

A GIS-based Habitat Model for Wood Thrush, *Hylocichla mustelina*, in Great Smoky Mountains National Park

Susan A. Shriner, Theodore R. Simons, and George L. Farnsworth

Management of wildlife populations requires knowledge of the ecology and habitat preferences of the species of interest. Typically, the majority of habitat information available for Neotropical migratory birds has been collected at a microhabitat scale. However, managers interested in identifying key areas for a given species rarely have the resources to collect detailed microhabitat data over broad regions. Therefore, identifying habitat variables that are both readily available over large areas and correlated to species occurrence is essential to effective management of large areas (Simberloff 1988). Topographic indexes based on readily available data (digital elevation models) have been successfully applied to models developed to predict plant occurrence (Wilds 1996; Wisser et al. 1998) but are not often used to predict animal occurrence. Nonetheless, these indexes may be useful predictors of animal occurrence over broad regions. In this chapter, we present a logistic regression model for wood thrush (*Hylocichla mustelina*) in Great Smoky Mountains National Park based on topographic and other readily available habitat variables.

As the result of several studies showing possible population declines (Robbins et al. 1989b; Askins et al. 1990), Neotropical migratory songbirds have become major research species in conservation biology. In particular, the wood thrush has been the focus of many studies examining the impact of fragmentation

and land-use change on population abundance, distribution, and productivity (Robinson 1988; Hoover and Brittingham 1993; Robinson et al. 1993; Hoover et al. 1995; Robinson et al. 1995a). Most efforts to develop predictive models to assess the location of suitable habitat over large areas for Neotropical migratory songbirds have given rise to models based on patch size and configuration and/or habitat fragmentation indexes. Although these types of models may be a useful first attempt at identifying potential habitat, they do not provide information about which habitat features within large forested landscapes are important predictors of Neotropical migratory songbird occurrence. Therefore, these models may not be usefully applied to large areas of contiguous forests. On the other hand, models that do incorporate variables associated with habitat use are usually based on fine-grained microhabitat data that require extensive data collection and cannot feasibly be applied to large areas.

The southeastern United States, in particular, Great Smoky Mountains National Park, is a species-rich region with a complex biota. The park is the largest tract of contiguous forest within the southeastern United States and the second-largest national park in the eastern United States. As part of a network of protected areas in the southern Appalachians, the park is an important reserve for many Neotropical migratory land birds. Surveying the physiological, genetic, and

life-history characteristics of each species in such a region is a near-impossible task (Martin et al. 1993). Models based on variables readily available in a GIS may offer a solution to such an intractable problem. Therefore, we investigated the utility of using commonly available GIS data to assess several topographic indexes and other habitat variables as predictors of species presence or absence for wood thrush in Great Smoky Mountains National Park.

Great Smoky Mountains National Park represents a unique opportunity to examine eastern forest birds in a setting free from recent human disturbance. A relatively large portion of the park has been undisturbed, and most of the areas that have been previously altered have had sixty-five to one hundred years to recover. This lack of human disturbance enhances the likelihood that topographic indexes correlated with vegetation features will be useful predictors of bird presence or absence. In disturbed areas, variables such as stand age and landscape configuration may be more-important predictors of vegetation features, but in undisturbed areas, topographic indexes are likely to be highly correlated with vegetation. Several researchers (Austin et al. 1984; Yee and Mitchell 1991) have been able to successfully predict vegetation characteristics based on site conditions (Wiser et al. 1998). We hypothesized that for areas primarily undisturbed by recent human impacts, topographic variables may be useful predictors of wood thrush presence or absence, since topographic indexes are likely to act as predictors of vegetation features.

Methods

Great Smoky Mountains National Park comprises 205,665 hectares of primarily contiguous forest straddling the Appalachian Trail along the Tennessee–North Carolina border. The park serves as the nucleus of a group of protected areas in the southern Appalachians that includes national forests, federally designated wilderness areas, state lands, Tennessee Valley Authority reservoirs, and National Park Service lands (see Fig. 47.1 in color section). This region consists of 15.1 million hectares, including more than 2 million hectares of public land. More than 70 percent of the region is currently forested (SAMAB 1996), providing

the largest extent of forested landscape in the eastern United States.

