GIS-BASED NICHE MODELING FOR MAPPING SPECIES' HABITAT

John T. Rotenberry,[1,3] Kristine L. Preston,[1] and Steven T. Knick[2]

[1]*Department of Biology and Center for Conservation Biology, University of California, Riverside, California 92521 USA*
[2]*Snake River Field Station, USGS Forest and Rangeland Ecosystem Science Center Boise, Idaho 82706 USA*

Abstract. Ecological "niche modeling" using presence-only locality data and large-scale environmental variables provides a powerful tool for identifying and mapping suitable habitat for species over large spatial extents. We describe a niche modeling approach that identifies a minimum (rather than an optimum) set of basic habitat requirements for a species, based on the assumption that constant environmental relationships in a species' distribution (i.e., variables that maintain a consistent value where the species occurs) are most likely to be associated with limiting factors. Environmental variables that take on a wide range of values where a species occurs are less informative because they do not limit a species' distribution, at least over the range of variation sampled. This approach is operationalized by partitioning Mahalanobis D^2 (standardized difference between values of a set of environmental variables for any point and mean values for those same variables calculated from all points at which a species was detected) into independent components. The smallest of these components represents the linear combination of variables with minimum variance; increasingly larger components represent larger variances and are increasingly less limiting. We illustrate this approach using the California Gnatcatcher (*Polioptila californica* Brewster) and provide SAS code to implement it.

Key words: California Gnatcatcher; geographical information systems; GIS; habitat relationships; Mahalanobis D^2; niche modeling; Polioptila californica; *principal-components analysis.*

INTRODUCTION

Spatially explicit habitat suitability models provide powerful tools for ecologists and conservation biologists (Scott et al. 2002). Improved Geographical Information Systems (GIS) software and digital environmental layers permit development of new modeling techniques that create multivariate species' "niche models" encompassing large geographic areas. These regional niche models incorporate hypotheses about a species' occurrence relative to various environmental variables that are available as GIS spatial layers. Digital environmental layers such as elevation, slope aspect, precipitation, temperature, soil type, land use, and especially vegetation type (often used as a surrogate for "habitat type") may be incorporated into regional niche models.

Such models can have direct relevance to the ecology and conservation of targeted species. First, most modeling techniques identify the relative "importance" of individual variables (or combinations of variables) in influencing the distribution of a species. Although these "importances" are often more statistical than biological, they nonetheless serve as working hypotheses that can guide further, perhaps more experimental, investigation, as well as assist in implementing and evaluating adaptive management decisions. Second, they provide a spatially

explicit assessment of habitat suitability. It is one thing to know what variables are important; knowing where the appropriate combination of variables occurs can be equally valuable. Third, if the model is robust, predictions about habitat suitability can be extended into areas where there is currently no information about the occurrence of a particular species. Such predictions may help to focus additional survey effort or guide the design of more efficient species' preserves (e.g., Raxworthy et al. 2003).

Ecological modelers are faced with several challenges in producing useful predictive models. For example, habitat suitability models typically are created using abundance, density, or presence–absence data collected during surveys for the species of interest (Guisan and Zimmerman 2000, Brotons et al. 2004). However, creation of models encompassing large geographic areas (such as a county, state, or even larger area) generally requires using multiple sources of data, often collected with different survey methodologies. Although large-scale databases for sensitive plant and animal species are available (e.g., government-supported endangered species databases, various regional- or state-based natural diversity databases, and museum collections' databases), these typically provide information on the presence of a target species at a point, but rarely document the absence of a species from a surveyed area. To further complicate matters, obtaining "true absence" data even with focused surveys can be problematic especially for species that are rare or difficult to detect (Knick and Rotenberry 1998, Dunn and Duncan 2000, Hirzel et al.

