

2001

Modeling Reptile and Amphibian Range Distributions from Species Occurrences and Landscape Variables

Geoffrey M. Henebry

University of Nebraska-Lincoln, geoffrey.henebry@sdstate.edu

Brian C. Putz

University of Nebraska - Lincoln

James W. Merchant

University of Nebraska - Lincoln, jmerchant1@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/usgspubs>



Part of the [Earth Sciences Commons](#)

Henebry, Geoffrey M.; Putz, Brian C.; and Merchant, James W., "Modeling Reptile and Amphibian Range Distributions from Species Occurrences and Landscape Variables" (2001). *Publications of the US Geological Survey*. 30.

<https://digitalcommons.unl.edu/usgspubs/30>

This Article is brought to you for free and open access by the US Geological Survey at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications of the US Geological Survey by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



| [GAP home](#) | [USGS home](#) |

Volume No. 10, 2001

Animal Modeling

Modeling Reptile and Amphibian Range Distributions from Species Occurrences and Landscape Variables

Geoffrey M. Henebry^{1,2}, Brian C. Putz¹, and James W. Merchant^{1,2}

¹Center for Advanced Land Management Information Technologies (CALMIT), University of Nebraska, Lincoln

²School for Natural Resource Sciences (SNRS), Institute for Agriculture and Natural Resources (IANR), University of Nebraska, Lincoln

Introduction

An international symposium in October 1999 demonstrated the state of the art in modeling species occurrences (Scott et al. 2001).

One clear message from the symposium was the broad diversity of approaches that constitute the state of the art.

No single method excels, largely because of the very particular and local nature of the problem. Organisms both influence and respond to their local environment; thus, the same species may key in on different resources in different landscapes. Furthermore, modeling methods vary widely in their "transparency," which can inhibit transportability or robustness.

In order to provide an analytical modeling framework that is transparent and durable, we have chosen to use recursive partitioning methods to develop "objective" semi-empirical models of wildlife-habitat relationships for the Nebraska Gap Analysis Project. Recursive partitioning aims to predict membership of individual cases (here, species occurrences) in classes of a categorical dependent variable from measurements of one or several independent variables (here, land cover, soils, climate, etc.). The motivation for using this strategy is twofold: (1) the resulting trees of decision points and values that form the models are readily understandable, debatable, and tunable; and (2) its non-parametric modeling handles the multimodality likely to be found in species occurrence data.

A recent review (Guisan and Zimmerman 2000) notes that although dichotomous trees are commonly employed in systematic biology for keys to species identification, regression techniques to generate these trees have rarely been used to model occurrences of vertebrate species. Several recent papers have used CART (Classification and Regression Trees: Breiman et al. 1984) to develop habitat models. Iverson and Prasad (1998) used CART models to predict tree species distributions under climate change scenarios. Rejwan et al. (1999) used CART to model smallmouth bass (*Micropterus dolomieu*) habitat. McKenzie et al. (2000) used CART to estimate regional fire return intervals across the Columbia River Basin from local data sets. De'ath and Fabricius (2000) provided a tutorial of CART modeling using habitat relationships of soft coral taxa in Australia. Anderson et al. (2000) used CART to develop a habitat model for the desert tortoise (*Gopherus agassizii*). They found that the CART method could handle complicated interactions between variables that stem from spatial autocorrelations and spatial associations. They argued that while the CART model was phenomenological and not mechanistic, it provided valuable insight into the organism's habitat requirements and laid the foundation for further studies.

A drawback of the CART algorithm is computational complexity and thus computer time. A recent improvement on the CART algorithm is QUEST (Quick, Unbiased, and Efficient Statistical Trees: Loh and Shih 1997), which greatly speeds up searching of the data space and which is more robust in the face of categorical variables with many levels. A comparative study of 33 classification algorithms has shown that QUEST ably combines speed with accuracy (Lim et al. 2000).

Amphibians and reptile occurrence data were used to develop, test, and refine objective semi-empirical models. The paper illustrates the modeling procedure, the model tree and resulting range distribution for an amphibian species (*Eumeces multivirgatus*), and discusses the weaknesses and strengths of the framework.

Data

Numerous environmental variables were calculated and tessellated statewide using a hexagonal coverage produced by the EPA EMAP program. The resolution of the hexagons is approximately 40 km² within Nebraska. Each variable was rescaled from a raster format (30 m or 1500 m) to the coarser "modeling" hexagonal coverage by performing calculations within each unique hexagon. The variables were expressed as a percent composition, an average, a weighted average, or a categorical class.

