

Ecography

ECOG-03900

Warren, D. L., Beaumont, L., Dinnage, R. and Baumgartner, J. 2018. New methods for measuring ENM breadth and overlap in environmental space. – Ecography doi: 10.1111/ecog.03900

Supplementary material

Appendix 1. Calculating niche breadth and overlap in environmental space.

The methods outlined here are implemented in two functions in the new ENMTools R package: “env.overlap” and “env.breadth”. Each will work with any ENM that is capable of producing suitability values using a data frame of predictor values and the “predict” function from the *dismo* package (Hijmans et al. 2017). At present this includes Bioclim, Domain, GLM, GAM, Maxent, ppmlasso, and random forest models. As input, all functions require an ENMTools model object containing the model(s) being evaluated and a Raster* object (Hijmans and van Etten 2016) containing the predictor variables used to construct the model. The bounds for each environmental axis are set by default to be the minimum and maximum value of the predictor variable from their respective layers in the Raster* object if one is passed, but bounds can also be passed manually as a list. We note that some modeling approaches are capable of extrapolating indefinitely along an environmental gradient, and as such overlaps calculated using these models may be sensitive to the choice of boundaries for the environment space. In these situations it may be useful to explore the sensitivity of overlap or breadth metrics to the choice of boundaries or, alternatively, to interpret and discuss these metrics in the context of the chosen boundaries for a given study. Further, we caution against comparing inter-model overlap across environment spaces with differing numbers of predictors. An increase in dimensionality typically implies a decrease in niche overlap (unless niches are identical), and as such overlap must always be interpreted in the context of the environmental dimensions considered (Blonder et al. 2014).

The process is demonstrated in Figure S1. An initial environmental space of user specifiable size (argument “*chunk.size*”, which defaults to 100,000 points) is constructed using random Latin hypercube sampling (see next section for details), implemented in the R package *lhs* (Carnell 2016). Predictions from the model(s) are made onto these environmental conditions, which are used to calculate B2 for “*env.breadth*” (Levins 1968; Mandle et al. 2010), or D, I, and Spearman rank correlation (ρ) for “*env.overlap*” (Schoener 1968; Warren, Glor, and Turelli 2008). These metrics are calculated as:

$$B2 = \frac{1}{n-1} \left(\frac{1}{\sum_i^n x_i^2} - 1 \right)$$

$$D = 1 - \frac{\sum_i^n |x_i - y_i|}{2}$$

$$I = 1 - \frac{\sum_i^n \left(\sqrt{x_i} - \sqrt{y_i} \right)^2}{2}$$

where x_i and y_i are the standardized suitability values predicted from models x and y in the i th combination of environmental variables and n is the total number of environmental combinations. After values are calculated for the initial sample, the algorithm continues to add points in environmental space in increments of “*chunk.size*”, recalculating the metrics in question with each addition. The difference between successive values for each metric is recorded, and sampling of environmental space stops when the difference between successive measurements falls below a user-specifiable tolerance.

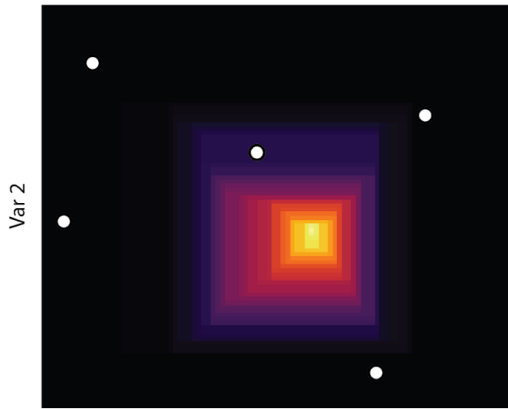
This approach allows users to manage the tradeoff between execution time and precision; lower tolerance values result in higher precision of estimates of niche overlap or breadth, while higher tolerance values allow for faster run

times. The rate-limiting step in these calculations is typically the time it takes *dismo* to make predictions from models, and as such execution time varies with algorithm. Additionally, convergence of the estimate will typically take longer for models with more predictors and lower environmental niche breadth. For a four-dimensional environmental space with a tolerance of 0.0001, execution times are typically on the order of one to five seconds. For very high dimensional environmental spaces or lower tolerance values, execution may take longer. We note that the minimum number of points used to calculate metrics will be $2 \times \text{"chunk.size"}$, and that users can use this argument accordingly. The default chunk size and tolerance level were chosen based on preliminary experimentation with a variety of model types as a good compromise between precision and execution time.