2002, Rotenberry et al. 2002). Another challenge to modeling is to predict a species' occurrence outside of the original study area or in a situation where the environment is undergoing change. In such cases, the particular combination of habitat characteristics present where the original data were collected may not exist (Knick and Rotenberry 1998, Rotenberry et al. 2002). To meet these challenges, new modeling techniques have been developed to create regional models that predict habitat suitability based solely on locations where a species is present, and that are relatively robust to the inadvertent inclusion of nonrelevant environmental variation (Clark et al. 1993, Knick and Rotenberry 1998, Dettmers and Bart 1999, Stockwell and Peters 1999, Dunn and Duncan 2000, Hirzel et al. 2002, Petersen et al. 2002, Rotenberry et al. 2002). Our objective is to make one of these techniques widely available and easily implementable.

Mahalanobis D^2

Concisely, Mahalanobis D^2 is simply the standardized difference between the values of a set of environmental variables for any point (or rasterized cell or pixel in a GIS layer) and the mean values for those same variables calculated from all points at which a species was detected (Clark et al. 1993, Dunn and Duncan 2000, Rotenberry et al. 2002, Browning et al. 2005). Thus, the more similar in environmental conditions a point is to the species' mean, the smaller the D^2 and the more "suitable" the habitat at that point:

$$D^2(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \qquad (1)$$

where **H** is "occupied habitat," an $n \times p$ matrix of p variables measured at n points where a species was detected; $\boldsymbol{\mu}$ is the $p \times 1$ vector of means based on **H** (i.e., the centroid); and **y** is the $p \times 1$ vector of measurements on any point (it may or may not be taken from **H**). Thus, $\mathbf{y} - \boldsymbol{\mu}$ is a vector of deviations of a point from a species' mean vector; $\boldsymbol{\Sigma}$ is the $p \times p$ variance–covariance matrix based on **H**; and D^2 is a squared scalar distance, standardized in the $\boldsymbol{\Sigma}$ metric.

Because $D^2(\mathbf{y})$ approximately follows a $\chi^2_{(\mathbf{y})}$ distribution under multinormal assumptions, it can be rescaled to range from 0 to 1 (called a "p-value"). This rescaling is desirable, as D^2 values can otherwise range from near zero to infinity. These p-values may be interpreted as analogous to a posterior probability resulting from a Bayes discriminant function or logistic regression (Dunn and Duncan 2000).

Use of D^2 to characterize a species' habitat relationship assumes that the original sample reflects the optimal habitat distribution of the animals in the sampled landscape. As a corollary, it assumes that the selection response has been fully characterized (at least in the vicinity of the mean), or in other words, that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ fully characterize the species' response to habitat. This implies two additional assumptions: (1) the sampled area contains the full range of habitat variation to which the

species responds, and (2) we have identified and measured the appropriate variables (i.e., we have not left out any that are important, and we have not included any that are irrelevant). These assumptions are not always justified. Although D^2 performs quite well in many circumstances (e.g., Clark et al. 1993, Knick and Dyer 1997), it may perform poorly when applied to areas not included in the original sample or if applied to landscapes that exhibit nonstationarity in space or time, such as those that are prone to disturbance (whether natural or anthropogenic) or are undergoing restoration or succession (Knick and Rotenberry 1998, Rotenberry et al. 2002).

Partitioning Mahalanobis D^2

Modeling techniques based on dissimilarity to an *optimum* configuration may not be ideal for predicting animal occurrence because of the uncertainty associated with defining a biological optimum from distributional data. Instead, identifying a *minimum* set of basic habitat requirements for a species may be more appropriate for predicting potential animal use in changing environments (Dunn and Duncan 2000, Rotenberry et al. 2002).

The performance of D^2 is improved by "partitioning" it into separate components, each representing an independent set of relationships between a species' distribution and environmental variables (Dunn and Duncan 2000, Rotenberry et al. 2002). Partitioned D^2 for any point **y** is given as

$$D^2(\mathbf{y}) = \sum_{j=1}^{p} d_j^2 / \lambda_j \qquad (2a)$$

where $\lambda_1 \geq \ldots \lambda_k \ldots \geq \lambda_p$ are the eigenvalues of $\boldsymbol{\Sigma}$, and

$$d_j = (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\alpha}_j$$

where **y** and $\boldsymbol{\mu}$ are as previously defined, and $\boldsymbol{\alpha}_j$ is the eigenvector associated with λ_j. This result arises from the spectral decomposition of $\boldsymbol{\Sigma}$ (e.g., Seber 1984; see Rotenberry et al. 2002).

