Urban environments aid invasion of brown widows (Theridiidae: *Latrodectus geometricus*) in North America, constraining regions of overlap and mitigating potential impact on native widows

– ODMAP Protocol –

Melissa Sadir, Katharine Marske

2021-06-11

Overview

Authorship

Contact: melissa.sadir@ou.edu

Model objective

Model objective: Mapping and interpolation

Target output: continuous habitat suitability

Focal Taxon

Focal Taxon: Latrodectus

Location

Location: North America

Scale of Analysis

Spatial extent: (-140, -53, 14, 63)

Spatial resolution: 5x5 km (2.5 arc-min)

Temporal extent: 1960-2020

Boundary: rectangle

Biodiversity data

Observation type: community science, field survey

Response data type: presence-only

Predictors

Predictor types: climatic, anthropogenic

Hypotheses

Hypotheses: Based on previous studies that suggest a close relationship between brown widows and human-built structures, we expect urbanization to have the strongest effect on the invasive species' distribution.

Assumptions

Model assumptions: 1) Sampling is representative and adequate (after correction) 2) Species are in equilibrium 3) Relevant ecological drivers (or proxies) for species distributions are accounted for 4) Detectability does not change across habitat gradients

Algorithms

Modelling techniques: maxent

Model complexity: We reduced the number of dimensions, minimizing the effects of collinearity, using a Principal Component Analysis and further balanced model complexity by running models across several feature classes ('L', 'LQ', 'LQP') and regularization multipliers (1, 1.5, 2, 2.5, 3) and selecting optimal models from the Corrected Akaike Information Criterion.

Model averaging: 10-fold cross-validation

Workflow

Model workflow: Occurrences were collected from GBIF.org (2019), field collections, and community science efforts on Facebook Groups, Flickr, and outreach presentations. Data was manually cleaned, and questionable occurrences from iNaturalist and BugGuide were visually verified and eliminated if they could not be confirmed. Clustered occurrences were then subsampled to a single point per cell on a 10x10km grid to reduce the effects of sampling bias on model performance. CHELSA environmental variables were downloaded from Paleoclim.org (2019), where they had been rescaled to 2.5 arc-min, and selected based on biological relevance and a PCA to reduce the dimensions of the environmental space of models, which cause potential over-fitting. NASA (2019) human population density data for the years 2000 and 2020 (projected) were downloaded. We created a raster for change in human population between 2000 and 2020 for a population growth variable. All variables were clipped and aligned to (-140, -53, 14, 63) extent and set to WGS-84 projections on ArcMap. We created minimum convex polygons with a 250km buffer for each species' background space on QGIS. We ran two models per species on MaxEnt via ENMeval on R. The first model was run using climatic variables alone, and a second model was run using both climatic and anthropogenic human population density variables. All models were partitioned into 10-fold cross-validation for performance assessment and variable importance for each predictor was assessed with jackknife analyses.

Software

Software: All analyses were conducted on R version 3.6.1 (R Core Team, 2019) using packages raster (Hijmans et al. 2020), ENMeval (Muscarella et al. 2014), dismo (Hijmans et al., 2017), QGIS version 3.10, ArcMap, ENMTools (Warren et al. 2020)

Code availability: GitHub (ecosadir/brownwidow)

Data availability: Supplementary material

Data

Biodiversity data

Taxon names: Latrodectus geometricus, Latrodectus hesperus, Latrodectus mactans, Latrodectus variolus

Taxonomic reference system: N/A

Ecological level: species

Data sources:

GBIF.org (September 2019)

Community science points were collected from volunteers on spider identification Facebook groups and Flickr

Field specimens were collected by trained collectors (Evan Waite, Laura Gatch, Melissa Sadir, Cayley Buckner, Madison Heisey)

Sampling design: Unknown for museum specimens. Community science and survey occurrences were opportunistically sampled

Sample size:

Latrodectus geometricus - 1389 total / 605 subsampled

Latrodectus hesperus - 2433 total / 1154 subsampled

Latrodectus mactans - 2373 total / 1107 subsampled

Latrodectus variolus - 316 total / 239 subsampled

Clipping: North America

Scaling: Variable rasters and spatial point thinning have a 2.5 arc-min resolution. Spatial points were thinned to one point per 10x10km to address potential sampling biases.

