

reprinting/republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.

[http://www.ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=1026537](http://www.ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1026537)

## Use of Datasets Derived from Time-Series AVHRR Imagery as Surrogates for Land Cover Maps in Predicting Species' Distributions

Stephen L. Egbert<sup>1</sup>, Enrique Martínez-Meyer<sup>2</sup>, Miguel Ortega-Huerta<sup>2</sup>, and A. Townsend Peterson<sup>3</sup>

<sup>1</sup>Department of Geography and Kansas Applied Remote Sensing Program

<sup>2</sup>Department of Geography, Natural History Museum, and Kansas Applied Remote Sensing Program

<sup>3</sup>Natural History Museum and Biodiversity Research Center

University of Kansas

Lawrence, KS 66045-7612

Telephone: 785.864.7719 FAX: 785.864.0392

Email: [s-egbert@ku.edu](mailto:s-egbert@ku.edu)

**Abstract** – We hypothesized that NDVI time-series composite imagery or clustered data derived from the NDVI time series could serve as effective surrogates for land cover data in predictive modeling of species' ecological niches and potential geographic distributions. Using two Mexican bird species, we examined our hypothesis with GARP, the *Genetic Algorithm for Rule-set Prediction*. Inputs included topographic and climate data, as well as the NDVI and clustered NDVI datasets. We used a land cover map previously derived from the NDVI dataset for comparison testing. Considering only topographic factors, we found that the NDVI or clustered NDVI data performed as well as or better than the land cover data. When climate data were added, the land cover data performed better than the NDVI data, but improvements were slight.

### I. INTRODUCTION

The use of ecological niche modeling is increasingly important in such diverse areas as gap analysis, habitat preservation, and the monitoring of invasive plant and animal species [1]. This modeling typically proceeds by linking known species occurrences (generally from natural history museum data) with geo-registered environmental datasets such as elevation, precipitation, temperature, and land cover to create a model of a species' ecological niche. This niche model is then projected back onto the landscape to predict a potential geographic distribution for the species.

Although all environmental datasets are subject to weaknesses (e.g., time-series discontinuities, poor spatial distribution of measurement sites), land cover maps offer particular challenges in ecological niche modeling: (1) production of land cover maps is time-consuming and costly, (2) thematic accuracy of land cover maps, especially those derived from coarse-scale, wide-area imagery such as AVHRR, is generally either low or inadequately reported, and (3) land cover maps derived from satellite imagery frequently fail to map classes that are relevant and necessary for predicting distributions of particular species.

We examined the use of datasets derived from time-series NOAA AVHRR imagery over Mexico as a surrogate for land cover maps in ecological niche modeling. Our

underlying hypothesis was that one or more of the AVHRR-derived datasets would be at least as effective in modeling species distributions as a land cover map derived from the same imagery. If such were to prove to be the case, it may be possible to use AVHRR, MODIS, or similar imagery, either in raw form or as easily and cheaply derived datasets, as direct inputs to models that predict species' distributions.

### II. METHODS

In this pilot analysis, we selected two bird species with well-known distributions in Mexico: *Aphelocoma californica* and *Ergaticus ruber*. *A. californica* inhabits relatively arid regions with scrub oak or pine woodlands, while *Ergaticus ruber* is found in mountainous areas with pine or pine-oak forest.

Environmental datasets included three topographic datasets (elevation, slope, and aspect) and two climatic datasets (mean annual temperature and mean annual precipitation). Topographic data were from the Hydro 1K dataset obtained from the EROS Data Center, Sioux Falls, South Dakota [2]; climate data were obtained from CONABIO, Mexico City, Mexico [3]. For land cover, we used a land cover map of Mexico created previously by us using time-series AVHRR imagery and supervised classification [4] and two surrogate datasets derived from the same imagery: (1) seasonal maximum NDVI composite images for March, May-June, August-September, and December – four datasets in all, and (2) a clustered dataset created by applying ISODATA clustering to 12 monthly maximum NDVI composites. Grid cells measured 0.01° on a side.

Ecological niche modeling was performed using GARP, the *Genetic Algorithm for Rule-set Prediction*, which uses atomic rules, range rules, negated range rules, and logistic regression to develop rule-sets for predicting species' distributions [5]. We employed DesktopGarp software (a beta version developed at the University of Kansas Natural History Museum). GARP is designed to identify correlations

between known species occurrences and environmental parameters through an iterative process of rule selection, evaluation, testing, and incorporation or rejection.

