A technique to locate isolated populations using satellite imagery

L. Mike Conner

Abstract Locating isolated populations is important to conservation efforts. Here I describe a technique using satellite imagery to predict species presence across large areas (e.g., watershed or state). The technique requires knowledge of species presence at training sites within the area of interest. These training sites are used to develop a supervised classification, using Mahalanobis distances as a classification rule, to identify similar sites. The technique should provide better prediction than a typical habitat classification (i.e., a habitat map derived from satellite image interpretation) because all data in the image can be used to develop the model. As an example, I developed and tested a model to predict presence of Bachman's sparrows (*Aimophila aestivalis*) within the 1.48×10^6 ha lower Flint River basin, Georgia. As predicted, the model was better than a habitat classification to identify presence of Bachman's sparrow. Using satellite imagery to develop species-presence models might provide more accurate models of species occurrence, thus reducing time needed to search for isolated populations.

Key words *Aimophila aestivalis,* Bachman's sparrow, Georgia, habitat model, Mahalanobis distance, monitoring, satellite imagery

Wildlife habitat models are created for a variety of purposes. Models may be used to predict species presence or absence (Scott et al. 1993) or to increase understanding of habitat requirements of a species (Clark et al.1993, Kopp et al. 1998). This latter group of models might require intensive habitat sampling (e.g., Jorgensen et al. 1998, Kopp et al. 1998) or a detailed geographical information system (GIS) for model development (e.g., Clark et al. 1993, Roseberry and Sudkamp 1998, Miller et al. 2000). When the goal of modeling is merely to determine species presence, intensive sampling or development of a detailed GIS might not be required.

Creation of habitat maps (i.e., layers) is a common use of satellite imagery, and these habitat layers are often incorporated into a GIS and used to predict species presence (Scott et al. 1993, Roseberry and Sudkamp 1998). Satellite images have multiple bands (statistically, a band is analogous to a variable) of data, and each habitat could provide a "signature" that can be detected using the imagery. Because different types of vegetation reflect and absorb radiation differently (e.g., coniferous trees absorb more radiation in the near-infrared region than hardwoods do), the amount of radiation reflected is often used to assign each pixel to a single vegetation or land-cover (i.e., habitat) class. In developing a habitat layer, multivariate statistics, usually discriminate function or cluster analysis, are used to identify similar habitat types. The classification process is designed to replace multiple bands of data with a single value, the habitat class. Thus, information about vegetation conditions represented by the pixel is lost in the classification process.

Habitat associated with presence of a species often requires more information than provided from a typical habitat layer as developed from satellite imagery (i.e., a land-cover layer). For example, Bachman's sparrows (*Aimophila aestivalis*, hereafter BACS) are associated with open-canopied pine

Author's address: Joseph W. Jones Ecological Research Center, Route 2, Box 2324, Newton, GA 31770, USA; e-mail: mconner@jonesctr.org.

stands or early successional habitats with a dense understory (Dunning 1993). Most land-cover layers do not provide sufficient detail to develop a useful "search strategy" for locating BACS populations at landscape scales. Rather, a typical land-cover layer delineates only very crude habitat types (e.g., deciduous forest, coniferous forest, agriculture, etc.). Despite obvious limitations, a land-cover layer might be helpful for ruling out areas where BACS are not likely to occur (e.g., hardwood habitats, agriculture, etc.). However, the land-cover layer might be too general to determine suitability of remaining habitats (e.g., incapable of discriminating between suitable and unsuitable coniferous forests).

Biologists need not be restricted to standard habitat classifications when predicting species occurrence. In fact, unclassified satellite images could actually prove better to locate isolated populations because unclassified images contain more data to predict species occurrence (i.e., all bands are available for model development). The objective of this paper is to describe a technique for using satellite imagery to locate isolated populations. This accomplished, I then provide an example using BACS to illustrate the procedure.

Modeling procedure

Satellite imagery consists of cells (i.e., pixels) of a given size covering an area of landscape. For each pixel, there are multiple bands of data representing amount of electromagnetic radiation either reflected or emitted, within a given wavelength. For example, LANDSAT Thematic Mapper (EROS Data Center, United States Geological Survey, Sioux Falls, S. D.) data have approximately 30×30 m pixels and contain 3 bands within the visible spectrum (i.e., approximately 0.4–0.7 μm), 1 band within the near infrared region (i.e., approximately 0.7–1.3 μm), 2 bands in the mid-infrared region (i.e., approximately 1.3–3.0 μm), and 1 band in the thermal infrared (approximately 3.0–14.0 μm; pixels in this band are 120×120 m). Thus, for each pixel there are 7 values describing the amount of electromagnetic radiation reflected or emitted at each monitored wavelength (Lillesand and Kiefer 1994).