Great Smoky Mountains National Park is characterized by a wide elevation range (575–1,830 meters) and complex topography, creating a rich diversity of habitat and vegetation types. MacKenzie (1993) used Landsat imagery to develop a vegetation map of the park based on thirteen major forest types. MacKenzie patterned his vegetation scheme after the original classification system of Whittaker (1956). The MacKenzie vegetation types range from pine and pine oak forests at the lowest elevations of the park to northern hardwoods and spruce fir at the highest elevations. Seven of these vegetation types are regularly used by breeding wood thrush. In general, these habitats occur along a wet-to-dry moisture gradient and include cove hardwood, mixed mesic hardwood, tulip poplar, mesic oak, xeric oak, pine-oak, and pine.

The Great Smoky Mountains was established as a national park in 1934. At that time, more than half of the land had been disturbed, primarily by different methods of logging. Other disturbances were caused by large-scale fires and clearing of the land for homesteads. Current disturbances include extensive loss of Fraser fir (*Abies fraseri*) due to infestation by the exotic balsam woolly adelgid (*Adelges piceae*) and decline of many pine species (*Pinus* spp.) due to southern pine beetle (*Dendroctonus frontalis*) infestations. The hemlock woolly adelgid (*Adelges tsugae*) is also expected to spread to the park within the next decade (SAMAB 1996). This exotic pest will likely affect eastern hemlock (*Tsuga canadensis*) populations and therefore impact many of the cove hardwood and mixed mesic hardwood forests. Bird species such as the wood thrush may be seriously impacted if the park experiences major losses of eastern hemlock. More than 90 percent of wood thrush nests found in the park have been located in eastern hemlocks (Farnsworth and Simons 1999).

Field Data

We conducted variable circular plot point counts (Reynolds et al. 1980) at more than four thousand locations throughout our study area during May and June of 1996–1999. Due to the large number of observers ($n = 40$), we employed several strategies to

minimize observer variability. We trained and tested all observers prior to the initiation of fieldwork. We also provided a review training session at the midpoint of the field season. In addition, we rotated observers throughout the different areas of the park so that the efforts of a single observer were not restricted to a particular area of the park. We conducted counts for a ten-minute interval between dawn and 10:15 A.M. and only in good weather (no rain or excessive wind). Our count protocol is consistent with the majority of the recommendations for point count methodology detailed by Ralph et al. (1995b).

During each ten-minute count, we recorded the number of breeding pairs of each bird species present. A single observer collected data at any given point. Before each count, observers estimated a 50-meter radius circle by spotting landmarks using a laser range finder and began the count immediately thereafter. At each point, we recorded bird detections in all directions for an unlimited radius plot. We mapped the location and movement of all individual birds detected in order to avoid double counting.

We established the majority of points systematically along trails with some points located on minor roads and some points located off-trail along transects. The majority of trails in Great Smoky Mountains National Park are low impact and do not result in a gap in the tree canopy. Comparison of counts from points located on- and off-trail showed no significant differences (S. Shriner unpublished data). We spaced points 250 meters apart to avoid double-counting birds. We sited points primarily by pacing and occasionally by using a laser range-finder. The meandering nature of the trails resulted in an average horizontal distance between points of approximately 175 meters. In areas where we could discern that the trail was very windy, we paced an additional 50 to 500 meters between points. We sited a small number of points in Cades Cove, a large, open area. We spaced these points 500 meters apart to account for an increase in detectability due to the sparse vegetation. In the event that an individual bird was heard at more than one point, we only recorded the bird at the point with the smaller detection distance.

We stratified points throughout the park with respect to the availability of each MacKenzie vegetation

type. We used Trimble GeoExplorer II global positioning system (GPS) units to collect the geographic coordinates of each point.

