RESEARCH ARTICLE

•

Habitat Models to Assist Plant Protection Efforts in Shenandoah National Park, Virginia, USA

Frank T. van Manen[1]

U.S. Geological Survey, Southern Appalachian Field Branch, University of Tennessee, 274 Ellington Hall, Knoxville, TN 37996, USA

John A. Young

U.S. Geological Survey, Aquatic Ecology Branch, Leetown Science Center, 11649 Leetown Road, Kearneysville, WV 25430, USA

Cindy A. Thatcher

University of Tennessee, Department of Forestry, Wildlife and Fisheries, 274 Ellington Hall, Knoxville; TN 37996, USA

Wendy B. Cass

National Park Service, Shenandoah National Park
3655 U.S. Highway 211 East
Luray, VA 22835, USA

Chris Ulrey

National Park Service, Blue Ridge Parkway, 199 Hemphill Knob Road, Asheville, NC 28803, USA

•

Natural Areas Journal 25:339–350

ABSTRACT: During 2002, the National Park Service initiated a demonstration project to develop science-based law enforcement strategies for the protection of at-risk natural resources, including American ginseng (*Panax quinquefolius* L.), bloodroot (*Sanguinaria canadensis* L.), and black cohosh (*Cimicifuga racemosa* (L.) Nutt. [syn. *Actaea racemosa* L.]). Harvest pressure on these species is increasing because of the growing herbal remedy market. We developed habitat models for Shenandoah National Park and the northern portion of the Blue Ridge Parkway to determine the distribution of favorable habitats of these three plant species and to demonstrate the use of that information to support plant protection activities. We compiled locations for the three plant species to delineate favorable habitats with a geographic information system (GIS). We mapped potential habitat quality for each species by calculating a multivariate statistic, Mahalanobis distance, based on GIS layers that characterized the topography, land cover, and geology of the plant locations (10-m resolution). We tested model performance with an independent dataset of plant locations, which indicated a significant relationship between Mahalanobis distance values and species occurrence. We also generated null models by examining the distribution of the Mahalanobis distance values had plants been distributed randomly. For all species, the habitat models performed markedly better than their respective null models. We used our models to direct field searches to the most favorable habitats, resulting in a sizeable number of new plant locations (82 ginseng, 73 bloodroot, and 139 black cohosh locations). The odds of finding new plant locations based on the habitat models were 4.5 (black cohosh) to 12.3 (American ginseng) times greater than random searches; thus, the habitat models can be used to improve the efficiency of plant protection efforts, (e.g., marking of plants, law enforcement activities). The field searches also indicated that the level of occupancy of the most favorable habitats ranged from 49.4% for ginseng to 84.8% for black cohosh. Given the potential threats to these species from illegal harvesting, that information may serve as an important benchmark for future habitat and population assessments.

Index terms: habitat analysis, illegal harvest, Mahalanobis distance, medicinal herbs, plant protection

INTRODUCTION

All native plant species in United States national parks are protected by law, but some species are illegally harvested because they are valued for their medicinal or ornamental qualities. Numerous plant species are considered at risk of illegal harvesting, but three species have received particular attention in recent years: American ginseng (*Panax quinquefolius* L.), bloodroot (*Sanguinaria canadensis* L.), and black cohosh (*Cimicifuga racemosa* (L.) Nutt. [syn. *Actaea racemosa* L.]). The growing popularity of herbal remedies has commanded high prices, particularly for American ginseng, resulting in increased harvest pressure on these species (Robbins 2000, McGraw et al. 2003). The primary markets for these plant products are in North America and East Asia (Robbins 2000; J. Chamberlain, U.S. Forest Service, pers. comm.). Large, gnarled roots of wild-harvested American ginseng are particularly valued in Asia. American ginseng is listed in Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), and global trade of plant roots and rhizomes is strictly regulated to avoid use incompatible with survival of the species (CITES 1973). Black cohosh and bloodroot currently are under consideration for inclusion in CITES Appendix II by the Division of Scientific Authority of the U.S. Fish and Wildlife Service (USFWS). International trade of American ginseng is monitored by the USFWS under Appendix II of the CITES treaty in order to identify detrimental effects of harvest on wild ginseng populations.

There is more habitat information available for American ginseng because of its economic importance and long history of presumed medicinal value. However, black cohosh and bloodroot are considered good indicators of American ginseng habitat (Kauffman 2002). All three plants are herbaceous perennials found in rich woods of the eastern deciduous forest of North America, with black cohosh being the most common. Although little is known of the soil requirements of ginseng, it is reported to grow best in well-drained soils with a rocky substrate of moderate pH (5-6) with adequate calcium and a humus layer (Das et al. 2001). Overstory canopy cover also seems to be an important habitat requirement; 70% canopy cover has been reported as an important feature of optimal habitat (Das et al. 2001). Besides bloodroot and black cohosh, herbaceous plants thought

---

[1] Corresponding author: vanmanen@utk.edu; 865-974-0200

to be associated with American ginseng in our study area include northern maidenhair (*Adiantum pedatum* L.), rattlesnake fern (*Botrychium virginianum* [L.] Sw.), blue cohosh (*Caulophyllum thalictroides* [L]. Michx.), false solomon's seal (*Smilacina racemosa* L.), jack in the pulpit (*Arisaema triphyllum* [L.] Schott), and mayapple (*Podophyllum peltatum* L.; Anderson et al. 1993; Michigan Natural Features Inventory 1996; J. Chamberlain, U.S. Forest Service, pers. comm.).

Although harvesting of the three plants is permitted on other federal lands (e.g., national forests) and private properties, illegal harvest, particularly in the case of American ginseng, has become a primary concern in several eastern national parks (e.g., Shenandoah National Park and Great Smoky Mountains National Park; National Park Service, unpubl. data). Such poaching activities may be an indication that national park areas are becoming a refuge for the last significant populations of American ginseng (Gagnon 1999). Enforcement efforts have not been adequate to protect these plants from illegal harvest (K. Johnson, National Park Service, pers. comm.). Moreover, little is known about the distribution and population abundance of wild populations of these plants (Robbins 2000). Most information exists for American ginseng, and recent studies suggest that harvesting is a substantial threat to this species (McGraw 2001, McGraw et al. 2003). Browsing by abundant white-tailed deer (*Odocoileus virginianus*) populations is an additional source of concern (McGraw and Furedi 2005). Therefore, resource managers with the National Park Service sought tools to delineate favorable habitat areas, which could then be used to assess potential threats due to illegal harvesting and to focus law enforcement activities to better protect plant resources. The objectives of our study were to: (1) develop and test predictive habitat models for the three plant species and (2) determine the relative frequency with which these plants currently exist in their optimal habitats.