If there are sites where a species of interest is known to occur, it is a simple matter to delineate these sites on a satellite image and then use these areas as training sites to develop a model to locate similar areas. The number of training sites must exceed the number of bands of data by at least 1, but, as with any statistical procedure, better results are obtained with larger samples (Lillesand and Kiefer 1994). This approach is analogous to standard biometric habitat modeling procedures (Brennan et al. 1986, Clark et al. 1993). However, the approach differs in that pixel reflectance values, instead of traditional habitat variables (e.g., vegetation type, land use, percentage groundcover, etc.), are used with the satellite image (Lovallo et al. 2001).

Standard statistical modeling approaches can be used to develop models of species presence. However, there are appealing attributes associated with using Mahalanobis distances (Clark et al. 1993, Lovallo et al. 2001) to develop a satellite image-based model. First, the use of Mahalanobis distances does not require using random sites or sites where the animal is absent to develop a model, a problem with other procedures such as discriminant function analysis or logistic regression (Clark et al. 1993, Conner and Leopold 1998). Second, Mahalanobis distances are distributed as χ^2 with df=(number of bands) - 1; thus, *P*-values can be calculated to represent an index of the probability that a cell belongs to the group of interest (Clark et al. 1993). For example, if we used 3 bands of data, the output distances would be compared to the χ^2 distribution with 2 df to derive *P*-values. Finally, because Mahalanobis distances are sometimes used to create habitat classifications using satellite imagery (ERDAS, Inc. 2001), most image-processing packages contain Mahalanobis distance routines.

Essentially, the modeling procedure treats areas where the species of interest is known to occur as training sites. The training sites are then used to develop a supervised classification (Lillesand and Kiefer 1994), using the Mahalanobis distance classification rule (ERDAS, Inc. 2001), to locate other areas where the species might occur.

Example application

Training-site selection

I conducted fixed-radius (50 m) point counts (Ralph et al. 1995) on Ichauway, the research and conservation landholding entrusted to the Joseph W. Jones Ecological Research Center, Baker County, Georgia. I selected 60 sample sites where BACS were detected during June 1999. I used a differentially corrected global positioning system (GPS) to locate the center of each sample station and used

ARC/INFO (Environmental Systems Research Institute, Inc. 2001) to place a 50-m buffer around these points. I treated the resulting circles as training sites to create a model to locate BACS habitat in the lower Flint River basin (FRB), Georgia.

Satellite images

I used 8 SPOT multispectral (MS) scenes (SPOT Image Corporation, Chantilly, Va.) covering the lower FRB for the remotely sensed data in this example. I purchased images obtained during leaf-off conditions (February) of 1999. I georegistered each scene using ground control points digitized from 7.5-minute United States Geologic Survey (USGS) topographic maps. After georegistration, I compiled scenes into a mosaic to form a continuous image and then clipped this image to create a single image containing only the lower FRB (Lillesand and Kiefer 1994). I normalized the image (Hall et al. 1991) to account for differences among scenes and used the normalized image for model development. I performed all image processing using IMAGINE (ERDAS, Inc. 2001) image-processing software.

I used 3 bands of data, 2 bands in the visible spectrum and 1 band in the near-infrared spectrum. Thus, when treating each band as a variable, I had 3 predictor variables available for model development. Spatial resolution of SPOT MS images was approximately 20 × 20 m (Lillesand and Kiefer 1994).

Model creation

I calculated a mean vector and covariance matrix from reflectance values of pixels within training sites. I then calculated the Mahalanobis distance between this mean vector and the reflectance values of each pixel in the lower FRB. Last, I derived *P*-values for each distance as an index of the probability of a pixel containing habitat similar to the habitat observed training sites. I used the supervised classification tool in IMAGINE (ERDAS, Inc. 2001) to perform all calculations.