Cleaning: Points without coordinates that could not be georeferenced were discounted. Museum occurrences with applicable locality data were georeferenced. Questionable points from iNaturalist and BugGuide were visually verified using photographs. Any occurrences that could not be confirmed were discarded.

Background data: Due to the wide range sample sizes between species, we weighed background points based on sample size. Background points were calculated by multiplying the sample sizes by 5 for consistency. We created minimum convex polygons with a 250km buffer for each species' background space on QGIS.

Errors and biases: For community science data, there is a sampling bias that favors urban areas where the community is more likely to detect and report spiders.

Data partitioning

Training data: 10-fold cross-validation

Validation data: 10-fold cross-validation

Predictor variables

Predictor variables: Bio3 (Isothermality), Bio4 (Temperature Seasonality), Bio5 (Max Temperature of Warmest Month), Bio6 (Min Temperature of Coldest Month), Bio12 (Annual Precipitation), Bio16 (Precipitation of Wettest Quarter), Bio17 (Precipitation of Driest Quarter), NASA 2020 human population density, NASA 2000 human population density

Data sources: PaleoClim (2019), NASA (2019)

Spatial extent: (-140, -53, 14, 63)

Spatial resolution: 2.5 arc-min

Coordinate reference system: WGS-84

Temporal extent: PaleoClim: December 1979 - January 2013 NASA: 2000, 2020

Data processing: Climatic and anthropogenic variable rasters were aligned and clipped on ArcMap. We subtracted NASA 2020 from NASA 2000 rasters to create a new variable: change in human population growth between 2000 and 2020.

Model

Variable Pre-selection

Variable pre-selection: We hypothesized variables that would be most ecologically relevant to the species.

Multicollinearity

Multicollinearity: In order to minimize collinearity, we initially ran a principal component analysis (PCA) to reduce the pool of relevant climatic variables and subsequently selected the most biologically significant variables from the pool (Saupe et al. 2011, Wang et al. 2018). Climatic niche models were constructed using seven CHELSA variables from PaleoClim: isothermatlity, temperature seasonality, maximum temperature of warmest month, minimum temperature of coldest month, annual precipitation, precipitation of the warmest quarter, and precipitation of the coldest quarter. Anthropogenic niche models included climatic variables in addition to two human population density variables from NASA Socioeconomic Data and Applications Center (SEDAC): human population density in 2020, and change in human population density between 2000-2020.

Model settings

maxent: featureSet (L, LQ, LQP) regularizationMultiplierSet (1, 1.5, 2, 2.5, 3), targetGroupSampleSize (L. geometricus - 604; L. hesperus - 1154; L. mactans - 1107; L. variolus - 239)

Model estimates

Parameter uncertainty: Spatial autocorrelation was assessed using Moran's I

Variable importance: Univariate variable importance was assessed using jackknife analyses on maxent

Analysis and Correction of non-independence

Spatial autocorrelation: Spatial autocorrelation for each species was assessed using Moran's I

Assessment

Performance statistics

Performance on training data: AUC

Performance on validation data: AICc, AUC

Performance on test data: AICc, AUC

Plausibility check

We checked model plausibility by assessing partial dependence plots

Prediction

Prediction output

Prediction unit: Models outputs were used to predict habitat suitability by mapping occurrence probabilities.

Post-processing: We produced a clog-log map that rescales occurrence probability on a scale from 0 to 1. The Caribbean and Hawaiian islands were clipped from the output.

Uncertainty quantification

Input data uncertainty: variable collinearity, sampling bias