Alternatively,

$$D^2(\mathbf{y}) = d_1^2 / \lambda_1 + \ldots + d_k^2 / \lambda_k + \ldots + d_p^2 / \lambda_p. \qquad (2b)$$

These distance partitions are additive, and each is associated with an eigenvalue and eigenvector arising from a principal-components analysis (PCA) of the data set **H** containing the values of the environmental variables from the points at which the species occurred. Unlike regular PCA, however, biological significance is attached to those components with the *smallest*, rather than the largest, eigenvalues (which in PCA are measures of variance). Dunn and Duncan (2000) and Rotenberry et al. (2002) show the relationship between the partition with the smallest eigenvalue and Pearson's "plane of closest fit," that plane for which the sums of squares of the perpendiculars from a set of points to the plane is a minimum (Pearson 1901). The variance of these projections of points on a vector normal to such a plane will be a minimum, the same as the variance of
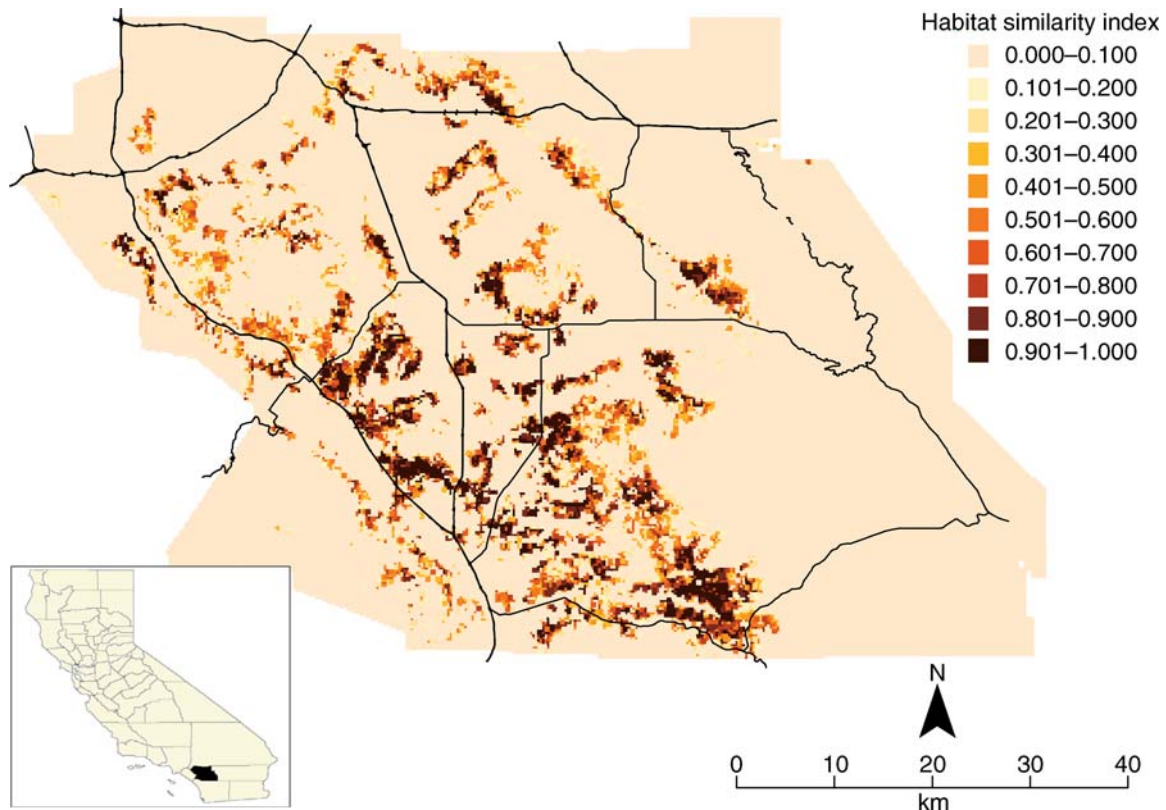
FIG. 1.　Full-rank ecological niche model [$D^2(\mathbf{y}; 1)$] for California Gnatcatchers based on 21 environmental variables. Habitat similarity to known, occupied locations increases as the color darkens.