Latin Hypercube Sampling

Latin hypercube sampling (LHS) is a Monte Carlo method developed to improve the evenness of randomly sampled points in a high-dimensional space (McKay, Beckman, and Conover 1979). Ordinary Monte Carlo sampling generally uses samples drawn from a uniform distribution along each dimension of a multi-dimensional space. The problem with uniform random samples, however, is that they tend to form clumps (randomly) along each axis. This is typically not too extreme along a single dimension, but in high-dimensional space the problem tends to be exacerbated, leading to many areas of the space with no sample coverage. LHS attempts to correct for this issue, by using an algorithm to spread samples evenly in every dimension. Briefly, the method involves dividing the multi-dimensional space into a set of n equal-sized hypercubes (n being the

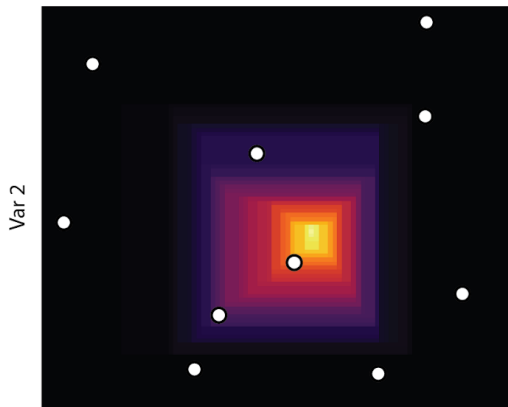
desired number of samples), and then drawing a single sample uniformly from within each hypercube. This guarantees that each region of space across all dimensions gets at least one sample. LHS has been shown to improve the efficiency of numerical integration, allowing convergence with fewer samples and reducing the variance of estimates made through Monte Carlo methods (McKay, Beckman, and Conover 1979; Stein 1987).