STUDY AREA

Because of available data and a documented history of plant poaching, we chose Shenandoah National Park and the northern portion of the Blue Ridge Parkway as the focal area for our study (Figure 1). The study area is in the Blue Ridge Mountain section of the southern Appalachian Mountains (Bailey 1980), which extend in a northeast-southwest direction. The study area is almost entirely forested, with the exception of high-elevation balds and rocky

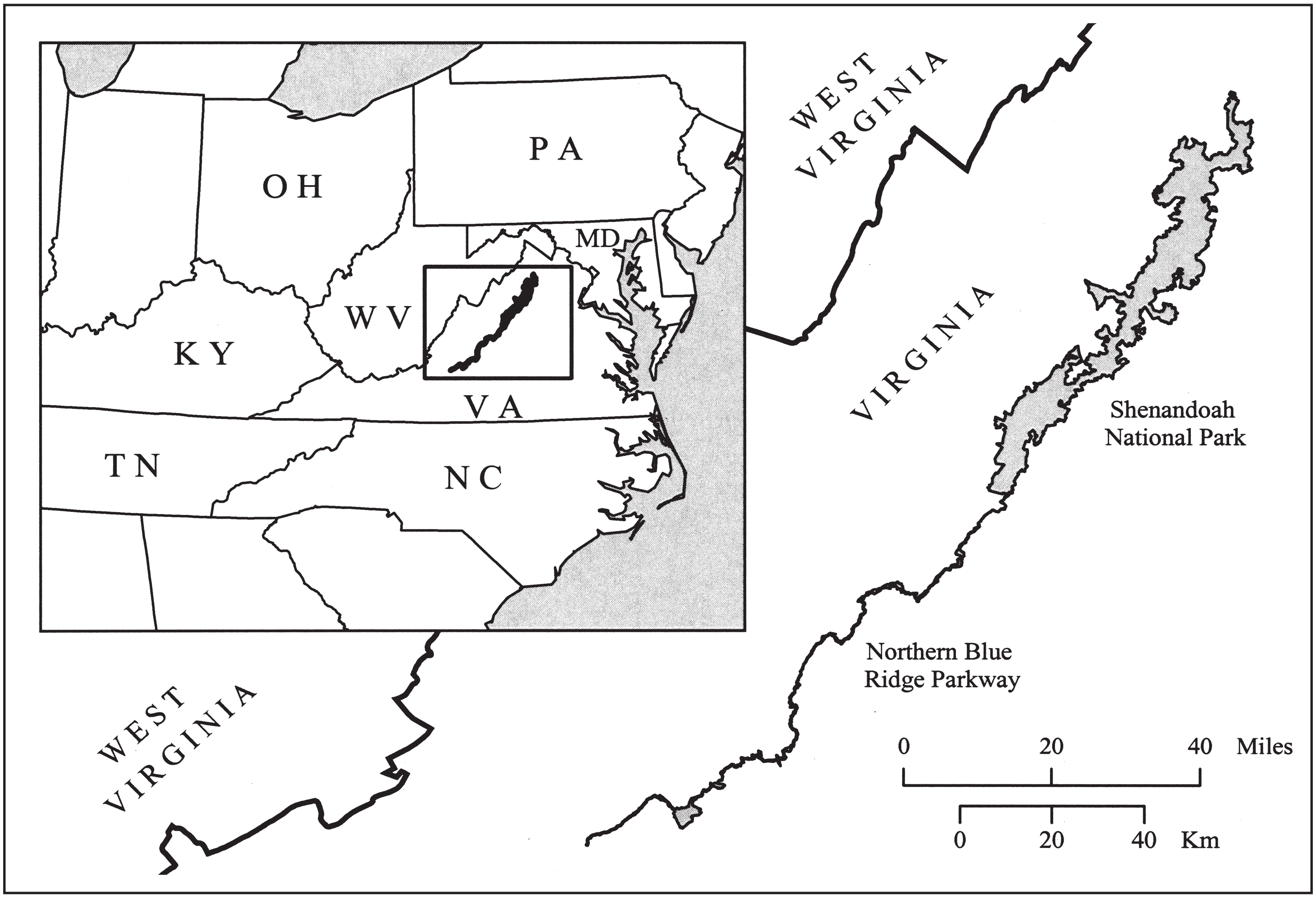

Figure 1. Study area to determine favorable habitat areas for American ginseng, bloodroot, and black cohosh in Shenandoah National Park and northern Blue Ridge Parkway, Virginia, USA.

outcrops and some areas maintained as cultural heritage sites. Deciduous forests dominate the study area, except for minor areas in eastern hemlock and pine and small stands of spruce and fir at the highest elevations. Recreational use of the study area was high in 2002, with 1.1 million visits to Shenandoah National Park and almost 20 million visits to the Blue Ridge Parkway (National Park Service 2003).

MODELING SPECIES DISTRIBUTIONS

Guisan and Zimmerman (2000) reviewed influences on plant growth and distribution and divided these into three gradients: (1) “resource”, (2) “direct”, and (3) “indirect.” Resource gradients are taken up directly by plants (e.g., water, photosynthetic active radiation, soil nutrients). Direct gradients influence the availability of resource gradients, such as the type of soil, or light, water, and temperature regimes. Indirect gradients represent relatively large-scale influences, such as geology, topography, climate, and position on earth (e.g., latitude), which create the conditions governing the formation of direct gradients (e.g., amount of sunlight, water availability, weathering of geologic substrates to produce soil) and, ultimately, the provision of resources to plants. Because of limitations due to data availability and current mapping technology, only indirect or direct gradients typically can be assessed over large areas with a geographic information system (GIS). For example, although few data exist to assess photosynthetic active radiation levels over broad areas, existing digital terrain models can be used to derive surrogates of solar radiation potential.

In eastern hardwood forests, the distribution and productivity of many plant species tend to be associated with local topography and soil type (Iverson et al. 1997). In the southern Appalachian Mountains, the complex topography, variable geology, and high rainfall result in some of the highest levels of forest ecosystem diversity in North America (Odom and McNab 2000). The distribution of plant communities in the study area is correlated with environmental gradients, which, in turn, are mainly a function of topographic variation (Odom and McNab 2000). Therefore, landform characteristics provide indirect measures of the environmental gradients that affect the distribution of plant species.

In the highly variable topography of the Appalachian Mountains, landform characteristics vary widely at local scales and have a strong influence on site productivity (Iverson et al. 1997). Terrain affects microclimate characteristics, such as moisture availability, solar radiation, wind exposure, and temperature. Local-scale variations in terrain also affect soil erosion and deposition. Consequently, plant species in the southern Appalachians often respond to terrain patterns that can be effectively measured with high-resolution (0.01 ha) GIS layers. For example, American ginseng tends to occur within cove areas, where cooler and moister conditions typically prevail (Anderson et al. 1993). It is that response of plant species to the distinct variations in local-scale topography that permits the development of statistical models to predict their occurrence.

METHODS

Plant Location Data

We obtained location databases for the three plant species from National Park Service botanists and compiled the data by species. We only included locations that were recorded or verified with a global positioning system (GPS) receiver. Before the Department of Defense turned off the selective availability of the GPS satellite signals, GPS locations were collected with an unassisted, military Y-code signal from the U.S. Department of Defense (PLGR +96, Rockwell International, Cedar Rapids, Iowa; average horizontal displacement error <12 m) GPS receivers. After the discontinuation of selective availability, locations were obtained with standard GPS receivers (GPS3, GPS3+, or GPS5; Garmin, Olathe, Kansas), resulting in average positional errors approximately 5-10 m. For each species, we created a dataset with geographic coordinates of the plant locations, which served as the point of reference for our habitat analyses.

Habitat Variables

We developed a database of GIS variables to identify landscape conditions in the study area that were similar to those associated with the sample locations of the three plant species. We used a 10-meter resolution digital elevation model from the National Elevation Dataset (Gesch et al. 2002) to produce new terrain variables. The variables we calculated included the Beers’ transformation of aspect (Beers et al. 1966), slope, relative solar insolation (i.e., hillshading), terrain shape index (McNab 1989), topographic convergence index (Wolock and McCabe 1995), and topographic relative moisture index (Parker 1982; Table 1). All terrain variables were calculated in ArcGIS® 8.2 (ESRI, Redlands, California).

In addition to the topographic variables, we used National Land Cover Data (Vogelmann et al. 2001), geology (Gathwright 1976), and ecoregion data (U.S. Environmental Protection Agency 2002) as categorical variables in the habitat models. We resampled the National Land Cover Data from a resolution of 30 m to 10 m to match the resolution of the topographic variables. Available geology data for the Blue Ridge Parkway were not compatible in scale and classification system with the data available for Shenandoah National Park. For the northern portion of the Blue Ridge parkway, we used ecoregion as a surrogate for geology by matching it with the geology layer of Shenandoah National Park. We also considered the use of soil type as a potentially important habitat variable. However, consistent GIS data of soil type for the study area were limited to State Soil Geographic (STATSGO) data (U.S. Department of Agriculture 1994), and the mapping scale of those data (1:250,000) was too coarse for use in our models. All categorical variables were transformed into binary design variables (“dummy” variables).

