

ENMTools: a toolbox for comparative studies of environmental niche models

Dan L. Warren, Richard E. Glor and Michael Turelli

*D. L. Warren (dan.l.warren@gmail.com) and M. Turelli, Dept of Evolution and Ecology, Univ. of California, Davis, CA 95616, USA.
– R. E. Glor, Dept of Biology, Univ. of Rochester, Rochester, NY 14627, USA.*

We present software that facilitates quantitative comparisons of environmental niche models (ENMs). Our software quantifies similarity of ENMs generated using the program Maxent and uses randomization tests to compare observed similarity to that expected under different null hypotheses. ENMTools is available online free of charge from <<http://purl.oclc.org/enmtools>>.

Environmental niche models are used to describe the ecological tolerances of populations or species. A range of methods generate ENMs from georeferenced occurrence data (i.e. sample localities associated with latitude and longitude coordinates) and environmental data in the form of geographic information system (GIS) data layers (reviewed in Elith et al. 2006). These ENMs can predict the suitability of habitats across a landscape and are routinely used in conservation planning and biodiversity inventories (Kremen et al. 2008). A growing number of studies compare predicted habitat suitabilities generated by ENMs from different species or populations to aid in species delimitation or test hypotheses about niche conservation (Buermann et al. 2008, Lentz et al. 2008, Warren et al. 2008). Comparative analyses of ENMs also permit studies of niche evolution in a phylogenetic framework. Warren et al. (2008) introduced a set of comparative similarity measures and statistical tests that permit quantitative comparison of ENMs. Here we provide an implementation of those methods called ENMTools. ENMTools interacts with the maximum entropy niche modeling program Maxent (Phillips et al. 2006, Phillips and Dudík 2008), allowing users to automate generation of ENMs, calculate similarity measures, and implement various statistical comparisons of ENMs. Although ENMTools is designed to interact directly only with Maxent, the pseudoreplicate data sets that it generates for hypothesis testing can be used with other methods of ENM construction. The similarity measures and tests discussed here are used on numerous examples in Warren et al. (2008). Other tools for making measurements on ENMs and comparing them are in development.

Implementation and required input

ENMTools is a Perl script with a graphical user interface written with the Tk package. It is also available as an executable file for Windows and Macintosh. In theory, the ENMTools Perl script may be implemented on any platform capable of running Perl and the associated Tkx package with minor adjustments. In practice, ENMTools has been implemented only on the Microsoft Windows and Mac OS X platforms. ENMTools requires working installations of Maxent and Java. Input for ENMTools is similar to that required by the widely used Maxent niche modeling software and consists of simple .asc and .csv-formatted text files containing locality data for two or more populations and associated ASCII-formatted GIS raster layers of environmental data.

Tests implemented in ENMTools

Quantifying niche similarity

Measurement of similarity among ENMs is central to comparative analyses. ENMTools quantifies niche similarity using the two measures introduced by Warren et al. (2008): Schoener's (1968) D and a measure derived from Hellinger distance called I . These similarity measures are obtained by comparing the estimates of habitat suitability calculated for each grid cell of a study area using a Maxent-generated ENM, after normalizing each ENM so that all suitability scores within the geographic space sum to 1. Schoener's D is defined as

$$D(p_X, p_Y) = 1 - \frac{1}{2} \sum_i |p_{X,i} - p_{Y,i}|,$$

where $p_{X,i}$ and $p_{Y,i}$ are the normalized suitability scores for species X and Y in grid cell i . I is defined as

$$I(p_X, p_Y) = 1 - \frac{1}{2} \sqrt{\sum_i \left(\sqrt{p_{X,i}} - \sqrt{p_{Y,i}} \right)^2}.$$