I recoded *P*-values into 10 classes (0.0-0.1, 0.1-0.2...0.9-1.0). Because each pixel represented a 20 × 20-m area on the earth, an area too small to support BACS, I used a 5 × 5 majority filter to "smooth" the image. This filter resulted in a moving window that was 1 ha in size, representing approximately half of a typical BACS territory (Dunning 1993). The center pixel within this 5 × 5 window was assigned the value associated with the majority of the cells within this window (Image Interpreter tool in IMAGINE; ERDAS, Inc. 2001). Thus, if the center pixel was classified as relatively poor habitat (i.e., had a low *P*-value) while surrounding habitat was excellent (i.e., had a high *P*-value), this pixel would have been reclassified as excellent. Similarly, if the center pixel was classified as excellent habitat while surrounding pixels were classed as poor, the center pixel would have reclassified as poor. I converted the resulting image from cells to polygons using the GRIDPOLY command in ARC/INFO (Environmental Systems Research Institute, Inc. 2001). Each polygon contained the *P*-value class assigned to the cells following the 5 × 5 majority filter.

I considered all polygons with *P*-values <0.5 as unsuitable. Further, because BACS are territorial during the breeding season and typically maintain territories of about 2 ha (Dunning 1993), I assumed that suitable BACS habitat must also be ≥2 ha. Therefore, I coded all polygons <2 ha as unsuitable. This model predicted that approximately 30,000 ha of BACS habitat existed in the lower FRB. These efforts resulted in a layer representing potential BACS habitat in the lower FRB, hereafter referred to as the Mahalanobis habitat model (MHM).

Model testing

One common problem with wildlife habitat models is a lack of model validation using independent data (Marcot et al. 1984, Verbyla and Litvaitis 1989). To assess validity of the BACS model, I selected all polygons >50 ha classified as suitable BACS habitat. I set a minimum size of 50 ha to ensure the site could be located using a road map and there was a reasonable chance the site was large enough to support a population of BACS. There were 58 patches throughout the basin that met the criteria for sampling as areas where BACS were predicted to occur, and these sites made up approximately 20% of the total area classed as suitable within the basin. Twelve of these sites were >100 ha. Because habitat occupancy is area-specific (MacArthur and Wilson 1967, Martin 1980, Blake and Karr 1987), I chose all 12 sites >100 ha and randomly chose an additional 12 sites ranging from 50-100 ha. I selected all sites such that no site was within 1 km of another site where BACS were predicted to occur. This resulted in a selection of 24 sites where BACS were predicted to occur.

I also estimated errors of omission (i.e., determined the percentage of sites where BACS occurred but were predicted absent by MHM). A land-cover layer was developed for the lower FRB using the 1999 SPOT satellite imagery. This layer classified land cover into 9 categories: 1) cultivated or bare ground, 2) pasture, 3) shrub-scrub, 4) coniferous, 5) mixed coniferous and deciduous, 6) deciduous, 7) open water, 8) wetland, and 9) urban. Because much recent attention has been directed toward testing "silly" null hypotheses (Johnson 1999, Anderson et al. 2001), I eliminated all land-cover classes that were obviously unsuitable to BACS. Instead, I focused selection of sample sites only on those areas classed as either pine or shrub-scrub (Dunning 1993). There were approximately 304,000 ha of pine and shrub-scrub habitat within the lower FRB. I excluded sites that occurred within 250 m of a site where BACS were predicted to occur by the MHM and randomly chose 24 of the remaining sites (all >100 ha) for evaluation efforts. I considered this set of sites as representing a simple habitat model (SHM) for determining presence of BACS. Because the SHM excluded obviously unsuitable habitats, I expected errors of omission in the MHM to be liberal (i.e., I expected greater errors of omission because obviously unsuitable habitats were unavailable for evaluation).

I visited all sites between 17 May and 23 June 2001. I ensured that all sampling for BACS presence occurred between dawn and 1,000 hrs, during periods with wind speed <8 km/hr and periods without precipitation. Upon arriving at a sample site, I listened for 10 min for BACS. If BACS were not detected during this initial listening phase, I played a tape recording of a BACS for 1 min and then listened for a response for 2 minutes (Johnson et al. 1981). I determined presence of BACS by visual or auditory detection during the monitoring period. If BACS were not detected, I considered them absent from the site. I made no effort to count singing birds.

I treated percentage of MHM sites with no BACS detections as an estimate of the error of commission and percentage of SHM sites with BACS present as an estimate of the error of omission. I used 2×2 contingency table and a χ^2 test to determine whether presence of BACS was independent of MHM and SHM sites (Zar 1996). I calculated kappa (Lillesand and Kiefer 1994) as a measure of the MHM performance relative to the SHM. Because most habitat models are compared against null (i.e., random) models during the validation process (e.g., site), I interpreted the results of the χ^2 test and the kappa estimate as being conservative relative to values that would have been obtained from a comparison of MHM to a null model.

