



OPEN

DATA DESCRIPTOR

VenomMaps: Updated species distribution maps and models for New World pitvipers (Viperidae: Crotalinae)

Rhett M. Rautsaw¹✉, Gustavo Jiménez-Velázquez^{2,3}, Erich P. Hofmann⁴,
Laura R. V. Alencar¹, Christoph I. Grünwald⁵, Marcio Martins⁶, Paola Carrasco^{7,8},
Tiffany M. Doan⁹ & Christopher L. Parkinson^{1,10}✉

Beyond providing critical information to biologists, species distributions are useful for naturalists, curious citizens, and applied disciplines including conservation planning and medical intervention. Venomous snakes are one group that highlight the importance of having accurate information given their cosmopolitan distribution and medical significance. Envenomation by snakebite is considered a neglected tropical disease by the World Health Organization and venomous snake distributions are used to assess vulnerability to snakebite based on species occurrence and antivenom/healthcare accessibility. However, recent studies highlighted the need for updated fine-scale distributions of venomous snakes. Pitvipers (Viperidae: Crotalinae) are responsible for >98% of snakebites in the New World. Therefore, to begin to address the need for updated fine-scale distributions, we created VenomMaps, a database and web application containing updated distribution maps and species distribution models for all species of New World pitvipers. With these distributions, biologists can better understand the biogeography and conservation status of this group, researchers can better assess vulnerability to snakebite, and medical professionals can easily discern species found in their area.

Background & Summary

Knowing where a species occurs is critical for understanding numerous aspects of biology including evolution, biogeography, ecology, and conservation. Species distributions are also important to applied disciplines outside biology. The general public and citizen scientists utilize published distributions in field guides to assist with identification, government agencies utilize known distributions to better plan management strategies, and medical practitioners must maintain a working knowledge of dangerous taxa in their area to better treat patients afflicted during an encounter. One group which highlights the importance of having well-described distributions is venomous snakes. Snake venom can vary tremendously both between and within a single species^{1–4} and envenomation by venomous snakes (hereafter, snakebite) is regarded as a priority neglected tropical disease by the World Health Organization due to the nearly 100,000 deaths and 400,000 disablements that occur globally every year⁵. Therefore, knowing which venomous snake species occur in a given area can inform snakebite risk analyses and inform medical treatment of snakebite^{6–8}.

Venomous snakes have a cosmopolitan distribution, occurring on every continent but Antarctica. Most medically significant species (*i.e.*, those resulting in hospitalization, permanent injury, or death to humans)

¹Clemson University, Department of Biological Sciences, Clemson, SC, USA. ²Universidad Nacional Autónoma de México, Posgrado en Ciencias Biológicas, Mexico City, Mexico City, Mexico. ³Vida Silvestre Coatl AC, Mexico City, Mexico City, Mexico. ⁴Cape Fear Community College, Science Department, Wilmington, NC, USA. ⁵HERP.MX, Villa del Álvarez, Colima, Mexico. ⁶Universidade de São Paulo, Departamento de Ecologia, Instituto de Biociências, São Paulo, São Paulo, Brazil. ⁷Universidad Nacional de Córdoba, Facultad de Ciencias Exactas, Físicas y Naturales, Centro de Zoología Aplicada, Córdoba, Argentina. ⁸Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Instituto de Diversidad y Ecología Animal (IDEA), Córdoba, Argentina. ⁹New College of Florida, Division of Natural Sciences, Sarasota, FL, USA. ¹⁰Clemson University, Department of Forestry and Environmental Conservation, Clemson, SC, USA. ✉e-mail: rrautsa@g.clemson.edu; viper@clemson.edu

fall into one of two families: Viperidae (vipers) and Elapidae (elapids). Other families such as Atractaspididae, Colubridae, and Dipsadidae also contain species capable of inflicting medically significant bites; however, they make up a small proportion of envenomations compared to vipers and elapids^{9–12}. In general, vipers contribute the most snakebites globally^{9–12}, particularly in North and South America (*i.e.*, the New World) where pitvipers (subfamily Crotalinae) such as rattlesnakes (*Crotalus* and *Sistrurus*), cantils (*Agkistrodon*), and lanceheads (*Bothrops*) are responsible for more than 98% of envenomations^{9,11,12}. Many species of New World pitvipers exhibit functional venom variation both between species and within species across geographic space^{1–4} which impacts snakebite treatment. As such, clear delimitation of species' ranges can inform medical treatment and antivenom use to better account for interspecific venom variation.

Unfortunately, the distributions of many species of venomous snakes remain unrefined, impacting our biological understanding of these organisms and limiting snakebite epidemiology. The need for refined species distributions of venomous species is well-documented, even in thorough studies synthesizing species distributions with epidemiological data. Hansson *et al.*⁷ used snake distributions to identify areas in need of improved accessibility to antivenom in Costa Rica⁷. Similarly, Yañez-Arenas *et al.*⁶ demonstrated that species distribution models were able to explain up to 35% of the variation in the incidence of snakebites across Veracruz, Mexico and could be used to infer potential areas of high snakebite risk. A more recent study used venomous snake distributions and occurrence records to map global vulnerability to snakebite envenoming based on antivenom availability, hospital accessibility, and the Healthcare Access and Quality (HAQ) index⁸. Each of these studies was limited by the available digitized distribution information^{6–8}. For example, Longbottom *et al.*⁸ excluded nine species due to an absence of geographical information, identified 216 species that require distributional assessments, and emphasized the importance of and need for fine-scale ($\leq 10 \text{ km}^2$) distributional information for venomous snake species⁸.

