

RESEARCH ARTICLE

Integrating Multiple Distribution Models to Guide Conservation Efforts of an Endangered Toad

Michael L. Treglia^{1*}, Robert N. Fisher², Lee A. Fitzgerald¹

1 Department of Wildlife and Fisheries Sciences, Biodiversity Research and Teaching Collections, Applied Biodiversity Science Program, Texas A&M University, College Station, Texas, United States of America, **2** U. S. Geological Survey, Western Ecological Research Center, San Diego Field Station, San Diego, California, United States of America

✉ Current Address: Department of Biological Science, University of Tulsa, Tulsa, Oklahoma, United States of America

* mike-treglia@utulsa.edu



OPEN ACCESS

Citation: Treglia ML, Fisher RN, Fitzgerald LA (2015) Integrating Multiple Distribution Models to Guide Conservation Efforts of an Endangered Toad. PLoS ONE 10(6): e0131628. doi:10.1371/journal.pone.0131628

Editor: Christian Andrew Hagen, Oregon State University, UNITED STATES

Received: November 24, 2014

Accepted: June 5, 2015

Published: June 30, 2015

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: The data used in this study are available at the sources referenced in the manuscript.

Funding: MLT was supported by a National Science Foundation Integrative Graduate Education and Research Traineeship (Division of Graduate Education Grant #0654377), a Tom Slick Doctoral Fellowship (Texas A&M University), an E.E. Williams Research Grant (Herpetologists' League), and a Sylvia Scholarship (Mensa Foundation). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Species distribution models are used for numerous purposes such as predicting changes in species' ranges and identifying biodiversity hotspots. Although implications of distribution models for conservation are often implicit, few studies use these tools explicitly to inform conservation efforts. Herein, we illustrate how multiple distribution models developed using distinct sets of environmental variables can be integrated to aid in identification sites for use in conservation. We focus on the endangered arroyo toad (*Anaxyrus californicus*), which relies on open, sandy streams and surrounding floodplains in southern California, USA, and northern Baja California, Mexico. Declines of the species are largely attributed to habitat degradation associated with vegetation encroachment, invasive predators, and altered hydrologic regimes. We had three main goals: 1) develop a model of potential habitat for arroyo toads, based on long-term environmental variables and all available locality data; 2) develop a model of the species' current habitat by incorporating recent remotely-sensed variables and only using recent locality data; and 3) integrate results of both models to identify sites that may be employed in conservation efforts. We used a machine learning technique, Random Forests, to develop the models, focused on riparian zones in southern California. We identified 14.37% and 10.50% of our study area as potential and current habitat for the arroyo toad, respectively. Generally, inclusion of remotely-sensed variables reduced modeled suitability of sites, thus many areas modeled as potential habitat were not modeled as current habitat. We propose such sites could be made suitable for arroyo toads through active management, increasing current habitat by up to 67.02%. Our general approach can be employed to guide conservation efforts of virtually any species with sufficient data necessary to develop appropriate distribution models.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Habitat loss and degradation are major causes of biodiversity loss in terrestrial and freshwater ecosystems [1]. Urbanization and agricultural expansion are among the primary drivers of alterations to natural habitats, causing land cover change. Indirect effects of these disturbances manifest in myriad ways: invasive vegetation can displace native species and alter physical habitat structure [2]; changes in hydrology can impact riparian conditions [3]; and introduced animals can alter ecosystems through trophic interactions [4]. Site-specific actions can improve habitats for individual species, though identifying appropriate locations for conservation efforts poses a challenge [5,6].

Within the ever-expanding toolkit for conservation biologists, species distribution models (SDMs) have become commonly employed for various purposes [7,8]. Though species distribution modeling can have various connotations and meanings, we follow a convention of using it to encompass the concept of habitat suitability models, environmental niche models, and others [7]. The principle behind species distribution modeling is that relationships among species' locality data and associated environmental variables can be used to make inferences of where else suitable conditions exist [8].

Common applications of SDMs include predicting how climate change may contribute to species extinctions and range shifts [9,10], identifying locations with undescribed species and new localities or habitat of known species [11,12,13], and projecting distributions of invasive species [14,15]. SDMs can also be used to estimate recent habitat loss for individual species [16], and to predict future habitat loss given changes in particular variables [17]. Although SDMs can be employed to explicitly guide conservation efforts, the published literature is lacking examples [18].

We focused this study on the arroyo toad (*Anaxyrus californicus*), which is endemic to southern California, USA and northern Baja California, Mexico [19,20]. It is a habitat specialist, closely tied to open, sandy streams and surrounding floodplains [20,21]. The species is listed as endangered by the U.S. Fish and Wildlife Service [22,23] and by the IUCN [19], facing threats of habitat destruction, habitat degradation, and invasive predators [20]. Anthropogenic alterations to hydrologic regimes and wildfire frequency have contributed to these threats, though it is possible to improve habitat through site-specific actions. For example, decreases of an introduced predator, the American bullfrog (*Lithobates catesbeianus*) can increase arroyo toad occupancy and abundance [24], and clearing of riparian vegetation may benefit breeding habitat. An active proposal to downlist the arroyo toad to threatened status cites ongoing, albeit decreased threats to the species [25], and corrects for previously erroneous locality records [26].

We developed and integrated multiple SDMs for the arroyo toad using static and dynamic environmental datasets [17] with the objective of identifying sites that can be used to create additional habitat for the species. We had three main goals: 1) develop a model of potential habitat for arroyo toads, based on long-term, static environmental variables (hereafter, the “potential model”); 2) develop a model of the species' current habitat by incorporating time-sensitive remote sensing data and only using locality data since 2005 (hereafter, the “current model”); and 3) integrate results of both models to identify sites that may be used for arroyo toad conservation. In addition to using SDMs to map habitat for arroyo toads, our techniques allowed us to further elucidate environmental characteristics associated with presence of the species.

SDMs exist for other amphibians in arid environments [27], and an early SDM was developed for arroyo toads in a portion of our study area [28]. Our models cover a large spatial extent at high resolution, and help identify sites with potential to be improved as habitat,

through actions such as thinning of stream-side vegetation to create more open riparian areas. Our model results may also inform future translocation efforts, and surveys for unknown populations of arroyo toads. Furthermore, our general methodology can be employed to guide conservation of other species in different systems.

Methods

Ethics Statement

The following institutions and organizations authorized field work associated with this study and issued the respective permits: The Texas A&M University Animal Care and Use Committee: AUP 2012–125; U.S. Fish and Wildlife Service: Recovery Permits TE-045994 and TE-73366A-0; California Department of Fish and Wildlife, Scientific Collecting Permits 00090 and 11723.

Study Area

Our study focused on five coastal watersheds of southern California (based on HUC-8 classification; [29]): the Aliso-San Onofre; the San Luis Rey-Escondido; the San Diego; the Santa Margarita; and the U.S. portion of the Cottonwood-Tijuana watershed. This area has undergone significant anthropogenic land cover changes in recent decades [30], and further development is projected into the future [31]. However, this region has active conservation programs (e.g., the Multiple Species Conservation Plan), with local stakeholder groups working to restore native ecosystems [32], thus our results can be readily adopted to guide on-the-ground actions. Furthermore, this area may comprise a reasonable management unit for the species, as range-wide genetic analyses showed arroyo toad populations from these drainages were more closely related to each other than to other populations [33].

Units of Analysis

We focused on streams and stream-side areas, corresponding to primary habitats used by arroyo toads [20,21,34] (Fig 1). For the best spatial accuracy we used stream data from the 1:24,000 scale National Hydrography Dataset (NHD; <http://nhd.usgs.gov>). We excluded extremely small stream segments, generally not used by arroyo toads, by eliminating sections that were not assigned an order in the 1:100,000 scale NHDPlus dataset (<http://www.horizon-systems.com/nhdplus>; 1:24,000 scale NHD data do contain stream order data). We accomplished this using a spatial overlay with a 50 m buffer of the streams to account for differences in spatial accuracy between these two datasets, using Manifold GIS version 2.0.28 (Manifold Software Limited). We converted remaining stream segments to a spatial layer of 200 m x 200 m square sample units (as pixels of a raster dataset), which allowed us to better incorporate metrics characterizing streams and stream-side areas. We removed sample units identified as water in the STATSGO2 Soil Database (<http://soildatamart.nrcs.usda.gov>) to mask out large water bodies, known not to serve as habitat for arroyo toads. Higher-resolution data (1:100,000 scale) in the SSURGO soil dataset were unavailable for parts of our study area, thus we used the STATSGO2 dataset in this study (1:250,000 scale). All of these spatial data sources were accessed in August and September 2013.