Note that I is $1 -$ “Hellinger’s distance” from probability theory; similarly, D is $1 -$ “total variation distance”. Both similarity measures range from 0, when species predicted environmental tolerances do not overlap at all (i.e. $p_{X,i}p_{Y,i} = 0$ for all i), to 1, when all grid cells are estimated to be equally suitable for both species (i.e. $p_{X,i} = p_{Y,i}$ for all i). We note that an actual suitability of 0 is rarely produced by Maxent for most real data sets unless a minimum suitability threshold is applied, as Maxent ENMs predict a nonzero suitability score at every grid cell. However, values arbitrarily close to 0 are possible. While these two similarity measures are calculated in a very similar fashion and are both simply related to standard measures of “distance” from probability theory, Schoener’s D carries an historical ecological interpretation that is not attached to I . The ecological interpretation of D implicitly assumes that the suitability scores $p_{X,i}$ produced by Maxent are proportional to species abundance, whereas I simply treats the two ENMs as probability distributions.

If localities for only two populations are loaded into ENMTools, single values of D and I will be produced. If three or more populations are loaded into ENMTools, pairwise D and I values will be calculated for each possible comparison. Pairwise similarity values or tables of pairwise similarity values can be opened and viewed in a simple text editor or in commercially available spreadsheet software such as Microsoft Excel.

Tests of niche similarity

ENMTools implements two quantitative tests of niche similarity introduced by Warren et al. (2008). These tests ask whether the ENMs generated from two populations are identical, or, at the other extreme, merely more similar than expected by chance.

Identity test

The first comparative hypothesis tested by ENMTools is whether the ENMs produced by two populations are identical. This hypothesis is addressed with the niche identity test. The test begins by pooling the georeferenced data points for a pair of populations, randomizing the population identities of these data points, and extracting two new population samples with the same sizes as the two original samples. This process is repeated to generate a user-specified number of pseudoreplicated data sets. For each pseudoreplicate, ENMTools calls on Maxent to generate an ENM from each of the two populations. ENMTools then uses the model and predicted suitability scores generated by Maxent from each population to calculate niche similarity indices (e.g. I and Schoener’s D). In this manner, ENMTools obtains a distribution of overlap scores between populations drawn from a shared distribution, assuming that the populations are interchangeable in their use of niche space. The actual

observed measures of niche similarity between populations can be compared to this null distribution to test whether they are significantly different from those generated assuming no niche differentiation: the hypothesis of niche identity is rejected when the empirically observed value for I and/or D is significantly lower than the values expected from the pseudoreplicated data sets.

As noted in Warren et al. (2008), the identity test will accurately reflect the probabilities under the null hypothesis if sampling is unbiased with respect to the species environmental tolerances (i.e. spatial sampling is such that the relationship between environmental variables and species occurrence accurately represents the suitability of habitat). However, biases can be introduced by many factors, including differences in sampling effort between populations, spatial bias in sampling, and differences in the habitat available to populations in geographic regions where they do not overlap. The identity test should be interpreted with these caveats in mind.

Background test

In addition to testing whether the ENMs generated from two species are identical, one may also want to determine whether ENMs are more similar than expected by chance, based on the geographical regions in which they reside (as opposed to the exact locations in which specimens are collected). This type of analysis is particularly important when allopatric populations are being compared, because some niche differentiation may follow inevitably from the fact that distinct geographic regions rarely encompass identical distributions of environmental variables. The background test in ENMTools may be used to ask whether the ENMs obtained from two allopatrically distributed populations are more different than would be expected given the underlying environmental differences between the regions in which they occur. The background test addresses this hypothesis by generating a null distribution for the ENM difference expected between one population and occurrence points placed at random within the range of another population (i.e. not the specific locations at which the other population was sampled, but within a user-specified area that includes all of those samples). The points to be treated as environmental background can either be specified as a set of comma-separated coordinates, or can be sampled from an ASCII raster file. If the observed values of the niche similarity measures obtained from the two original populations are significantly higher (or lower) than expected from this null distribution, the null hypothesis that similarity (or divergence) between species is no more than expected based on the availability of habitat is rejected.