Here, we focus on enhancing fine-scale distributional information for venomous snakes by compiling novel distribution maps and species distribution models (SDMs) for all 158 species of pitvipers in North, Central, and South America in an easy to use, publicly available web interface. We use occurrence records, published distribution maps, ecoregion maps, and a relief map to manually reconstruct the distributions of these species. Species Distribution Modeling (SDM) and Ecological Niche Modeling (ENM) are also often used to estimate the geographic ranges. Although there are theoretical differences between these two tools, both utilize niche theory to model processes that shape distributions based on statistical associations between environmental predictors and records of species presence^{13,14}. Therefore, in addition to our curated distribution maps, we perform species distribution modeling to capture fine-scale (1 km^2) distribution information that may be missed in our hand-curated, generalized distributions. We chose SDM because our goal is to map the species distribution rather than understand the processes/factors underlying the niche. This dataset includes approximately 74 of the 216 species of venomous snakes identified by Longbottom *et al.*⁸ as needing reassessment and adds 69 additional species that were either recently described or did not have prior distribution information and were thus excluded in these analyses⁸. These distribution maps can be used for a variety of purposes including snakebite vulnerability assessment, informing medical treatment via species identification, assessing access to appropriate antivenom, biogeographic analyses, conservation assessments, and for general information on species ranges. Finally, we provide all the code and methodologies used for range estimation and the final maps in an intuitive user-friendly, publicly-accessible GitHub repository (github.com/RhettRautsaw/VenomMaps) and Shiny application (rhettrautsaw.app/shiny/VenomMaps) available from any computer with an internet connection. Stable releases are archived on Zenodo¹⁵.

Methods

The custom code used to clean occurrence records and construct SDMs is available at (github.com/RhettRautsaw/VenomMaps). We used the following R¹⁶ packages for data cleaning, manipulation, species distribution modeling, and Shiny app creation: *tidyverse*¹⁷, *readxl*¹⁸, *data.table*¹⁹, *sf*²⁰, *sp*^{21,22}, *rgdal*²³, *raster*²⁴, *smoothr*²⁵, *ape*²⁶, *phytools*²⁷, *argparse*²⁸, *parallel*¹⁶, *memuse*²⁹, *dismo*³⁰, *rJava*³¹, *conconveman*³², *spThin*³³, *usdm*³⁴, *ENMeval*³⁵, *kuenm*³⁶, *shiny*³⁷, *leaflet*³⁸, *leaflet.extras*³⁹, *leaflet.extras2*⁴⁰, *RColorBrewer*⁴¹, *ggpubr*⁴², *ggtext*⁴³, and *patchwork*⁴⁴.

Updating occurrence record taxonomy. Our goal was to update and reconstruct the distributions of New World pitvipers. We used the Reptile Database⁴⁵ (May 2021) as our primary source for current taxonomy which included the following genera: *Agkistrodon*, *Atropoides*, *Bothriechis*, *Bothrocophias*, *Bothrops*, *Cerrophidion*, *Crotalus*, *Lachesis*, *Metlapilcoatlus*, *Mixcoatlus*, *Ophryacus*, *Porthidium*, and *Sistrurus*. However, to ensure we captured all New World pitvipers records, we incorporated all members of the family Viperidae (all vipers and pitvipers) into our pipeline for updating occurrence record taxonomy (*i.e.*, to account for errors in the recorded latitude, longitude, or if subfamily was not recorded).

First, we collected global occurrence records for “Viperidae” from GBIF (downloaded 2021-08-19)⁴⁶, Bison (downloaded 2021-08-19)⁴⁷, HerpMapper (only New World taxa; downloaded 2021-08-19)⁴⁸, Brazilian Snake Atlas⁴⁹, BioWeb (downloaded 2021-07-07)⁵⁰, unpublished data/databases from RMR, GJV, EPH, LRVA, MM, and CLP, and georeferenced literature records totaling 373,673 species-level records, 292,425 of which are New World pitvipers. Given the fluidity of taxonomy, records were often associated with outdated names. For example, *Crotalus mitchelli pyrrhus* was elevated to *Crotalus pyrrhus*⁵¹, but may still be recorded as the former in a given repository (*e.g.*, GBIF). To correct taxonomy in our database, we checked records against a list of synonyms found on the Reptile Database and compared them to current taxonomy. If species and subspecies columns matched the same taxon (or no subspecies was recorded), then species IDs were not altered. If species and subspecies IDs did not match the same taxon, we updated taxonomy by minimizing the number of changes required to a given character string. We then manually checked all changes.

Constructing distribution maps. Next, we collected preliminary distribution maps from the International Union for Conservation of Nature (IUCN; downloaded 2018-11-27)⁵², Global Assessment of Reptile Distributions

(GARD) v1.1⁵³, Heimes⁵⁴, Campbell and Lamar⁵⁵, and unpublished maps. We manually curated distribution maps for all New World pitvipers in QGIS using the occurrence records, previous distribution maps, and recent publications for each taxon (note that distributions for Old World Viperidae have not yet been updated). We used a digital relief map (maps-for-free.com) and The Nature Conservancy Terrestrial Ecoregions (TNG.org)⁵⁶ to identify clear distribution boundaries (e.g., mountains). We then clipped the final distributions to a land boundary (GADM v3.6)⁵⁷ and smoothed the distribution using the “chaikin” method in the R package *smoothr*²⁵.

Occurrence-distribution overlap. Our initial taxonomy check was only concerned with records for which a subspecies was recorded and had since been elevated to species status. Therefore, many records with no assigned subspecies likely remained associated with an incorrect or outdated generic and/or specific identification. Fortunately, taxonomic changes are typically associated with changes in the species’ expected distribution. For example, when *Crotalus simus* was resurrected from *C. durissus*, the distribution of *C. durissus* was split: the northern portion of its range in Central America now represented the resurrected species (*C. simus*) and the southern portion of its range remained *C. durissus*⁵⁵. Yet, occurrence records in Central America often remain labelled as *C. durissus* in data repositories. Therefore, we spatially joined records with the newly reconstructed species distribution maps to determine if they overlapped with their expected distribution (Old World taxa were joined with the GARD 1.1 distributions⁵³).

Briefly, we developed a custom function (`occ_cleaner.R`) to perform the spatial join and update taxonomy. First, we calculated the distance for each record to the 20 nearest distributions within 50 km (full overlap resulted in a distance of 0 m). Next, we calculated the phylogenetic distance between the recorded species ID and each species with which that record overlapped using the tree from Zaher *et al.*⁵⁸ and adding taxa based on recent clade-specific publications (`bind.tip2.R`; see github.com/RhettRautsaw/VenomMaps for full list of references and details). If records overlapped with their expected species, no changes were made. If records fell outside of their expected distribution, we filtered the potential overlapping and nearby species (within 50 km) to minimize phylogenetic distance. If multiple species were equally distant (i.e., share the same common ancestor), we attempted to minimize geographic distance. If multiple species remained equally distant in both phylogenetic and geographic distance, we flagged the record to be manually checked. We also flagged records if a species’ taxonomy had changed and records were additionally flagged as potentially dubious if the taxonomic change had a phylogenetic divergence greater than 5 million years. We manually checked all flagged records and returned records to their original species ID if species identity remained uncertain. We flagged these records as potentially dubious, along with records that fell outside of their expected distribution (within 50 km), and removed all flagged records for species distribution modeling. Our final cleaned database contained 344,998 global records, of which 275,087 were New World pitvipers.