Predicting Plant Occurrence

We used Mahalanobis distance (D^2) as the primary method to predict species occurrence (Clark et al. 1993). D^2 is a multi-

Table 1. Variables considered for inclusion in Mahalanobis distance models to predict occurrence of American ginseng, bloodroot, and black cohosh, Shenandoah National Park and northern Blue Ridge Parkway, Virginia, USA, 2002-2003.

Variable	Description	Value range	Source
Aspect	Beers' transformation of aspect: 1 + cos(45 - aspect)[a]	0–2.0	Calculated from aspect based on Beers et al. (1966)
Geology	Geology type		National Park Service data, based on U.S. Geological Survey geologic maps
Land-cover type	Land-cover type		U.S. Geological Survey 30-m resolution National Land Cover data
Ecoregion	Polygons containing areas of general similarity in ecosystems and environmental resources		U.S. Environmental Protection Agency Level IV ecoregion data
Elevation	Elevation (m)	211–1,236	U.S. Geological Survey 10-m resolution digital elevation model
Relative slope position	Indicates where on a slope a grid cell is located	0–100	Calculated from elevation based on Wilds (1997)
Slope	Slope steepness (degrees)	0–63	Calculated from elevation with the SLOPE command (ArcGIS® 8.2)
Solar insolation	Index of exposure to sunlight; approximated for the solar equinox	1–227	Calculated from elevation with the HILLSHADE command (ArcGIS® 8.2)
Terrain shape index	Measure of local topographic variability as a continuous variable indicating convex (<-0.05) or concave (>0.05) landforms	-121–86	Calculated based on McNab (1989)
Topographic convergence index	Simulates the flow accumulation of water; TCI = ln(A/tan B), where A is drained surface area and B is drained surface slope	-5.3–12.2	Calculated based on Wolock and McCabe (1995)
Topographic relative moisture index	Index of moisture considering the effects of slope position, aspect, and elevation	0–59	Calculated based on Parker (1982)

[a] For this transformation, we represented a gradient of aspect conditions by using northeastern aspects (45 degrees) to represent optimal conditions (relatively cool and moist slopes with northeastern aspects), receiving the maximum transformed value of 2.0. Transformed aspect values diminished in either direction of the aspect gradient towards a value of 0 for southwestern aspects (relatively hot and dry slopes).

variate statistic that represents a measure of dissimilarity (Rao 1952). This method provides an effective and proven approach to predict species occurrence based on habitat data by combining multivariate statistics with GIS technology. For example, Boetsch et al. (2003) used Mahalanobis distance to predict the occurrence of mountain bittercress (*Cardamine clematitis* Shuttlw. ex. Gray) in Great Smoky Mountains National Park, and van Manen et al. (2002) used this method to delineate habitats of surviving butternut trees (*Juglans cinerea* L.) in the Blue Ridge Mountains. We calculated this statistic for each GIS grid cell in the study area by combining the information from all habitat layers using the following equation:

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

where $\underline{x}$ is a vector of habitat characteristics based on the GIS grids, $\underline{\hat{u}}$ is the mean vector of habitat characteristics of known plant locations ("training locations"), and Σ^{-1} is the inverse of the variance-covariance matrix calculated from these training locations. The D^2 statistic represents the standard squared distance between a set of sample variates, $\underline{x}$, and 'ideal' habitat characterized by the plant locations and represented by $\underline{\hat{u}}$. The D^2 statistic provides a dimensionless index of similarity to the multivariate habitat conditions associated with the locations of the target species (Knick and Rotenberry 1998). Small values of D^2 represent habitat conditions similar to those of known locations, whereas larger distance values represent increasingly different conditions. A variety of habitat combinations can result in identical distance values. Mahalanobis distance is the sum of squares of standardized scores; therefore, correlations among variables are accounted for. However, to reduce duplication of habitat variables, we excluded those variables whose correlation coefficients were >0.65. Mahalanobis distance is appropriate for predicting rare plant occurrence because it requires only presence locations for input, rather than both presence and absence data, thus avoiding the potential difficulties involved in classifying available habitats as unused (Clark et al. 1993) as is required for other techniques (e.g., logistic regression). For example, because of poaching or deer browsing, the plant species we studied may be absent at a particular location despite ideal habitat conditions.

We calculated $\underline{\hat{u}}$ and Σ^{-1} with SAS® software (Proc DISCRIM; SAS® Institute 2000) based on the habitat characteristics of the training locations. We used this information and the Mahalanobis distance equation to calculate D^2 in ArcGIS® based on the habitat conditions of each grid cell in the study area. Habitat modeling is an iterative process (Stormer and Johnson 1986). We improved model reliability by including or excluding variables based on data quality and biological criteria (e.g., geology) and by using stricter criteria for the precision of GPS data. For each plant species, we tested each model iteration using methods we describe subsequently and chose the model with the best test results.

Model Testing

Cumulative Frequency Graphs

We developed the plant habitat models by randomly selecting 75% of the locations for each species (training set). We then compared the D^2 values of each model with those of the remaining 25% of plant locations (validation set). We also generated 150 random locations within the study area to serve as a null model reference (i.e., the distribution of D^2 values if plants were randomly distributed).

We compared cumulative frequency graphs of D^2 values at locations from the training set, the validation set, and the null model set. We first examined whether the cumulative frequency curves of the training and validation locations were different; similar distributions would indicate model consistency. Secondly, we assessed model effectiveness by selecting the D^2 value (cutoff value) that maximized the difference between the cumulative frequency graphs of the validation and random locations. We chose that D^2 value as a threshold because it included the greatest percentage of plant locations within the smallest percentage of the study area, and, as such, provided a meaningful measure to delineate favorable habitats. Any grid cells with D^2 values below that value may be considered more favorable habitat, whereas grid cells above that value may be less favorable (Pereira and Itami 1991). If little or no differences exist between the cumulative frequency graphs of the validation and null model locations, the model would be no better than one based on random chance.

Independent Field Test

We used data from 148 field plots in Shenandoah National Park (J. Young and D. Walton, U.S. Geological Survey, unpubl. data) collected during 2001-2002 and 103 plots previously sampled by the Virginia Natural Heritage Program (G. Fleming, Virginia Natural Heritage Program, unpubl. data) for our independent field tests. Those data (total $n = 251$) were collected using the relevé method (sensu Peet et. al 1998) for a separate study of vegetation distribution, and vegetation plots were located based on a stratified random design that attempted to capture the range of potential habitats in Shenandoah National Park based on a classification of ecological land units. Because plots were distributed throughout Shenandoah National Park, we were able to test the full range of Mahalanobis distance predictions for the plant models. We used logistic regression (Hosmer and Lemeshow 1989) with SAS® statistical software (Proc LOGISTIC; SAS® Institute 2000) to determine whether the presence or absence of each of the three plant species in those 251 plots (dependent binomial variable) was associated with the D^2 values (independent variable) of the corresponding grid cells. We determined the fit of the logistic regression models with the Hosmer-Lemeshow goodness-of-fit statistic (Hosmer and Lemeshow 1989).