points projected onto an axis defined by the eigenvector associated with the smallest eigenvalue. Emphasizing the plane(s) associated with the smallest eigenvalue(s) is based on the notion (consistent with the idea of the species' niche; Hutchinson 1957, Pulliam 2000) that we want to identify the *constant relationships* in a species' distribution. These *constant relationships* are defined by those variables that maintain a *consistent value* where the species occurs (and as a consequence have low variance), and thus are those most likely to be associated with limiting factors. Environmental variables that take on a wide range of values where a species occurs (and which will therefore be associated with principal components with larger eigenvalues) are less likely to be informative because they do not appear to limit a species' distribution, at least over the range of variation sampled. Axes, or partitions of D^2, associated with increasingly larger eigenvalues represent combinations of variables that are increasingly less consistent (i.e., more variable) where a species occurs. This concept has also been presented by Collins (1983) and Knopf et al. (1990), although in substantially different forms.

Not all of the p components of $D^2(\mathbf{y})$ as partitioned above define limiting combinations of habitat variables. Some $p - k$ of these do not define habitat suitability, but rather are included in $D^2(\mathbf{y})$ simply because the

investigator decided a priori to measure p habitat variables. Certainly the first principal component cannot be considered a limitation, because its variance is λ_1, the maximum possible. Thus, habitat suitability for a p-dimensional $\mathbf{y}$ is

$$D^2(\mathbf{y}; k) = d_k^2/\lambda_k + ... + d_p^2/\lambda_p \qquad (3)$$

for some $1 \leq k \leq p$. Accordingly, suitability of a particular habitat configuration $\mathbf{y}$ for a species at a location would be measured in terms of the location's deviations from k basic requirements for that species, to the extent that we are able to know k.

The partitioned D^2's can be considered sequentially, beginning with that associated with the single smallest eigenvalue, then the two smallest, the three smallest, and so on. If we add all of the partitions together, we have the original D^2 model.

The choice of k is likely to be somewhat qualitative. Dunn and Duncan (2000) suggest that one examine the magnitude and relative spacings among the eigenvalues, the interpretability of the partitions, and the credibility of predicted-use areas that result from particular choices of k. In this respect, use of partitioned D^2 does not differ from other principal-components applications where interpretability often dictates the choice of the number of "ecologically significant" dimensions. Browning et al. (2005) used bootstrapping and internal cross-validation
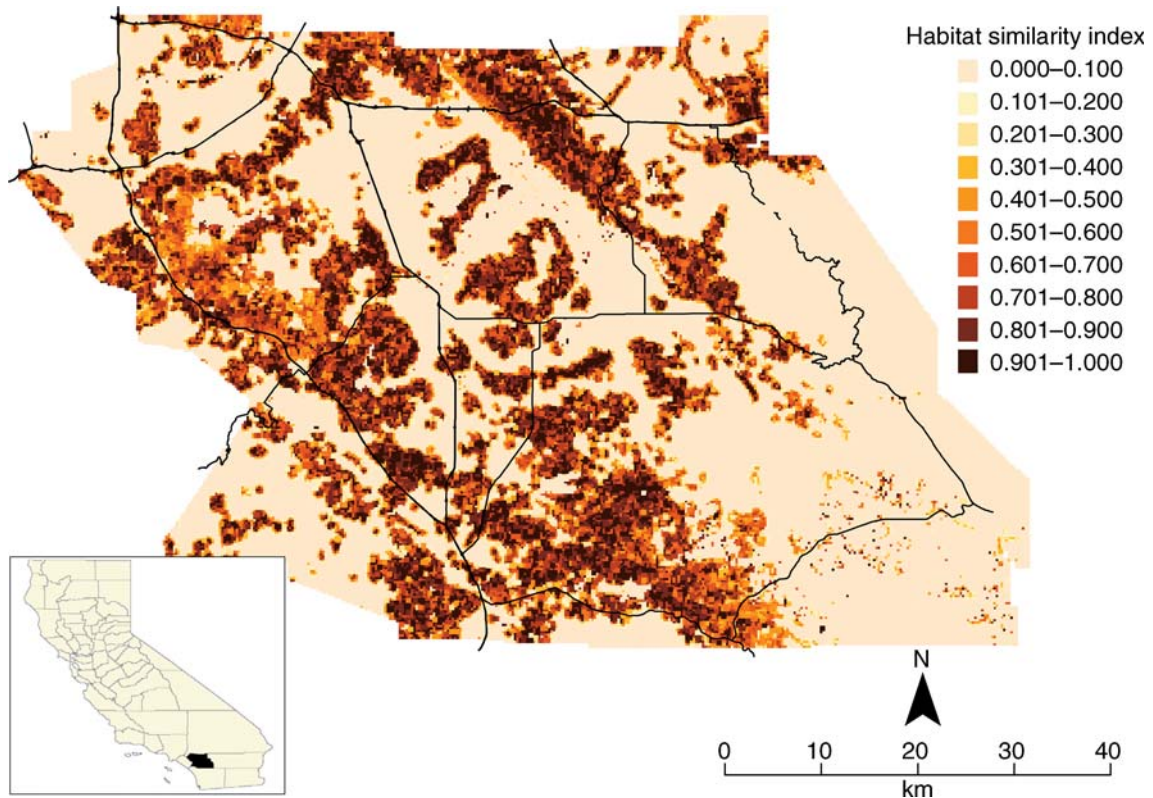
Fig. 2. Reduced-rank ecological niche model for California Gnatcatchers using same variables as in Fig. 1, but using the smallest partitioned D^2 [$D^2(\mathbf{y}; 21)$]. Habitat similarity to known, occupied locations increases as the color darkens.