Species distribution modeling. We attempted to infer SDMs for the 158 species of New World pitvipers currently recognized by the Reptile Database (May 2021) and additionally modeled the three subspecies of *Crotalus ravus* separately based on recommendations for species status elevation by Blair *et al.*⁵⁹ for a total of 160 species. We developed a unix-executable R script (`autokuenm.R`) designed to take occurrence records, distribution maps, and environmental data and prepare these data for species distribution modeling with *kuenm*³⁶. We chose to use *kuenm* – and MaxEnt v3.4.4⁶⁰ – because it has been shown to have good predictive power⁶¹ and fine-tuning of this algorithm has performance comparable to more computationally intensive ensembles^{62,63}. Additionally, MaxEnt allows for flexibility in parameter selection⁶⁴ and can function entirely with presence data¹⁴.

Prior to `autokuenm`, to account for sampling/spatial bias during SDM, we created a bias file by using the pooled New World pitviper occurrence records as representative background data^{65–68}. Specifically, we converted occurrence records to a raster and performed two-dimensional kernel density estimation (`kde2d`) with the *MASS* package with default settings⁶⁹ and rescaled the kernel density by a factor of 1000 and rounded to three decimal places. This was then used as input to factor out sampling bias by MaxEnt. We then ran `autokuenm`, which is designed to subset/partition the cleaned occurrence records for a given species and prepare additional files for SDM. We first defined M-areas – or areas accessible to a given species – using the World Wildlife Fund Terrestrial Ecoregions⁷⁰. Biogeographic regions represent distributional limits for many species and are reasonable hypotheses for the areas accessible to a given species^{71,72}. To do this, we created alpha hulls from the subset of occurrence records for a given species using `concaveman`³² with default settings. We then identified regions with at least 20% of the region covered by the alpha hull and merged these regions together to form our final M-area. All environmental layers and the bias file were cropped to this M-area which was used as the geographic extent for modeling. We then randomly selected 5% of records to function as an independent test set for final model evaluation. Next, we generated 2000 random background points across the cropped environmental layers and used ENMeval to partition occurrence records into four sets using the checkerboard2 pattern³⁵. Note that the background points here were not used in MaxEnt. One of the four partitions was selected at random to be used as the testing set; the remaining three partitions were used for training the MaxEnt models. If the number of occurrence records in the independent test set was less than five, then we used the training partition for final model creation and used the testing partition for final model evaluation.

We tested the top-contributing variables from three sets of environmental layers: (1) bioclimatic variables, (2) EarthEnv topographic variables⁷³, and (3) a combination of these variables. To select the top-contributing variables in each set, we wrote a custom function (`SelectVariables`) which used a combination of MaxEnt permutation importance and Variable Inflation Factors (VIF) to remove collinearity while keeping the variables that contributed the most to the model. Compared with variable selection via principal component analysis loadings, the permutation importance and VIF methodology demonstrated significant improvement in MaxEnt model fit. First, we designed `SelectVariables` to run MaxEnt using `dismo::maxent` with default settings and then extracted the permutation importance. We removed variables if they had 0% permutation importance. Next, we calculated

