Hard edges, soft edges, and species range evolution: A genomic analysis of the Cumberland Plateau salamander

– ODMAP Protocol –

Emily Watts, Brian Waldron, Shawn Kuchta

2024-05-06

## Overview

#### Authorship

Contact : [watts.emily.f@gmail.com](mailto:watts.emily.f@gmail.com)

Study link: Available upon publication

#### Model objective

Model objective: Inference and explanation

#### Focal Taxon

Focal Taxon: Plethodon kentucki

#### Location

Location: Kentucky, West Virginia, and Virginia, USA

#### Scale of Analysis

Spatial extent: 36, 39.5, -84.5, -80.5 (xmin, xmax, ymin, ymax)

Spatial resolution: 1 km^2

Temporal extent: 1970-2000

Temporal resolution: Monthly

Boundary: rectangle

#### Biodiversity data

Observation type: GPS tracking, range map

Response data type: presence-only

#### Predictors

Predictor types: climatic

#### Hypotheses

Hypotheses: We expected that habitat across the Ohio and Kanawha Rivers would be suitable for Plethodon kentucki, despite its inability to access that habitat to the east and north of its range. We expected to find decreasingly suitable habitat in the western and southern portions of the range.

#### Assumptions

Model assumptions: We assume that this dataset represents present-day conditions.

#### Algorithms

Modelling techniques: maxent

Model complexity: We removed variables with a correlation coefficient < ±0.7 using the ‘Remove Highly Correlated Variables’ tool in SDMTOOLBOX v. 2.5 and retained seven climatic variables for use in niche modeling. To parameterize the ENM, we evaluated the performance of various combinations of five feature classes (linear; linear, and quadratic; hinge; linear, quadratic, and hinge; and linear, quadratic, hinge, product, and threshold) and 10 regularization multipliers (from 1 to 5, in increments of 0.5) using the Maxent algorithm in the R package ‘SDMTune.’ We evaluated the performance of many ENMs built under each combination of model parameters through inspections of the omission rate, area under the curve (AUC), and model feature class complexity. We used the ENM with the lowest omission rate, highest AUC, and lowest complexity, in the order listed.

Model averaging: We chose the best fit model for our study and published it.

#### Workflow

Model workflow: We downloaded raster coverages of the 19 bioclimatic variables from WorldClim version 2/1 database at 30 arc-seconds resolution (~1 km^2). We clipped the coverages to our study region using the Extract by Mask tool in Arc-GIS v. 10.8 Spatial Analyst Tools.We removed any variables with a correlation coefficient < ±0.7 using the ‘Remove Highly Correlated Variables’ tool in SDMToolbox v. 2.5. We retained BIO1, BIO3, BIO7, BIO8, BIO9, BIO12, and BIO15 for modeling. We used GPS coordinates for 908 individuals from 86 localities that represent all known localities of P. kentucki from museum specimens and data from our own field work. We then removed duplicate occurrences and spatially rarefied these points to 1 km using ‘Spatially Rarefy Occurrence Data’ in SDMToolbox. This reduced our dataset to 65 occurence pounts. Background pounts were sampled from a minimum convex polygon defined as a 20 km buffer drawn around occurrence points. To parameterize the ENM, we evaluated the performance of various combinations of five feature classes (linear; linear, and quadratic; hinge; linear, quadratic, and hinge; and linear, quadratic, hinge, product, and threshold) and 10 regularization multipliers (from 1 to 5, in increments of 0.5) using the Maxent algorithm in the R package ‘SDMTune.’ We evaluated the performance of many ENMs built under each combination of model parameters through inspections of the omission rate, area under the curve (AUC), and model feature class complexity.

#### Software

Software: Arc-GIS v. 10.8, SDMToolbox v. 2.5, and SDMTune in R v. 4.2.2

Code availability: <https://github.com/efwatts/Plethodonkentucki_RangeLimits>

Data availability: <https://www.worldclim.org/data/worldclim21.html>

## Data

#### Biodiversity data

Taxon names: Plethodon kentucki

Taxonomic reference system: Notes on species taxonomy can be found in Watts, E. F., Waldron, B. P., Hantak, M. M., Lemmon, E. M., Lemmon, A. R., & Kuchta, S. R. (2024). Cryptic Species within a Cryptic Species? Species Delimitation in the Cumberland Plateau Salamander, Plethodon kentucki. Herpetologica.