Suitability
0
0
0
0
0.1
0

N = 5
B2 = 0

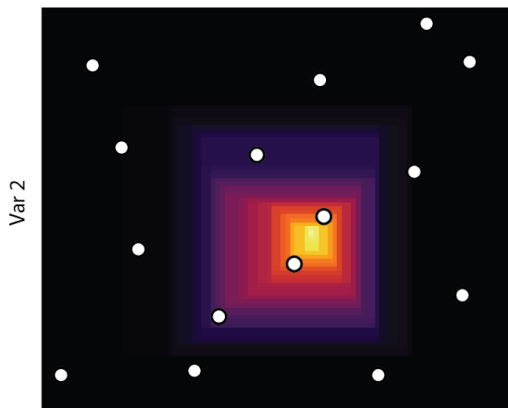
Var 1



Suitability
0
0
0
0.1
0
0.8
0
0.2
0
0

N = 10
B2 = .08
Diff = .08

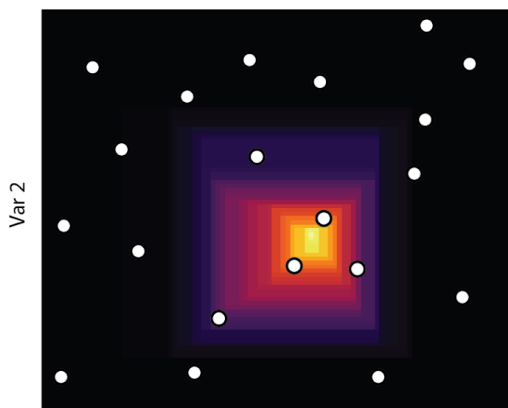
Var 1



Suitability
0
0
0
0.1
0
0.8
0
0.2
0
0
0
0
0
0
0
0

N = 15
B2 = .13
Diff = .05

Var 1



Suitability
0
0
0
0.1
0
0.8
0
0.2
0
0
0
0
0
0
0
0.5
0.9
0
0.7
0

N = 20
B2 = .13
Diff = 0

Var 1

Figure A1. Calculating environmental breadth in a two dimensional environmental space (Bioclim variables 2 and 4, for species *Iberolacerta monticola*). For demonstration purposes, “chunk.size” in this example is set to 5 and tolerance to 0.01. In the first iteration (top row), an initial sampling of five points is performed and the metric (in this case B2) is calculated. A second sample of 5 points is added (second row), and B2 is recalculated. Since the absolute difference between the two samples is greater than the tolerance ($|B2 \text{ from the second row} - B2 \text{ from the top row}| > \text{tolerance}$), a third sample is taken and the metric recalculated (third row). These steps are repeated until the new B2 does not differ from the previous one by more than the tolerance setting (bottom row, where $\text{diff} < \text{tolerance}$), at which point the calculation stops.

The process for calculation of niche overlap is the same, except that overlap metrics are designed to compare two models constructed with a consistent set of predictors. For each iteration, both models are used to predict the suitability of habitat across the sampled environmental conditions. The suitabilities from the two models are then used to calculate similarity using overlap (D and I, as per the equations given above) and correlation (Spearman’s ρ) metrics.

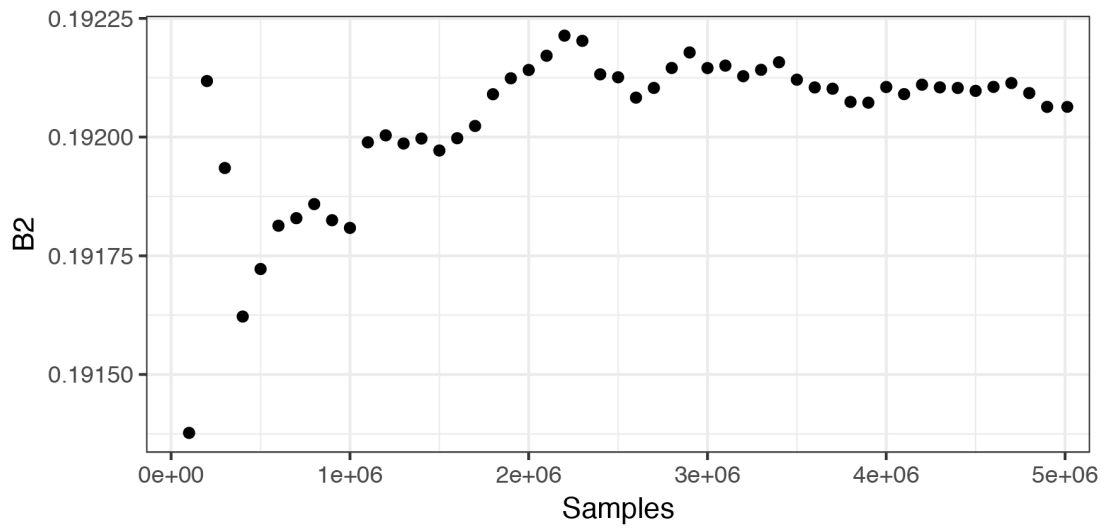


Figure A2. Calculation of niche breadth from the model above. In this case chunk size was 100,000, and tolerance was set to 0.00001. Elapsed system time was 1.997 seconds.