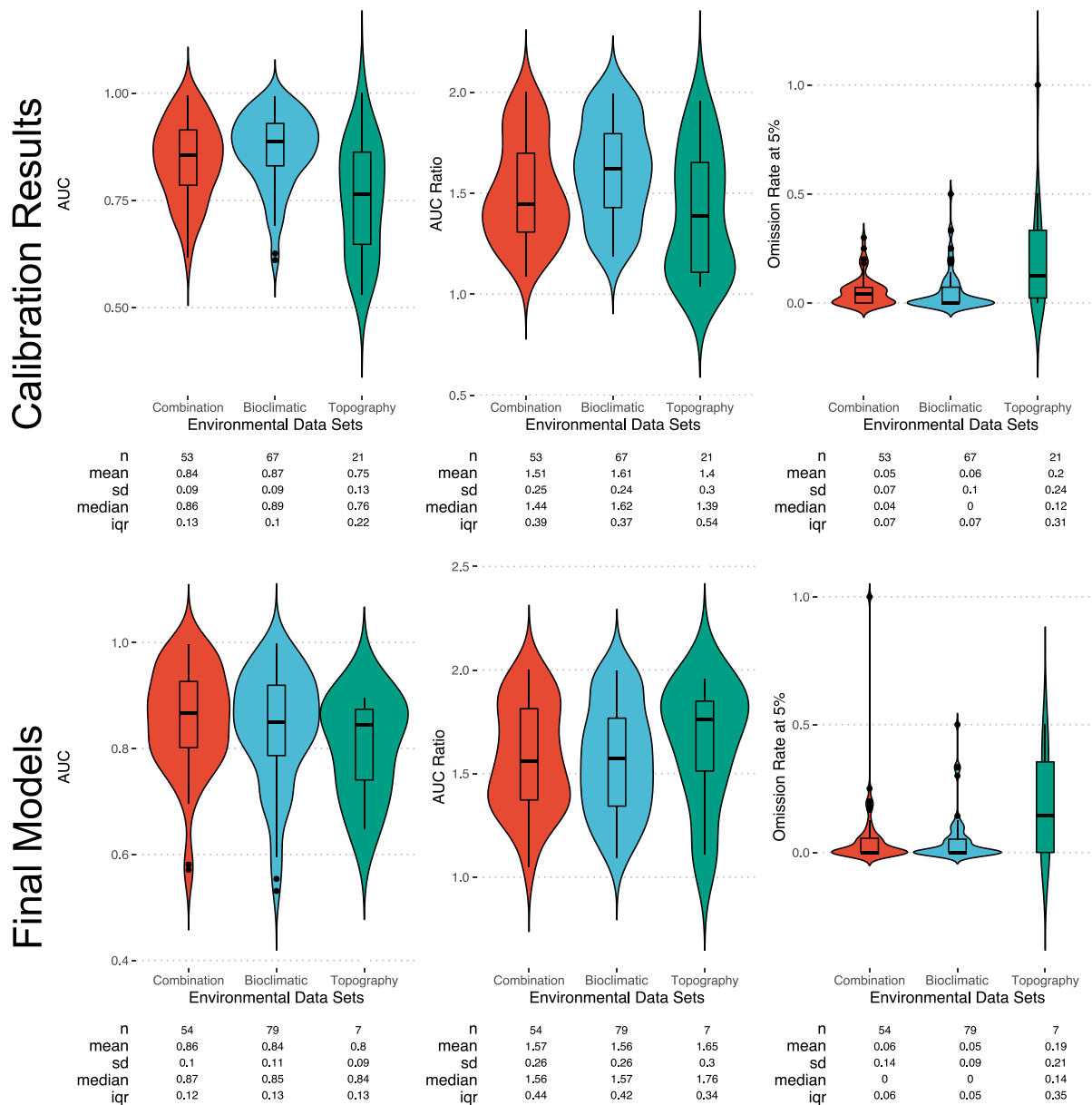


Fig. 1 Area Under the Curve (AUC), AUC ratios, and Omission Rates for SDMs. Top row represents the best models selected after *kuenm* calibration of 2,958 models. Bottom row represents the selected final or averaged models for each species.

VIF with *usdm::vif* and then iteratively removed variables by selecting the variables with two highest VIF values and removing whichever variable had the lowest permutation importance. We then recalculated VIF and repeated the process until the maximum VIF value was less than 10. Finally, we recalculated permutation importance with the remaining variables using *dismo::maxent* with default settings and removed variables with less than 1% permutation importance to create the final variable sets. This process was done for each species independently.

With the final environmental variable, testing, and training sets, we generated SDMs using *kuenm*. First, we created candidate calibration models with multiple combinations of regularization multipliers (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 8, 10), feature classes (l, q, h, lq, lp, lt, lh, qp, qt, qh, pt, ph, th, lqp, lqt, lqh, lpt, lph, lth, qpt, qph, qth, pth, lqpt, lqph, lqth, lpth, qpth, lqpth), and sets of environmental predictors (bioclimatic, topographic, combination) totaling 2,958 candidate models per species. We then ran each model in parallel using GNU Parallel⁷⁴. Next, we evaluated the candidate models and selected the best models using statistical significance (partial ROC), prediction ability (omission rates; OR), and model complexity (AICc) with the “*kuenm_ceval*” function with default settings. Specifically, models were only considered if they were statistically significant and had an OR less than 5%. If no models passed the OR criteria, the models with the minimal OR were considered. Finally, any remaining models were filtered to those within 2 AICc of the top model (Supplementary Table 1). In addition to evaluating and comparing all models together, we evaluated bioclimatic-only and combination-only

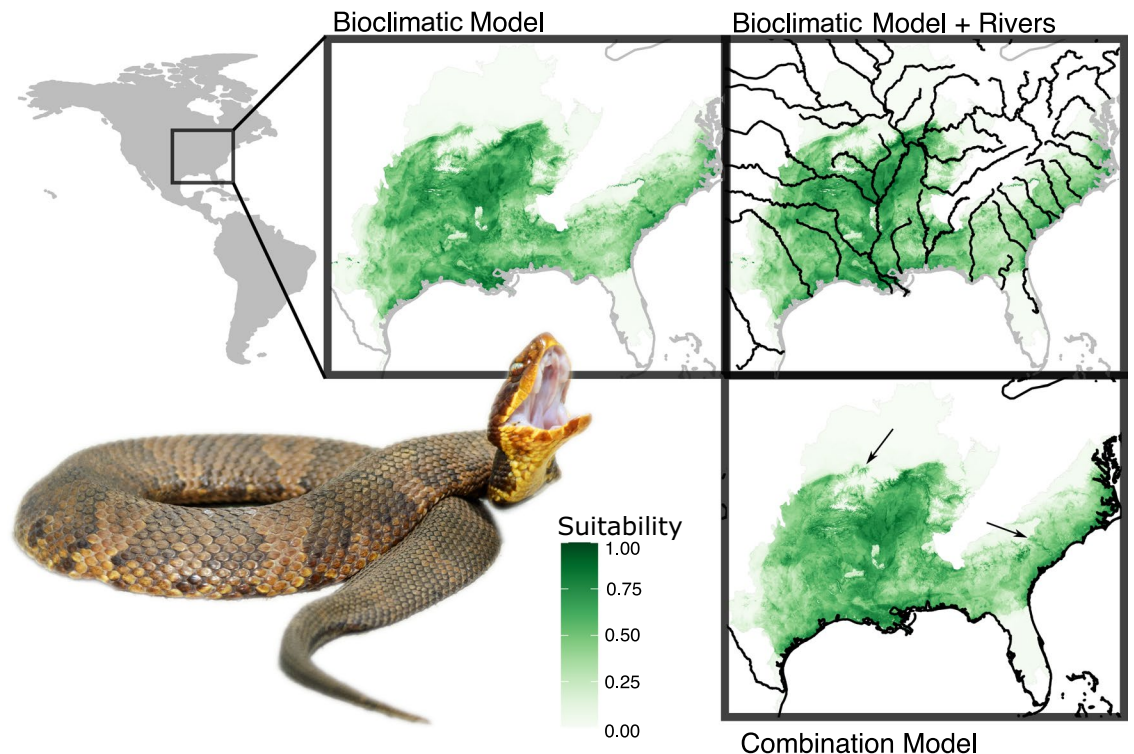


Fig. 2 Bioclimatic and Combination SDMs for *Agkistrodon piscivorus* demonstrate how inclusion of additional variables (other than bioclimatic) can refine SDMs to more precisely produce a more precise map of expected habitats for a given species, such as rivers. Arrows in the combination model point to areas where the combination model refines the bioclimatic model. Photo Credit: *A. piscivorus* (No copyright; Jesus Moreno, United States Fish and Wildlife Service).

models separately since these two sets of environmental variables were expected to be the best performing models given the ubiquity of bioclimatic variables in species distribution modeling (Supplementary Table 1).