Percent composition of land cover classes was derived from the Nebraska Gap Analysis Project land-cover data set (see Henebry et al. 2000). Soil data were derived from the Nebraska State Soil Geographic Database (STATSGO) and map. Soil texture groups were cross-walked into five classes: coarse, moderately coarse, medium, moderately fine, and fine. The previously mentioned data and hydric soils were then calculated as a percentage.

Terrain data used in the data set were calculated from United States Geological Society Digital Elevation Models (DEMs). Elevation averages were calculated within each hexagon. Slope data was divided into six percentage classes: 0-2, 2-5, 5-10, 10-15, 15-20, and >20. These classes

were expressed as a percent composition. A buffered stream data set was developed to create a binary class variable (presence/absence).

Climate data were acquired from weather stations throughout the state of Nebraska and selected stations from surrounding states. Means and coefficients of variation (CV%) were calculated for monthly average precipitation and monthly average, minimum, and maximum temperatures. Total average quarterly and growing season precipitation, growing degree days, and frost-free days were also calculated.

These data were submitted to a robust interpolation algorithm (nngridr; Watson 1994) and output as raster coverages. These data sets were then averaged within each modeling hexagon.

Voucher specimens of amphibians and reptiles collected in Nebraska since 1969 were obtained from the Nebraska State Museum and used for the occurrence data. Older legal descriptions were translated into latitude and longitude with a spatial accuracy of approximately one quarter-section (ca. 65 ha).

Methods

Voucher specimen data sets were queried from a database and converted to a point coverage ([Figure 1](#)). The observation points and modeling hexagonal coverage were intersected and the associated hexagon values attributed to the intersecting point coverage. Variables for each specimen point were submitted to the QUEST software program. An inversion for each species was developed from the output classification tree ([Figure 2](#)). Trimming of the classification tree leaves was done through a query of the modeling hexagonal coverage to determine appropriate tree splits for each species ([Figure 3](#)).

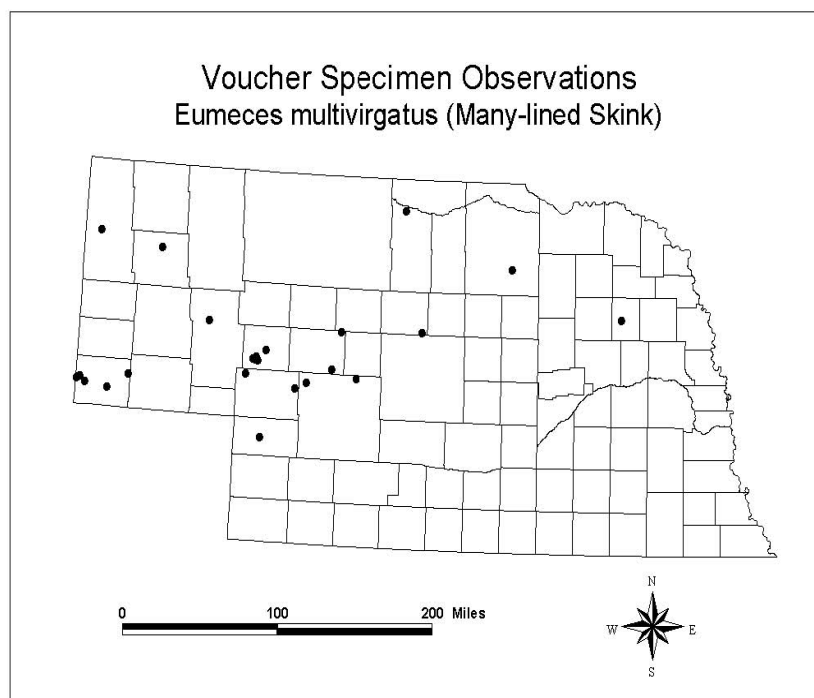


Figure 1. Occurrence data from georeferenced voucher specimens



Figure 2. Classification tree for three skink species in Nebraska

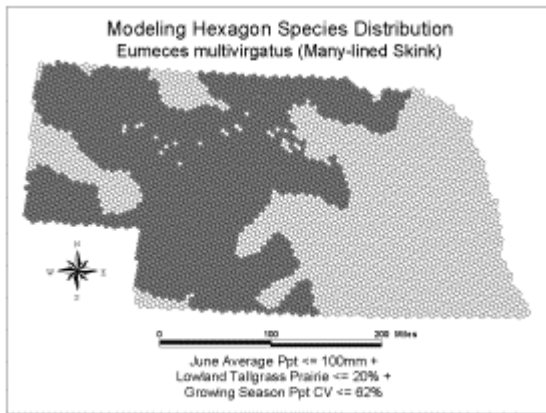


Figure 3. Model inversion produces the habitat distribution map

The queried modeling hexagons were intersected with a coarser resolution (ca. 650 km²) "reporting" hexagonal coverage. Percent probability was determined by the percent area of the modeling hexagons within each unique reporting hexagon. The reporting hexagonal coverage expresses the probability of finding suitable habitat within each particular hexagon (Figure 4).

