

Management Implications

The technique demonstrated here may be applied to other rare plant species and, as such, can be a valuable aid to finding specimens for genetic conservation or for developing plant conservation strategies by identifying core habitat areas or predicting the effects of change. Automation of modeling could also lead to the simultaneous development and testing of models for multiple species within the same preserve. Despite the limitations inherent in landscape modeling, the development of habitat models for rare species is a crucial step toward providing the spatial information required by land managers and resource specialists to locate and assess potential habitat.

Some species are likely to be more amenable to landscape-scale habitat modeling than others. The following characteristics may help identify good candidates for modeling: (1) species that are relatively common in their respective habitat (i.e., high habitat saturation); such species are much more likely to be captured by stratified sampling, and will typically provide a larger data set for model development and testing; (2) species that respond to large-scale gradients that can be simulated using GIS surrogate predictors; species that respond to small-scale factors (e.g., mi-

cro-scale disturbance, soil resource ratios, etc.) may be much more difficult to model across large spatial areas because the information on predictive conditions is often lacking at that scale; and (3) species that are reliably associated with specific environmental conditions across the range of application of the model; species associated with heterogeneous conditions across their range will be more difficult to model, especially if the information used to build the model does not represent that heterogeneity.

ACKNOWLEDGMENTS

We acknowledge the logistic support provided by the staff of Great Smoky Mountains National Park throughout the duration of this project, with special appreciation for Janet Rock and Keith Langdon for their encouragement and assistance. Many thanks to Peter White, Bob Peet, and Dean Urban for their helpful comments and review of an earlier draft of this manuscript. We also acknowledge Dr. Charles Williams and an anonymous reviewer for their valued contributions. We thank staff of the Natural Heritage Programs of Tennessee, North Carolina, and GSMNP for providing preliminary location data. We are indebted to the field technicians and volunteers at GSMNP who cheerfully endured the long hours and wet boots: Eric Nielsen, Jennifer Dean, Erica Choberka, Jennifer Tietjen, Thomas Moore, and Brian Yahn. Henry McNab and Don von Blaricomb generously provided the code for two GIS indices (TSI and LFI). This project was funded in part by U.S. Geological Survey, Biological Resources Division agreement #1445-1A09-96-0027.

John Boetsch is an Ecologist at Olympic National Park where he manages ecological inventory and monitoring data. His diverse interests include the biogeography, autecology, and demography of rare plants, plant community ecology, and ecological modeling.

Frank van Manen is a Research Ecologist with the U.S. Geological Survey's Southern Appalachian Field Branch in Knoxville, Tennessee. His primary research interest is in landscape ecology as it relates to large carnivores, rare plants and tree species.

Joe Clark is Branch Chief of the U.S. Geological Survey Southern Appalachian Field Branch. He specializes in population dynamics.

Joe Clark is Branch Chief of the U.S. Geological Survey Southern Appalachian Field Branch. He specializes in population dynamics.

LITERATURE CITED

- Al-Shehbaz, I.A. 1988. The genera of Arabideae (Cruciferae; Brassicaceae) in the southeastern United States. *Journal of the Arnold Arboretum* 69:85-166.
- Beers, T.W., P.E. Dress, and L.C. Wensel. 1966. Aspect transformation in site productivity research. *Journal of Forestry* 64:691-692.
- Beven, K.J., and M.J. Kirkby. 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Science Bulletin* 24:43-69.
- Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. *Classification and Regression Trees*. Wadsworth International Group, Belmont, Calif. 358 pp.
- Clark, J.D., J.E. Dunn, and K.G. Smith. 1993a. A multivariate model of female black bear habitat use for a geographic information system. *Journal of Wildlife Management* 57:519-526.
- Clark, J.D., K.G. Smith, and J.E. Dunn. 1993b. Modelling wildlife habitat requirements from environmental and radiotelemetry data using GIS. *Annual Symposium on Geographic Information Systems in Forestry, Environment and Natural Resources Management* 7:523-528.
- Corsi, F., E. Dupre, and L. Boitani. 1999. A large-scale model of wolf distribution in Italy for conservation planning. *Conservation Biology* 13:150-159.
- Dudley, T.R. 1974. The correct authority for *Cardamine clematitis* (Cruciferae). *Rhodora* 76:53-57.
- Guerro, V.M., and R.A. Johnson. 1982. Use of the Box-Cox transformation with binary response models. *Biometrika* 69:309-314.
- Halpin, P.N. 1995. A cross-scale analysis of environmental gradients and forest patterns in the giant sequoia-mixed conifer forest of the Sierra Nevada. Ph.D. diss., University of Virginia, Charlottesville. 277 pp.
- Hanski, I. 1998. Metapopulation dynamics. *Nature* 396:41-49.
- Hanski, I., and D. Simberloff. 1997. The metapopulation approach, its history, conceptual