We generated 10 bootstrap replicates for each of the “best” calibration models using the “*kuenm_mod*” function. We also performed jackknifing to assess variable importance and models were output in raw format. We evaluated the final models using “*kuenm_feval*” with default settings. To select the best model for each comparative set (*i.e.*, all, bioclimatic-only, and combination-only sets), we filtered the final evaluation results to minimize the OR and maximize the AUC ratio (Supplementary Table 2). If multiple models remained and were considered equally competitive, we averaged these models together (Supplementary Table 3). Because we performed three different sets of comparisons, there were three “best” models per species, so we again aimed to minimize the OR and maximize the AUC ratio to select a final model for each species (Supplementary Table 4). We then converted our final models into cloglog format for visualization and thresholded the models using a 10th percentile training presence cutoff (Fig. S2). Both conversion and thresholding functions are provided as R functions (*raw2log*, *raw2clog*, *raster_threshold* in functions.R; github.com/RhettRautsaw/VenomMaps).

Data Records

A project repository containing all data is available on GitHub (github.com/RhettRautsaw/VenomMaps) with stable releases archived in Zenodo¹⁵. All distribution maps are available as vector-based geojson files in the project repository under the `data/distributions` directory. All cleaned occurrence records are available as Microsoft Excel files in the project repository under the `data/occurrence` directory. All final SDMs are available as raster .tif files under the `data/enm` directory. This work is licensed under a [CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/).

Updates. Distributions for Old World Viperidae are currently provided via GARD 1.1⁵³; however, the methodology and code used here (*i.e.*, *VenomMaps*) can be made the standard for range estimation and understanding species distributions. We hope to expand this dataset in the future to update the distributions of Old World Viperidae and incorporate other medically significant snake families. We plan to push updates annually and welcome contributions from other experts in the field.

Technical Validation

SDM evaluation. Of the 160 species of New World pitvipers (including recommendations from Blair *et al.* 2018), 20 had insufficient occurrence records (1–26 records) or occupied a limited geographic area (*e.g.* insular taxa) and species distribution modeling could not be performed (Supplementary Table 4). When comparing all possible model sets, *kuenm* selected a combination of variables as the best model set for 53 species,

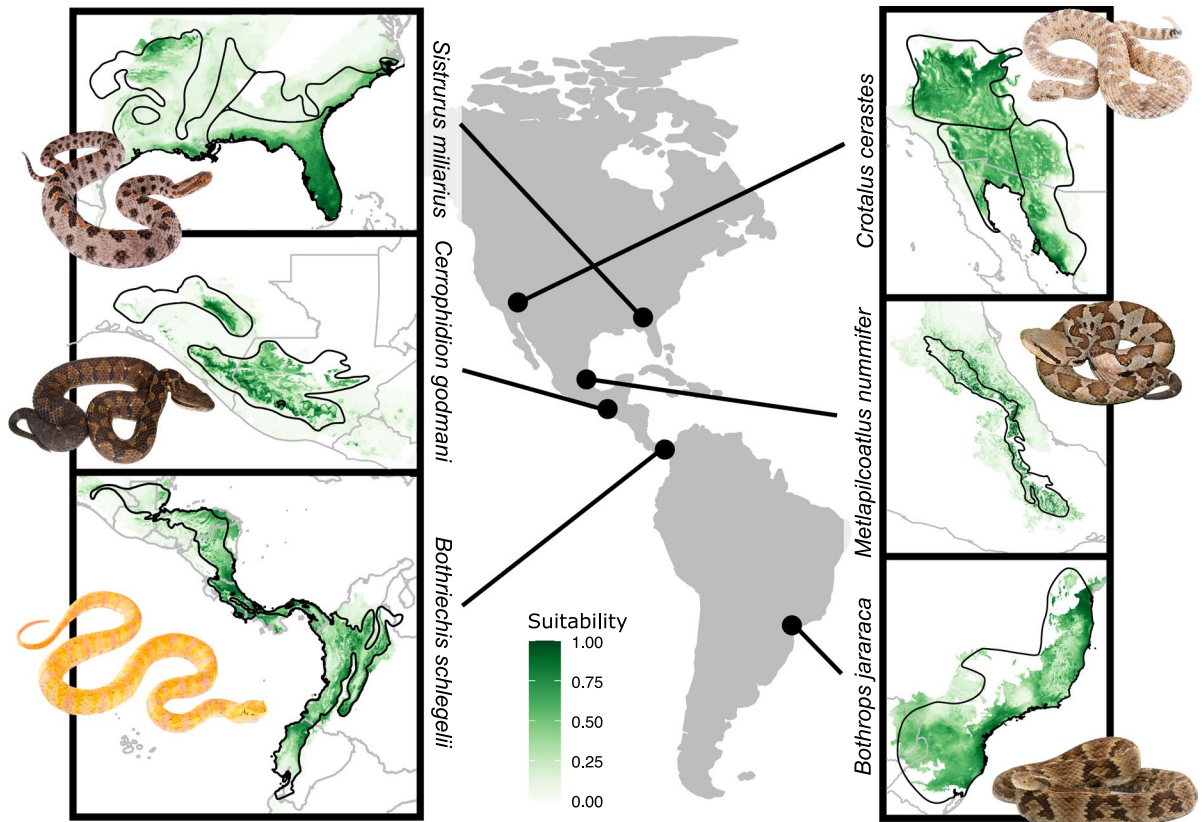


Fig. 3 Distribution maps and SDMs of nine representative species. Distribution maps are outlined with black (subspecies boundaries displayed in *Crotalus cerastes* and *Sistrurus miliarius*). Photo Credit: *Sistrurus miliarius* and *Crotalus cerastes* (Tristan Schramer, Clemson University); *Cerrophiidion godmani* and *Metlapilcoatlus nummifer* (Jason Jones, Herp.MX); *Bothriechis schlegelii* (Tropical Herping, tropicalherping.com); *Bothrops jararaca* (Wellington Coelho).

bioclimatic-only for 67 species, and topography-only for 21 species (Supplementary Table 2). Model Area Under the Curve (AUC) ratios ranged from 1.03–1.99 with a median of 1.53 (Fig. 1; Supplementary Table 2). Topography-only models generally had lower AUC ratios (median = 1.39), compared to bioclimatic-only and combination models (median = 1.62 & 1.44, respectively; Supplementary Table 2). Additionally, when topography was selected, models generally had a high suitability across the entire modeling extent (Fig. S1; Supplementary Table 2). Omission rates were similarly lower for bioclimatic and combination models (median = 0.00 & 0.04, respectively) compared to topographic models (median = 0.12). Topography-only models were generally only chosen when species had a low number of occurrence records ($\bar{x}_{\text{topographic}} = 295$; $\bar{x}_{\text{bioclimatic}} = 386$; $\bar{x}_{\text{combination}} = 892$; Supplementary Table 2). Given the poorer fit of topography-only models despite being selected in the comparison of all models, we also performed comparisons of bioclimatic-only and combination-only models separately.