Model Application

Targeted Field Searches

We used the predictions of the plant habitat models to direct National Park Service personnel to locations where the three plant species were most likely to occur. This application of the model accommodated National Park Service goals of establishing covert plots and marking American ginseng plants (e.g., fluorescent dye, silicon granules). Secondly, it provided data to test the efficacy of the models in locating new plant locations. We used stratified random sampling to maximize the potential number of new plant locations. We proportionally located our test plots in areas where the species were more likely to occur, according to model predictions, using the following geometric equation (van Manen et al. 2002):

$$100 \text{ (percent total grid cells)}$$
$$= n + 2n + 4n + 8n + 16n + 32n.$$

This equation doubles the amount of area in each of the six successive classes (i.e., $n = 1.59$). Class 1 contained 1.59% of the grid cells in the study area that were associated with the lowest range of D^2 values, class 2 represented 3.17% of the study area associated with the next lowest range of D^2 values, and so on. For these field searches, we only sampled the best habitat areas, as represented by the first two stratifications. Furthermore, because the habitat requirements for the three species are similar, we only generated random locations within areas where model predictions for all three species were in the first or second stratification, thereby increasing the overall efficiency of the search efforts. We restricted the location of test plots to areas 100-500 m from trails or roads to facilitate sampling. The proportion of habitat areas in the first two stratifications was similar between those areas (6.0%) and the entire study area (4.4%), so we assumed that restricting the sampling area would not introduce a bias. The 100-m minimum distance was incorporated to reduce possible bias due to disturbance effects associated with trails and roads (e.g., light penetration).

National Park Service personnel searched 227 sites for the three species during the summers of 2002 (n = 142) and 2003 (n = 85). The field searches during 2002 were based on the initial model iterations; the 2003 field searches were based on the final habitat models. All field searches followed standardized sampling procedures (W. Cass, J. Rock, and C. Ulrey, National Park Service, unpubl. report). Sampling positions were established in the field with standard GPS receivers or WAAS-enabled (Wide Area Augmentation System; average horizontal displacement error <3 m) GPS receivers (Garmin, Olathe, Kansas). Once the sample position was established, an area representing a 30-m x 30-m GIS grid cell centered on the sampling position was surveyed and the presence or absence of each of the three plant species was recorded, along with information to characterize the site. Although we used 10-m x 10-m GIS grid cells for our analyses, we used a 30-m x 30-m search area to account for the potential effects of GPS and GIS errors. To assess the effectiveness of the habitat models to find new locations of each species, we used chi-square tests to compare the frequency of occurrence of the targeted field searches with the frequencies observed for the 251 independent field test plots. We also calculated odds ratios to determine the degree of effectiveness.

RESULTS

Mahalanobis Distance Model

GPS coordinates were available for 137 ginseng sites, 37 bloodroot sites, and 66 black cohosh sites. The training dataset represented 75% of these locations (n = 103, 28, and 50, respectively). The initial models we developed included seven topographic variables and land-cover type. For our final models, we included the same variables but added geology/ecoregion and an index to account for potential flow accumulation effects (topographic convergence index). Finally, we excluded relative slope position because it was highly correlated with the topographic relative moisture index (r = -0.74) and because it showed inconsistencies in its spatial pattern. The final models included land-cover type, geology, elevation, slope, Beers' transformation of aspect, relative solar insolation, terrain shape index, topographic convergence index, and topographic relative moisture index.

Mean D^2 values in the study area ranged from 22.3 (SD = 16.0, range = 1.6-426.4) for American ginseng, 28.0 (SD = 17.4, range = 1.6-663.5) for bloodroot, and 26.8 (SD = 18.7, range = 1.7-663.5) for black cohosh. Based on the training set locations, mean D^2 values were 10.1 (SD = 5.2, range = 1.8-27.0) for American ginseng, 9.7 (SD = 6.0, range = 2.9-30.3) for bloodroot, and 10.7 (SD = 5.2, range = 2.0-31.5) for black cohosh.

Model Testing

Cumulative Frequency Graphs

The cutoff value to define favorable habitats for American ginseng was based on $D^2 \leq 12.7$ (Figure 2). Using that cutoff value, we correctly classified 65% of the validation locations while restricting predictions of favorable habitat to 29% of the study area (null model; Figure 2). For bloodroot, a cutoff value of $D^2 \leq 13.1$ correctly classified 44% of the validation data; 19% of the study area was predicted to contain favorable habitat (Figure 3). Finally, for black cohosh, a cutoff value of $D^2 \leq 13.2$ correctly classified 70% of the validation locations, representing 28% of the study area (Figure 4).

Independent Field Test

The logistic regression analysis based on species presence or absence at the 251 vegetation plots indicated significant negative associations between D^2 values and occurrence of American ginseng, bloodroot, and black cohosh (Table 2). Thus, lower D^2 values tended to be associated with presence of the species in the field test plots, whereas greater D^2 values were associated with species absence. Model fit was good for the ginseng and bloodroot analyses (Hosmer and Lemeshow goodness-of-fit statistic = 7.88, P = 0.445 and 6.51, P = 0.590, respectively), but marginal for the black cohosh model (Hosmer and Lemeshow goodness-of-fit statistic = 15.37, P = 0.052); the poor model fit for black cohosh seemed to be related to poor predictability of observations in a small range of sub-optimal habitats.

Model Application

Targeted Field Searches

During 2002, field searches at 142 sites in Shenandoah National Park and northern Blue Ridge Parkway were based on the results from the initial model iterations. Those searches resulted in 40 new ginseng locations at 28.2% of the sites, 29 bloodroot locations (20.4%), and 67 black cohosh locations (47.2%). During 2003, field searches were conducted at 85 sites identified based on the final plant habitat models: 42 of those sites contained American ginseng (49.4%), 44 sites contained bloodroot (51.8%), and 72 sites contained black cohosh (84.8%). The 251 vegetation plots independently sampled for the ecological land unit classification resulted in 11 occurrences of American ginseng (4.4%), 19 occurrences of bloodroot (7.6%), and 65 occurrences of black

