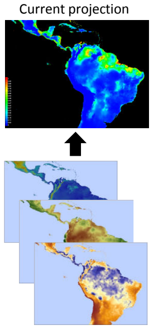
Optimizing Ecological Niche Models – Preliminary Insights

Matthew R. Lichty[1,2], Aleksandar Radosavljevic[2,3], and Patrick S. Herendeen[2]

[1]Knox College, Galesburg IL, 61401 USA
[2]Department of Plant Science, Chicago Botanic Garden, Glencoe, Illinois 60022 USA
[3]Plant Biology and Conservation, Northwestern University, Evanston, Illinois 60208 USA;

Introduction

- •Correlative ecological niche models (ENM) combine environmental , climatic, and/or biotic variables with species occurrence data to generate an approximation of a species' abiotically suitable habitat (Fig 1).
- •These models have seen increasing application as publically accessible occurrence databases and climate data have proliferated (www.gbif.org, Global Biodiversity Information Facility; www.worldclim.org, Hijmans, 2005).
- •ENMs have shown great promise, but there is still methodological uncertainty regarding best practices (Elith et al., 2010; Merow et al., 2013).
- •The goal of this project is to observe the effect that occurrence data quality has on model output.



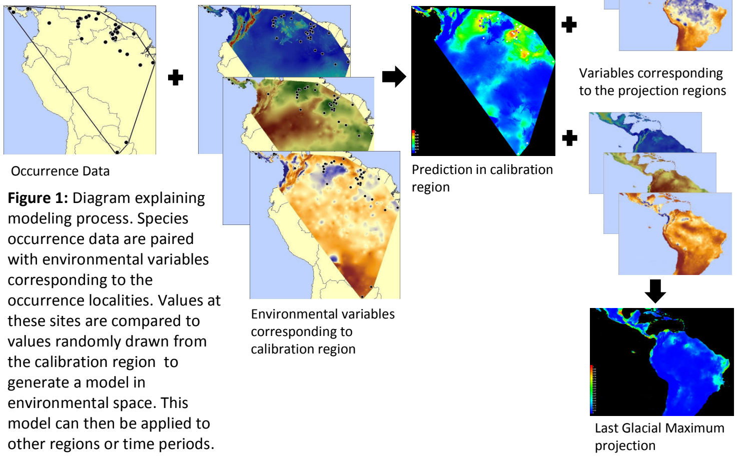

Occurrence Data
Environmental variables corresponding to calibration region
Prediction in calibration region
Variables corresponding to the projection regions
Last Glacial Maximum projection

Figure 1: Diagram explaining modeling process. Species occurrence data are paired with environmental variables corresponding to the occurrence localities. Values at these sites are compared to values randomly drawn from the calibration region to generate a model in environmental space. This model can then be applied to other regions or time periods.

Methods

Occurrence Data

- •We used two occurrence datasets to generate our models (Fig 2). One, downloaded from the Global Biodiveristy Information Facility (GBIF) consisted of all records of the neotropical legume *Cynometra bauhiniifolia* contained in their database (n=46). The other dataset consisted only of occurrence records whose identifications we were able to verify and which we carefully georeferenced using multiple sources (n=39).
- •We spatially filtered the points, removing all points that were less than 10 km from another point, in an effort to reduce the effects of sampling bias (Pearson et al., 2007).

Environmental Data

- •Environmental data consisted of two sets of 19 bioclimatic variables downloaded from WorldClim; one corresponded to current climatic conditions while the other corresponded to conditions at the Last Glacial Maximum (Hijmans et al., 2005).
- •To create our calibration regions, we used ArcMap 10.1 to draw a minimum convex hull enclosing either set of occurrence data, and then buffered them by 0.5° (ca. 50 km; Anderson & Raza, 2010; ESRI, 2011).
- •Our projection extents correspond to an area that includes what we consider to be the maximum potential habitat given the species dispersal abilities.

Modeling algorithm

- •We used Maxent (Phillips, et al, 2006) 3.3.1 to create the models; settings were left at default except regularization multiplier, which was set to 2.0, and the number of replicates, which was set to 4.

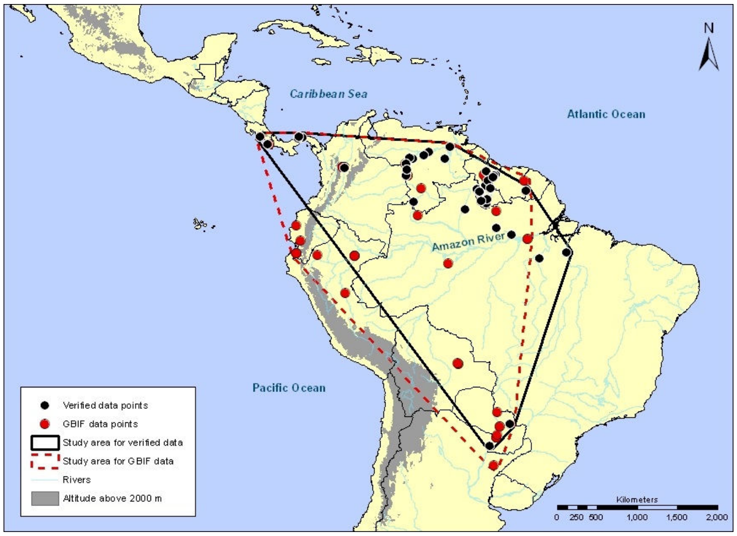


Figure 2: Map displaying the occurrence data and study areas for both the verified and GBIF data sets.