Digital Data

We obtained habitat and topographic variable data from a GIS database provided by the Inventory and Management Division of Great Smoky Mountains National Park. All queries were performed using ArcInfo (ESRI 1997) software. The Great Smoky Mountains National Park GIS database contains 90×90-meter grid data for vegetation type, bedrock geology, and disturbance history. Vegetation type information is based on the groundtruthed analysis of satellite imagery produced by MacKenzie (1993); it includes thirteen vegetation types. The bedrock geology data include twenty-four classes of bedrock. The disturbance history data are based on an analysis of park records and include five categories of human disturbance: undisturbed, selective cut, light cut, heavy cut, and settlement (Pyle 1985).

Data for elevation and several topographic indexes (including topographic convergence index, terrain shape index, landform index, topographic complexity, relative moisture, and relative slope position) are available in the database as 30×30-meter grids. Elevation for each point was calculated from a digital elevation model (DEM) using an interpolation function based on the nearest-neighbor cells. Slope and aspect were computed in ArcInfo using the DEM. Aspect was transformed into north/south and east/west components using $\sin(\text{aspect})$ and $\cos(\text{aspect})$, respectively. The different topographic variables available in the park GIS were developed by several different researchers to characterize the shape of the landscape and local moisture regimes at various spatial scales. These indexes are coded in Arc Macro Language (AML) and are available as coverages in the park GIS database. The topographic convergence index (TCI) is an index of potential soil moisture developed by Beven and Kirkby (1979) and was developed to simulate runoff saturation and infiltration. This index has been successfully used in spatial models of vegetation distribution. The underlying formula is $TCI = \ln[(A/\tan B)]$ where A is the surface area of each grid cell providing drainage and B is the surface slope of the grid cell.

The terrain shape index (TSI) was developed by McNab (1989) to characterize the topographical curvature of the landscape. The TSI distinguishes between ridges/exposed areas and coves/protected areas by calculating the average difference in elevation between the center of a plot and its boundary. The landform index (LFI) was also developed by McNab (1993) and is a large-area parameter that describes general classes of protection at a site, or in other words, cove, slope, or ridge. The LFI is the mean of eight different slope gradients (N, NE, E, SE, S, SW, W, NW) calculated from the center of a plot to the skyline.

Topographic complexity was computed as the Shannon-Weaver Index of Topographic Complexity (SWI), an index developed by Miller (1986) to explain the distribution of rare and endemic plant species. The SWI index is a fine-scale index of the topographic diversity of a 150×150-meter plot. Topographic complexity is calculated for elevation, slope, and aspect and then combined into a single measure (Wilds 1996). The SWI is calculated as $SWI = -\sum(p_i \cdot \log_2(p_i))$, where p = the proportion of area for each elevation/slope/aspect category.

Relative moisture is described by the topographic relative moisture index (TRMI) developed by Parker (1982) to model vegetation distribution in the western United States. The TRMI was developed to describe local moisture regimes for areas with diverse topography. It is based on aspect, steepness, topographic position, and curvature. Relative slope is also a measure of relative moisture and is the distance from the point to the bottom of the slope divided by the total distance from the bottom of the slope to the nearest ridge. The distance to the bottom of the slope is measured to the nearest stream or the nearest topographic concavity from the point, perpendicular to contour lines. The distance to the top of the slope is measured to the nearest ridge or topographic convexity, perpendicular to contour lines.

The Model

The model is a logistic regression model that returns the probability of detecting a wood thrush as a function of habitat and topographic variables. Logistic re-

gression is a statistical technique used to predict the probability of an event occurrence (P_0) and has been used to predict the probability that an organism will occur based on the conditions present at a particular site (e.g., Margules and Stein 1989; Wiser et al. 1998). Probability values are constrained to range between 0 and 1 and therefore cannot be modeled using linear functions. In logistic regression, explanatory variables that can range from positive to negative infinity are transformed to a probability using a logit link function. The logistic regression equation models the logit transformation as a linear function of the explanatory variables (Christensen 1997). The result is a probability value in the 0 to 1 range.