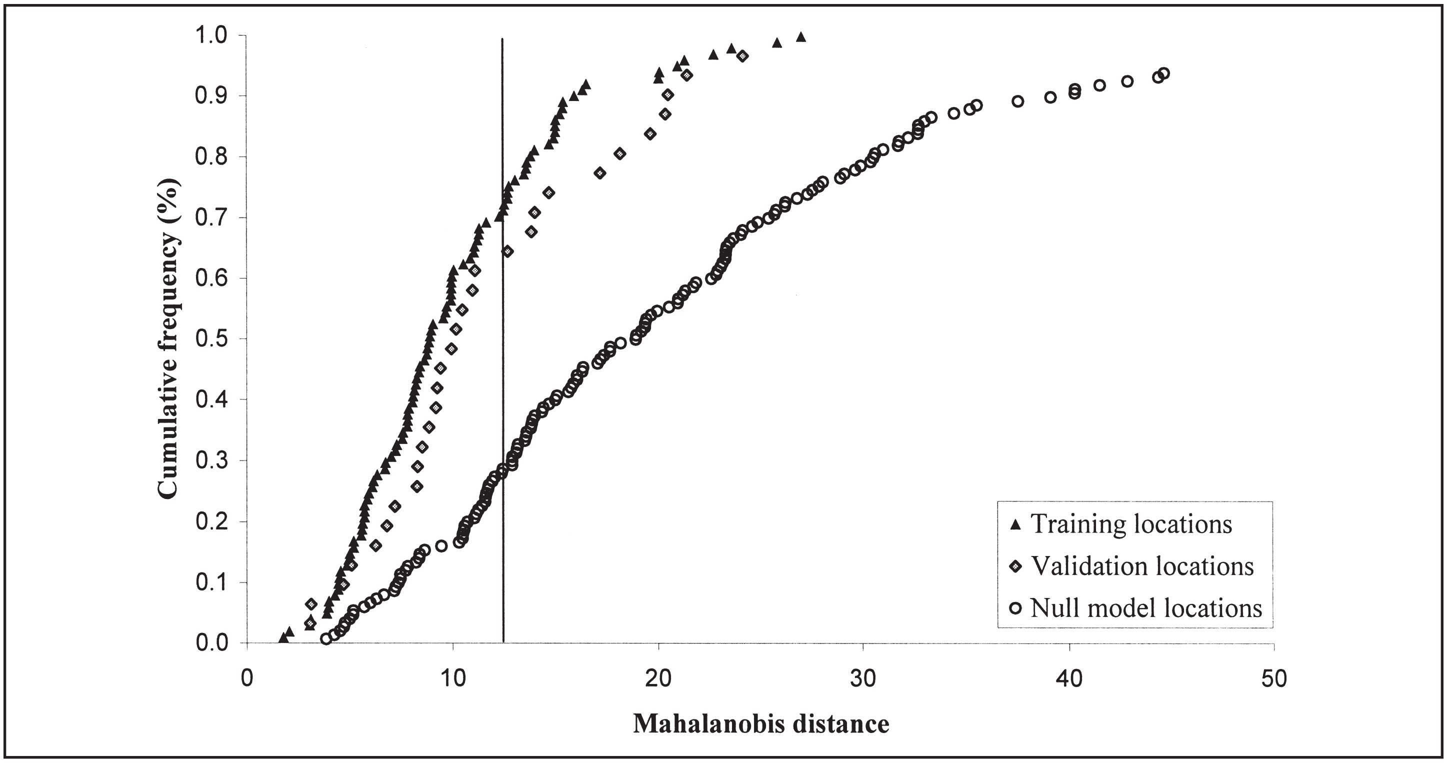

Figure 2. Cumulative frequency distribution of Mahalanobis distance values for American ginseng (training and validation locations) and null model (random) locations, Shenandoah National Park and northern Blue Ridge Parkway, Virginia, USA, 2002-2003. Vertical line indicates cut-off value of Mahalanobis distance to define favorable habitat areas.

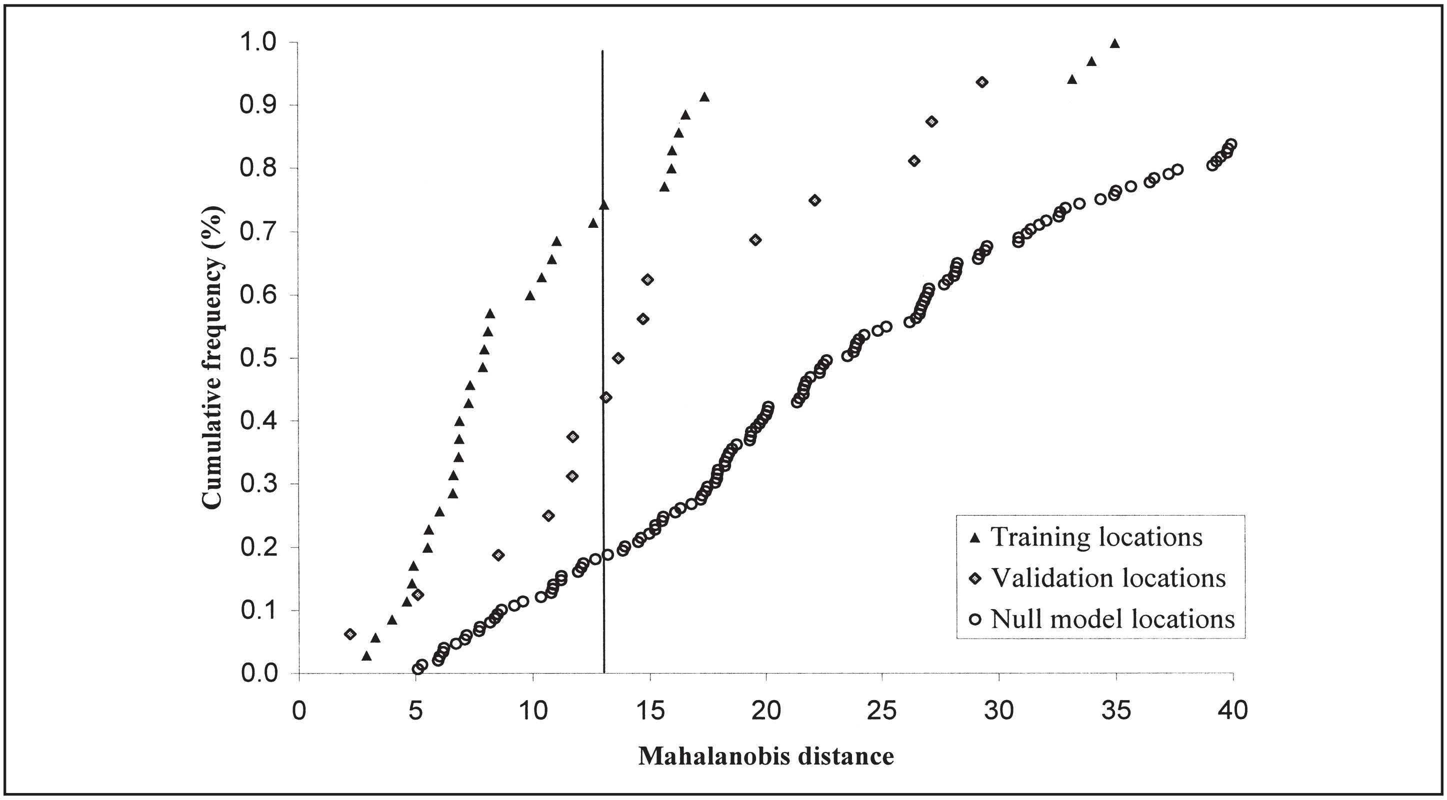

Figure 3. Cumulative frequency distribution of Mahalanobis distance values for bloodroot (training and validation locations) and null model (random) locations, Shenandoah National Park and northern Blue Ridge Parkway, Virginia, USA, 2002-2003. Vertical line indicates cut-off value of Mahalanobis distance to define favorable habitat areas.