Environmental Data

In both models we used static variables [17], characterizing climate, soil, topography, and geomorphology (Table 1). In the current model we also used 2010 Landsat TM imagery [35] to derive indices of brightness, greenness, and wetness (i.e., Tasseled Cap bands), as dynamic

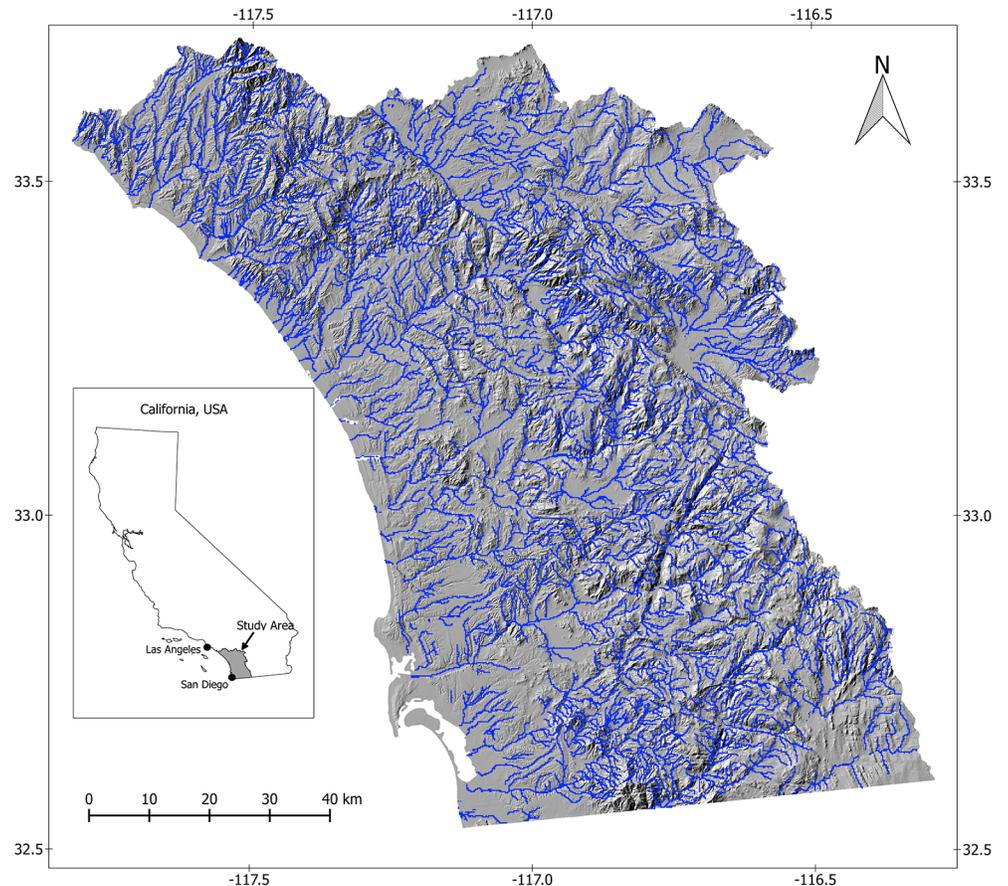


Fig 1. Map of streams and topography of southwestern California. This map illustrates streams (in blue), overlaid on a hillshade layer of southwestern California, USA, covering five focal watersheds: the Aliso-San Onofre, the San Luis Rey-Escondido, the San Diego, the Santa Margarita, and the U.S. portion of the Cottonwood-Tijuana.

doi:10.1371/journal.pone.0131628.g001

variables. These metrics characterize various aspects of land cover including vegetation, substrate, and moisture [36]. We assumed relatively little land cover change along riparian areas in our focal period (2005–2013), and by using imagery from a fairly central year, we minimize potential effects of drastic land cover differences from the beginning to the end of our focal period. Additionally, climate conditions for 2010 were nearly average (based on annual climate reports for San Diego County queried from <http://www.ncdc.noaa.gov>, accessed September 2013), which is important in that weather can influence vegetation and hydrologic patterns detected in imagery, ultimately influencing activity of arroyo toads. We obtained cloud-free imagery for 27 March and 3 September, representing wet and dry seasons, respectively. We converted raw data to top of atmosphere reflectances, atmospherically corrected them using dark object subtraction (DOS1; [37]), derived Tasseled Cap bands using the Tasseled Cap Transformation [36] in GRASS GIS version 6.4.4 [38], and extracted the median and variance per sample unit. We prepared environmental data for analyses using SAGA GIS version 2.1.1 [39] and Manifold GIS version 8.0.28 (Manifold Software Limited).

Given the large number of predictor variables, we used principal component analyses (PCAs) to derive reduced variable-sets [40]. For each model, we conducted a PCA on the correlation matrix of predictor variables, and used principal components (PCs) with eigenvalues greater than one in place of the original data, following Kaiser’s rule [40]. We conducted PCAs

Table 1. Description of environmental data layers used in models of arroyo toad habitat.

Name (Abbreviation) ¹	Description	Value Used	Source
Climate Data			
Avg. Monthly. and Annual: Precipitation (Ppt [1–13]); Maximum Temperature (TMx [1–13]); and Minimum Temperature (TMn [1–13])	Average values from 1981–2010; original pixel size of 800 m	Majority value per analysis pixel	PRISM Climate Group, Oregon State University ²
Soil Data			
% Clay (Clay); % Sand (Sand); % Silt (Silt); Soil Water Storage Capacity (WaterSt)	Weighted average of values per soil type across all soil layers, obtained from 1:250,000 scale soil data	Average, weighted by area of each soil type per analysis pixel	Derived from STATSGO2 Soil Data, produced by the Natural Resources Conservation Service, U.S. Dept. of Agriculture ³
Topography and Geomorphology			
Elevation along Stream Segment (Elev)	Estimated as lowest elevation value within analysis pixels	Calculated value per analysis pixel	10 m National Elevation Dataset (NED; [70]) ⁴
% Stream Slope (Slope)	Estimated within each analysis pixel using GIS data for elevation and streams	Calculated value per analysis pixel	Derived from 10 m NED overlaid on 1:24,000 National Hydrologic Dataset ⁵
Multiresolution Index of Valley Bottom Flatness (MRVBF)	Measure of how flat and wide a valley is.	Maximum value per analysis pixel	Derived from 10 m NED using methodology described by Gallant and Dowling [71]
Vector Ruggedness Measure (VRM03 and VRM18)	Measure of how rugged terrain is, based on, analysis windows of 3 and 18 pixels from 10 m NED	Minimum values per analysis pixel	Derived from 10 m NED using methodology described by Sappington et al. [72]
Catchment Area (CatchArea)	Total area draining into a given analysis pixel	Maximum value per analysis pixel	Derived from sink-filled 10m NED using methodology described by Gruber and Peckham [73]
Remotely Sensed Data			
Brightness (Brt[3,9].Med, Brt[3,9].Var); Greenness (Grn[3,9].Med, Grn[3,9].Var); Wetness (Wet[3,9].Med, Wet[3,9].Var)	Indices of “brightness,” “greenness,” and “wetness” for 27 March and 9 Sept. 2010.	Median (Med) and Variance (Var) within analysis pixel	Derived from Landsat TM imagery ⁶ using the Tasseled Cap Transformation [36] for Landsat data [35].

¹ Bracketed numbers with abbreviations denote corresponding months layers were from (1–12) or indicate that it is the annual average (13)

² Available from: <http://www.prism.oregonstate.edu/>

³ Available from: <http://soildatamart.nrcs.usda.gov>

⁴ Available from: <http://viewer.nationalmap.gov/viewer/>

⁵ Available from: <http://nhd.usgs.gov>

⁶ Available from: http://landsat.usgs.gov/Landsat_Search_and_Download.php

doi:10.1371/journal.pone.0131628.t001

using the package ‘vegan’ [41] in R version 3.0.2 [42]; we present tables with variable loadings as supporting information (S1 File).

Arroyo Toad Locality Data

We obtained locality data for arroyo toads from the U.S. Fish and Wildlife Service (<http://www.fws.gov/carlsbad/GIS/CFWOGIS.html>) and we also used museum records from the following institutions, accessed through the HerpNet data portal (<http://herpnet.org/>): California Academy of Sciences; Natural History Museum of Los Angeles County; San Diego Natural History Museum; Smithsonian National Museum of Natural History; and University of California, Berkeley, Museum of Vertebrate Zoology. We included data from Cleveland National Forest [43,44], and our own surveys conducted as part of the U.S. Geological Survey (USGS) Amphibian Research and Monitoring Initiative (ARMI; <http://armi.usgs.gov/>) and the San Diego Monitoring and Management Program (SDMMP; <http://sdmmp.com>). Datasets were accessed in September 2013.

