Supplementary Materials 1

Establishment potential for the two gecko species adapted to different climates, *Gekko japonicus* and *G. swinhonis*, introduced to South Korea

Il-Kook Park, Yucheol Shin, Hae-Jun Baek, Jongsun Kim, Dae-In Kim, Minjeong Seok, Yaechan Oh, Daesik Park

ODMAP protocol niche modeling report protocol

Overview

Authorship

Contact: parkda@kangwon.ac.kr

Model objective

Model objective: Forecast and transfer

Target output: Suitable vs. unsuitable habitat

Focal Taxon

Focal Taxon: *Gekko japonicus* and *Gekko swinhonis*

Location

Location: Northeast Asia (Korean Peninsula, China, Japan), with the geographic extent covering the entire known distribution of *G. japonicus* and *G. swinhonis* (including both native and nonnative ranges).

Scale of Analysis

Spatial extent: 103.7725473777321525, 146.6975473777321213, 21.5184765574970207, 45.7851432241636758 (xmin, xmax, ymin, ymax)

Spatial resolution: 1km

Boundary: rectangle

Biodiversity data

Observation type: citizen science, field survey, range map

Response data type: presence-only, point occurrence

Predictors

Predictor types: climatic, topographic, habitat

Assumptions

Model assumptions: The inclusion of occurrence points from nonnative ranges should not change model prediction outputs significantly.

Algorithms

Modelling techniques: maxent

Model complexity: Selection from an extensive testing of model hyperparameters

Model averaging: No

Workflow

Model workflow: 1) Data preparation 2) Data partitioning 3) Model fitting 4) Model evaluation 5) Model prediction 6) Model transfer

Software

Software: The modeling was conducted using the maximum entropy (MaxEnt) algorithm via the R package *SDMtune* (Vignali et al. 2020), in R version 4.2.2.

Code availability: The data and codes for niche modeling and associated analyses are available from the GitHub repository of Yucheol Shin (https://github.com/yucheols/Gekko).

Data availability: The data and codes for niche modeling and associated analyses are available from the GitHub repository of Yucheol Shin (https://github.com/yucheols/Gekko).

Data

Biodiversity data

Taxon names: *Gekko japonicus*, *Gekko swinhonis*

Ecological level: species

Data sources: From the published literature and iNaturalist Research-Grade observations (https://www.inaturalist.org/) [Accessed on 16 November 2023].

Sampling design: Targeted surveys and opportunistic observations

Sample size: *G. japonicus*: *n* = 985 for initial data, *n* = 309 after spatial thinning. *G. swinhonis*: *n* = 64 for initial data, *n* = 51 after spatial thinning.

Scaling: The occurrence data for both species were spatially thinned by selecting only one coordinate within a 10 km radius of each location data.

Background data: We randomly sampled 10,000 background points across the study extent using the "randomPoints" function of the dismo package (Hijmans et al. 2022).

Errors and biases: For iNaturalist data, the chance of including potentially misidentified observations was minimized by only using "Research-Grade" observations with attached photographs.

Data partitioning

Validation data: 4-fold spatial block cross-validation, with spatial folds generated using the *ENMeval* v2.0 package (Kass et al. 2021).

Predictor variables

Predictor variables:

- Climate: Annual mean temperature, temperature seasonality, and annual precipitation
- Habitat characteristics: Distance to urban areas and distance to forests
- Topography: Altitude

Data sources: The raster layers for climate and altitude data were obtained from WorldClim 2.1 (https://www.worldclim.org/) (Fick and Hijmans 2017). The layers for the distance to forests and urban areas were produced based on the digital land cover map (Copernicus Global Land Service; https://land.copernicus.eu).

Spatial extent: 103.7725473777321525, 146.6975473777321213, 21.5184765574970207, 45.7851432241636758 (xmin, xmax, ymin, ymax)

Spatial resolution: 1km (= 0.008333 dd)

Coordinate reference system: EPSG:4326 (WGS84)

Temporal extent: Climatic data derived from WolrdClim span 1970 to 2000

Transfer data

Data sources: Future climate data was downloaded from WorldClim 2.1 (https://www.worldclim.org/).

Spatial extent: 103.7725473777321525, 146.6975473777321213, 21.5184765574970207, 45.7851432241636758 (xmin, xmax, ymin, ymax)

Spatial resolution: 1km (= 0.008333 dd)

Temporal extent: Climatic averages between 2041 and 2060

Models and scenarios: HadGEM3-GC31 climate model under the Shared Socioeconomic Pathways scenario 2-4.5 (SSP245)

Model

Variable pre-selection

Variable pre-selection: No variable pre-selection was done as the variables were selected based on previous studies.

Multicollinearity

Multicollinearity: No multicollinearity assessment was done as the variables were selected based on previous studies.

Model settings

maxent: Maxent feature classes (L, Q, H, P, LQ, LP, QH, QP, HP, LQH, LQP, LQHP, LQHPT), Regularization multipliers (From 0.5 to 5 at an increment of 0.5)

Model settings (extrapolation): Clamping was applied

Model estimates

Variable importance: Based on permutation importance and percent contribution computed with the "varImp" function of the SDMtune package.

Model selection - model averaging - ensembles

Model selection: We evaluated a total of 130 candidate models per species using spatial block cross-validation, and calculated AUC from the training (AUC_{TRAIN}) and testing data (AUC_{TEST}). To select a model with a low degree of overfitting and high predictive performance, we retained the optimal hyperparameter combinations for each species based on the highest AUC_{TEST} and the lowest AUC_{DIFF} (AUC_{TRAIN} - AUC_{TEST}) (Warren and Seifert 2011).

Threshold selection

Threshold selection: We selected a maximum sum of model sensitivity and specificity (maxSSS) threshold.

Assessment

Performance statistics

Performance on training data: AUC, TSS

Performance on validation data: AUC, TSS

Plausibility check

Response shapes: We examined the response curves of the final model to assess the plausibility of modeled species-environment relationships given the ecology of our target taxa.

Expert judgement: We examined the prediction map to verify the overall consistency with known ranges of the two study taxa.

Prediction

Prediction output

Prediction unit: Relative habitat suitability on a scale of 0 to 1.

Post-processing: The output continuous habitat suitability maps were thresholded to produce binary presence/absence maps

Uncertainty quantification

Novel environments: Clamping was used for all variables during the modeling step, and novel environmental conditions for model transfer were visualized using Multivariate Environmental Similarity Surface (MESS).