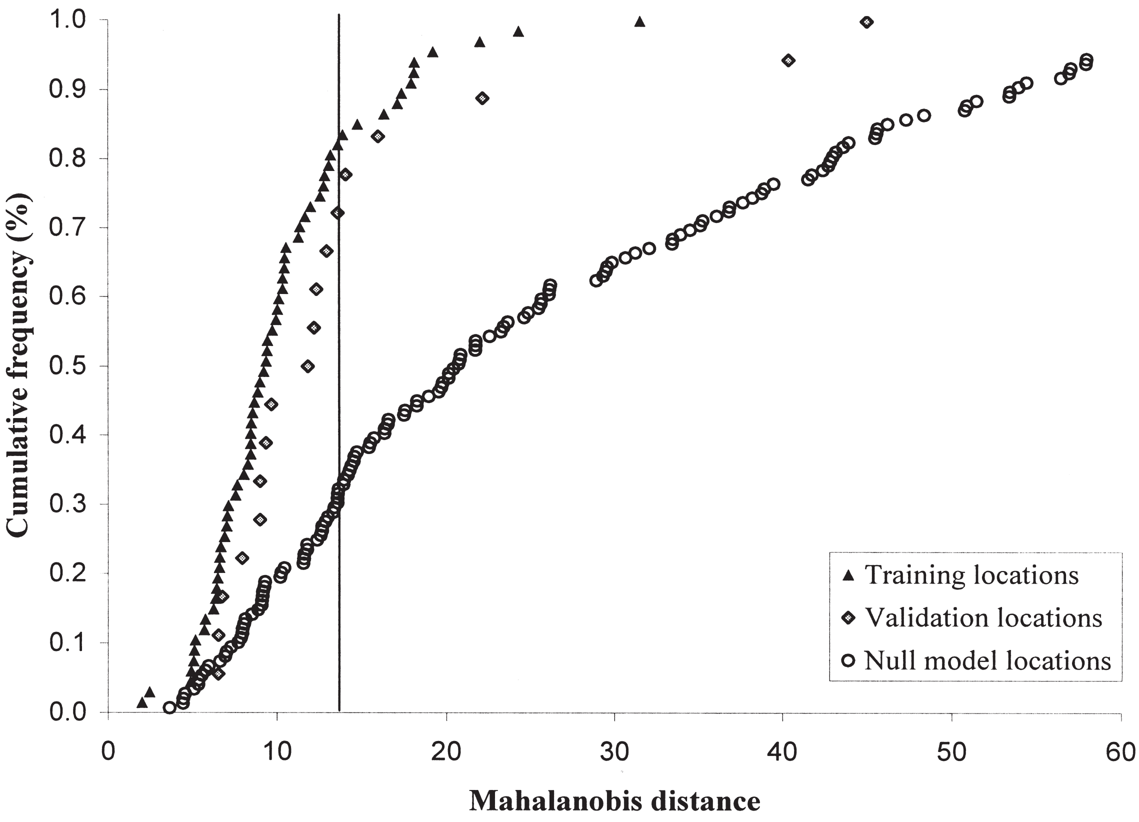

Figure 4. Cumulative frequency distribution of Mahalanobis distance values for black cohosh (training and validation locations) and null model (random) locations, Shenandoah National Park and northern Blue Ridge Parkway, Virginia, USA, 2002-2003. Vertical line indicates cut-off value of Mahalanobis distance to define favorable habitat areas.

cohosh (25.9%; Figure 5). Comparisons of the frequencies of species occurrences between the targeted field searches and the 251 independent vegetation plots indicated that the habitat model was highly effective at finding new plant locations (American ginseng: $\chi^2 = 76.6$, 1 df, $P < 0.001$; bloodroot: $\chi^2 = 46.4$, 1 df, $P < 0.001$; black cohosh: $\chi^2 = 60.8$, 1 df, $P < 0.001$). The odds ratios indicated that the odds of finding the species based on model predictions rather than random searches was 12.3 times greater for ginseng (95% CI = 6.4-23.9), 5.8 greater for bloodroot (95% CI = 3.4-10.0), and 4.5 greater for black cohosh (95% CI = 3.1-6.7).

DISCUSSION

Plant Habitat Models

Our analyses indicate that the Mahalanobis distance statistic provided an effective method to identify potential habitat for the three plant species across a large area and to facilitate the search for new locations. The cumulative frequency graphs indicated that Mahalanobis distance values were greater for the null model locations than those associated with occurrences of the three plant species (Figures 2-4). We tested for model consistency by withholding 25% of the locations for use as a validation dataset. The cumulative frequency graphs for American ginseng and black cohosh

Table 2. Logistic regression results to test the relationship between Mahalanobis distance and occurrence of three plant species in Shenandoah National Park and northern Blue Ridge Parkway, Virginia, USA, 2003.

Species	No. presence locations	No. absence locations	Parameter estimate	*P*-value	Max. rescaled R^2
American ginseng	11	240	-0.1785	0.002	0.24
Bloodroot	19	232	-0.0497	0.014	0.087
Black cohosh	65	186	-0.0623	<0.001	0.235

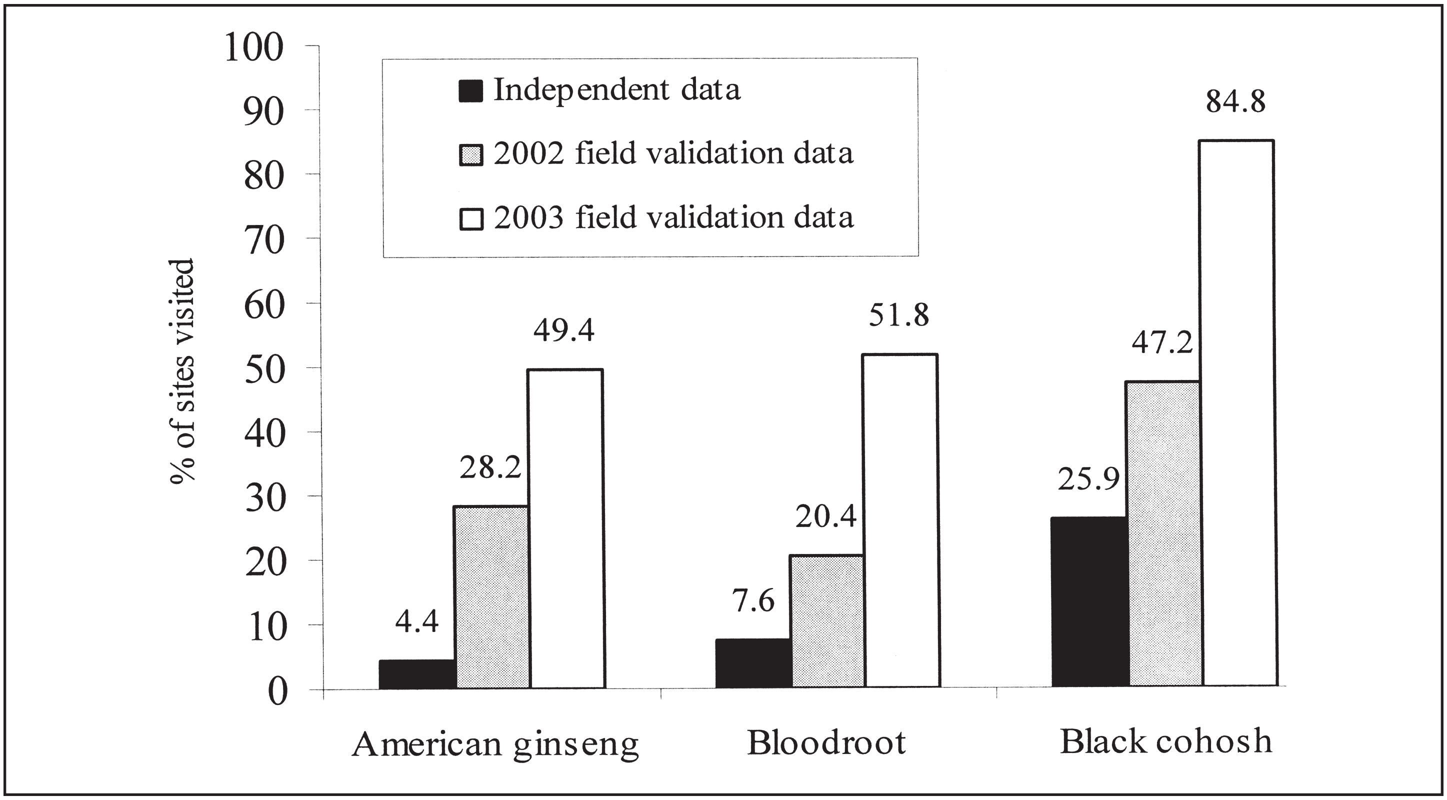

Figure 5. Percentage of sites visited in Shenandoah National Park containing targeted plant species. Independent data refer to a database of 148 random vegetation plots (J. Young and D. Walton, U.S. Geological Survey, unpublished data) sampled during 2001-2002 and 103 previously sampled plots (G. Fleming, Virginia Natural Heritage Program, unpublished data). The 2002 field validation data refer to targeted field searches of 142 sites located within the most favorable habitats according to the initial iteration of the plant habitat models. The 2003 field validation data refer to targeted field searches of 89 sites located within the most favorable habitats according to the final iteration of the plant habitat models.