Results

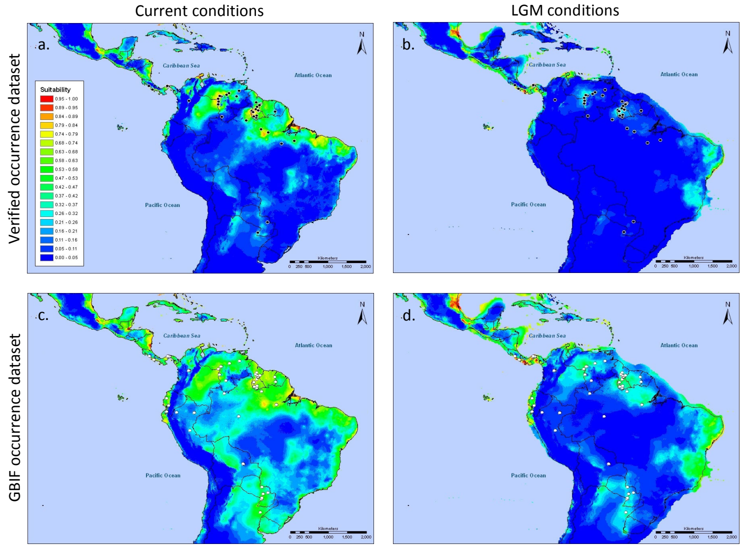


Figure 3: Maps showing suitability predictions for models made with verified (a, b) and GBIF data sets (c, d), projected onto current (a, c) and Last Glacial Maximum environmental variables (b, d). We display modern day country borders on the LGM maps for viewing reference and to show the difference in land area due to lower global sea level.

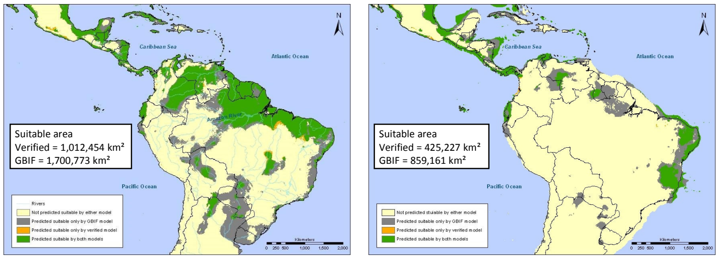


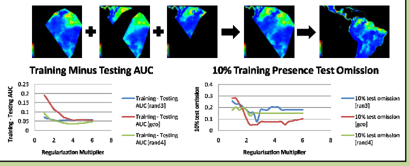
Figure 4: Maps showing the difference in predicted suitable area between the verified and GBIF data sets for current environmental conditions and environmental conditions at the Last Glacial Maximum. Predicted suitable area is based on thresholds for 10 percentile training presence omission rates. For both time periods, the models calibrated with the GBIF data set predicted an area that was roughly twice the extent of the area predicted as suitable by models calibrated with the verified dataset.

Conclusions

- •Natural history databases can contain many misidentified specimens, even from well curated collections; specimens mistakenly identified as *Cynometra bauhiniifolia* accounted for more than 20% of unique occurrence localities in our GBIF dataset.
- •Including misidentified specimens can lead to models that at best over-predict suitable habitat, and at worst, lead to models that are not representative of the species niche.
- •Modelers should be careful when using un-verified occurrence data, especially for taxonomically confusing species and taxa they are unfamiliar with.

Future Research

- •Tuning model parameters using spatially independent evaluation data (Radosavljevic & Anderson, accepted).
- •Determining an optimized geographic partitioning scheme of replicates.
- •Comparing strategies for model evaluation (Warren & Seifert, 2011; Hijmans, 2012).



Acknowledgements

This project was funded by the National Science Foundation through their Research Experience for Undergraduates program. We thank the Missouri Botanical Garden, the New York Botanical Garden, the Field Museum, and the Smithsonian for herbarium records.

References

- •Anderson R.P. & Raza A. (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela. *Journal of Biogeography*, **37**, 1378–1393.
- •Elith J., Kearney M., & Phillips S. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.
- •ESRI 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.
- •Hijmans R.J. (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, **93**, 679–88.
- •Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G., & Jarvis A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- •Merow, C., M. Smith, J. A. SilanderJr. 2013.A practical guide to Maxent: what it does, and why inputs and settings matter. *Ecography,* **36,** 1-12.
- •Pearson R.G., Raxworthy C.J., Nakamura M., & Townsend Peterson A. (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34**, 102–117.
- •Phillips S.J., Anderson R.P., & Schapire R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- •Warren D.L. & Seifert S.N. (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, **21**, 335–342.