While the null hypothesis addressed by the background test may be more biologically meaningful than that of the identity test for allopatric populations, the outcome and interpretation of the test may be sensitive to the definition of the environmental background from which each species is presumed to be selecting habitat. Whenever possible, a biological justification should be made for the definition of “background” used for a particular study. In the absence of a compelling biological reason for selecting background regions to sample from, background samples can be taken

from buffer zones around known occurrences or from range maps estimated using one of the many algorithms available for this purpose, such as minimum convex polygon (e.g. as implemented in ArcGIS (ESRI 2006)) or localized convex hull (Getz and Wilmers 2004). When there are questions about the definition of environmental background for a

particular study, users are advised to assess the sensitivity of their conclusions to selection of alternative background regions.

Empirical examples of both tests are provided in Fig. 1 and 2, using data from *Anolis ahli* and *Anolis allogus* (Knouft et al. 2006). The tests show that the two species are

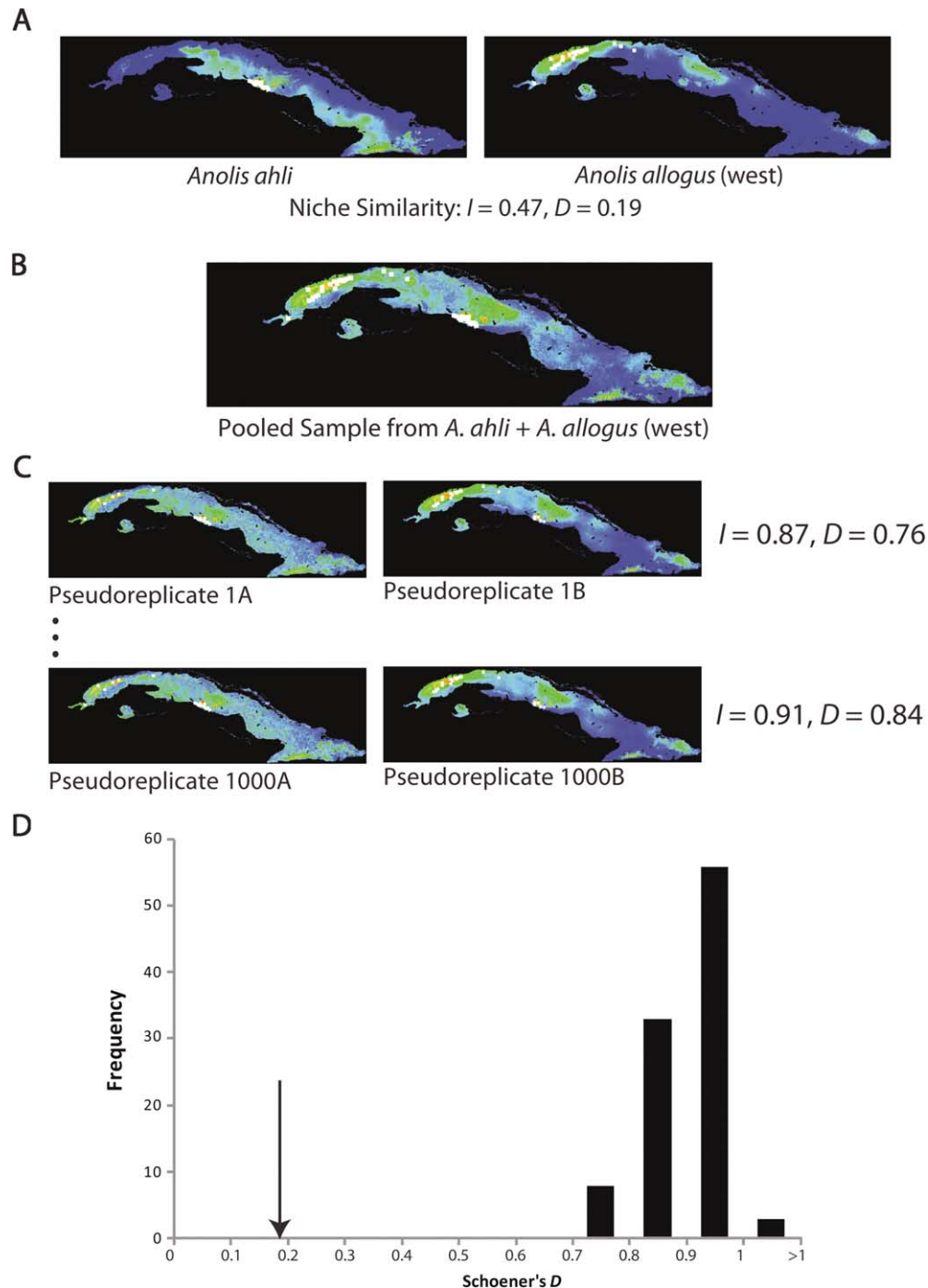


Figure 1. Sampling for identity test. In the identity test, the similarity score for ENMs built from known occurrences of two species (panel A) is compared to the distribution of similarity scores between ENMs built from occurrences drawn randomly from the pooled occurrences for the two species (panels B–D). The entirety of Cuba was treated as the study area for ENM construction. This example compares the environmental niches of *Anolis allogus* (west) and *Anolis ahli*. As seen in panel D, the similarity score for ENMs built for the actual occurrences of the two species (black arrow) is much lower than expected based on the null hypothesis of niche equivalency, indicating that the two species' environmental niches are not equivalent.

not drawn from identical distributions of environmental variables (identity test), and are in fact more different from each other than would be expected under the hypothesis that the differences between the two species are due only to differences in the habitat available to them. Users should

note that the background test can yield different results when the comparison is reversed; it is possible for species to be more similar than expected based on the habitat available for one species, but less than expected based on the habitat available for the other.