(Figures 2 and 4) indicated that the validation data closely followed the pattern of the training data upon which the models were based. Therefore, we suggest that predictions were consistent for the range of habitat conditions represented by the original datasets. The validation data for bloodroot showed some differences with the training set (Figure 3), and we attribute that to the relatively small sample size for this species ($n = 28$ training locations). This observation suggests that larger sample sizes may be needed for model development to better represent varying habitat conditions.

Identifying model bias also is an important aspect of model testing. Model bias may be introduced, for example, because the training locations were not collected based on systematic field searches. Withholding data from model development would not allow detection of such biases, so we used independent field data from the 251 vegetation plots in Shenandoah National Park to further test our predictions. The logistic regression analysis based on those data confirmed that the plant habitat models were effective in identifying both suitable and unsuitable habitat types, although factors we could not examine likely played a role as well (range of maximum rescaled R^2 = 0.087-0.235; Table 2).

Finally, we tested model performance based on targeted field searches. We recognize that the areas we targeted for field searches (100-500 m from trails or roads) may be more accessible and experience greater harvest pressure than more remote areas. This sampling regime may have influenced our test results if plants already had been harvested from some areas predicted to be favorable habitat. Consequently, the targeted field searches likely provided a conservative assessment of model performance. The greater success rate of field searches conducted during 2003 compared with 2002 indicates that the iterative process of model development and testing was effective. The success rate was greatest for black cohosh, which likely reflects the fact that this species is common and, because of its large size, relatively easy to find. Bloodroot also is common, but it is a small and ephemeral plant that can easily be overlooked, possibly contributing to the lower success rate. The success rate for American ginseng was lowest, likely because of the rarity of the species, which, in turn, may be a result of harvest pressure.

Comparisons of the occurrence statistics from the field searches with the 251 random vegetation plots in Shenandoah National Park provided a useful measure of model efficacy. Compared with random searches, the odds of finding plants were 4.5 to 12.3 times greater when we targeted the best habitat areas based on the habitat model. The level of occupancy of the most favorable habitats was greatest for black cohosh (84.8%) and lowest for ginseng (49.4%; Figure 5). All three species, particularly

American ginseng, were frequently absent in areas where the model predicted favorable habitat conditions. There may be numerous reasons for these high false positive rates. For example, limited dispersal ability can prevent plant species from occupying suitable habitats (Boetsch et al. 2003), whereas deer browsing could reduce abundance of aboveground plants (McGraw and Furedi 2005), particularly in areas with high deer densities, such as national parks. Nonetheless, excessive harvesting of species such as American ginseng species has been documented (U.S. Fish and Wildlife Service 2003), and we speculate that this may have contributed to the low level of occupancy we observed for this species.

Performance of the habitat models may have been reduced by spatial errors associated with the location data. We mapped the habitat variables at a relatively high resolution of 10 m, which allowed us to incorporate microtopographic features that may be important for the establishment and growth of the plant species we studied. However, with greater resolution of GIS data, GPS errors may increase the likelihood that plant locations are mapped in incorrect grid cells. We reduced that error by only using plant locations for model development with mean positional errors <10 m. Data limitations also affect model performance. For example, some indirect gradients associated with plant responses may not be well represented with GIS data. Despite spatial errors and data limitations, however, our models clearly improved the efficiency of finding new plant locations. Model improvements may be possible once more accurate and consistent regional data for geology and soil type become available.

Model Application

The field searches by National Park Service personnel were designed to sample areas that were predicted to have favorable habitat for American ginseng, bloodroot, and black cohosh combined. All field searches resulted in a sizeable number of new location records for the plants. The targeted searches also may be useful to estimate potential species abundance, thereby providing a baseline for future studies. For example, 181,806 grid cells in Shenandoah National Park were within the two most favorable habitat classes for ginseng, bloodroot, and black cohosh combined. If 49.4% of those cells (2003 field search; Figure 5) were occupied by at least one ginseng plant, the minimum number of plants would be almost 90,000. Such information should be treated with extreme caution, however, because more independent field data would be needed to test the full range of model predictions. Furthermore, in the case of ginseng, a relatively high population estimate does not necessarily reflect a healthy, stable population. Because harvesting tends to remove the largest and oldest individuals, population estimates may reflect mostly plants that are younger than the reproductive age class of individuals (>5 years or 3-leaved).

MANAGEMENT AND RESEARCH IMPLICATIONS

Protection of vulnerable resources only can be effective if land managers know where those resources are, how abundant they are, and where they are the most vulnerable to exploitation. The plant habitat models we developed provided baseline data on the area and distribution of favorable habitats of the three plant species within Shenandoah National Park and northern Blue Ridge Parkway. Areas targeted for additional field searches based on the model results indicate that the potential level of occupancy of the most favorable habitats ranges from 84.8% for black cohosh to 49.4% for American ginseng. Thus, the results from our study may be used to gain insights into the potential area of occupied habitat and to provide a more quantitative assessment of threats due to illegal harvesting. The data on occupancy levels of favorable habitat areas, for example, may serve as an important benchmark for future habitat and population assessments. If similar searches were conducted in the future, marked reductions in the percentage of sites where the species are present could indicate possible population declines.

The second knowledge base needed to protect these plant species is to identify where they are most vulnerable to illegal harvest. Law enforcement personnel can use that information to target their actions (e.g., plant marking, covert plots, field monitoring of access points) to areas where enforcement efficacy can be maximized. Indeed, field searches using our habitat models as a guide led to the placement of covert plots to monitor removal of American ginseng from Shenandoah National Park. Subsequent enforcement activities revealed that roots from those plots were discovered as far away as South Korea (W. Cass, National Park Service, unpubl. report), demonstrating the reality of the threat to these resources due to poaching and the applicability of habitat modeling efforts to plant protection. Future research should focus on developing methods, such as expert-assisted models (Saaty 1977), to characterize and map potential harvest pressure based on terrain components (e.g., difficulty in traversing terrain due to topography or understory vegetation) and by incorporating elements of poaching behavior (e.g., known locations of illegal harvesting, traditional entrance and exit routes across park boundaries, the plant ‘search image’ used by poachers). That information may be combined with our habitat models to delineate areas where resources are most vulnerable to exploitation.

The habitat modeling procedures we used are made possible by the collection of accurate coordinates of species locations. This modeling approach may be applied to other plant species as a tool for identifying core habitat areas, developing management plans (van Manen et al. 2002), identifying new populations of rare plants, or documenting illegal harvesting activities. Efforts to collect location records of at-risk species would provide crucial data for future habitat modeling. Because location accuracy may affect model performance (Young et al. 2003), establishing data collection and metadata protocols for such databases would be important, particularly regarding the accuracy and precision of GPS locations.

ACKNOWLEDGMENTS

The National Park Service, Natural Resources Preservation Program (NRPP), provided funding for this study. We are very grateful to the National Park Service personnel who conducted the field searches for new locations based on our models. Janet Rock and Alan Williams provided technical and data management assistance. We owe many thanks to Ken Johnson for developing the concepts behind the Appalachian Chain Demonstration Project. We thank Gary Fleming of the Virginia Natural Heritage program for use of field data for an independent evaluation of our models. Finally, we thank Jay Clark for review of this manuscript.

Frank van Manen is a Research Ecologist with the Southern Appalachian Field Branch of the U.S Geological Survey Leetown Science Center. He holds an Adjunct Associate Professor appointment with the Department of Forestry, Wildlife, and Fisheries at the University of Tennessee. Frank uses principles of landscape ecology, wildlife science, and botany to address applied research questions related to a diverse array of species, ranging from large mammalian carnivores to rare plants and trees.