Given undocumented spatial accuracy for some sources, and our focus on stream habitats, we excluded data outside a 50 m buffer of the streams to minimize potential error, and we removed data that had accuracy documented as >160 m in the U.S. Fish and Wildlife Service dataset. For analyses, we used the final locality data as presence per sample unit. Locality records ranged from 1927 to 2013 and occurred among 1,037 sampling units. All but nine of these sampling units had at least one record from 1990 or later, and those that did not were generally in the vicinity of more recent records. Given our potential model is based on long-term, relatively stable environmental variables, we consider locality records from the entire temporal range as informative in understanding where the species could occur without constraints associated with land cover characteristics. For the current model, we used presence records from 2005 to 2013, among 791 sample units.

We incorporated absence data into our current model, attained through standardized daytime and nighttime surveys [24,45], conducted as part of ARMI and SDMMMP. Low detection probabilities of focal species can result in false-absences, in which species are recorded as not present at sites where they actually occur, contributing error in subsequent analyses. In distribution modeling studies, inaccurate absence records could result, erroneously, in a more constrained distribution of focal species. Thus, based on previous work investigating detectability of arroyo toads [45], we considered the species absent from sample units where it was not detected at least once in eight nighttime surveys or five daytime surveys since 2005, minimizing our chances of including false-absence records. If a sample unit had presence and absence records since 2005, the presence record was given priority. Based on these criteria, we had absence records among 89 sample units for the current model.

Species Distribution Models

Model Development. We used Random Forests [46] to develop the potential and current models. This is a machine-learning technique that merges classification and regression trees with a bootstrap resampling procedure to create an optimal model [47]. Random Forests reduces problems of over-fitting and does not rely on assumptions of parametric methods [48,49], and it has been implemented in a variety of ecological studies [47,50,51].

Random Forests is generally considered a presence/absence method [7], but has successfully been used with presence/pseudoabsence data [52,53]. Pseudoabsences are used when true absence data are unavailable, and they are acquired by sampling locations from the study region that lack locality records [8]. Given that imbalanced training datasets can decrease model fit and contribute to bias [54], we generated sufficient numbers of pseudoabsences for both models to create balanced datasets. Thus, we used presence/pseudoabsence data for our potential model, and a combination of presence/absence and pseudoabsence data in our current model. To account for spatial biases in our arroyo toad data that could influence model results we selected pseudoabsences with the same spatial biases as our data based on kernel density surfaces [55,56]. We ran models 10 times with different pseudoabsence points and averaged the results [57], effectively reducing weight of individual pseudoabsence points relative to verified presences and absences, and ultimately minimizing potential error due to the location of individual pseudoabsences, for which we cannot verify absence of arroyo toads.

We implemented Random Forests using the 'randomForest' package [58] in R [42]. We set the number of bootstrapped trees (k) based on the point at which the error rate for withheld (out-of-bag [OOB]) samples stabilized. Given that variable interaction may stabilize at a slower rate than the OOB error [48], we used twice that number, setting $k = 10,001$. In each tree, the OOB sample was 36.8%, and the number of variables permuted at each branching node was set to the square root of the number of variables. We used preliminary model comparisons to

investigate whether removal of any PC-transformed variables would yield more parsimonious results [50,59], though we found inclusion of all variables was optimal. We used the final, averaged models to predict habitat in our study area in terms of “probability of occurrence” [8], and we used the mean decrease in accuracy for randomized permutations of input variables as a measure of variable importance [58]. Figures representing per-sample unit coefficient of variation across the 10 runs for potential and current models are available in supporting information (S2 File).

Model Evaluation. We evaluated model performance by comparing probabilities of occurrence with presence/pseudoabsence data (potential model) and presence/absence data (current model). As a threshold-independent metric of model performance, we used the area under the receiver operating curve (AUC), which ranges 0.5–1.0; models with values >0.9 are considered to have high performance [7,8]. As a measure of model significance, we also compared our models to 1,000 models of randomized presence/absence data, calculating the p-value as the proportion of times that the OOB error in randomized models was less than that of our models [50,59]; we set α to 0.05.

We also used threshold dependent measures of model performance [7] in which probabilities of occurrence were converted to binary predictions of presence/absence and compared to the original data. We set the cutoff value for binary predictions to the lowest modeled probability of occurrence for a sample unit with confirmed arroyo toad presence [12,60]. We calculated the True Skill Statistic (TSS; [61]) and misclassification rates for our models. TSS ranges -1–1 and misclassification rates range 0–1; higher values of TSS and lower values of misclassification rates indicate better model performance.

Identification of Conservation Opportunities and Comparison of Model Results

We integrated results of the potential and current models to identify areas that may be employed in conservation of the arroyo toad by subtracting the binary presence/absence predictions of the current model from those of the potential model. This yielded a map illustrating the differences in prediction of habitat between the two models, with three possible values at each sample unit: 1 –sample unit was predicted as habitat in the potential model, but not the current model; 0 –no change in predictions; and -1 –sample unit was predicted as habitat in the current model, but not the potential model. We consider sample units with a value of 1 as representing sites that should be considered for habitat management and restoration, as the intrinsic conditions are predicted to be suitable for arroyo toads. We anticipated sample units with values of -1 would be rare, but possible given the current model may include interactions between dynamic and static variables. To compare the amount of habitat predicted among the potential and current models, we calculated the percentage of sample units predicted to be habitat in each.

We also calculated how much modeled habitat is on lands precluded from development by overlaying our results with boundaries for conserved lands in our study area. We obtained the most comprehensive and accurate boundaries of conserved lands available for our study area as follows: for San Diego County, from the San Diego Monitoring and Management Program (http://sdmmp.com/Online_Map.aspx, accessed March 2014); for Riverside County, from the Public and Quasi-Public Conserved Lands dataset (<http://gis.rivcoit.org/GISData.aspx>, accessed May 2014); and for Orange County, from the California Protected Areas Database (version 2014a; <http://projects.atlas.ca.gov/projects/cpad/>, accessed May 2014). For San Diego County, comprising the majority of our study area, these included all areas legally required to maintain open space of various forms, including natural habitats, and for specific species,

through easements, habitat conservation plans, and fee title ownership designations. For Riverside County, these included all lands owned or managed by the county for purposes of long-term conservation, including those contributing to multi-species conservation programs. For Orange County, these boundaries designated lands protected in various forms of open space through fee title ownership designations. We conducted this analysis using Manifold GIS version 8.0.28 (Manifold Software Limited).

Results

Model Evaluation and Summary

Our models performed well based on all fit metrics. All runs for both models were significant based on permutation tests ($p < 0.001$), and AUC values for models were 0.957 for the potential model and 1.000 for the current model. For threshold dependent measures of fit, the cutoff values for binary predictions were 0.435 and 0.492 for potential and current models, respectively, resulting in TSS values of 0.809 and 1.000. Back-predictions to our presence/pseudoabsence and presence/absence data had 9.60% and 0.00% misclassification rates in the potential and current models, respectively. Maps illustrating the presence/absence predictions are presented in Figs 2 and 3.

Distinct variables contributed to the potential and current models (Table 2). In the potential model, PCs representing soil and topography were most influential, though in the current model the most important PCs represented aspects of climate, elevation, and wetness (Table 2). Given that machine learning techniques are optimized for predictive performance and they implicitly include complexities such as variable interactions, relationships among variables can be difficult to interpret [47], though we provide some interpretation of our model results in Table 2.

Both models showed similar relationships between habitat suitability and relatively static variables (Table 2). For example, both models indicate positive relationships between habitat suitability and percent sand in the soil, valley bottom flatness (MRVBF), catchment area draining into the site, maximum temperatures in the summer, and minimum temperatures in the winter and spring. Our models also indicated negative relationships between habitat suitability and percent silt and percent clay in the soil, slope, elevation, and late-summer precipitation. In addition, our current model indicated positive relationships between habitat suitability and brightness and per-sample unit variance in brightness, and negative relationships between habitat suitability and greenness, wetness, and variance in wetness.

Identification of Conservation Opportunities and Comparison of Model Results

Of the 46,305 sample units in our study area, 3,260 were modeled as potential habitat, but not current habitat (Fig 4). This represents 7.04% of our focal area, which has potential to be employed in habitat improvement and conservation efforts, but is not currently suitable for the species. An additional 1,467 sample units were modeled as current but not potential habitat. Individually, our models predict potential habitat in 14.37% and current habitat in 10.50% of the sample units in our study area. Thus, we estimate a net decrease of 26.93% in modeled habitat, resulting from constraints associated with dynamic variables in our current model, representing land cover characteristics.