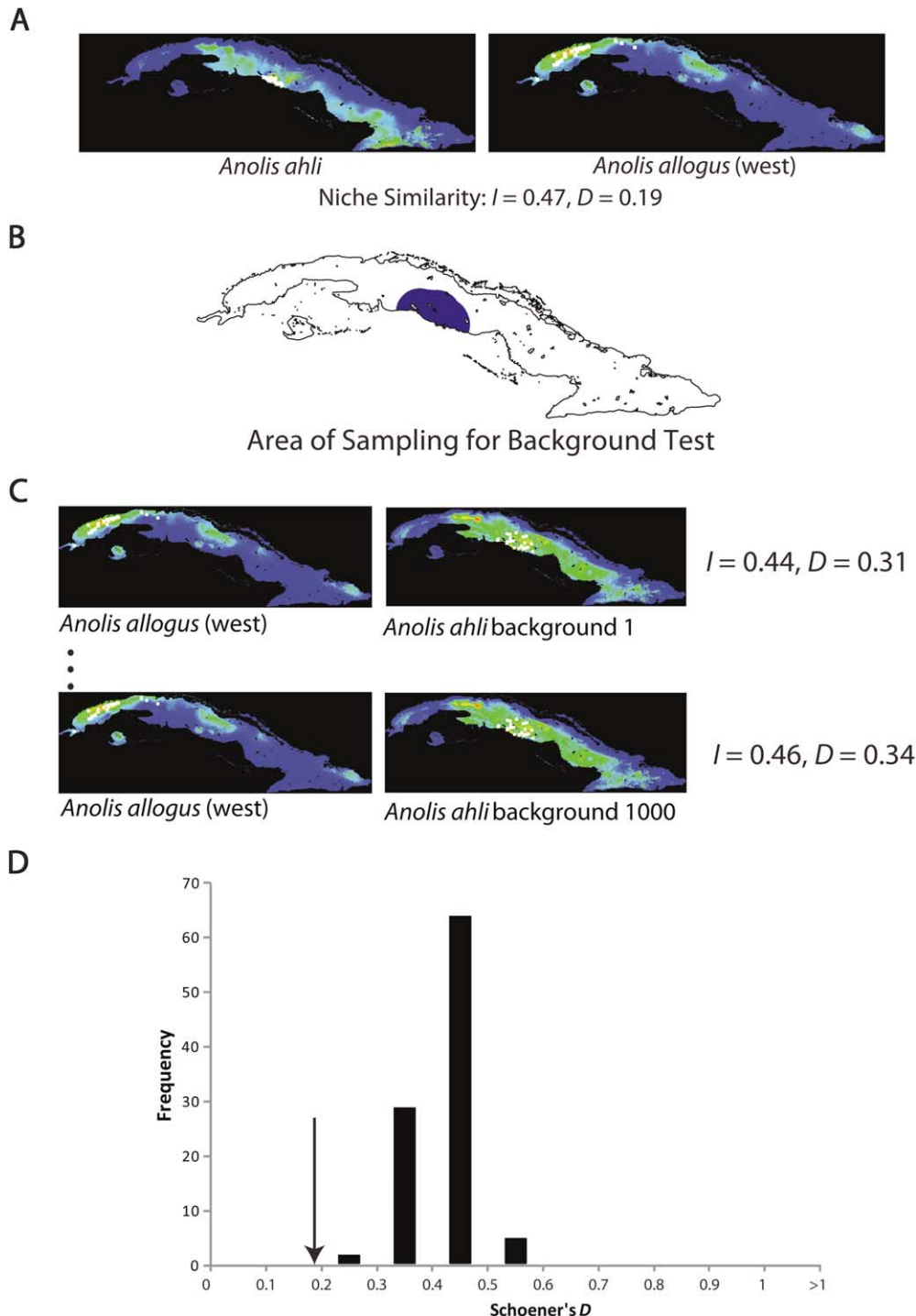


Figure 2. Sampling for background test. In the background test, the similarity score for ENMs built from known occurrences of two species (panel A) is compared to similarity scores for ENMs constructed using points drawn at random from the region defined as “environmental background” for one of the species (panels B–D). In this example, occurrence points for *Anolis allogus* (west) are being compared to points drawn at random from the environmental background that has been chosen for *Anolis ahli*. In this case, the background (blue region, panel B) was chosen arbitrarily to be all points within 0.5 decimal degrees of a known occurrence of *A. ahli*. As seen in panel D, the observed overlap between the two species (black arrow) is lower than expected under the null hypothesis, indicating that the two species are more divergent than expected based on the habitat available to *A. ahli*.

The ENMTools software, manual, and sample data are available at the ENMTools web site <<http://purl.oclc.org/enmtools>>.

To cite ENMTools or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for “Version 0”:

Warren, D. L., Glor, R. E. and Turelli, M. 2009. ENMTools: a toolbox for comparative studies of environmental niche models. – *Ecography* 33: 607–611 (Version 0).