Test results

I detected BACS at 17 of the 24 MHM sites (error of commission = 29.2%), and at 7 of 24 SHM sites (error of omission = 29.2%). Presence of BACS was dependent ($\chi^2_1 = 8.33$, $P < 0.01$) on model predictions. Kappa was 0.42, indicating that MHM predictions were 42% better than SHM predictions.

Discussion and management implications

Using satellite imagery to develop a model of species occurrence is limited to special applications. If a goal of model development is to better understand the habitat requirements of a species (Kopp et al. 1998, Miller et al. 2000), the approach suggested here is not appropriate. However, if the goal of the modeling effort is to identify sites where a species might occur (Scott et al. 1993), developing models from unclassified imagery could be beneficial.

Success of an image-based approach to locate isolated populations will be based on several factors. First, the species of interest should be a habitat specialist. If the species is a habitat generalist, representative sites used to develop the model will have pixels with great variability in reflectance as a result of the variation among suitable habitat patches. This will result in many patches being classified as suitable. Although such a model might indeed be correct, it would be of little use because it would predict a ubiquitous distribution of the species. However, this is an unlikely problem because locating populations of a habitat generalist is usually of little conservation concern.

Second, spatial resolution associated with animal needs and spatial resolution of satellite imagery should be reconciled. Habitat suitability is influenced by habitat area (MacArthur and Wilson 1967, Martin 1980, Blake and Karr 1987). For most wildlife species, the spatial resolution of satellite imagery is likely greater than needed (e.g., few wildlife populations can persist in a 30×30-m area). This requires the biologist to determine a minimum relevant habitat block and thus rely on published information and experience to determine the minimum area deemed suitable. One possible solution to this issue is to assess model validity on the largest blocks available and then attempt to determine the relationship between area and habitat occupancy.

Third, the estimation of a mean vector and covariance matrix from the reflectance values requires sites where the species of interest is known to occur (Clark et al. 1993). In the absence of such sites, biologists must rely on published information and experience, necessitating a more traditional approach to the modeling process (e.g., using land-use classification and natural history of the species). In some cases, however, it may be possible to use historic data on species presence in conjunction with historic satellite data to develop a reasonable model. Unfortunately, this latter approach could be profoundly affected by land-use changes occurring after image acquisition.

Finally, habitat features of importance must be visible from above. For example, if a species requires habitat features that exist entirely underneath a dense vegetation canopy, a satellite image would be unable to detect such features. In this case, the modeling approach suggested here could at best be expected to locate dense canopy, which could be of little discriminatory value.

As with all modeling efforts, habitat models based on satellite imagery should be evaluated before they are implemented (Marcot et al. 1984). This evaluation is relatively simple when the species of interest is not particularly rare, as was the case with BACS. However, a model developed for an endangered species might perform well, yet test results could indicate poor performance because the species was lost from that particular area. When a species is too rare for proper model evaluation, biologists will be forced to compare vegetation at training sites to vegetation at predicted sites to evaluate the model.

The application of the model to BACS within the lower FRB provided an example of potential advantages of this approach over developing a typical land-cover classification and using habitat association to identify isolated populations. Model validation results indicated that MHM correctly predicted BACS presence in 71% of cases. Moreover, the model suggested that only 10% of the pine and shrub-scrub habitats (i.e., 30,000 ha predicted suitable of 304,000 ha of pine or shrub-scrub as identified in the SHM) within the lower FRB represented suitable BACS habitat. Thus, greater monitoring effort would have been expended in unproductive habitats if a traditional land-cover classification had been used to create a simple habitat association model for BACS.

Acknowledgments. I thank D. D. Jones for assistance with validation. D. A. Miller and G. R. Conner provided valuable editorial assistance. I. Godbois, R. Gray, B. Plowman, and T. Story assisted with point counts. J. Brock provided technical support. The Georgia Department of Natural Resources, the Joseph W. Jones Ecological Research Center, and the Robert W. Woodruff Foundation provided financial support for this project. M. J. Chamberlain and 2 anonymous reviewers provided valuable technical and editorial advice.

Literature cited

Anderson, D. R., W. A. Link, D. H. Johnson, and K. P. Burnham. 2001. Suggestions for presenting the results of data analyses. Journal of Wildlife Management 65:373-378.

Blake, J. G., and J. R. Karr. 1987. Breeding birds of isolated woodlots: area and habitat relationships. Ecology 68:1724-1734.

Brennan, L. A., W. M. Block, and R. J. Gutierrez. 1986. The use of multivariate statistics for developing habitat suitability index models. Pages 177-182 *in* J. Verner, M. L. Morrision, and C. J. Ralph, editors. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. University of Wisconsin Press, Madison, USA.