After comparing bioclimatic-only and combination-only models, the selected models in these comparisons often had lower omission rates and higher AUC ratios than the comparison of all models. Therefore, we selected the model from across the three comparative sets which minimized omission rate and maximized AUC Ratio as our final model. In the final model set, a combination of bioclimatic and topographic variables was selected as the best model for 54 species, bioclimatic-only variables was selected as the best model for 79 species, and 7 species remained supporting topographic-only models (Supplementary Table 3; Supplementary Table 4). AUC ratio ranged from 1.04–1.99 with a median of 1.58 and omission rates ranged from 0.00 to 0.14 with a median of 0.00 (Fig. 1; Supplementary Table 3; Supplementary Table 4).

Although topography variables often resulted in worse models, they likely facilitated micro-habitat refinement of the models. For example, including elevation variables into the final combination model for *Agkistrodon piscivorus* – an aquatic specialist – produced a model which closely followed riverine habitats and reduced presence in high-elevation habitats (Fig. 2). The combination models, therefore, more accurately traced the expected distribution and habitats for a given species. Finally, the SDMs demonstrated that the constructed distribution maps closely matched the optimal habitat for many species (Fig. 3; Fig. S2). Final SDM statistics are available in Supplementary Table 4.

Usage Notes

Data, distributions, SDMs, and code can be accessed from the VenomMaps GitHub repository (github.com/RhettRautsaw/VenomMaps) with stable releases archived on Zenodo¹⁵. To aid in public accessibility and utility, we developed a R Shiny app to view distribution maps, occurrence records, and SDMs (rhetttrautsaw.app/shiny/VenomMaps). A user guide for the Shiny App is available on the GitHub repository. In the Shiny app, distributions can be filtered by country or available SDM. Additional information on maximum size of each species was compiled from Feldman *et al.*⁷⁵ and is easily viewed in secondary tab for “General Information”.

Code availability

All code, including custom scripts such as `occ_cleaner`, `bind.tip2`, `SelectVariables`, and `autokuenm` discussed above, are available as R scripts and summarized in Markdown format under the code directory in the GitHub project repository (github.com/RhettRautsaw/VenomMaps). We also provide functions to convert SDM outputs (R function: `raw2clog`, `raw2log`, `log2raw`) and threshold models (R function: `raster_threshold`) in the functions.R script.

Received: 2 November 2021; Accepted: 19 April 2022;