We used presence/absence data for wood thrush derived from the point counts as the dependent variable in the model and the eleven habitat and topographic variables derived from the GIS as the explanatory variables. We ran the model using the PROC LOGISTIC procedure in SAS (SAS Institute 1990b). We used the backward elimination and forward selection procedures (threshold p -values = 0.1) in SAS to compare different logistic regression models based on the eleven variables, the squared values of the topographic variables, and interactions between the topographic variables. The backward elimination procedure analyzes the full model and then removes variables one at a time as they fail to meet the specified significance level for staying in the model. The forward selection procedure begins with an intercept-only model and adds variables one at a time based on adjusted chi-square statistics. A variable is added to the model if it meets the specified significance level for staying in the model.

The topographic variables (but not the squared terms) were standardized (mean = 0, variance = 1) prior to analysis to aid in the interpretation of the parameter estimates. If a squared variable or an interaction variable was significant (p -value less than or equal to 0.1), then the main variable was included in the model without regard to its significance level. A significant interaction term indicates that one of the variables has a modifying effect on the impact of another variable.

We treated the thirteen vegetation types, twenty-

three bedrock geology types, and five disturbance classes as class variables. We only included type variables that were represented by a minimum of thirty observations such that the final analysis was performed on ten vegetation types and eighteen geology types. All counts with missing data or type variables not included in the analysis were deleted from the data set, which resulted in 3,743 points available for analysis. Each of these data points was randomly assigned to either a model development data set ($n = 1,833$) or a validation data set ($n = 1,910$); only the model development data set was used for model selection.

Final model selection was based on concordance scores (SAS Institute 1990a,b). Concordance is a measure of rank correlation between the predicted probability that a wood thrush is present at a site and the actual presence/absence of wood thrush at that site (Bolger et al. 1997). Concordance is calculated as a percentage of all possible pairs of observations where one member of the pair represents the presence of a wood thrush and the other member of a pair represents an absence of a wood thrush. A pair is concordant if an observation representing an absence of wood thrush has a lower predicted event probability than does the observation representing the presence of wood thrush. A pair is discordant if the observation representing an absence of wood thrush has a higher predicted event probability than the observation representing the presence of wood thrush. If a pair is neither concordant nor discordant, then it is a tie. A model with a higher concordance score is more likely to correctly classify the presence or absence of wood thrush at a particular location. We compared the concordance scores for the models determined by the backward elimination and forward selection procedures to choose the final model.

Validation and Model Assessment

We applied the final model to the validation data set to compare the performance of the validation data with the model development data. Although model selection was based on concordance scores, we also evaluated the final model according to its ability to correctly classify presence and absence for the ob-

served data. We classified points with a probability greater than or equal to 0.3 as present. We chose this probability because it represents a balance between model sensitivity and specificity (see Dettmers et al., Chapter 54).

The Probability Map

We used the logistic regression model to develop a probability map for Great Smoky Mountains National Park. The probability map represents the probability of detecting a wood thrush at any location in the Park. We created a 90×90-meter grid cell coverage of the park by programming the logistic regression model in ArcInfo map algebra language. Because the vegetation type, bedrock geology, and disturbance history coverages are only available as 90×90-meter grids, we coded the probability map at that same grain. For each grid cell, the values of each of the explanatory variables were determined by querying the appropriate GIS coverage. The probability for each grid cell was then calculated based on the logistic regression equation to create a probability coverage of detecting a wood thrush.

Results and Discussion

Two of the topographic indexes (SWI and LFI) tested were significantly correlated with wood thrush occurrence (Table 47.1). The model with the highest concordance score includes these two indexes as well as elevation, disturbance history, and geology type. Several squared and interaction terms were also significantly associated with wood thrush presence/absence. All of the variables included in the model were significant at the p less than 0.05 level. The overall model had a relatively high concordance of 77.9 percent and correctly classified 83 percent of the observed data points. Application of the best-fit model to the validation data set resulted in a concordance of 78.9 percent and correct classification of 86 percent of the observed data.

The parameter estimates indicate that elevation had the strongest predictive power of the variables included in the model. This result is consistent with the elevation range of the wood thrush, which is primarily re-

TABLE 47.1.

