Ecological Applications, 31(1), 2021, e02228

Improving area of occupancy estimates for parapatric species using distribution models and support vector machines

Jamie M. Kass,[1,2,3,6] Sarah I. Meenan,[2] Nicolás Tinoco,[4] Santiago F. Burneo,[4] and Robert P. Anderson[1,2,5]

[1]*Ph.D. Program in Biology, The Graduate Center, CUNY, New York, New York 10016 USA*
[2]*Department of Biology, City College of New York (CUNY), New York, New York 10031 USA*
[3]*Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Tancha, Onna-son, Kunigami-gun, Okinawa 904-0495 Japan*
[4]*Museo de Zoología, Pontificia Universidad Católica del Ecuador, Avenida 12 de Octubre, 1076 y Roca,170517Quito, Ecuador*
[5]*Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, New York, New York 10024 USA*

Citation: Kass, J. M., S. I. Meenan, N. Tinoco, S. F. Burneo, and R. P. Anderson. 2021. Improving area of occupancy estimates for parapatric species using distribution models and support vector machines. Ecological Applications 31(1):e02228. 10.1002/eap.2228

Abstract. As geographic range estimates for the IUCN Red List guide conservation actions, accuracy and ecological realism are crucial. IUCN's extent of occurrence (EOO) is the general region including the species' range, while area of occupancy (AOO) is the subset of EOO occupied by the species. Data-poor species with incomplete sampling present particular difficulties, but species distribution models (SDMs) can be used to predict suitable areas. Nevertheless, SDMs typically employ abiotic variables (i.e., climate) and do not explicitly account for biotic interactions that can impose range constraints. We sought to improve range estimates for data-poor, parapatric species by masking out areas under inferred competitive exclusion. We did so for two South American spiny pocket mice: *Heteromys australis* (Least Concern) and *Heteromys teleus* (Vulnerable due to especially poor sampling), whose ranges appear restricted by competition. For both species, we estimated EOO using SDMs and AOO with four approaches: occupied grid cells, abiotic SDM prediction, and this prediction masked by approximations of the areas occupied by each species' congener. We made the masks using support vector machines (SVMs) fit with two data types: occurrence coordinates alone; and coordinates along with SDM predictions of suitability. Given the uncertainty in calculating AOO for low-data species, we made estimates for the lower and upper bounds for AOO, but only make recommendations for *H. teleus* as its full known range was considered. The SVM approaches (especially the second one) had lower classification error and made more ecologically realistic delineations of the contact zone. For *H. teleus*, the lower AOO bound (a strongly biased underestimate) corresponded to Endangered (occupied grid cells), while the upper bounds (other approaches) led to Near Threatened. As we currently lack data to determine the species' true occupancy within the post-processed SDM prediction, we recommend that an updated listing for *H. teleus* include these bounds for AOO. This study advances methods for estimating the upper bound of AOO and highlights the need for better ways to produce unbiased estimates of lower bounds. More generally, the SVM approaches for post-processing SDM predictions hold promise for improving range estimates for other uses in biogeography and conservation.

Key words: area of occupancy; biotic interaction; competition; extent of occurrence; parapatric; range limits; Red List; rodent; species distribution model; support vector machine.

Introduction

Estimates of species' geographic ranges, derived from expert information, statistical models, or a combination of both, represent essential sources of information that guide conservation actions. Range estimates have a variety of uses in conservation biology, from prioritizing reserve networks (Urbina-Cardona and Flores-Villela 2010) to monitoring population trends (Noon et al. 2012). In particular, assessments of species' extinction risk by the IUCN Red List rely on geographic range estimates (IUCN 2019), which remain the principal sources of information for the vast majority of species (Gaston 2009). Red List range estimates are separated into two categories. Extent of occurrence (EOO) is defined as the "spatial spread of the areas currently occupied by the taxon" and is not intended as an estimate of occupied areas but as an indication of the spread of extinction risks to the taxon (IUCN 2019). Area of occupancy (AOO) represents the "area of suitable habitat currently

Manuscript received 12 September 2019; revised 11 May 2020; accepted 13 July 2020. Corresponding Editor: Marissa L. Baskett.
[6] E-mail: jamie.m.kass@gmail.com

occupied by the taxon" within the EOO at a defined reference scale of 2 × 2 km (4 km^2; IUCN 2019). The simplest approaches are to estimate EOO based on a convex hull around occurrence localities, and AOO by totaling the area of occupied grid cells at the defined resolution that hold occurrences (IUCN 2019). However, calculating these geographic range metrics accurately is especially difficult for poorly known species with few occurrence records and little natural history knowledge (Roberts et al. 2016).

With the advent of correlative statistical modeling techniques such as species distribution models (SDMs; also called ecological niche models) and documentation of best practices for their use (Peterson et al. 2011, Araújo et al. 2019), the potential geographic ranges of most species can be estimated. Modeled range estimates can be considered in assessments of EOO and AOO given data on occurrence localities and the environment; section 4.10.7 of the IUCN Red List Guidelines specifies conditions for the use of SDMs in indirectly estimating these metrics (IUCN 2019). Some approaches instead derive species' range estimates using expert maps and/or habitat associations with existing geographic data (Ocampo-Peñuela and Pimm 2014). For example, the newly proposed "area of habitat" methodology uses remotely sensed data such as elevation and land cover maps to estimate an upper bound for AOO (Brooks et al. 2019). However, approaches using statistical models can improve range estimates, especially in areas with low sampling effort, as well as predict heterogeneity within the known range (Peterson et al. 2018). These predicted areas may represent significant parts of the true species' range that otherwise would not be included in range estimates (Marcer et al. 2013, Fivaz and Gonseth 2014, Zhang and Vincent 2018). Alternatively, they may also represent suitable areas that are unoccupied for other reasons such as anthropogenic factors, dispersal barriers, Allee effects, or lack of suitable land cover, so such factors must be taken into account in post-processing.