Of the sample units that our models identify as potential and current habitat, 35.61% and 33.48%, respectively, are on conserved lands. Furthermore, of the sample units identified as potential but not current habitat, 33.37% are on conserved lands. Thus, approximately two-

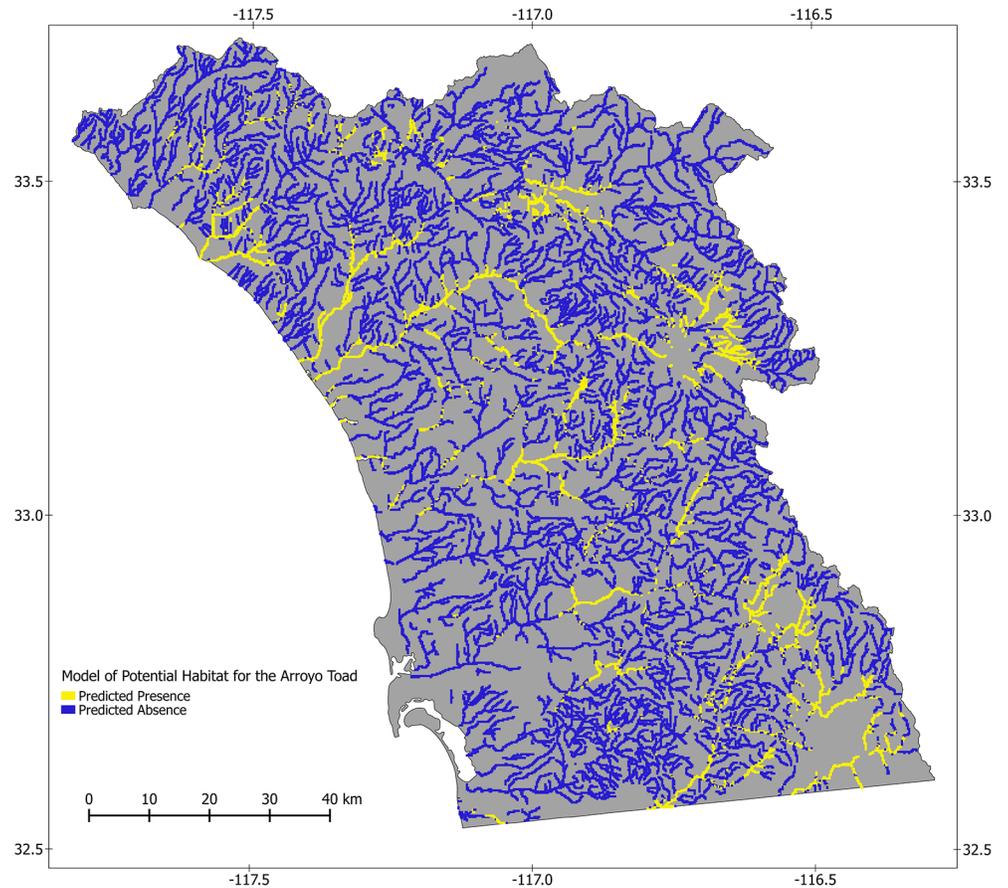


Fig 2. Modeled potential distribution of the arroyo toad in southwestern California. This map depicts the modeled potential distribution of the arroyo toad in streams and stream-side areas of southwestern California. Input data for the model included presence/pseudoabsence data and relatively stable, long-term environmental data representing characteristics of topography, soil, and climate. The Random Forests algorithm was used to develop the model, from which we predicted the probability of arroyo toad presence throughout our study area. The model performed well, with an Area Under the Receiver Operating Curve of 0.957 and a True Skill Statistic of 0.809. The lowest modeled probability of arroyo toad presence for a site known to have arroyo toads was 0.435. Sites with modeled probability of presence lower than this value were designated as not habitat (blue) and sites with probabilities of presence greater than or equal to this value were designated as habitat (yellow). Based on this model, of our 46,305 sample units, arroyo toads were predicted to occur in 14.37%.

doi:10.1371/journal.pone.0131628.g002

thirds of sites with modeled habitat and the potential to serve in conservation efforts may be susceptible to future development pressures.

Discussion

Our study illustrates how integration of multiple SDMs can yield valuable results for guiding conservation efforts. Individually, our potential and current SDMs for the endangered arroyo toad indicate that 14.37% of our study area has intrinsic characteristics likely suitable for the species, and 10.57% of the area may currently be suitable. By integrating results of these models, we identified 7.04% of our study area as potential habitat but not current habitat; we propose local management of these sites could render them suitable for arroyo toads, resulting in an increase of current habitat by up to 67.02%. Relevant management strategies suggested by our results may include thinning of stream-side vegetation to create more open riparian

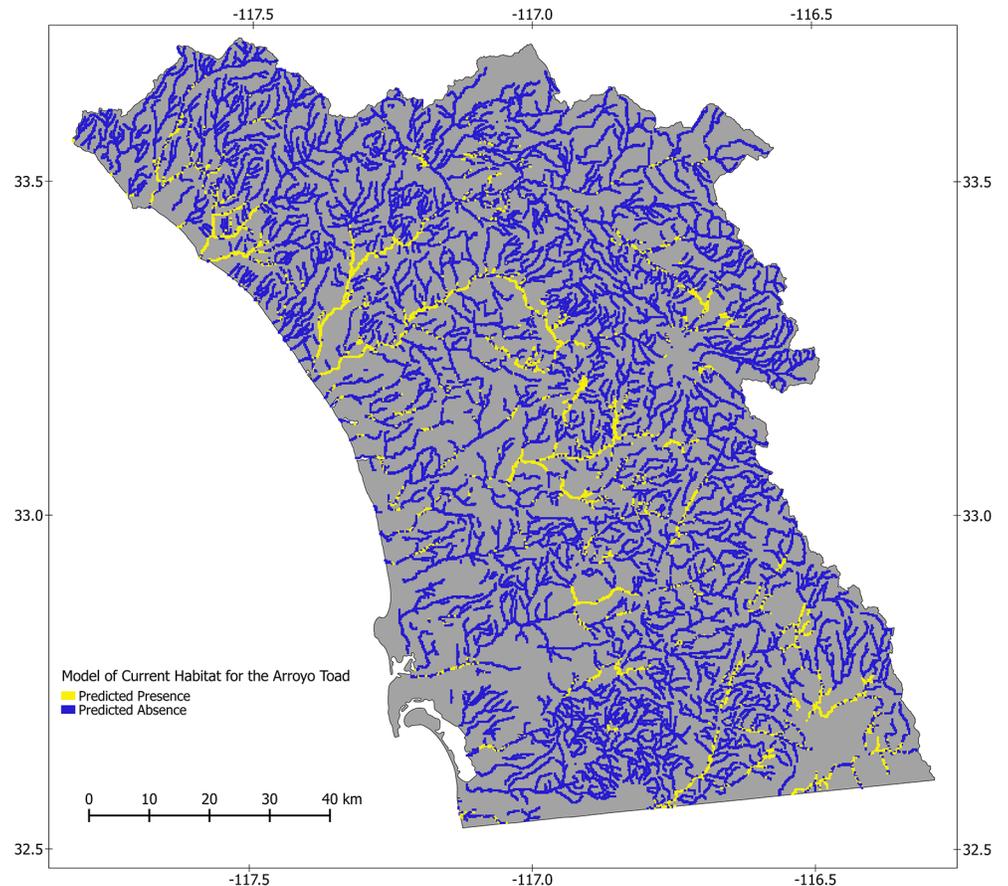


Fig 3. Modeled current distribution of the arroyo toad in southwestern California. This map depicts the modeled current distribution of the arroyo toad in streams and stream-side areas of southwestern California. Input data included presence/absence and pseudoabsence data, long-term environmental data representing characteristics of topography, soil, and climate, and indices of Brightness, Greenness, and Wetness, which represent more dynamic characteristics of land cover, derived from 2010 Landsat TM imagery. The Random Forests algorithm was used to develop the model, from which we predicted the probability of arroyo toad presence throughout our study area. The model performed well, with an Area Under the Receiver Operating Curve and True Skill Statistic of 1.0. The lowest modeled probability of arroyo toad presence for a site known to have arroyo toads was 0.492. Sites with modeled probability of presence less than this value were designated as not habitat (blue) and sites with probabilities of occurrence greater than or equal to this value were designated as habitat (yellow). Of 46,305 sample units, arroyo toads were predicted to occur in 10.57% based on relatively static landscape characteristics.

doi:10.1371/journal.pone.0131628.g003

habitats, particularly in areas with sandy soils and larger floodplains. Given the remotely-sensed variables we used in our current model, this model should reflect such management at individual sites, given updated satellite imagery. Though studies that employ SDMs typically are not well-linked to on-the-ground conservation efforts [18], our models can directly inform management actions, helping answer the difficult question of where to focus limited resources [5,6].