Acknowledgements – The authors would like to thank the reviewers for useful comments and early users of ENMTools who have been indispensable in suggesting features and bug fixes. This work was funded in part by the Center for Population Biology at UC Davis, the John Muir Inst. for the Environment, the California Dept of Fish and Game, NSF grant DEB 0920892 to REG, and NSF grant DEB 0815145 to MT.

References

Buermann, W. et al. 2008. Predicting species distributions across the Amazonian and Andean regions using remote sensing data. – *J. Biogeogr.* 35: 1160–1176.

- Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – *Ecography* 29: 129–151.
- ESRI 2006. ArcGIS 9.2. – Environmental Systems Research Inst., Redlands, CA.
- Getz, W. and Wilmer, C. 2004. A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. – *Ecography* 27: 489–505.
- Knouft, J. H. et al. 2006. Phylogenetic analysis of the evolution of the niche in lizards of the *Anolis sagrei* group. – *Ecology* 87: S29–S38.
- Kremen, C. et al. 2008. Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. – *Science* 320: 222–226.
- Lentz, D. L. et al. 2008. Ecological niche modeling and distribution of wild sunflower (*Helianthus annuus* L.) in Mexico. – *Int. J. Plant Sci.* 169: 541–549.
- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Phillips, S. J. and Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. – *Ecography* 31: 161–175.
- Schoener, T. W. 1968. The *Anolis* lizards of Bimini: resource partitioning in a complex fauna. – *Ecology* 49: 704–726.
- Warren, D. L. et al. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. – *Evolution* 62: 2868–2883.