Ecological level: individuals

Data sources: <https://naturalhistory.si.edu/research/vertebrate-zoology/amphibians-reptiles> and own field observations

Sampling design: All known populations of Plethodon kentucki

Sample size: 906 individuals from 86 localities

Clipping: Cumberland Plateau Region of Appalachia, USA

Scaling: Background points were sampled from a minimum convex polygon defined as a 20 km buffer drawn around the occurrence localities.

Cleaning: We removed duplicate coordinates and spatially rarefied these points to 1 km to avoid areas of high spatial autocorrelation using the ‘Spatially Rarefy Occurrence Data’ tool in SDMTOOLBOX. This reduced our 86 localities to 65 unique occurrence points.

Absence data: Presence-only

Background data: Background points were sampled from a minimum convex polygon defined as a 20 km buffer drawn around the occurrence localities.

Errors and biases: It is possible that individuals from museums were misidentified individuals of Plethodon glutinosus, a species with which Plethodon kentucki is nearly morphologically indistinguishable. All samples collected by us were genetically confirmed.

#### Predictor variables

Predictor variables: BIO1 Annual Mean Temperature BIO3 Isothermality BIO7 Temperature Annual Range BIO8 Mean Temperature of Wettest Quarter BIO9 Mean Temperature of Driest Quarter BIO12 Annual Precipitation BIO15 Precipitation Seasonality

Data sources: World Clim version 2.1, accessed 5/2/2024

Spatial extent: 36, 39.5, -80.5, -84.5 (xmin, xmax, ymin, ymax)

Spatial resolution: 30 arc-seconds/1km^2

Coordinate reference system: Geographic/WGS 84

Temporal extent: 1970-2000

Temporal resolution: Monthly

Data processing: n/a

Errors and biases: n/a

Dimension reduction: We chose the finest scale available from WorldClim

#### Transfer data

<Spatial extent>

<Spatial resolution>

<Temporal extent>

## Model

#### Variable pre-selection

Variable pre-selection: We included all 19 WorldClim variables available in our test for multicollinearity.

#### Multicollinearity

Multicollinearity: We removed variables with a correlation coefficient < ±0.7 using the ‘Remove Highly Correlated Variables’ tool in SDMTOOLBOX v. 2.5.

#### Model settings

maxent: featureSet (lqpht), regularizationMultiplierSet (1.5), targetGroupSampleSize (908 individuals from 86 localities ), notes (65 unique occurrence points)

#### Model estimates

<Coefficients>

Variable importance: Create response curves, create prediction pictures, and spatial jackkifing

#### Model selection - model averaging - ensembles

Model selection: To parameterize the ENM, we evaluated the performance of various combinations of five feature classes (linear; linear, and quadratic; hinge; linear, quadratic, and hinge; and linear, quadratic, hinge, product, and threshold) and 10 regularization multipliers (from 1 to 5, in increments of 0.5) using the Maxent algorithm (Phillips et al., 2006) in the R package ‘SDMTune.’ We evaluated the performance of many ENMs built under each combination of model parameters through inspections of the omission rate, area under the curve (AUC), and model feature class complexity. We used the ENM with the lowest omission rate, highest AUC, and lowest complexity, in the order listed. The best-fit model was parameterized with linear feature class and regularization multiplier of 1.5.

Model averaging: Crossvalidate setting with 20 percentage random test points

#### Analysis and Correction of non-independence

Spatial autocorrelation: We removed variables with a correlation coefficient < ±0.7 using the ‘Remove Highly Correlated Variables’ tool in SDMTOOLBOX v. 2.5

## Assessment

#### Performance statistics

Performance on training data: AIC, TSS, AUC

#### Plausibility check

Response shapes: Response curves and jackknife to measure variable importance