(Efron and Tibshirani 1993) to support examination of map credibility in selecting k in an analysis of den site selection in timber rattlesnakes (*Crotalus horridus*). Alternatively, if one has external, independent validation data, one can examine the predictive value obtained by successive changes in k. We will provide an example of this approach. An additional consideration is that increasing the number of partitions used results in an increasingly spatially restrictive model, analogous to increasing the fit (and potentially reducing the generality) of a multiple regression model by adding variables. When all of the partitions are added up, of course, it yields the original D^2 model [$D^2(\mathbf{y}; 1)$ in this notation] that Knick and Rotenberry (1998) criticized as overly restrictive.

Assessing which environmental variables are associated with likelihood of occurrence is based on examination of the PCA's eigenvector values associated with each partition of D^2; variables with larger absolute eigenvector values are considered more "important" (Dunn and Duncan 2000). Variables with small absolute eigenvector values may vary considerably, while allowing habitat to remain close to the plane of closest fit. Moreover, if two eigenvector values are large in absolute value, the variables effects will tend to cancel if they have opposite signs, and the habitat will still remain close to the plane. This is a perceived advantage of this modeling approach because it allows the possibility of

detecting that a species can make a trade-off, balancing different habitat values yet still maintaining habitat utility (Dunn and Duncan 2000). For these interpretations to be valid, environmental variables should be in identical units. This is effectively achieved by performing the PCA on a correlation matrix (i.e., a variance–covariance matrix of standardized variables). A major advantage of partitioning is that less distributionally relevant variables (assuming that some are unknowingly included in the original variable set) are shifted to components with larger eigenvalues, and thus may not contribute to the final, reduced-rank model (Rotenberry et al. 2002). As with selecting k, there is no simple numerical or statistical criterion for distinguishing "important" from "unimportant" eigenvector values. However, in practice there often appears to be a sharp demarcation between zero (or near-zero) and nonzero values (Dunn and Duncan 2000).

Once a satisfactory model is obtained for a species, it may be used to calculate a "*p*-value" (representing habitat similarity on an increasing 0–1 scale, with 1 representing environmental conditions identical to the species' mean) for any point for which one has values for the environmental variables:

$$p\text{-value for } D^2(\mathbf{y}; k) = 1 - \text{prob}(\chi^2_{(p+1-k)}) \qquad (4)$$

where k may range from p to 1, the latter representing a

PLATE 1. Female California Gnatcatcher (*Polioptila californica*) at the University of California's Motte Rimrock Reserve, near Riverside, California, USA. Photo credit: Mark A. Chappell.