John Young is a Research Biologist with the U.S. Geological Survey Leetown Science Center, where he focuses on issues relating to landscape ecology, species distribution, and spatial modeling. He has been conducting spatial analysis on ecological issues for over 15 years, and his work has supported research for the National Park Service, the U.S. Fish and Wildlife Service, the U.S. Forest Service, and state agencies.

Cindy Thatcher is a Research Associate at the University of Tennessee's Department of Forestry, Wildlife, and Fisheries. Her research interests include biogeography and habitat modeling. She applies her experience with geographic information systems to address information needs of natural resource managers.

Wendy Cass is the Botanist for Shenandoah National Park. She manages the Park's long-term forest monitoring program, and works to monitor and protect rare plant species and communities throughout the 80,000-hectare Park.

Chris Ulrey is a plant ecologist with the Blue Ridge Parkway, National Park Service. He works on several issues such as rare plant monitoring and conservation, exotic plant management, forest pests and diseases, plant community research, grassy bald management, and wetland management.

LITERATURE CITED

Anderson, R.C., J.S. Fralish, J.E. Armstrong, and P.K. Benjamin. 1993. The ecology and biology of *Panax quinquefolium* L. (Araliaceae) in Illinois. American Midland Naturalist 129:357-372.

Bailey, R.G. 1980. Description of the ecoregions of the United States. Miscellaneous Publication No. 1391, U.S. Department of Agriculture, Washington, D.C.

Beers, T.W., P.E. Dress, and L.C. Wensel. 1966. Aspect transformation in site productivity research. Journal of Forestry 64:691-692.

Boetsch, J, F.T. van Manen, and J.D. Clark. 2003. Predicting rare plant occurrence in Great Smoky Mountains National Park. Natural Areas Journal 23:229-237.

Clark, J.D., J.E. Dunn, and K.G. Smith. 1993. A multivariate model of female black bear habitat use for a geographic information system. Journal of Wildlife Management 57:519-526.

CITES. 1973. Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). Part of the U.S. Endangered Species Act. Public law 93-205, Title 50.

Das, S., L. Shillington, and T. Hammett. 2001. Ginseng. Non-timber forest products, Fact Sheet no. 7. College of Natural Resources, Virginia Polytechnic Institute and State University, Blacksburg.

Gagnon, D. 1999. An analysis of the sustainability of American ginseng harvesting from the wild: the problem and possible solutions. Final report to the Division of Scientific Authority of the U.S. Fish and Wildlife Service. Available online <http://nps.gov/plants/medicinal/pubs/ginseng.htm>. Accessed 1 August 2003.

Gathwright, T.M. 1976. Geology of the Shenandoah National Park (map). Virginia Department of Conservation and Recreation, Division of Mineral Resources, Richmond.

Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler. 2002. The National Elevation Dataset. Photogrammetric Engineering and Remote Sensing 68:5-11.

Guisan, A., and N.E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modeling 135:147-186.

Hosmer, D.W., Jr., and S. Lemeshow. 1989. Applied Logistic Regression. J. Wiley, New York.

Iverson, L.R., M.E. Dale, C.T. Scott, and A. Prasad. 1997. A GIS-derived integrated moisture index to predict forest composition and productivity of Ohio forests (U.S.A.). Landscape Ecology 12:331-348.

Kauffman, G. 2002. Black cohosh (Cimicifuga racemosa) harvest monitoring 2002. Available online <http://www.nps.gov/plants/medicinal/pubs/2002bcohosh.htm>. Accessed 1 August 2003.

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.

McGraw, J.B. 2001. Evidence for decline in stature of American ginseng plants from herbarium specimens. Biological Conservation 98:25-32.

McGraw, J.B., and M.A. Furedi. 2005. Deer browsing and population viability of a forest understory plant. Science 307:920-922.

McGraw, J.B., S.M. Sanders, and M. Van der Voort. 2003. Distribution and abundance of *Hydrastis canadensis* L. (Ranunculaceae) and *Panax quinquefolius* L. (Araliaceae) in the central Appalachian region. Journal of the Torrey Botanical Society 130:62-69.

McNab, W.H. 1989. Terrain shape index: quantifying effect of minor landforms on tree height. Forest Science 35:91-104.

Michigan Natural Features Inventory. 1996. Special plant abstract for *Panax quinquefolius* (ginseng). Michigan Natural Features Inventory, Lansing, MI. Available online http://web4.msue.msu.edu/mnfi/abstracts/botany/panax_quinquefolius.pdf>. Accessed 25 September 2003.

National Park Service. 2003. National Park Service, Public Use Statistics Office. Available online <http://www2.nature.nps.gov/stats/>. Accessed 30 September 2003.

Odom, R.H., Jr., and W.H. McNab. 2000. Using digital terrain modeling to predict ecological types in the Balsam Mountains of western North Carolina. Research Note SRS-8, U.S. Forest Service Southern Research Station, Atlanta, Ga.

Parker, A.J. 1982. The topographic relative moisture index: an approach to soil-moisture assessment in mountain terrain. Physical Geography 3:160-168.

Peet, R.K., T.R. Wentworth, and P.S. White. 1998. A flexible, multipurpose method for recording vegetation composition and structure. Castanea 63:262-274.

Pereira, J.C., and R.M. Itami. 1991. GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham red squirrel. Photogrammetric Engineering and Remote Sensing 57:1475-1486.

Rao, C.R. 1952. Statistical Methods in Biometric Research. J. Wiley, New York.

Robbins, C.S. 2000. Comparative analysis of management regimes and medicinal plant trade monitoring mechanisms for American ginseng and goldenseal. Conservation Biology 14:1422-1434.

Saaty, T.L. 1977. A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology 15:234-281.

SAS® Institute, Inc. 2000. User's Guide, version 8.2: SAS® Institute Inc., Cary, N.C.

Stormer, F.A., and D.H. Johnson. 1986. Introduction: biometric approaches to modeling. Pp. 159-160 *in* J. Verner, M.L. Morrison, and C.J. Ralph, eds., Wildlife 2000: Modeling Habitat Relationships of Terrestrial Vertebrates. University of Wisconsin Press, Madison.

U.S. Department of Agriculture. 1994. State soil geographic (STATSGO) database data user's guide. Miscellaneous Publication 1492, U.S. Department of Agriculture, Natural Resource Conservation Service, Fort Worth, Tex.

U.S. Environmental Protection Agency. 2002. Level IV ecoregions of the continental United States (map). U.S. Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Corvallis, Ore.

U.S. Fish and Wildlife Service, Division of Scientific Authority. 2003. Convention permit applications for wild ginseng (*Panax quinquefolius*) harvested in 2003 and 2004. Available online http://international.fws.gov/ginseng/2003-2004ginsengfinding.htm>. Accessed 18 August 2003.

van Manen, F.T., J.D. Clark, S.E. Schlarbaum, K. Johnson, and G. Taylor. 2002. A model to predict the occurrence of surviving butternut trees in the southern Blue Ridge Mountains. Pp. 491-497 *in* J.M. Scott, P.J. Heglund, and M.L. Morrison, eds., Symposium on Predicting Species Occurrences: Issues of Scale and Accuracy. Island Press, Covelo, Calif.

Vogelmann, J E., S.M. Howard, L. Yang, C.R. Larson, B.K. Wylie, and N. Van Driel. 2001. Completion of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. Photogrammetric Engineering and Remote Sensing 67:650-652.

Wilds, S.P. 1997. Gradient analysis of the distribution of a fungal disease of *Cornus florida* in the Southern Appalachians, Tennessee. Journal of Vegetation Science 8:811-818.

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.

Young, J.A., F.T. van Manen, and C.A. Thatcher. 2003. Habitat modeling for protection of illegally harvested plants in national parks of the Blue Ridge Mountains. Final report to the National Park Service. U.S. Geological Survey, Leetown Science Center, Kearneysville, W.Va.