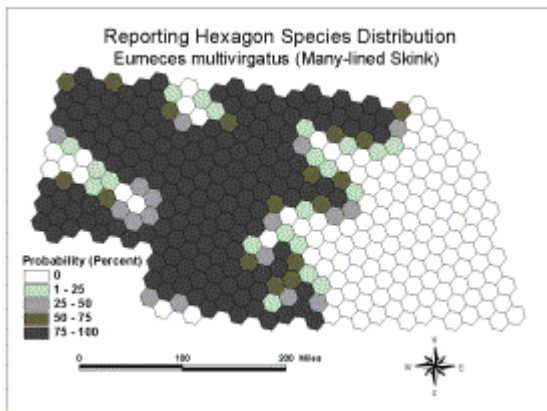


Figure 4. Probability of encountering species' modeled habitat

Discussion

The QUEST algorithm rapidly (within seconds) produced candidate models from groups of species occurrences, including model cross-validation calculations. The time-consuming step in the modeling process was trimming the leaves (or terminal nodes) to produce a model of sufficient generality and understandability. Recursive-partitioning algorithms allocate each occurrence to a terminal node. While this procedure can fit multimodal distributions, it can also lead to an overspecified model. Model refinement through leaf-trimming enables subjective ecological understanding to enhance the transparency and robustness of the model.

The models have frequently included temperature variability. The interannual variability (as CV%) of spring maximum and fall minimum temperatures enters into many of the models. This result is not surprising, given that reptiles and amphibians are ectotherms. Surficial soil texture, land cover, and proximity to streams are also important components of habitat. Elevation was found to be significant only for some snake species, and the number of frost-free days failed to provide any explanatory power. The models are undergoing expert review. Accuracy assessment will be conducted using other sources of occurrence data, including voucher specimens from other museums, data from theses and dissertations, species lists from natural areas, and county dot maps. Given the assumptions in the modeling methodology, we expect high but defensible rates of commission error and significantly lower rates of omission error.

These wildlife-habitat relationship models provide an objective framework from which to predict range distributions. They also provide a means through which to assess the gaps in knowledge about species habitat requirements, tolerances, and limits. Future work in modeling species occurrences and predicting range distributions must integrate the temporal dimension into geospatial data, but there are significant challenges in this task (Henebry and Merchant 2001).

Predicting species occurrences needs to be an iterative process that is performed periodically as new data, management tools, and policy objectives become available.

Literature Cited

Anderson, M.C., J.M. Watts, J.E. Freilich, S.R. Yool, G.I. Wakefield, J.F. McCauley, and P.B. Fahnestock. 2000. Regression-tree modeling of desert tortoise habitat in the central Mojave Desert. *Ecological Applications* 10(3):890-900.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. Classification and regression trees. Wadsworth and Brooks/Cole, Monterey, California. 358 pp.

De'ath, G., and K.E. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* 81:3178-3192.

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

Henebry, G.M., and J.W. Merchant. 2001. Geospatial data in time: Limits and prospects for predicting species occurrences. Pages 291-302 in Scott, J. M., P. J. Heglund, M. Morrison, editors. *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Covello, California.

Henebry, G.M., J.W. Merchant, J.W. Fischer, and D. Garrison. 2000. Expert review for land cover: Integrating information from specific comments and evaluating the results. *Gap Analysis Bulletin* 9:18-20.

Iverson, L.R., and A.M. Prasad. 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs* 68:465-485.

Lim, T.-S., W.-Y.Loh, and Y.-S. Shih. 2000. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning Journal* 40:203-228.

Loh, W.-Y., and Y.-S. Shih. 1997. Split selection methods for classification trees. *Statistica Sinica* 7:815-840.

McKenzie, D., D.L. Peterson, and J.K. Agee. 2000. Fire frequency in the interior Columbia River basin: Building regional models from fire history data. *Ecological Applications* 10:1497-1516.

Rejwan, C., N.C. Collins, L.J. Brunner, B.J. Shuter, and M.S. Ridgway. 1999. Tree regression analysis on the nesting habitat of smallmouth bass. *Ecology* 80:341-348.

Scott, J.M., P.J. Heglund, and M. Morrison, editors. 2001. *Predicting species occurrences: Issues of scale and accuracy*. Island Press, Covello, California. 868 pp.

Watson, D. 1994. nnggridr: An implementation of natural neighbor interpolation. David Watson, Claremont, Australia. 170 pp.