TABLE 1. Results of PCA on a correlation matrix of 21 environmental variables assessed at 566 points where California Gnatcatchers were detected in western Riverside County, California, USA.

Principal component (k)	Eigenvalue	Proportion of total variance
21	0.018	0.001
20	0.111	0.005
19	0.213	0.010
18	0.296	0.014
17	0.347	0.017
16	0.418	0.020
15	0.445	0.021
14	0.487	0.023
13	0.723	0.034
12	0.774	0.037
11	0.783	0.037
10	0.861	0.041
9	0.947	0.045
8	1.064	0.051
7	1.172	0.056
6	1.269	0.060
5	1.300	0.062
4	1.498	0.071
3	1.776	0.085
2	2.192	0.104
1	4.308	0.205

Note: Entries are arranged by increasing magnitude of eigenvalues.

full-rank model, or simply D^2 (Clark et al. 1993). The degrees of freedom is the number of principal components used in the partitioned $D^2(\mathbf{y}; k)$ calculation, which for k is $p + 1 - k$. Until this is validated with more extensive analyses, we do not advise treating these *p*-values in a statistical inference context. In most cases, *p*-values will be calculated and mapped for all points in a landscape of interest (Rotenberry et al. 2002, Browning et al. 2005), and/or for a set of independently derived validation points (i.e., points where the target species was detected, but that were not included in creating the model; Guisan and Zimmermann [2000]). One expects high *p*-values at known occupied points; however, because distances from the centroid are scaled by the sample variance, inevitably those occupied points most distant from the centroid will have low *p*-values.

The robustness of any principal-components analysis is affected by both the number of observations (which influences the robustness of the correlations) and the ratio of observations to variables (a low ratio leads to "overfitting" and reduces generalizability) (Osborne and Costello 2004). We suggest a minimum of 40 observations and an observations : variables ratio of 10:1. However, it may be challenging to obtain many observations for a rare or difficult-to-detect species. In such a case, Browning et al. (2005) employed resampling

TABLE 2. Eigenvector values associated with the 21st component ($k = 21$) resulting from PCA performed on a correlation matrix of 21 environmental variables assessed at 566 independent points where California Gnatcatchers were detected in western Riverside County, California, USA.

Environmental variable	Description	Eigenvector$_{21}$ value
ELEV	median elevation for 8×8 pixel neighborhood (m)	0.028
EAST_0	eastness = median sin(aspect) for 8×8 pixel neighborhood	0.010
NORTH_0	northness = median cos(aspect) for 8×8 pixel neighborhood	−0.002
SLOPE	median slope (%) for 8×8 pixel neighborhood	−0.017
PRECIP	median annual precipitation (mm)	−0.012
MINTJAN	minimum mean annual temperature (°F)	0.015
MAXTJUL	maximum mean annual temperature (°F)	−0.004
CSS_AREA	size (m^2) of coastal sage scrub patch closest to or at the point	0.028
FIN_DCSS	distance (m) from point to closest patch of coastal sage scrub (value is zero when point is within a coastal sage scrub patch)	0.057
DIST2_WAT	distance squared (m^2) from point to nearest body of open water	−0.001
CL1GS	sum of pixels classified as "agriculture" within 8×8 pixel neighborhood	0.226
CL2GS	sum of pixels classified as "developed" within 8×8 pixel neighborhood	0.178
CL3GS	sum of pixels classified as "riparian" within 8×8 pixel neighborhood	0.104
CL4GS	sum of pixels classified as "coastal sage scrub" within 8×8 pixel neighborhood	0.662
CL5GS	sum of pixels classified as "chaparral" within 8×8 pixel neighborhood	0.459
CL6GS	sum of pixels classified as "nonnative grassland" within 8×8 pixel neighborhood	0.497
CL7GS	sum of pixels classified as "woodlands" within 8×8 pixel neighborhood	0.054
EDGE	amount of edge (m) between developed and natural habitats within 75×75 pixel neighborhood	−0.013
PERGRA	percentage of pixels classified as "nonnative grassland" within 75×75 pixel neighborhood	−0.016
PERCSS	percentage of pixels classified as "coastal sage scrub" within 75×75 pixel neighborhood	−0.031
PERDEV	percentage of pixels classified as "developed" within 75×75 pixel neighborhood	0.000