In GARP, a method is first chosen from a set of possibilities (logistic regression, BIOCLIM rules, etc.), applied to the data, and a rule developed. Then, based on 1250 points resampled from the test dataset and 1250 points sampled randomly from the study region as a whole, predictive accuracy is calculated as the sum of points actually present that were predicted as present and those actually absent that were predicted as absent, divided by the total number of points in the map. Change in predictive accuracy from one iteration to the next is used to evaluate whether a particular rule should be incorporated into the model. The algorithm runs either 1000 iterations or until addition of rules has no appreciable effect on the accuracy measure (convergence) [6].

Two groups of experiments were run for each species. The first experiments included one control group using topographic variables only (elevation, slope, and aspect) and three experimental groups that combined the topographic variables with one of the NDVI datasets or the land cover map: topography plus NDVI composite images, topography plus clustered NDVI data, and topography plus land cover (Table 1).

The second set of experiments included a control group that combined topographic variables (elevation, slope, and aspect) and two climate variables (precipitation and temperature). Again, three experimental groups combined the topographic/climate variables with one of the NDVI datasets or the land cover map: topography and climate plus NDVI composite images, topography and climate plus clustered NDVI data, and topography and climate plus land cover (Table 2). For each experiment, we generated 100 models per species.

To evaluate model performance, we used random 50-50 partitions of the available occurrence data to create data sets for model creation (training data) and model testing (test data). The independent test data were used to measure omission rates, and  $100 - \text{omission}$  was used as a measure of model quality. For each species, Kruskal-Wallis ANOVA tests were used to test differences between control and experimental groups and among the experimental groups.

**Table 1. Topographic Data Only + Imagery Datasets**

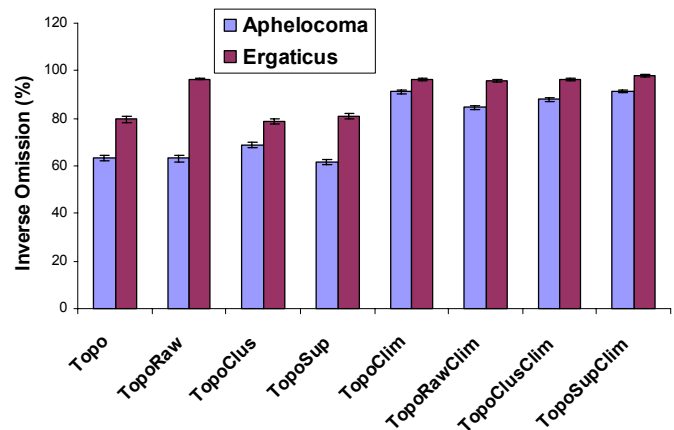
| Dataset Name | Dataset Contents  |
|--------------|---|
| Topo         | Topographic variables                                       |
| TopoRaw      | Topographic variables<br>Seasonal NDVI composite images (4) |
| TopoClus     | Topographic variables<br>Clustered 12-date NDVI data        |
| TopoSup      | Topographic variables<br>Land cover map classes             |

**Table 2. Topographic and Climate Data + Imagery Datasets**

| Dataset Name | Dataset Contents   |
|--------------|--|
| TopoClim     | Topographic variables<br>Climate variables                                       |
| TopoRawClim  | Topographic variables<br>Seasonal NDVI composite images (4)<br>Climate variables |
| TopoClusClim | Topographic variables<br>Clustered 12-date NDVI data<br>Climate variables        |
| TopoSupClim  | Topographic variables<br>Land cover map classes<br>Climate variables             |

### III. RESULTS AND DISCUSSION

Model results can be seen graphically in Figure 1. Models that included only topographic variables, in conjunction with NDVI or land cover data, are in the four groups of bars on the left side of the graph. Models that included both topography and climate variables, in conjunction with NDVI or land cover, are shown in the four groups of bars on the right side of the graph. Models including climate variables generally performed better than models including topography only. Furthermore, for models using both topography and climate, differences among the land cover and land cover surrogate (NDVI and clustered NDVI) datasets are minor.



**Figure 1. Results of Model Runs – Inverse Omission.** Bar heights indicate the inverse of the omission error. Error bars depict standard error. For explanation of model names (X axis), see Tables 1 and 2.

For *Aphelocoma californica*, the topography-only and the topography plus climate groups of experiments were significantly different (Kruskal-Wallis ANOVA:  $\chi^2 =$

[http://www.ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=1026537](http://www.ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1026537)

503.107, d.f. = 7,  $P < 0.001$ ; Figure 1). Comparing the experiments within the non-climate group, we found that topography plus clustered NDVI had statistically smaller omission (larger inverse omission) than other combinations (Kruskal-Wallis ANOVA:  $\chi^2 = 21.549$ , d.f. = 3,  $P < 0.001$ ). In contrast, in the topography plus climate group, topography plus climate alone and topography and climate plus land cover performed statistically better than topography and climate plus NDVI composite images, and topography and climate plus clustered NDVI data (Kruskal-Wallis ANOVA:  $\chi^2 = 29.535$ , d.f. = 3,  $P < 0.001$ ).

