The importance of climatic data sources and limitations of ecological niche models for estimating historical ranges and niche overlaps in distantly related lungless salamanders

– ODMAP Protocol –

Yucheol Shin, Amaël Borzée, Daesik Park

2024-10-29

## Overview

#### Authorship

Contact : Yucheol Shin ([yshin@amnh.org](mailto:yshin@amnh.org)); Daesik Park ([parkda@kangwon.ac.kr](mailto:parkda@kangwon.ac.kr))

Study link: N/A

#### Model objective

Model objective: Forecast and transfer

Target output: Continuous habitat suitability

#### Focal Taxon

Focal Taxon: Korean clawed salamander (*Onychodactylus koreanus*), Korean crevice salamander (*Karsenia koreana*)

#### Location

Location: Korean Peninsula

#### Scale of Analysis

Spatial extent: 120, 135, 33, 44 (xmin, xmax, ymin, ymax)

Temporal extent: Calibration: current, mid-Holocene (MH; 6 Ka), Last Glacial Maximum (LGM; 21 Ka), Last Interglacial (LIG; 130 Ka), Marine Isotope Stage 19 (MIS19; 787 Ka), mid-Pliocene Warm Period (mPWP; ca. 3.2 Ma)

Boundary: rectangle, political

#### Biodiversity data

Observation type: citizen science, field survey

Response data type: point occurrence, presence-only

#### Predictors

Predictor types: climatic, habitat, topographic

#### Hypotheses

Hypotheses: Given their lunglessness and habitat requirements, the current geographic distributions of the two target species are strongly determined by temperature and precipitation.

#### Assumptions

Model assumptions: Stable ecological niche between model calibration timeframe (current) and model transfer timeframe (Plio-Pleistocene)

#### Algorithms

Modelling techniques: maxent

Model complexity: Selection of optimal model parameter combinations through extensive testing of candidate models

Model averaging: Model averaging across the cross-validation runs

#### Workflow

Model workflow: 1) Collection and processing of relevant environmental variables 2) Collection and processing of species occurrence data 3) Collection and processing of target group occurrence points, conversion of target group occurrence points into a kernel density raster, and sampling of background points 4) Removal of highly correlated environmental variables 5) Candidate model testing and selection of the optimal model parameter combinations 6) Model evaluation 7) Examination of variable importance and response plots 8) Model prediction under the current environmental conditions 9) Model transfer 10) Extrapolation risk assessment

#### Software

Software: R programming language, mainly utilizing ENMeval and dismo packages for model testing and transfer

Code availability: <https://github.com/yucheols/TwoSalDist>

Data availability: All data available from the code repository (<https://github.com/yucheols/TwoSalDist>) or through public databases with details provided in the main text.

## Data

#### Biodiversity data

Taxon names: Korean clawed salamander (*Onychodactylus koreanus*), Korean crevice salamander (*Karsenia koreana*)

Taxonomic reference system: Amphibian Species of the World version 6.2

Ecological level: species

Data sources: 1. Global Biodiversity Information Facility (GBIF) - DOI not available, downloaded through the megaSDM package in R. 2. Field survey data - based on the 4th National Ecosystem Surveys (NES) - Accessed through the Ecobank platform (<https://www.nie-ecobank.kr/cmmn/Index.do>?) 3. From previous studies (see details in text)

Sample size: 187 occurrence points for O. koreanus and 137 Occurrence points for K. koreana

Background data: Sampled within the calibration range. 10,000 points sampled based on a kernel density raster representing the amphibian sampling effort across the Korean Peninsula to compensate for the bias of occurrence points toward South Korea.

#### Data partitioning

Training data: Hierarchical spatial checkerboard cross-validation (implemented using the “checkerboard2” function of the ENMeval package)

Validation data: Hierarchical spatial checkerboard cross-validation (implemented using the “checkerboard2” function of the ENMeval package)

#### Predictor variables

Predictor variables: Climatic variables (temperature and precipitation), topography (elevation, slope), consensus land cover (forest cover)

Data sources: 1) Climate: WorldClim 2.1 (<https://www.worldclim.org/>) and CHELSA (<https://chelsa-climate.org/>) 2) Topography: EarthEnv (<https://www.earthenv.org/>) 3) Land cover: EarthEnv (<https://www.earthenv.org/>)

Spatial extent: 124.1833, 130.9417, 33.10833, 43.00833 (xmin, xmax, ymin, ymax)

Spatial resolution: 1km (0.0083333 dd)

Coordinate reference system: WGS 84

Temporal extent: WorldClim: climatic averages spanning 1970-2000 CHELSA: climatic averages spanning 1979-2013

#### Transfer data

Data sources: Paleoclimatic data downloaded from the PaleoClim database (<http://www.paleoclim.org/>)

Spatial extent: 120, 135, 33, 44 (xmin, xmax, ymin, ymax)

Spatial resolution: 5km (0.0416666 dd), statistically downscaled to 1km (0.0083333 dd) using the bilinear interpolation. Downscaling was done using the “disaggregate” function of the raster package.

Temporal extent: Four time periods of the Plio-Pleistocene, bracketing the divergence times of the two species estimated from genetic data. These time periods are as follows: mid-Holocene (MH; 6 Ka), Last Glacial Maximum (LGM; 21 Ka), Last Interglacial (LIG; 130 Ka), Marine Isotope Stage 19 (MIS19; 787 Ka), mid-Pliocene Warm Period (mPWP; ca. 3.2 Ma)

Models and scenarios: Paleoclimate models were obtained from the PaleoClim database (<http://www.paleoclim.org/>)

Quantification of Novelty: Multivariate Environmental Similarity Surface (MESS)

## Model

#### Multicollinearity

Multicollinearity: Pearson’s correlation test, followed by the removal of highly correlated variables (collinearity cutoff: | r | > 0.7)

#### Model settings

maxent: featureSet (L, Q, H, P, LQ, LP, QH, QP, HP, LQH, LQP, LQHP, LQHPT), regularizationMultiplierSet (A range between 0.5 and 5 at a 0.5 increment)

Model settings (extrapolation): Clamping

#### Model estimates

Coefficients: Mean

#### Threshold selection

Threshold selection: 10% presence threshold, used for the estimation of current habitat suitability only.

## Assessment

#### Performance statistics

Performance on training data: AUC, Continuous Boyce Index (CBI), 10% omission rate

Performance on validation data: AUC, AUC difference, Continuous Boyce Index (CBI)

<Performance on test data>

#### Plausibility check

Response shapes: Response plots

Expert judgement: Visual check of model prediction outputs and comparisons with known geographic distributions

## Prediction

#### Prediction output

Prediction unit: Continuous habitat suitability

#### Uncertainty quantification

Scenario uncertainty: Direct quantification of prediction differences is the core goal of this study

Novel environments: Multivariate Environment Similarity Surface (MESS)