Our general approach can be applied to virtually any taxa with sufficient locality information, in regions with relevant environmental datasets necessary for distribution models. We classified predictor variables as static or dynamic based on our objectives and focal time period. Future studies may incorporate additional variables in either category, and reclassify variables we used if deemed appropriate. Specific modeling techniques employed can also be adjusted, though our use of a transition map to integrate results of multiple models is valuable in

Table 2. Importance of the PCA-transformed variables in the potential and current models.

Principle Component	Highest-Loading Environmental Variables ¹ (Positive and Negative)	Relationship with Habitat Predictions	Mean Decrease in Accuracy
<i>Potential Model</i>			
PC4	(+) MRVBF; WaterSt; Sand; CatchArea; Ppt11(-) VRM18; Slope; VRM03; Silt; Clay; Ppt07	+	0.1078
PC2	(+) TMx05; TMx09; TMx08; TMx06; TMx13(-) TMn07; TMn08; Ppt06; TMn06; TMn09	+	0.0766
PC1	(+) Elev; Ppt09; Ppt08; Ppt07; Ppt13(-) TMn04; TMn03; TMn05; TMn02; TMn10	-	0.0738
PC7	(+) Slope; Ppt06; Sand; TMn12; TMn01(-) CatchArea; VRM03; WaterSt; VRM18; MRVBF	-	0.0727
PC3	(+) MRVBF; Ppt08; Ppt07; Sand; WaterSt(-) Ppt06; Ppt02; Ppt01; Ppt11; Ppt10	+	0.0689
PC6	(+) VRM03; Ppt06; TMx12; TMx01; TMx11(-) TMn07; TMn08; TMx06; TMx07; TMn09	+	0.0628
PC5	(+) Silt; Clay; WaterSt; MRVBF; Ppt06(-) Sand; VRM18; VRM03; Slope; CatchArea	-	0.0580
<i>Current Model</i>			
PC1	(+) Elev; Ppt09; Ppt08; Ppt07; Ppt13(-) TMn04; TMn03; TMn05; TMn02; TMn10	-	0.0611
PC2	(+) TMx05; TMx09; TMx06; TMx08; TMx13(-) Wet09.Var; TMn07; TMn08; Ppt06; Wet03.Var	+	0.0540
PC7	(+) Silt; Clay; Grn03.Med; Wet03.Med; Grn09.Med(-) Sand; Ppt01; Brt09.Var; CatchArea; TMn07	-	0.0457
PC3	(+) Ppt06; TMx09; Ppt02; Ppt01; VRM18(-) Brt09.Var; MRVBF; Brt03.Var; Ppt08; Ppt07	-	0.0447
PC10	(+) Grn03.Var; Wet09.Med; Grn09.Var; Brt09.Med; Slope(-) CatchArea; VRM03; WaterSt; VRM18; Brt09.Var	-	0.0432
PC4	(+) Wet09.Med; Wet03.Med; Brt09.Med; Brt03.Med; Grn03.Med(-) Slope; VRM18; VRM03; Silt; Clay	+	0.0310
PC6	(+) Wet09.Var; Grn09.Var; Wet03.Var; Grn09.Med; Sand(-) Silt; Brt04.Var; Brt09.Var; Clay; Brt03.Med	+	0.0288
PC8	(+) Grn03.Var; VRM03; VRM18; Slope; Sand(-) MRVBF; TMn07; TMn08; Silt; TMx06	+	0.0255
PC9	(+) Brt09.Med; Brt03.Med; Wet09.Var; Ppt03; TMx11(-) Brt03.Var; Grn09.Var; Grn03.Var; TMn07; TMn08	+	0.0216
PC5	(+) Brt03.Med; Brt09.Med; VRM18; Wet09.Var; Slope(-) Grn03.Var; Brt09.Var; Brt03.Var; MRVBF; WaterSt	-	0.0108

¹Five highest-loading variables for each PC are listed, ordered by decreasing importance.

doi:10.1371/journal.pone.0131628.t002

identification of sites with conservation potential. The transition map also aids in visualization of results and helps convey information to stakeholders.

Subsequent steps necessary for conservation of arroyo toads based on our results may involve surveys to document unknown populations, habitat improvement actions such as removal of riparian vegetation and exotic predators, and translocation of the species to unoccupied habitat. Of the sample units we identify as targets for habitat improvement efforts, only one-third are on conserved lands, thus long-term conservation will require working closely with stakeholders in the region. For example, Marine Corps Base Camp Pendleton is currently managed for arroyo toads, and contains 13.12% and 10.97% of the modeled potential and current habitat, respectively. However, long-term persistence of arroyo toads will depend on cooperative efforts.

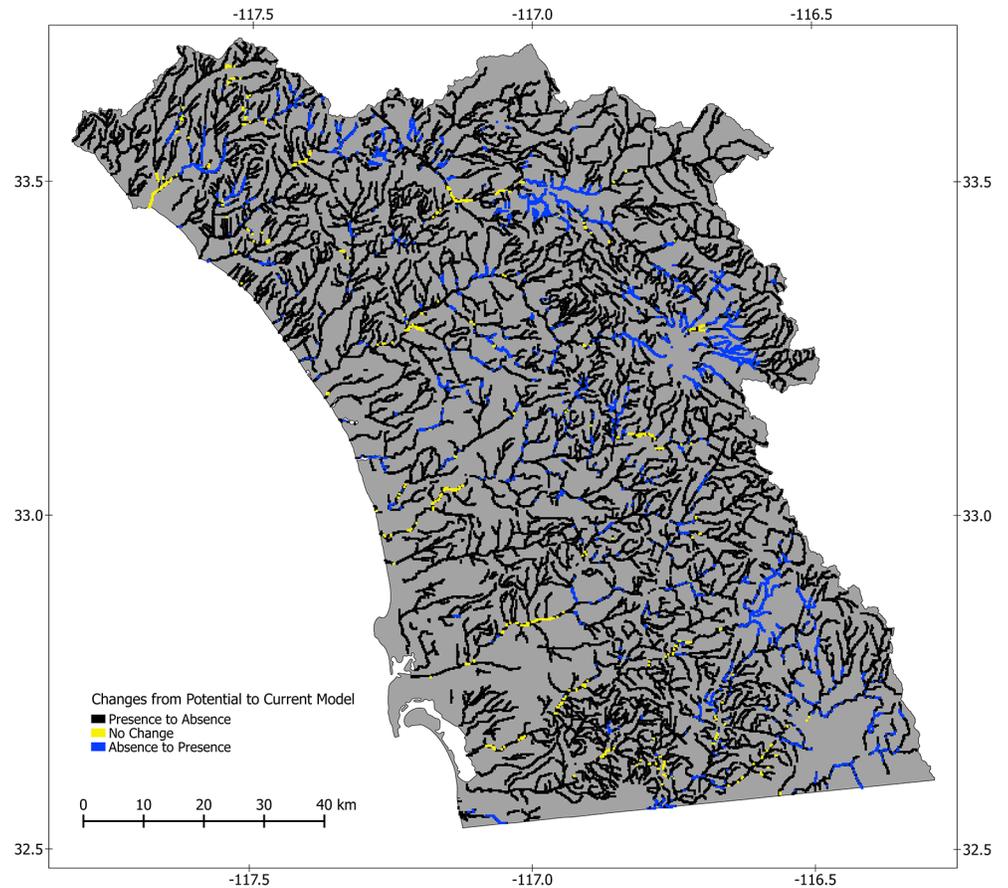


Fig 4. Comparison of two models of the distribution of the arroyo toad in southwestern California. This map was derived from two models for the distribution of the arroyo toad in southwestern California. Both models focused on streams and stream-side areas, and used relatively stable, long-term predictor variables characterizing aspects of soil, topography, and climate. The first model (potential model) only used those predictor variables and was designed to identify areas that may be suitable for the species based on intrinsic characteristics of the landscape. The second model (current model) also integrated more dynamic variables associated with current land cover conditions, and was designed to identify sites that may be suitable for the species, given constraints of land cover characteristics. This map represents the differences in predictions among the two models: black areas represent sites for which prediction of habitat did not change from the potential to the current model; blue represents sites predicted as potential but not current habitat, and yellow represents sites predicted as current but not potential habitat.

doi:10.1371/journal.pone.0131628.g004

Our models performed well, and general associations we identified between static variables and arroyo toad habitat (Table 2) are corroborated in [28] and other work summarized by the U.S. Fish and Wildlife Service [62]. For example, those studies documented associations between arroyo toads and third and higher-order streams. We used several continuous geomorphological measures in place of stream order to more precisely represent conditions [63], but found comparable relationships, with habitat identified in areas with high MRVBF, low Slope, and low VRM. Similarly, our models and the earlier studies all document associations between arroyo toads and sandy soils. Though these variables are characteristics of the landscape and it may not be possible to alter them to improve habitat, they are informative in identifying sites where suitable habitat may exist.