Published online: 25 May 2022

References

- Margres, M. J., Bigelow, A. T., Lemmon, E. M., Lemmon, A. R. & Rokyta, D. R. Selection to increase expression, not sequence diversity, precedes gene family origin and expansion in rattlesnake venom. *Genetics* **206**, 1569–1580, <https://doi.org/10.1534/genetics.117.202655> (2017).
- Strickland, J. L. *et al.* Evidence for divergent patterns of local selection driving venom variation in Mojave Rattlesnakes (*Crotalus scutulatus*). *Scientific Reports* **8**, 17622, <https://doi.org/10.1038/s41598-018-35810-9> (2018).
- Mason, A. J. *et al.* Trait differentiation and modular toxin expression in palm-pitvipers. *BMC Genomics* **21**, 147, <https://doi.org/10.1186/s12864-020-6545-9> (2020).
- Holding, M. L. *et al.* Phylogenetically diverse diets favor more complex venoms in North American pitvipers. *Proceedings of the National Academy of Sciences* **118**, e2015579118, <https://doi.org/10.1073/pnas.2015579118> (2021).
- Gutiérrez, J. M. *et al.* Snakebite envenoming. *Nature Reviews Disease Primers* **3**, 17063, <https://doi.org/10.1038/nrdp.2017.63> (2017).
- Yañez-Arenas, C., Peterson, A. T., Mokondoko, P., Rojas-Soto, O. & Martínez-Meyer, E. The use of ecological niche modeling to infer potential risk areas of snakebite in the Mexican State of Veracruz. *PLoS ONE* **9**, <https://doi.org/10.1371/journal.pone.0100957> (2014).
- Hansson, E., Sasa, M., Mattisson, K., Robles, A. & Gutiérrez, J. M. Using geographical information systems to identify populations in need of improved accessibility to antivenom treatment for snakebite envenoming in Costa Rica. *PLoS Neglected Tropical Diseases* **7**, <https://doi.org/10.1371/journal.pntd.0002009> (2013).
- Longbottom, J. *et al.* Vulnerability to snakebite envenoming: a global mapping of hotspots. *The Lancet* **392**, 673–684, [https://doi.org/10.1016/S0140-6736\(18\)31224-8](https://doi.org/10.1016/S0140-6736(18)31224-8) (2018).
- Gutiérrez, J. M. In *Handbook of Venoms and Toxins of Reptiles*, 1 Mackessy, S. P. (ed.) chap. 24, 491–508 (CRC Press, 2010).
- Chippaux, J.-P. In *Handbook of Venoms and Toxins of Reptiles*, 1 Mackessy, S. P. (ed.) chap. 22, 453–474 (CRC Press, 2010).
- Smith, J. & Bush, S. In *Handbook of Venoms and Toxins of Reptiles*, 1 Mackessy, S. P. (ed.) chap. 23, 475–490 (CRC Press, 2010).
- Mackessy, S. P. (ed.) *Handbook of Venoms and Toxins of Reptiles* 1 edn. (CRC Press, 2010).
- Peterson, A. T. & Soberón, J. Species distribution modeling and ecological niche modeling: Getting the concepts right. *Natureza e Conservacao* **10**, 102–107, <https://doi.org/10.4322/natcon.2012.019> (2012).
- Yañez-Arenas, C., Castaño-Quintero, S., Rioja-Nieto, R., Rodríguez-Medina, K. & Chiappa-Carrara, X. Assessing the relative role of environmental factors that limit the distribution of the Yucatan Rattlesnake (*Crotalus tzabcan*). *Journal of Herpetology* **54**, 216, <https://doi.org/10.1670/19-055> (2020).
- Rautsaw, R. M. RhettRautsaw/VenomMaps: VenomMaps v1.2. *Zenodo* <https://doi.org/10.5281/zenodo.5637094> (2022).
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. (2020).
- Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* **4**, 1686, <https://doi.org/10.21105/joss.01686> (2019).
- Wickham, H. & Bryan, J. *readxl: Read Excel Files* <https://CRAN.R-project.org/package=readxl> (2019).
- Dowle, M. & Srinivasan, A. *data.table: Extension of 'data.frame'* <https://CRAN.R-project.org/package=data.table> (2021).
- Pebesma, E. Simple features for R: Standardized support for spatial vector data. *The R Journal* **10**, 439–446, <https://doi.org/10.32614/RJ-2018-009> (2018).
- Pebesma, E. J. & Bivand, R. S. Classes and methods for spatial data in R. *R News* **5**, 9–13 (2005).
- Bivand, R. S., Pebesma, E. & Gomez-Rubio, V. *Applied spatial data analysis with R, Second edition* (Springer, NY, 2013).
- Bivand, R., Keitt, T. & Rowlingson, B. *rgdal: Bindings for the 'Geospatial' Data Abstraction Library* <https://CRAN.R-project.org/package=rgdal> (2021).
- Hijmans, R. J. *raster: Geographic Data Analysis and Modeling* <https://CRAN.R-project.org/package=raster> (2021).
- Strimas-Mackey, M. *smoothr: Smooth and Tidy Spatial Features* <https://CRAN.R-project.org/package=smoothr> (2020).
- Paradis, E. & Schliep, K. *ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R*. *Bioinformatics* **35**, 526–528 (2019).
- Revell, L. J. *phytools: An R package for phylogenetic comparative biology (and other things)*. *Methods in Ecology and Evolution* **3**, 217–223, <https://doi.org/10.1111/j.2041-210X.2011.00169.x> (2012).
- Davis, T. L. *argparse: Command Line Optional and Positional Argument Parser* <https://CRAN.R-project.org/package=argparse> (2021).
- Schmidt, D. *memuse: Memory Estimation Utilities* <https://cran.r-project.org/package=memuse> (2020).
- Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. *dismo: Species Distribution Modeling* <https://CRAN.R-project.org/package=dismo> (2020).
- Urbanek, S. *rJava: Low-Level R to Java Interface* <https://CRAN.R-project.org/package=rJava> (2021).
- Gombin, J., Vaidyanathan, R. & Agafonkin, V. *concaveman: A very fast 2D concave hull algorithm* <https://CRAN.R-project.org/package=concaveman> (2020).
- Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B. & Anderson, R. P. *spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models*. *Ecography* **38**, 541–545, <https://doi.org/10.1111/ecog.01132> (2015).
- Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K. & Toxopeus, A. G. Where is positional uncertainty a problem for species distribution modelling. *Ecography* **37**, 191–203, <https://doi.org/10.1111/j.1600-0587.2013.00205.x> (2014).