Note: Spatial uncertainty is <200 m for the independent points where California Gnatcatchers were detected; pixel size is 30×30 m.

techniques (bootstrap and cross-validation) to investigate the stability of the correlation matrix, and to determine the influence of individual observations on $D^2(k)$ values. Decisions were made to eliminate certain PCs from further consideration, based on bootstrap and cross-validation results.

It is easy to calculate D^2, $D^2(\mathbf{y}; k)$, and their *p*-values. First, one performs a PCA on a correlation matrix derived from *p* environmental variables measured at *n* points (observations) where the species was detected. The resulting eigenvalues and eigenvectors are examined to (1) determine an appropriate *k*, and (2) interpret which variables contribute to which eigenvectors. The eigenvalues and eigenvectors are then used to compute $D^2(k)$ for $k = 1, ..., l$ (Eq. 2) where *l* is the number of nonzero eigenvalues (usually *p*, the number of habitat variables). $D^2(k)$ can be computed for any point at which the *p* environmental variables have been measured, including (1) the original observations where the species was detected, (2) a validation data set, or (3) all of the points in a study region. Any $D^2(k)$ can be converted to a *p*-value (Eq. 4). We provide a complete implementation of this process in SAS (Duncan and Dunn 2001, SAS Institute 2001; see Supplement).

An example

We show two niche models for California Gnatcatchers (*Polioptila californica* Brewster; see Plate 1) calculated for a roughly 490 800-ha area of western Riverside County, California, USA (Figs. 1 and 2). Both are based on an analysis of 21 environmental variables (several climatic and topographic variables, distances to certain landscape elements, plus proportional coverages of major vegetation types within a 250×250 m or 2250×2250 m neighborhood around a point) assessed at 566 points (80%) that were randomly selected from an original 706 spatially independent locations where gnatcatchers had been detected. One is a full-rank model based on *p*-values from the total D^2 (Fig. 1); the other is a reduced-rank model using the smallest partition of D^2, $D^2(\mathbf{y}; 21)$ (Fig. 2). Beginning with the smallest partition, we evaluated the relative change in eigenvalues between each partition and the next (Table 1). Examining the distribution of eigenvalues suggested that, at least as a preliminary cut, retaining only the last eigenvector might produce a satisfactory reduced-rank model. Other potential "breaks" in the distribution appear with eigenvalues associated with the 20th [$D^2(\mathbf{y}; 20)$] and the 14th [$D^2(\mathbf{y}; 14)$] components (Table 1). Environmental variables with relatively high absolute eigenvector values on the 21st component are CL4GS, CL5GS, and CL6GS (Table 2). We interpret these variables as defining the most "suitable" habitat for gnatcatchers in this region, and these are consistent with general observations on the distribution of California Gnatcatchers in southern California (Atwood and Bontrager 2001). Note that the increased precision (reduced generality) of the full-rank model is manifest in the identification of less area as potentially "suitable" (compare Fig. 1 with Fig. 2). The reduced-rank model using only the 21st component [i.e., $D^2(\mathbf{y}; 21)$] scored the independent validation data set quite well, yielding a median *p*-value of 0.828 for 140 points. Other suggested reduced-rank models did not perform quite as well; $D^2(\mathbf{y}; 20)$ yielded a median *p*-value

of 0.734 for the validation points, and $D^2(\mathbf{y}; 14)$ generated a median of 0.619.