Result of logistic regression model selection for wood thrush presence/absence using model selection data.

Variable ^a	DF	Parameter estimate	Standard error	Chi-square	
				Wald	Pr
INTERCPT	1	-12.3752	19.2466	4.9268	0.0264
ELEV	1	9.0385	1.9063	22.4803	< 0.0001
ELEV2	1	< -0.001	< 0.001	43.3424	< 0.0001
SWI	1	-2.8090	0.9866	8.1065	0.0044
SWI2	1	0.0231	0.0075	9.5938	0.0020
LFI	1	0.8238	0.3433	5.7578	0.0164
ELLFI	1	-1.2637	0.4531	7.7668	0.0053
GEOL	21	—	—	40.4717	0.0065
DISTURB	4	—	—	10.3836	0.0344

Concordance: Concordant = 77.9 percent; Discordant = 21.7; Tied = 0.4 percent; (356,040 pairs)

^aVariable abbreviations are as follows: ELEV = elevation, ELEV2 = elevation squared, SWI = Shannon-Wiener Index of Topographic Complexity, SWI2 = SWI squared, LFI = Landform Index, ELLFI = elevation * LFI, GEOL = bedrock type, DISTURB = disturbance history type.

stricted to elevations below 1,200 meters while elevations in the park range up to 1,800 meters. This elevation boundary is evident in the probability map (see Fig. 47.2 in color section), which shows low probability of detecting a wood thrush in the center (highest elevations) of the park. The Shannon-Weaver Index of Topographic Complexity (SWI) was negatively associated with the likelihood of detecting a wood thrush and had the next-greatest explanatory power. The SWI is a relatively fine-scale (150-meter) measure of land form that describes the diversity of elevations, slopes, and aspects around a center point. The model also includes a positive association with the coarse-scale Landform Index, which is a broad measure of site protection on the scale of kilometers. This variable indicates that the probability of detecting a wood thrush increases with increasing protection. The disturbance history and geology type class variables were also significantly associated with the probability of detecting a wood thrush. Parameter estimates (Table 47.2) for individual disturbance types were calculated relative to the undisturbed type. Negative parameter estimates for areas that experienced logging in the past indicate these areas may be less likely to be associated with wood thrush occurrence than undisturbed areas. It is interesting that the vegetation type variable was

not identified as being significantly correlated with wood thrush presence/absence. It is possible that the vegetation type variable did not have any explanatory power beyond the variation explained by the topographic variables.

The specific results of this model are unlikely to be applicable outside of Great Smoky Mountains National Park, which is largely undisturbed. However, this research highlights the potential for topographic indexes to be useful predictors of songbird occurrence and they should be more commonly tested in habitat models. In addition to their possible predictive value, topographic indexes are attractive because many of them can be easily calculated from DEM data, which is often readily available for large areas. These vari-

TABLE 47.2.

Parameter estimates for the disturbance history type class variable. Estimates are relative to the undisturbed type.

Type	Parameter estimate
Selective cut	-0.0222
Light commercial cut	-0.3675
Heavy cut	-0.3603
Settlement area	0.4152

ables are also appealing because they may be useful surrogates for vegetation information that might otherwise have to be collected in the field.

Conclusion

In order to develop a management strategy for the conservation of wood thrush, it is important to identify variables associated with species distribution. Traditionally, ecologists have sought to identify fine-scale microhabitat features associated with occurrence. Microhabitat data is typically only available for small areas and can be quite costly and time intensive to collect, making large-area assessments difficult. On the other hand, the data needed to develop

models based on topographic variables are often readily available in GIS databases and can be applied to large areas.

Acknowledgments

This work was funded by grants to Ted Simons from the Friends of Great Smoky Mountains National Park and the U.S. Geological Survey, Biological Resources Division. We would like to thank Jeremy Lichstein for helpful comments and statistical advice. We thank the staff at Great Smoky Mountains National Park for technical and logistical support. We also thank the many members of our field crew who made this work possible.