35. Muscarella, R. *et al.* ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution* **5**, 1198–1205, <https://doi.org/10.1111/2041-210x.12261> (2014).
36. Cobos, M. E., Townsend Peterson, A., Barve, N. & Osorio-Olvera, L. Kuenm: An R package for detailed development of ecological niche models using Maxent. *PeerJ* **2019**, 1–15, <https://doi.org/10.7717/peerj.6281> (2019).
37. Chang, W. *et al.* shiny: Web Application Framework for R <https://CRAN.R-project.org/package=shiny> (2021).
38. Cheng, J., Karambelkar, B. & Xie, Y. leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library <https://CRAN.R-project.org/package=leaflet> (2021).
39. Karambelkar, B. & Schloerke, B. leaflet.extras: Extra Functionality for 'leaflet' Package <https://CRAN.R-project.org/package=leaflet.extras> (2018).
40. Sebastian, G. leaflet.extras2: Extra Functionality for 'leaflet' Package <https://CRAN.R-project.org/package=leaflet.extras2> (2020).
41. Neuwirth, E. RColorBrewer: ColorBrewer Palettes <https://CRAN.R-project.org/package=RColorBrewer> (2014).
42. Kassambara, L. ggpubr: 'ggplot2' Based Publication Ready Plots. <https://CRAN.R-project.org/package=ggpubr> (2019).
43. Wilke, C. O. ggtext: Improved Text Rendering Support for 'ggplot2' <https://CRAN.R-project.org/package=ggtext> (2020).
44. Pedersen, T. L. patchwork: The Composer of Plots. <https://CRAN.R-project.org/package=patchwork> (2020).
45. Uetz, P. *et al.* A quarter century of reptile and amphibian databases. *Herpetol. Rev.* **52**, 246–255 (2021).
46. Occdownload Gbif.Org. Occurrence Download. *The Global Biodiversity Information Facility* <https://doi.org/10.15468/dl.6fg294> (2021).
47. Biodiversity Information Serving Our Nation. BISON Occurrence Download: Viperidae. <https://bison.usgs.gov/index.jsp?scientificName=Poa&ITIS=itis#home> (2021).
48. HerpMapper - A Global Herp Atlas and Data Hub. *HerpMapper Occurrence Download: New World Crotalinae*. <https://www.herpMapper.org/> (2021).
49. Nogueira, C. C. *et al.* Atlas of Brazilian snakes: Verified point-locality maps to mitigate the Wallacean shortfall in a megadiverse snake fauna. *South American Journal of Herpetology* **14**, 1–274, <https://doi.org/10.2994/SAJH-D-19-00120.1> (2019).
50. BioWeb Ecuador. *BioWeb Occurrence Download: Viperidae*. <https://bioweb.bio/> (2021).
51. Meik, J. M., Streicher, J. W., Lawing, A. M., Flores-Villela, O. & Fujita, M. K. Limitations of climatic data for inferring species boundaries: Insights from speckled rattlesnakes. *PLoS ONE* **10**, 1–19, <https://doi.org/10.1371/journal.pone.0131435> (2015).
52. International Union for Conservation of Nature. *IUCN Spatial Data Download: REPTILES*. <https://www.iucnredlist.org/resources/spatial-data-download> (2018).
53. Roll, U. *et al.* The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology and Evolution* **1**, 1677–1682, <https://doi.org/10.1038/s41559-017-0332-2> (2017).
54. Heimes, P. *Snakes of Mexico*, 1 edn. (Edition Chimaira, 2016)
55. Campbell, J. A. & Lamar, W. W. *The Venomous Reptiles of the Western Hemisphere: Volume II* (Cornell University Press, 2004).
56. The Nature Conservancy. Terrestrial Ecoregions <https://geospatial.tnc.org/datasets/b1636d640ede4d6ca8f5e369f2dc368b> (2019).
57. Global Administrative Areas. GADM Data Download: World. <https://gadm.org/data.html> (2018).
58. Zaher, H. *et al.* Large-scale molecular phylogeny, morphology, divergence-time estimation, and the fossil record of advanced caenophidian snakes (Squamata: Serpentes). *PLOS ONE* **14**, e0216148, <https://doi.org/10.1371/journal.pone.0216148> (2019).
59. Blair, C. *et al.* Cryptic diversity in the Mexican highlands: Thousands of UCE loci help illuminate phylogenetic relationships, species limits and divergence times of montane rattlesnakes (Viperidae: *Crotalus*). *Molecular Ecology Resources* **0–2**, <https://doi.org/10.1111/1755-0998.12970> (2018).
60. Phillips, S. J., Dudík, M. & Schapire, R. E. *Maxent software for modeling species niches and distributions (Version 3.4.4)* https://biodiversityinformatics.amnh.org/open_source/maxent/ (2022).
61. Elith, J. *et al.* Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151 (2006).
62. Hao, T., Elith, J., Lahoz-Monfort, J. J. & Guillera-Aroita, G. Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography* **43**, 549–558, <https://doi.org/10.1111/ecog.04890> (2020).
63. Kaky, E., Nolan, V., Alatawi, A. & Gilbert, F. A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants. *Ecological Informatics* **60**, <https://doi.org/10.1016/j.ecoinf.2020.101150> (2020).
64. Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. & Blair, M. E. Opening the black box: an open-source release of Maxent. *Ecography* **40**, 887–893, <https://doi.org/10.1111/ecog.03049> (2017).
65. Phillips, S. J. & Dudík, M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161–175, <https://doi.org/10.1111/j.0906-7590.2008.5203.x> (2008).
66. Elith, J. *et al.* A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* **17**, 43–57, <https://doi.org/10.1111/j.1472-4642.2010.00725.x> (2011).
67. Ranc, N. *et al.* Performance tradeoffs in target-group bias correction for species distribution models. *Ecography* **40**, 1076–1087, <https://doi.org/10.1111/ecog.02414> (2017).
68. Inman, R., Franklin, J., Esque, T. & Nussear, K. Comparing sample bias correction methods for species distribution modeling using virtual species. *Ecosphere* **12**, e03422, <https://doi.org/10.1002/ECS2.3422> (2021).
69. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* 4 edn. (Springer, 2002).
70. Olson, D. M. *et al.* Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience* **51**, 933–938 (2001).
71. Soberón, J. M. Niche and area of distribution modeling: A population ecology perspective. *Ecography* **33**, 159–167, <https://doi.org/10.1111/j.1600-0587.2009.06074.x> (2010).
72. Barve, N. *et al.* The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling* **222**, 1810–1819, <https://doi.org/10.1016/j.ecolmodel.2011.02.011> (2011).
73. Amatulli, G. *et al.* Data Descriptor: A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data* **5**, 1–15, <https://doi.org/10.1038/sdata.2018.40> (2018).
74. Tange, O. GNU Parallel. [zenodo https://doi.org/10.5281/zenodo.1146014](https://doi.org/10.5281/zenodo.1146014) (2018).
75. Feldman, A., Sabath, N., Pyron, R. A., Mayrose, I. & Meiri, S. Body sizes and diversification rates of lizards, snakes, amphisbaenians and the tuatara. *Global Ecology and Biogeography* **25**, 187–197, <https://doi.org/10.1111/geb.12398> (2016).

Acknowledgements

We thank the three anonymous reviewers and editorial board that provided helpful comments and suggestions that improved the quality of this work. We thank Michelle Gaynor for extensive useful discussions on occurrence record cleaning, SDMs, and environmental variable selection. We also thank Michael Belitz for useful discussion and code regarding environmental variable selection for SDMs. Some of these data are provided by iNaturalist (www.inaturalist.org) and HerpMapper (www.herpMapper.org) and their networks of citizen and community contributors. Gustavo Jiménez Velázquez is a doctoral student from Programa de Doctorado en Ciencias Biológicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship CVU No. 372020 from CONACYT. Funding was provided by the National Science Foundation (DEB 1822417 to CLP), Clemson University College of Science and Department of Biological Sciences (Professional Development Graduate

Research Assistantship and Harry & Catherine Findley Student Assistance Endowment to RMR), and Sao Paulo Research Foundation (#2020/12658-4 to MM).

Author contributions

Conceptualization: R.M.R.; Methodology: R.M.R.; Validation: G.J., E.P.H., C.I.G., L.R.V.A., M.M., P.C., T.M.D. and C.L.P.; Formal Analysis: R.M.R.; Investigation: R.M.R.; Resources: R.M.R., G.J., E.P.H., L.R.V.A., C.I.G., M.M.; Data Curation: R.M.R.; Writing – Original Draft: R.M.R.; Writing – Review & Editing: R.M.R., G.J., E.P.H., L.R.V.A., C.I.G., M.M., P.C., T.M.D., C.L.P.; Visualization: R.M.R.; Supervision: R.M.R., C.L.P.; Project Administration: R.M.R., C.L.P.; Funding Acquisition: C.L.P. Authors reviewed distributions based on their expertise with regard to geographic area and/or taxonomic group.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01323-4>.

Correspondence and requests for materials should be addressed to R.M.R. or C.L.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022