Tasseled Cap metrics served as effective dynamic variables, representing temporally-specific, continuous measures associated with land cover. Our current model indicates positive

relationships between habitat suitability and both median and variance in per-sample unit brightness. High brightness tends to be associated with bare ground [36]. Thus, this relationship likely characterizes the preference of arroyo toads for open, sandy streams [21,34]; high variance in brightness may be driven by streams passing through sample units, surrounded by brighter areas of bare ground. This aligns with the generally negative relationships between habitat suitability and high median greenness and wetness, as arroyo toads are generally associated with sparsely vegetated, sandy flats [21,34].

These results, in conjunction with inspection of high resolution 2010 aerial imagery (Fig 5), support that transitions from modeled potential habitat to not current habitat are attributable to land covers not likely suitable for arroyo toads. All sites presented in Fig 5 were modeled as potential habitat and had historic locality records at or nearby them. Sites A, at Marine Corps Base Camp Pendleton, and D, at Buckman Springs in the Descanso District of Cleveland National Forest, were also modeled as current habitat, with open sandy streams, but sites B, at Barker Valley in the Palomar District of Cleveland National Forest and C, along a golf course in El Cajon, CA were not, and contain dense riparian vegetation and considerable anthropogenic development, respectively. We are unaware of recent survey work at site C, though our own visits to site B, while not enough to definitively declare absence, have not yielded any detections. We note this site was previously more open, and arroyo toads were documented there as recently as the late 1990s, though it has experienced considerable vegetation encroachment.

Of the 4,862 sample units identified as current habitat, only 791 have recent records of arroyo toads, and of those, only 264 are within conserved lands. Multiple factors may contribute to the lack of confirmed presence in sites with modeled habitat. First, arroyo toads may be present in sites where no surveys have been conducted to document them. Second, sites may currently be suitable, but historic conditions caused local extirpations. Lastly, some errors may exist, stemming in part from the fact that it is impossible to encompass all habitat variables relevant to the persistence of arroyo toads in such an analysis. For example, we could not incorporate variables reflecting fine-scale hydrology, which may influence breeding success.

Although the relationships we identify between dynamic variables and habitat suitability are interpretable and supported by the literature, our models are best used for identifying potential and current habitat, rather than making prescriptive recommendations to restore habitat. Machine learning methods such as Random Forests are robust classification algorithms that can yield accurate predictions, though variable relationships can be difficult to interpret, particularly given potential for complex interactions [47]. Furthermore, though our use of Tasseled Cap metrics resulted in a well-fitting model of the current distribution of the arroyo toad, these metrics are not directly transferable to specific land cover or vegetation types. Using discrete land cover classifications may yield improved interpretability, at the risk of introducing classification error. For example, an accuracy assessment of the 2006 National Land Cover Dataset reported 80% accuracy [64]. Thus, our models are valuable in identifying sites that may be used in conservation of arroyo toad, but site-specific management recommendations may be best informed by existing studies on natural history, fine-scale habitat use, and occurrence patterns of the species [20,21,24,34].

Though our study focuses on identifying site-specific opportunities for arroyo toad conservation, future work should also consider large-scale processes that affect these habitats. Freshwater ecosystems are sensitive to environmental conditions across entire watersheds [65,66], and in an area slightly north of ours negative relationships between watershed-scale urbanization and abundance of native amphibians have been documented [67]. Complexities of factors influencing habitat at multiple scales can create unanticipated challenges for conservation [68]. However, integration of results from studies such as ours with information on species'

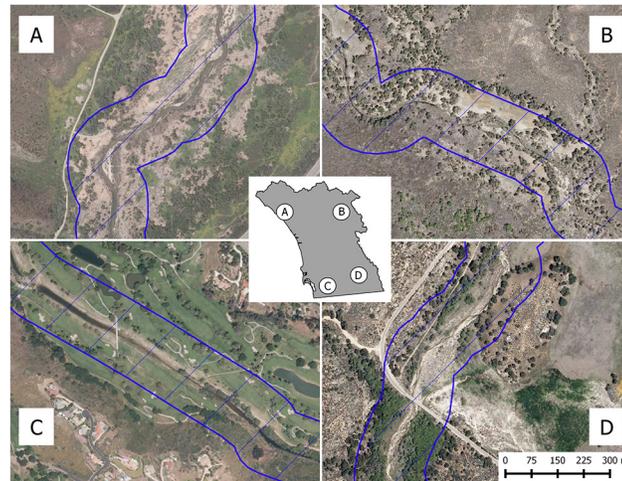


Fig 5. Aerial imagery of sites modeled as suitable and not suitable for arroyo toads based on current conditions. All four panels (A-D) depict a 100 m buffer of stream channels (outlined and hatched in blue), overlaid on 2010 aerial imagery. Sites presented in all panels were modeled to be suitable based on relatively static long-term environmental variables. Based on relatively dynamic variables associated with recent land cover, the sites in panels A and D were modeled to be currently suitable, with open, sandy habitats around the streams, but those in panels B and C were not, with considerable vegetation encroachment and anthropogenic development, respectively. The inset (middle) depicts the location of each site, within the focal study area of southwestern California, USA. The imagery is 1 m pixel resolution, and is public domain, courtesy of the U.S. Department of Agriculture, Farm Service Agency.

doi:10.1371/journal.pone.0131628.g005

ecologies, causes of decline, and larger-scale analyses [69] should yield the effective strategies to protect and restore species across broader landscapes into the future.

Supporting Information

S1 File. Tables of variable loadings from the principal component analyses, for principal components used in potential and current models.

(DOC)

S2 File. Figures representing the per-sample unit coefficient of variation among 10 runs for the potential and current models of the distribution of the arroyo toad.

(DOC)

Acknowledgments

Data used in this study were made available by the U.S. Fish and Wildlife Service, California Dept. of Fish and Wildlife, San Diego Association of Governments, Marine Corps Base Camp Pendleton, and the U.S. Forest Service. We thank Kirsten Winter (Cleveland National Forest) for sharing U.S. Forest Service documents, and USGS personnel, particularly Cheryl Brehme, Chris Brown, Denise Clark, and Carlton Rochester, who collected data used in this study, and Jeff Tracey and Kristine Preston for insights on analyses. Comments were received from Rich Inman and two anonymous reviewers. MLT also thanks the Herpetology Lab at Texas A&M for insight. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This paper is contribution number 511 of the U.S. Geological Survey's Amphibian Research and Monitoring Initiative and publication number 1504 of the Biodiversity Research and Teaching Collections at Texas A&M University.

Author Contributions

Conceived and designed the experiments: MLT RNF LAF. Analyzed the data: MLT. Contributed reagents/materials/analysis tools: MLT RNF LAF. Wrote the paper: MLT RNF LAF.