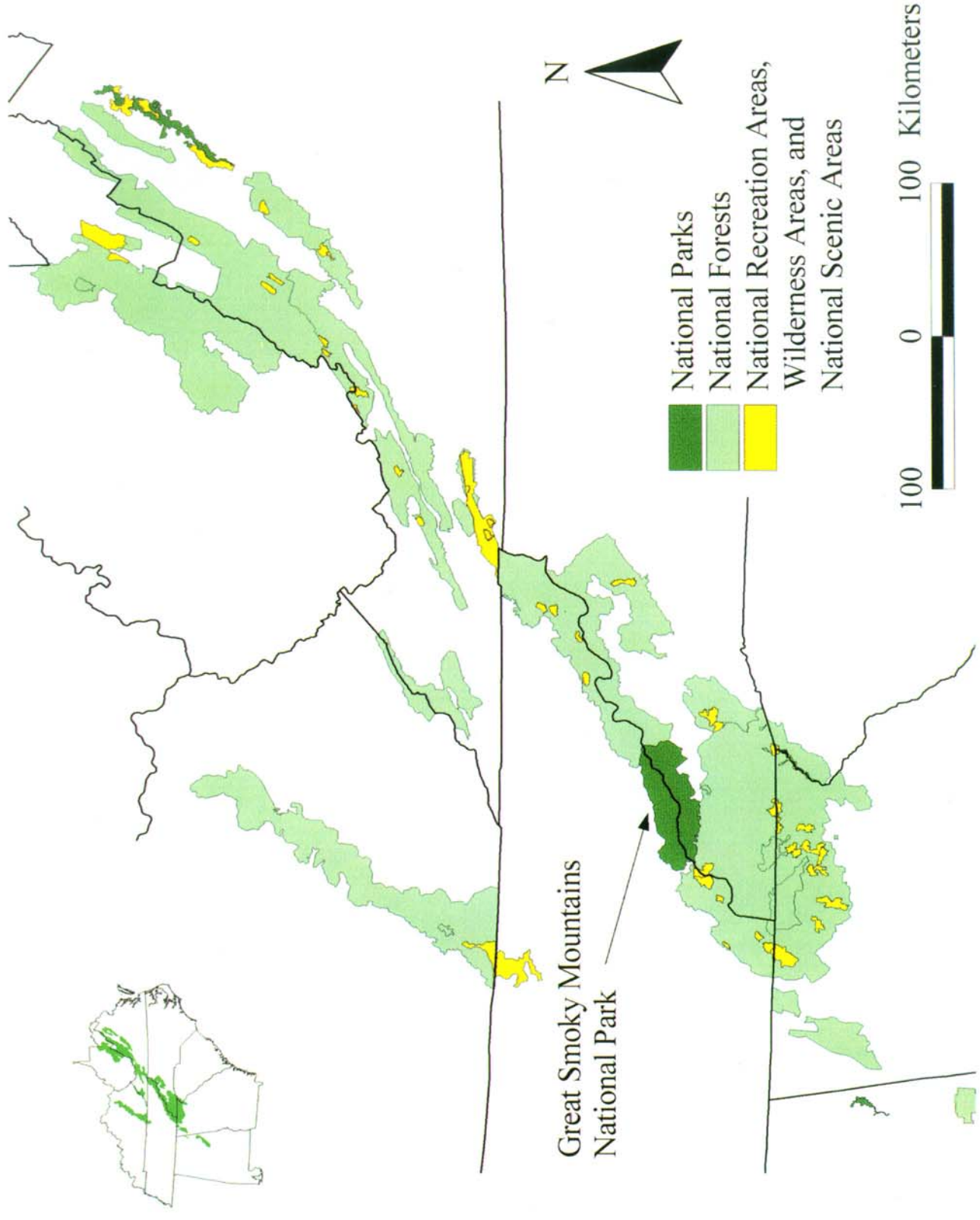


Figure 47.1. Great Smoky Mountains National Park as part of the network of protected areas in the southern Appalachians.