- domain, and application to conservation. Pp. 5-26 in I.A. Hanski and M.E. Gilpin, eds., *Metapopulation Biology*. Academic Press, San Diego, Calif.
- Hosmer, D.W., Jr., and S. Lemeshow. 1989. *Applied Logistic Regression*. John Wiley and Sons, New York. 307 pp.
- Kearney, T.H., Jr. 1892. Notes on the flora of Thunderhead. *Agricultural Science* 6:71-74.
- King, P.B., and A. Stupka. 1950. The Great Smoky Mountains—their geology and natural history. *Scientific Monthly* 71:31-43.
- Knick, S.T., and J.T. Rotenberry. 1998. Limitations to mapping habitat use areas in changing landscapes using the Mahalanobis distance statistic. *Journal of Agricultural, Biological, and Environmental Statistics* 3:311-322.
- Lachenbruch, P.A. 1975. *Discriminant Analysis*. Hafner Press, New York. 128 pp.
- MacKenzie, M.D. 1993. The vegetation of Great Smoky Mountains National Park: past, present and future. Ph.D. diss., University of Tennessee, Knoxville. 154 pp.
- McNab, W.H. 1989. Terrain shape index: quantifying effect of minor landforms on tree height. *Forest Science* 35:91-104.
- McNab, W.H. 1993. A topographic index to quantify the effect of mesoscale landform on site productivity. *Canadian Journal of Forest Research* 23:1100-1107.
- Miller, R. 1986. Predicting rare plant distribution patterns in the southern Appalachians of the southeastern USA. *Journal of Biogeography* 13:293-311.
- 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* 15:59-71.
- Parker, A.J. 1982. The topographic relative moisture index: an approach to soil-moisture assessment in mountain terrain. *Physical Geography* 3:160-168.
- Pyle, C. 1985. Vegetation disturbance history of Great Smoky Mountains National Park: an analysis of archival maps and records. U.S. Department of the Interior, National Park Service, Research/Resources Management Report SER-77, Gatlinburg, Tenn. 69 pp.
- Pyle, C. 1988. The type and extent of anthropogenic vegetation disturbance in the Great Smoky Mountains before National Park Service acquisition. *Castanea* 53:225-235.
- Rao, C.R. 1952. *Statistical Methods in Biometric Research*. Wiley and Sons, New York. 390 pp.
- Scott, J.M., P.J. Heglund, J.B. Haufler, M.L. Morrison, M.G. Raphael, W.B. Wall, and F. Samson (eds.). 2001. *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Covello, Calif. 868 pp.
- ter Braak, C.J.F., and C.W.N. Looman. 1995. Regression. Pp. 29-77 in R.H.G. Jongman, C.J.F. ter Braak, and O.F.R. van Tongeren, eds., *Data Analysis in Community and Landscape Ecology*. Cambridge University Press, Cambridge, U.K.
- Thornthwaite, C.W. 1948. An approach toward a rational classification of climate. *Geological Review* 38:55-94.
- Verner, J., M.L. Morrison, and C.J. Ralph (eds.). 1986. *Modeling Habitat Relationships of Terrestrial Vertebrates*. University of Wisconsin Press, Madison. 470 pp.
- Weakley, A.S. 1996. *Flora of the Carolinas and Virginia*, unpublished draft. The Nature Conservancy, Southeast Regional Offices, Chapel Hill, N.C.
- White, P.S. 1982. The flora of Great Smoky Mountains National Park: an annotated checklist of the vascular plants and a review of previous floristic work. Research/Resources Management Report SER-55, U.S. Department of the Interior, National Park Service, Gatlinburg, Tenn. 219 pp.
- Wilds, S.P. 1996. Gradient analysis of the distribution of flowering dogwood (*Cornus florida* L.) and dogwood anthracnose (*Discula destructiva* Redlin.) in western Great Smoky Mountains National Park. M.S. thesis, University of North Carolina, Chapel Hill. 151 pp.
- Wofford, B.E. 1989. *Guide to the Vascular Plants of the Blue Ridge*. University of Georgia Press, Athens. 384 pp.
- Wolock, D.M., and G.J. McCabe Jr. 1995. Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. *Water Resources Research* 31:1315-1324.