References

1. Millenium Ecosystem Assessment. Ecosystems and Human Well-Being: Synthesis Washington, DC: World Resources Institute; 2005.
2. Zedler JB, Kercher S. Causes and consequences of invasive plants in wetlands: opportunities, opportunists, and outcomes. *Crit Rev Plant Sci*. 2004; 23(5):431–52. doi: [10.1080/07352680490514673](https://doi.org/10.1080/07352680490514673)
3. Poff NL, Allan JD, Bain MB, Karr JR, Prestegard KL, Richter BD, et al. The natural flow regime. *Bioscience*. 1997; 47(11):769–84.
4. Zavaleta ES, Hobbs RJ, Mooney HA. Viewing invasive species removal in a whole-ecosystem context. *Trends Ecol Evol*. 2001; 16(8):454–9. [http://dx.doi.org/10.1016/S0169-5347\(01\)02194-2](http://dx.doi.org/10.1016/S0169-5347(01)02194-2).
5. Clewell A, Rieger JP. What practitioners need from restoration ecologists. *Restor Ecol*. 1997; 5(4):350–4. doi: [10.1046/j.1526-100X.1997.00548.x](https://doi.org/10.1046/j.1526-100X.1997.00548.x)
6. Miller JR, Hobbs RJ. Habitat restoration—do we know what we're doing? *Restor Ecol*. 2007; 15(3):382–90. doi: [10.1111/j.1526-100X.2007.00234.x](https://doi.org/10.1111/j.1526-100X.2007.00234.x)
7. Franklin J. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge, UK: Cambridge University Press; 2009.
8. Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, et al. Ecological Niches and Geographic Distributions (MPB-49): Princeton University Press; 2011.
9. Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, et al. Extinction risk from climate change. *Nature*. 2004; 427(6970):145–8. doi: [10.1038/nature02121](https://doi.org/10.1038/nature02121) PMID: [14712274](https://pubmed.ncbi.nlm.nih.gov/14712274/)
10. Loarie SR, Carter BE, Hayhoe K, McMahon S, Moe R, Knight CA, et al. Climate change and the future of California's endemic flora. *PLoS ONE*. 2008; 3(6):e2502. doi: [10.1371/journal.pone.0002502](https://doi.org/10.1371/journal.pone.0002502) PMID: [18648541](https://pubmed.ncbi.nlm.nih.gov/18648541/)
11. Raxworthy CJ, Martinez-Meyer E, Horning N, Nussbaum RA, Schneider GE, Ortega-Huerta MA, et al. Predicting distributions of known and unknown reptile species in Madagascar. *Nature*. 2003; 426(6968):837–41. PMID: [14685238](https://pubmed.ncbi.nlm.nih.gov/14685238/)
12. Pearson RG, Raxworthy CJ, Nakamura M, Townsend Peterson A. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J Biogeogr*. 2007; 34(1):102–17. doi: [10.1111/j.1365-2699.2006.01594.x](https://doi.org/10.1111/j.1365-2699.2006.01594.x)
13. Wang H-H, Wonkka CL, Treglia ML, Grant WE, Smeins FE, Rogers WE. Species distribution modelling for conservation of an endangered endemic orchid. *AoB Plants*. 2015. doi: [10.1093/aobpla/plv039](https://doi.org/10.1093/aobpla/plv039)
14. Smolik MG, Dullinger S, Essl F, Kleinbauer I, Leitner M, Peterseil J, et al. Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *J Biogeogr*. 2010; 37(3):411–22. doi: [10.1111/j.1365-2699.2009.02227.x](https://doi.org/10.1111/j.1365-2699.2009.02227.x)
15. Rodda G, Jarnevich C, Reed R. What parts of the US mainland are climatically suitable for invasive alien pythons spreading from Everglades National Park? *Biol Invasions*. 2009; 11(2):241–52. doi: [10.1007/s10530-008-9228-z](https://doi.org/10.1007/s10530-008-9228-z)
16. Barrows CW, Preston KL, Rotenberry JT, Allen MF. Using occurrence records to model historic distributions and estimate habitat losses for two psammophilic lizards. *Biol Conserv*. 2008; 141(7):1885–93. <http://dx.doi.org/10.1016/j.biocon.2008.05.006>.
17. Stanton JC, Pearson RG, Horning N, Ersts P, Reşit Akçakaya H. Combining static and dynamic variables in species distribution models under climate change. *Methods Ecol Evol*. 2012; 3(2):349–57. doi: [10.1111/j.2041-210X.2011.00157.x](https://doi.org/10.1111/j.2041-210X.2011.00157.x)
18. Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, et al. Predicting species distributions for conservation decisions. *Ecol Lett*. 2013; 16(12):1424–35. doi: [10.1111/ele.12189](https://doi.org/10.1111/ele.12189) PMID: [24134332](https://pubmed.ncbi.nlm.nih.gov/24134332/)
19. Hammerson G, Santos-Barrera G. *Anaxyrus californicus* 2004. IUCN 2013 IUCN Red List of Threatened Species Version 20131.
20. Sweet SS, Sullivan BK. *Bufo californicus* Camp, 1915 (Arroyo Toad). In: Lannoo M, editor. Amphibian Declines: The Conservation Status of United States Species. Berkley, CA: University of California Press; 2005. p. 396–400.
21. Griffin PC, Case TJ. Terrestrial habitat preferences of adult arroyo southwestern toads. *J Wildl Manage*. 2001; 65(4):633–44.

22. U.S. Fish and Wildlife Service. Endangered and Threatened Wildlife and Plants: Determination of the Endangered Status for the Arroyo Southwestern Toad. Federal Register. 1994; 59(241):64859–68.
23. U.S. Fish and Wildlife Service. Arroyo Toad (*Bufo californicus* (= *microscaphus*)) 5-Year Review: Summary and Evaluation. Ventura, California: Ventura Fish and Wildlife Office, 2009.
24. Miller DAW, Brehme CS, Hines JE, Nichols JD, Fisher RN. Joint estimation of habitat dynamics and species interactions: disturbance reduces co-occurrence of non-native predators with an endangered toad. J Anim Ecol. 2012; 81(6):1288–97. doi: [10.1111/j.1365-2656.2012.02001.x](https://doi.org/10.1111/j.1365-2656.2012.02001.x) PMID: [22702337](https://pubmed.ncbi.nlm.nih.gov/22702337/)
25. U.S. Fish and Wildlife Service. Endangered and threatened wildlife and plants; 12-month finding on a petition to downlist the arroyo toad (*Anaxyrus californicus*), and a proposed rule to reclassify the arroyo toad as threatened. Federal Register. 2014; 79(59):17106–25.
26. Ervin EL, Beaman KR, Fisher RN. Correction of locality records for the endangered Arroyo Toad (*Anaxyrus californicus*) from the desert region of southern California. Bulletin, Southern California Academy of Sciences. 2013; 112(3):197–205. doi: [10.3160/0038-3872-112.3.197](https://doi.org/10.3160/0038-3872-112.3.197)
27. Dayton GH, Fitzgerald LA. Habitat suitability models for desert amphibians. Biol Conserv. 2006; 132:40–9. doi: [10.1016/j.biocon.2006.03.012](https://doi.org/10.1016/j.biocon.2006.03.012)
28. Barto WS. Predicting potential habitat for the arroyo toad (*Bufo microscaphus californicus*) in San Diego County using a Habitat Suitability Model and Digital Terrain Data. San Diego, CA: San Diego State University; 1999.
29. U.S. Geological Survey. Watershed Boundary Dataset. 2012. Available: http://water.usgs.gov/GIS/wbd_huc8.pdf
30. Biggs TW, Atkinson E, Powell R, Ojeda-Revah L. Land cover following rapid urbanization on the US-Mexico border: Implications for conceptual models of urban watershed processes. Landscape Urban Plann. 2010; 96(2):78–87.
31. Syphard AD, Clarke KC, Franklin J, Regan HM, McGinnis M. Forecasts of habitat loss and fragmentation due to urban growth are sensitive to source of input data. J Environ Manage. 2011; 92(7):1882–93. doi: [10.1016/j.jenvman.2011.03.014](https://doi.org/10.1016/j.jenvman.2011.03.014) PMID: [21477919](https://pubmed.ncbi.nlm.nih.gov/21477919/)
32. Regan HM, Hierl LA, Franklin J, Deutschman DH, Schmalbach HL, Winchell CS, et al. Species prioritization for monitoring and management in regional multiple species conservation plans. Divers Distrib. 2008; 14(3):462–71. doi: [10.1111/j.1472-4642.2007.00447.x](https://doi.org/10.1111/j.1472-4642.2007.00447.x)
33. Lovich RE. Phylogeography and Conservation of the Arroyo Toad (*Bufo californicus*). Loma Linda, CA: Loma Linda University; 2009.
34. Mitrovich MJ, Gallegos EA, Lyren LM, Lovich RE, Fisher RN. Habitat use and movement of the endangered Arroyo Toad (*Anaxyrus californicus*) in coastal southern California. J Herpetol. 2011; 45(3):319–28. doi: [10.1670/10-160.1](https://doi.org/10.1670/10-160.1)
35. NASA Landsat Program. Landsat TM. Sioux Falls, SD: USGS; 2010. Available: <http://glovis.usgs.gov/>.
36. Crist EP, Cicone RC. A physically-based transformation of Thematic Mapper data- the TM tasseled cap. IEEE Transactions on Geoscience and Remote Sensing. 1984; GE- 22(3):256–63.
37. Song C, Woodcock CE, Seto KC, Lenney MP, Macomber SA. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? Remote Sens Environ. 2001; 75(2):230–44. [http://dx.doi.org/10.1016/S0034-4257\(00\)00169-3](http://dx.doi.org/10.1016/S0034-4257(00)00169-3).
38. GRASS Development Team. Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation Project. 2012. Available: <http://grass.osgeo.org>.
39. Böhner J. System for Automated Geoscientific Analyses (SAGA GIS), version 2.1.1. 2013. Available: www.saga-gis.org.
40. Jolliffe IT. Principal Component Analysis. New York, NY: Springer-Verlag; 2002
41. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. vegan: Community Ecology Package. R package version 2.0–4. 2012.
42. R Core Team. R: A language and environment for statistical computing.: R Foundation for Statistical Computing, Vienna, Austria. 2013. Available: <http://www.R-project.org>.
43. Hollingsworth BD. The Distribution of the Arroyo Toad (*Anaxyrus californicus*) at Cottonwood and Kitchen Creeks in the Cleveland National Forest, San Diego County, California. Report prepared for Cleveland National Forest, U.S. Forest Service, San Diego, CA; 2009.
44. Wells J, Turnbull J. 1997 Sensitive Species Survey Results for Pine Creek and Hauser Canyon Wilderness Areas, Descanso Ranger District, Cleveland National Forest. Report Prepared for Cleveland National Forest, U.S. Forest Service, San Diego, CA; 1998.
45. Atkinson AJ, Yang BS, Fisher RN, Ervin E, Case TJ, Scott N, et al. MCB Camp Pendleton arroyo toad monitoring protocol: 1. Summary of results from a workshop on August 27, 2002; 2. Monitoring protocol