For *Ergaticus ruber*, we also found significant differences between the topography-only and the topography plus climate groups of experiments (Kruskal-Wallis ANOVA:  $\chi^2 = 433.234$ , d.f. = 7,  $P < 0.001$ ; Figure 1). As with *Aphelocoma californica*, models that included climate variables performed better.

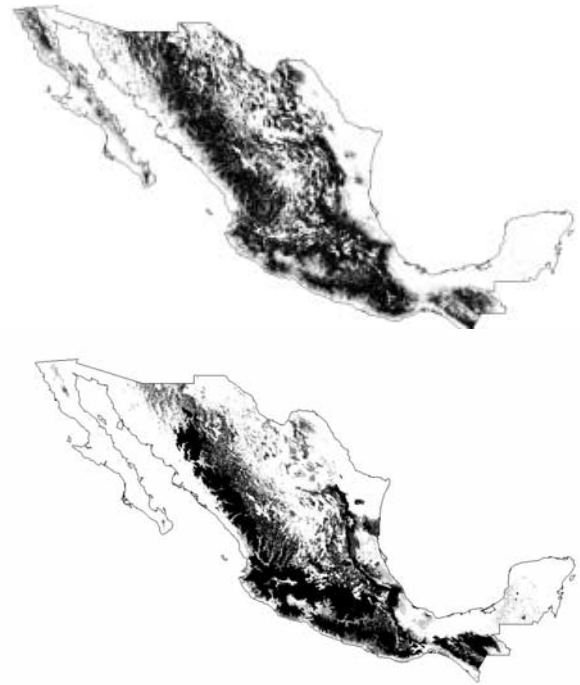
Examining experiments within the topo-only group, we found that the model that used topography plus NDVI composite images had statistically lower omission than the other experiments (Kruskal-Wallis ANOVA:  $\chi^2 = 167.788$ , d.f. = 3,  $P < 0.001$ ; Figure 1). On the other hand, in the topography plus climate group differences were more subtle (Kruskal-Wallis ANOVA:  $\chi^2 = 11.433$ , d.f. = 3,  $P < 0.05$ ): the model using land cover (TopoClim) performed slightly better than those using NDVI composite images and clustered NDVI data.

Spatial output from GARP also provides interesting insights. Since GARP generates distributional predictions based on random rule selection, and predictions vary from run to run, we generated 100 models per species. For visual comparisons, the 6 best-subsets models were selected from among these analyses based on omission-commission distributions [7]. Figure 2 compares one of the worst performing models (topography alone) with one of the best performing models (topography and climate plus clustered NDVI) for *Ergaticus ruber*. The visible difference in predicted distribution patterns graphically demonstrates the importance of good predictive datasets.

We plan to extend these analyses to broad suites of species in numerous regions of different scales and resolutions. Should further results validate these initial findings, datasets easily derived from AVHRR, MODIS, or similar sensors will be able to be used as direct input to ecological niche models, bypassing the laborious, costly, and error-prone process of creating land cover maps.

#### ACKNOWLEDGMENTS

This research was funded in part by a grant from the National Science Foundation (NSF) under its National Partnership for Advanced Computational Infrastructure, Earth System Science (NPACI/ESS) Thrust. E.M.-M. was supported by a graduate fellowship from the Direccion General de Asuntos del Personal Academico of the National University of Mexico (UNAM).



**Figure 2. Two distribution prediction maps for *Ergaticus ruber*.** Top: topography alone. Bottom: topography and climate plus clustered NDVI. Darker tones indicate higher levels of coincidence.

#### REFERENCES

- [1] A. Guisan and E. Zimmerman, "Predictive habitat distribution models in ecology," *Ecological Modelling* 135 (2000), pp. 147-186, 2000.
- [2] EROS Data Center. [edcwww.cr.usgs.gov](http://edcwww.cr.usgs.gov)
- [3] CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad). [www.conabio.gob.mx](http://www.conabio.gob.mx)
- [4] Ortega-Huerta, M.A., Martinez-Meyer, E., Egbert, S.L., and Price, K.P. 2000. Mapping the Land Cover of Mexico Using AVHRR Time-Series Data Sets. *Geocarto International* 15(3): 5-18.
- [5] Stockwell DRB, I.R. Noble. 1991. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation* 32: 249-254.
- [6] Stockwell D.R.B. and D. Peters. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographic Information Science* 13: 143-158.
- [7] Anderson, R.P., D. Lew, and A.T. Peterson. Submitted. Using intermodel variation in error components to select best subsets of ecological niche models. *Ecological Modelling*.