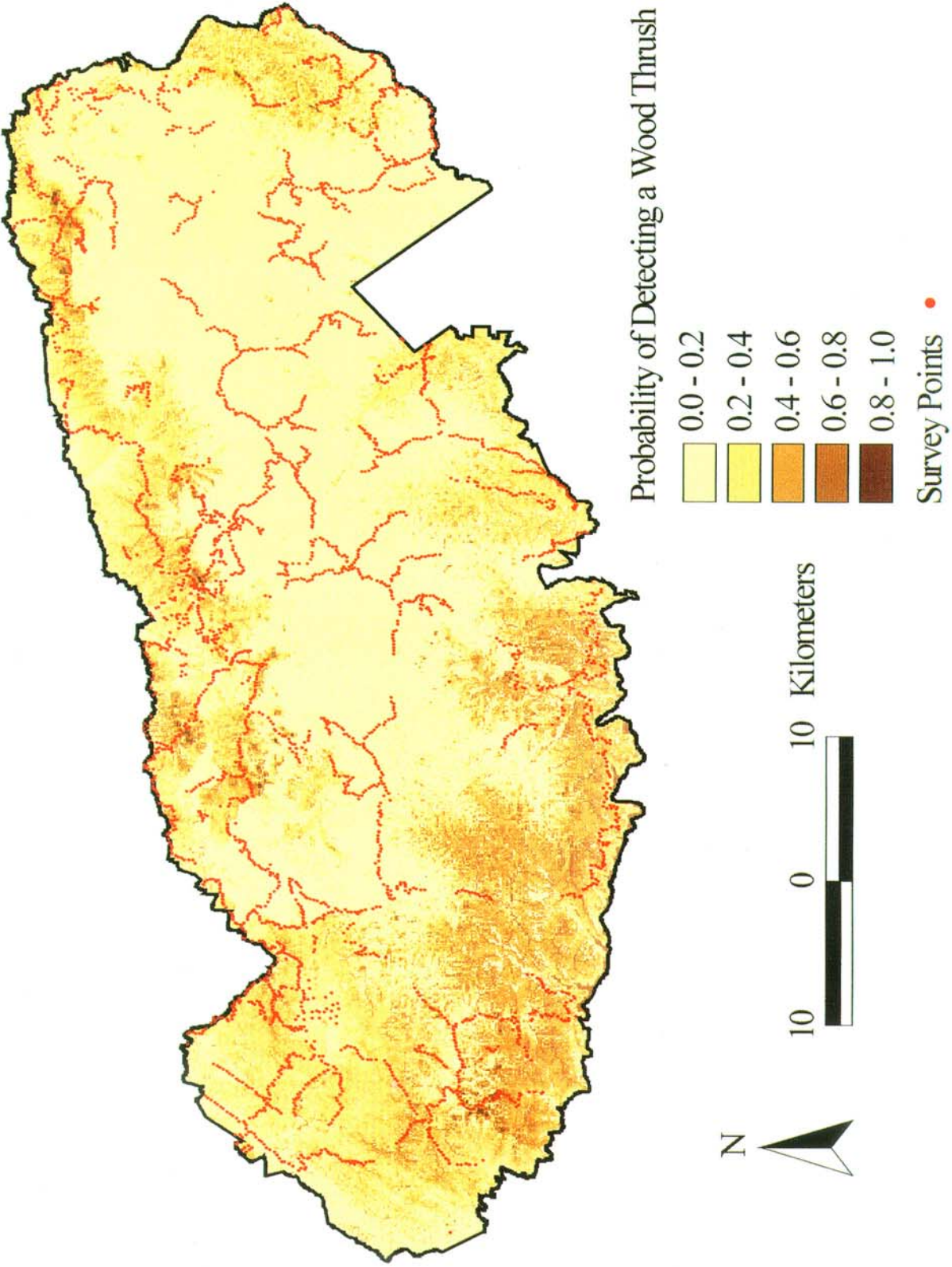


Figure 47.2. The probability of detecting a wood thrush in Great Smoky Mountains National Park.