Conclusions

To conclude, we concur with Raxworthy et al. (2003) that ecological niche modeling using readily available occurrence records coupled with GIS-based environmental data can provide a potentially powerful tool to assess spatial patterns of species' distributions at regional scales. We can compare among species to determine which environmental drivers appear most important in predicting a particular species distribution, and can assess the extent to which multiple species share common responses. A greater understanding of habitat relationships manifest at the regional scale has two additional direct benefits. First, it will allow us to design and refine future monitoring and sampling strategies to further test habitat relationships and to see how populations respond to changing environmental conditions (natural and anthropogenic). Second, it will permit identification of localities with potentially high conservation value for a target species, which can then be prioritized for acquisition and integration into a reserve network.

Literature Cited

Atwood, J. L., and D. R. Bontrager. 2001. California Gnatcatcher (*Polioptila californica*). Number 574 *in* A. Poole and F. Gill, editors. The birds of North America. Philadelphia, Pennsylvania, USA.

Brotons, L., W. Thuiller, M. B. Araújo, and A. H. Hirzel. 2004. Presence–absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27:437–448.

Browning, D. M., S. J. Beaupré, and L. Duncan. 2005. Using partitioned Mahalanobis $D^2(K)$ to formulate a GIS-based model of timber rattlesnake hibernacula. Journal of Wildlife Management 69:33–44.

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.

Collins, S. L. 1983. Geographical variation in habitat structure of the Black-throated Green Warbler (*Dendroica virens*). Auk 100:382–389.

Dettmers, R., and J. Bart. 1999. A GIS modeling method applied to predicting forest songbird habitat. Ecological Applications 9:152–163.

Duncan, L., and J. E. Dunn. 2001. Partitioned Mahalanobis D^2 to improve GIS classification. Proceedings of the SAS Users Group International Number 26, Paper 198–26. SAS Institute, Cary, North Carolina, USA.

Dunn, J. E., and L. Duncan. 2000. Partitioning Mahalanobis D^2 to sharpen GIS classification. Pages 195–204 *in* C. A. Brebbia and P. Pascolo, editors. Management information systems 2000: GIS and remote sensing. WIT Press, Southampton, UK.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, London, UK.

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

Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology 83:2027–2036.

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor Symposium on Quantitative Biology 22:415–427.

Knick, S. T., and D. L. Dyer. 1997. Spatial distribution of black-tailed jackrabbit habitat determined by GIS in southwestern Idaho. Journal of Wildlife Management 61:75–85.

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.

Knopf, F. L., J. A. Sedgwick, and D. B. Inkley. 1990. Regional correspondence among shrubsteppe bird habitats. Condor 92:45–53.

Osborne, J. W., and A. B. Costello. 2004. Sample size and subject to item ratio in principal components analysis. Practical Assessment, Research and Evaluation 9(11). ⟨http://PAREonline.net/getvn.asp?v=9&n=11⟩

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2:559–572.

Peterson, A. T., D. R. B. Stockwell, and D. A. Kluza. 2002. Distributional prediction based on ecological niche modeling of primary occurrence data. Pages 617–623 *in* J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: issues of accuracy and scale. Island Press, Washington, D.C., USA.

Pulliam, H. R. 2000. On the relationship between niche and distribution. Ecology Letters 3:349–361.

Raxworthy, C. J., E. Martinez-Meyer, N. Horning, R.A. Nussbaum, G. E. Schneider, M. A. Ortega-Huerta, and A. T. Peterson. 2003. Predicting distributions of known and unknown reptile species in Madagascar. Nature 426:837–841.

Rotenberry, J. T., S. T. Knick, and J. E. Dunn. 2002. A minimalist approach to mapping species' habitat: Pearson's planes of closest fit. Pages 281–289 *in* J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrences: issues of accuracy and scale. Island Press, Washington, D.C., USA.

SAS Institute. 2001. The SAS System for Windows. Version 8.02. SAS Institute, Cary, North Carolina, USA.

Scott, J. M., P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, , and F. B. Samson, editors. 2002. Predicting species occurrences: issues of accuracy and scale. Island Press, Washington, D.C., USA.

Seber, G. A. F. 1984. Multivariate observations. John Wiley, New York, New York, USA.

Stockwell, D. R. B., and D. P. Peters. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. International Journal of Geographical Information Systems 13:143–158.

SUPPLEMENT

SAS code to produce Mahalanobis D^2 and its partitions (*Ecological Archives* E087-084-S1).