- and targeted studies. U.S. Geological Survey Protocol prepared for Marine Corps Base Camp Pendleton; 2003.
46. Breiman L. Random forests. *Machine Learning*. 2001; 45(1):5–32. doi: [10.1023/a:1010933404324](https://doi.org/10.1023/a:1010933404324)
 47. Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, et al. Random forests for classification in ecology. *Ecology*. 2007; 88(11):2783–92. doi: [10.1890/07-0539.1](https://doi.org/10.1890/07-0539.1) PMID: [18051647](https://pubmed.ncbi.nlm.nih.gov/18051647/)
 48. Evans JS, Murphy MA, Holden ZA, Cushman SA. Modeling species distribution and change using Random Forest. In: Drew CA, Wiersma YF, Huettmann F, editors. *Predictive Species and Habitat Modeling in Landscape Ecology*. New York, NY: Springer; 2011. p. 139–59.
 49. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York, NY: Springer; 2009.
 50. Evans J, Cushman S. Gradient modeling of conifer species using random forests. *Landscape Ecol*. 2009; 24(5):673–83. doi: [10.1007/s10980-009-9341-0](https://doi.org/10.1007/s10980-009-9341-0)
 51. Oliveira S, Oehler F, San-Miguel-Ayanz J, Camia A, Pereira JMC. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *For Ecol Manage*. 2012; 275(0):117–29. <http://dx.doi.org/10.1016/j.foreco.2012.03.003>.
 52. Hernandez PA, Franke I, Herzog SK, Pacheco V, Paniagua L, Quintana HL, et al. Predicting species distributions in poorly-studied landscapes. *Biodivers Conserv*. 2008; 17(6):1353–66. doi: [10.1007/s10531-007-9314-z](https://doi.org/10.1007/s10531-007-9314-z)
 53. Senay SD, Worner SP, Ikeda T. Novel three-step pseudo-absence selection technique for improved species distribution modelling. *PLoS ONE*. 2013; 8(8):e71218. doi: [10.1371/journal.pone.0071218](https://doi.org/10.1371/journal.pone.0071218) PMID: [23967167](https://pubmed.ncbi.nlm.nih.gov/23967167/)
 54. Chawla NV, Lazarevic A, Hall LO, Bowyer KW. SMOTEBoost: Improving Prediction of the Minority Class in Boosting. In: Lavrač N, Gamberger D, Todorovski L, Blockeel H, editors. *Knowledge Discovery in Databases: PKDD 2003. Lecture Notes in Computer Science*. 2838: Springer Berlin Heidelberg; 2003. p. 107–19.
 55. Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, et al. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol Appl*. 2009; 19(1):181–97. doi: [10.1890/07-2153.1](https://doi.org/10.1890/07-2153.1) PMID: [19323182](https://pubmed.ncbi.nlm.nih.gov/19323182/)
 56. Fitzpatrick MC, Gotelli NJ, Ellison AM. MaxEnt versus MaxLike: empirical comparisons with ant species distributions. *Ecosphere*. 2013; 4(5):art55. doi: [10.1890/es13-00066.1](https://doi.org/10.1890/es13-00066.1)
 57. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol Evol*. 2012; 3(2):327–38. doi: [10.1111/j.2041-210X.2011.00172.x](https://doi.org/10.1111/j.2041-210X.2011.00172.x)
 58. Liaw A, Wiener M. Classification and regressoin by randomForest. *R News*. 2002; 2(3):18–22.
 59. Murphy MA, Evans JS, Storfer A. Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology*. 2010; 91(1):252–61. doi: [10.1890/08-0879.1](https://doi.org/10.1890/08-0879.1) PMID: [20380214](https://pubmed.ncbi.nlm.nih.gov/20380214/)
 60. Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Model*. 2006; 190(3–4):231–59. doi: [10.1016/j.ecolmodel.2005.03.026](https://doi.org/10.1016/j.ecolmodel.2005.03.026)
 61. Allouche O, Tsoar A, Kadmon R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol*. 2006; 43(6):1223–32. doi: [10.1111/j.1365-2664.2006.01214.x](https://doi.org/10.1111/j.1365-2664.2006.01214.x)
 62. U.S. Fish and Wildlife Service. Revised Critical Habitat for the Arroyo Toad (*Anaxyrus californicus*) Federal Register. 2009; 74(196):52612–64.
 63. Allan JD, Castillo MM. *Stream Ecology: Structure and Function of Running Waters*. Dordrecht, The Netherlands: Springer; 2007.
 64. Wickham JD, Stehman SV, Gass L, Dewitz J, Fry JA, Wade TG. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sens Environ*. 2013; 130(0):294–304. <http://dx.doi.org/10.1016/j.rse.2012.12.001>.
 65. King RS, Baker ME, Whigham DF, Weller DE, Jordan TE, Kazyak PF, et al. Spatial considerations for linking watershed land cover to ecological indicators in streams. *Ecol Appl*. 2005; 15(1):137–53. doi: [10.1890/04-0481](https://doi.org/10.1890/04-0481)
 66. King RS, Baker ME, Kazyak PF, Weller DE. How novel is too novel? Stream community thresholds at exceptionally low levels of catchment urbanization. *Ecol Appl*. 2011; 21(5):1659–78. doi: [10.1890/10-1357.1](https://doi.org/10.1890/10-1357.1) PMID: [21830709](https://pubmed.ncbi.nlm.nih.gov/21830709/)
 67. Riley SPD, Busted GT, Kats LB, Vandergon TL, Lee LFS, Dagit RG, et al. Effects of urbanization on the distribution and abundance of amphibians and invasive species in southern California streams. *Conserv Biol*. 2005; 19(6):1894–907.

68. Brown CJ, Saunders MI, Possingham HP, Richardson AJ. Interactions between global and local stressors of ecosystems determine management effectiveness in cumulative impact mapping. *Divers Distrib.* 2013; 20(5):538–46. doi: [10.1111/ddi.12159](https://doi.org/10.1111/ddi.12159)
69. Treglia ML. *Multi-scale Conservation in an Altered Landscape: The Case of the Endangered Arroyo Toad in southern California.* College Station, TX: Texas A&M University; 2014.
70. Gesch DB. The National Elevation Dataset. In: Maune D, editor. *Digital Elevation Model Technologies and Applications: The DEM Users Manual, 2nd Edition.* Bethesda, MD: American Society for Photogrammetry and Remote Sensing; 2007. p. 99–118.
71. Gallant JC, Dowling TI. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour Res.* 2003; 39(12):1347. doi: [10.1029/2002wr001426](https://doi.org/10.1029/2002wr001426)
72. Sappington JM, Longshore KM, Thompson DB. Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert. *J Wildl Manage.* 2007; 71(5):1419–26. doi: [10.2193/2005-723](https://doi.org/10.2193/2005-723)
73. Gruber S, Peckham S. Land-Surface Parameters and Objects in Hydrology. In: Tomislav H, Hannes IR, editors. *Developments in Soil Science.* Volume 33: Elsevier; 2009. p. 171–94.