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

Conner, L. M., and B. D. Leopold. 1998. A multivariate habitat model for female bobcats: a GIS approach. Proceedings of the Southeastern Association of Fish and Wildlife Agencies 52:232-243.

Dunning, J. B. 1993. Bachman's Sparrow (*Aimophila aestivalis*). Account No. 38 *in* A. Poole, P. Stettenheim, and F. Gills editors. The birds of North America. The American Ornithologists' Union, Washington D.C., and the Academy of Natural Sciences, Philadelphia, Pennsylvania, USA.

Environmental Systems Research Institute, Inc. 2001. ARC/INFO command references and user's guides 8.0, the geographic information system software. Redlands, California, USA.

ERDAS, Inc. 2001. ERDAS IMAGINE software and online help 8.5. Atlanta, Georgia, USA.

Hall, F. G., D. E. Strebel, J. E. Nickeson, and S. J. Goetz. 1991. Radiometric reflection: toward a common radiometric response among multidate multisensor images. Remote Sensing of Environment 35:11-27.

Johnson, D. H. 1999. The insignificance of statistical significance testing. Journal of Wildlife Management 63:763-772.

Johnson, R. R., B. T. Brown, L. T. Haight, and J. M. Simpson. 1981. Playback recordings as a special avian censusing technique. Studies in Avian Biology 6:68-75.

Jorgensen, E. E., S. Demarais, S. M. Sell, and S. P. Lerich. 1998. Modeling habitat suitability for small mammals in Chihuahuan Desert foothills of New Mexico. Journal of Wildlife Management 62:989-996.

Kopp, S. D., F. S. Guthery, N. D. Forrester, and W. E. Cohen. 1998. Habitat selection and modeling for northern bobwhites on subtropical rangeland. Journal of Wildlife Management 62: 884-895.

Lillesand, T. M., and R. W. Kiefer. 1994. Remote sensing and image interpretation.Third edition. Wiley and Sons, Inc., New York, New York, USA.

Lovallo, M. G. L. Storm, D. S. Klute, and W. M. Tzilkowski. 2001. Multivariate models of bobcat habitat suitability in Pennsylvania. Pages 4-17 *in* A.Woolf, C. K. Nielsen and R. D. Bluett, editors. Proceedings of a symposium on current bobcat research and implications for management. The Wildlife Society 2000 conference, Nashville,Tennessee, USA.

MacArthur, R. H., and E. O. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton, New Jersey, USA.

Marcot, B. G., M. G. Raphael, K. H. Berry. 1984. Monitoring wildlife habitat and validation of wildlife-habitat relationships models. Transactions of the North American Wildlife and Natural Resources Conference 48:315-329.

Martin, T. E. 1980. Diversity and abundance of spring migratory birds using habitat islands on the Great Plains. Condor 82: 430-439.

Miller, D. A., B. D. Leopold, G. A. Hurst, and P. D. Gerard. 2000. Habitat selection models for eastern wild turkeys in central Mississippi. Journal of Wildlife Management 64:765-776.

Ralph, C. J., S. Droege, and J. R. Sauer. 1995. Managing and monitoring birds using point counts: standards and applications. Pages 161-168 *in* C.J. Ralph, J. R. Sauer, and S. Droege, editors. Monitoring bird populations by point counts. United States Forest Service General Technical Report PSW-GTR-149.

Roseberry, J. L., and S. D. Sudkamp. 1998. Assessing the suitability of landscapes for northern bobwhite. Journal of Wildlife Management 62:895-902.

Scott, J. M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'Ercchia, T. C. Edwards, J. Ulliman, and R. G. Wright. 1993. Gap analysis: a geographic approach to protection of biological diversity. Wildlife Monographs 123.

Verbyla, D., and J. A. Litvaitis. 1989. Resampling methods for evaluating classification accuracy of wildlife habitat models. Environmental Management 13:783-787.

Zar, J. H. 1996. Biostatistical analysis. Third edition. Prentice Hall, Upper Saddle River, New Jersey, USA.

Mike Conner is an associate scientist at the Joseph W. Jones Ecological Research Center in Newton, Georgia and adjunct professor at the University of Georgia and Mississippi State University. He obtained his B.S. in natural resources management from the University of Tennessee at Martin and his M.S. and Ph.D. in wildlife ecology from Mississippi State University. His research interests include predator-prey relationships and land management influences on wildlife communities.

Associate editor: *Chamberlain*