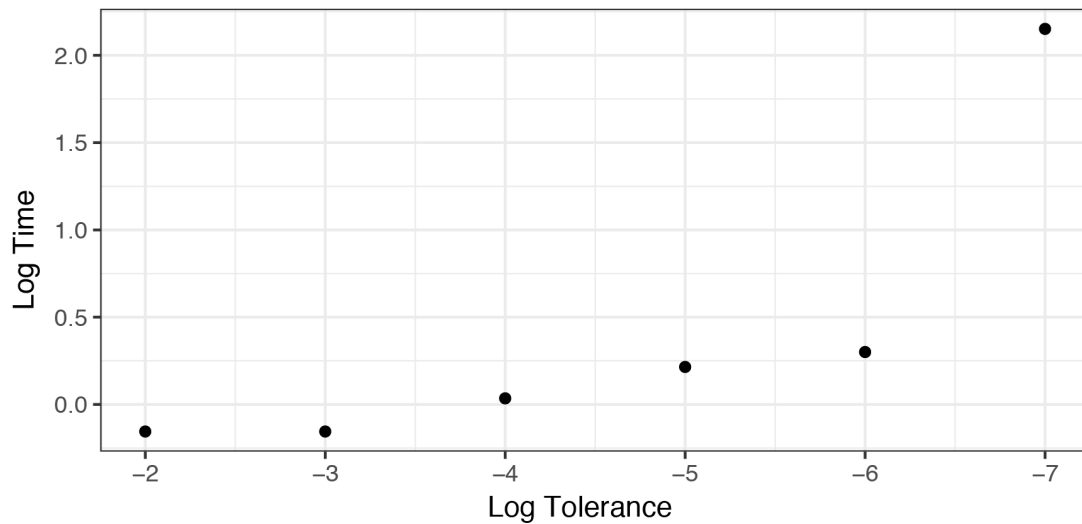


Figure A3. Execution time as a function of tolerance when calculating environmental breadth (B2), with chunk.size equal to 100,000. The log (base 10) of the number of seconds needed to produce a stable breadth estimate is plotted as a function of the log (base 10) tolerance argument passed to the “env.breadth” function. Minimum execution time was approximately 0.7 seconds, maximum was 141 seconds. The difference between minimum and maximum estimates for B2 across all tolerance values was approximately 3%, showing that the estimates were very stable across tolerance values. This example indicates the general performance of our approach, yet the time taken for metrics to converge will vary depending on specifics of the problem (e.g., dimensionality, the shape of the niche volume, the range along each dimension, etc.).

- Carnell, Rob. 2016. *Lhs: Latin Hypercube Samples* (version 0.14). <https://CRAN.R-project.org/package=lhs>.
- Hijmans, Robert J., and Jacob van Etten. 2016. *Raster: Geographic Data Analysis and Modeling* (version 2.5-8). <https://CRAN.R-project.org/package=raster>.
- Hijmans, Robert J., Steven Phillips, John Leathwick, Jane Elith, and Maintainer Robert J. Hijmans. 2017. *Package "dismo."* <http://www12.frugalware.org/mirrors/cran.r-project.org/web/packages/dismo/dismo.pdf>.
- Levins, Richard. 1968. *Evolution in Changing Environments: Some Theoretical Explorations*. Princeton University Press.
- Mandle, Lisa, Dan L. Warren, Matthias H. Hoffmann, A. Townsend Peterson, Johanna Schmitt, and Eric J. von Wettberg. 2010. "Conclusions about Niche Expansion in Introduced *Impatiens Walleriana* Populations Depend on Method of Analysis." *PloS One* 5 (12):e15297.
- McKay, Michael D., Richard J. Beckman, and William J. Conover. 1979. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences* 21 (2). Taylor & Francis Group:239–45.
- Schoener, Thomas W. 1968. "The Anolis Lizards of Bimini: Resource Partitioning in a Complex Fauna." *Ecology* 49 (4). Wiley Online Library:704–26.
- Stein, Michael. 1987. "Large Sample Properties of Simulations Using Latin Hypercube Sampling." *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences* 29 (2). Taylor & Francis:143–51.
- Warren, Dan L., Richard E. Glor, and Michael Turelli. 2008. "Environmental Niche Equivalency versus Conservatism: Quantitative Approaches to Niche Evolution." *Evolution; International Journal of Organic Evolution* 62 (11):2868–83.