Although often overlooked, additionally considering biotic interactions (e.g., competition, parasitism, or predation) can help refine AOO estimates. Biotic interactions are increasingly recognized as important drivers of species ranges at the macroscale, and many studies have recommended that they be considered in some way alongside environmental variables to improve range estimates (Wisz et al. 2013, Anderson 2017). In fact, the Red List Guidelines advocate employing "both biological and statistical considerations" to better ensure that each range estimate is an "accurate representation of the habitat requirements of the species" (section 4.10.7; IUCN 2019). Including biotic variables in a modeling analysis can help prevent overprediction of species' ranges that are estimated with SDMs (Freeman and Mason 2015). Alternatively, when using models to estimate AOO, post-processing to remove areas within the prediction that are unoccupied for biotic reasons should result in a better representation of "suitable habitat occupied by the taxon." Such considerations of biotic factors can directly benefit conservation actions, such as designing reserve networks better tailored for species of concern that face threats from climate change (Hof et al. 2012).

Delineating biotically unsuitable areas of a species' range may be a difficult process due to the lack of data on biotic interactions at the macroscale (Wisz et al. 2013), but this would constitute new and useful information for species of concern. Predictions from SDMs that use only abiotic variables (the vast majority of such studies) will likely fail to account for range limits due to biotic interactions. Other options are to include the occurrences or SDM predictions of interacting species as predictor variables (Bateman et al. 2012). Yet, theory indicates that interacting species should only be included as SDM predictor variables if they are not themselves affected by the focal species (Soberón 2007, Anderson 2017). Often, competitive relationships between closely related species are bidirectional in effect, disqualifying either species from inclusion as a predictor variable in an SDM for the other. Thus, addressing inferred biotically unsuitable area for such interactions is better done by post-processing model predictions of potential distribution. A good example for this would be closely related (e.g., congeneric) parapatric species that replace each other across space, presumably due to bidirectional effects on each other's distributions, a common phenomenon. If the geographic boundary between ranges is caused by competitive exclusion, the range of the focal species' congener would be defined as biotically unsuitable area and could be removed from the estimate of the focal species' potential distribution. Some studies have predicted range boundaries of congeneric parapatric species with SDMs under the assumption of competitive exclusion by removing areas from each species' estimated range with higher abiotic suitability for the congener (Anderson and Martínez-Meyer 2004, Gutiérrez et al. 2014). However, this purely environmental approach for predicting range boundaries may yield results that are not biogeographically plausible. For example, areas with high abiotic suitability predictions for species A outside its true range may exist in the true range of congener species B, possibly quite far from the zone of contact.

Support vector machines (SVMs) are models with multiple characteristics that make them attractive for predicting range boundaries in such circumstances. They are supervised learning models that classify data into groups and have tuning settings that control the complexity of fit; these settings can be optimized using cross validation. In the case of species occurrence data, SVMs can classify the study extent with regard to the species most likely to be present in each grid cell when trained on the coordinates of occurrence localities. SVMs can additionally use other predictor variables besides spatial coordinates to guide the classification, such as SDM

predictions of suitability. Even for species with low sample sizes, range boundaries might still be reasonably predicted given that occurrence data exist for both species in the boundary region. Further, if hybrid individuals or known sites of sympatry exist along a narrow contact zone, ecological realism can be evaluated by plotting the coordinates of such localities on the range boundary classifications.

In the case of congeneric parapatric species, SVMs can thus be used to predict the range boundary when the locations of the congener provide essential information that environmental variables alone cannot capture. This predicted boundary can then be compared with existing expert maps or SDM predictions and used to remove areas inferred to be biotically unsuitable. Additionally, SVMs could be used in a similar way for species with other kinds of interactions that affect range boundaries such as parasitism or predation. Although there are a few studies that have demonstrated various ecological applications for SVMs (Drake et al. 2006, Pouteau et al. 2011), to date none have used them for classifying range boundaries. It is important to note that although we focus here on SVMs, other machine learning methods are available for making similar classifications based on spatial predictors, and other methods exist for delineating species ranges based on spatial data (i.e., alpha hulls, kernel density estimators; Fortin et al. 2005). Our aim here is not to determine an optimal method, but rather to demonstrate the particular utility of SVMs for classifying range boundaries between species. In this study, we investigate whether (1) considering biotic interactions in SDMs can lead to real differences in IUCN Red List classifications when compared with standard approaches, and whether (2) the accuracy and ecological realism of SDM-derived AOO estimates can be improved for two data-poor parapatric species using SVMs.

To demonstrate these techniques, we focus on the spiny pocket mice *Heteromys australis* (Least Concern; Anderson et al. 2018) and *Heteromys teleus* (Vulnerable D2; Naylor and Roach 2018) in Ecuador and southwestern Colombia, west of the crest of the Andes. This region encompasses the southern part of the range of *H. australis* (the part with the highest quality occurrence data) and the full known range of *H. teleus*. Both species have poor sampling for this region (Anderson 2015), typical for tropical species and less-studied groups in other areas, and neither has a currently listed AOO estimate. For this region, both species have verified occurrence localities inside or proximal to (<2 km from) some protected areas. Sampling efforts span many years for both species, which are considered parapatric because of a thin region of possible sympatry. Competitive exclusion was inferred to explain biogeographic patterns of *H. australis* and a different parapatric congener (*H. anomalus*; Anderson et al. 2002), and no clear climatic demarcations were found to define the range boundary with *H. teleus* (Anderson and Martínez-Meyer 2004, Shcheglovitova and Anderson 2013). We thus interpret it is likely that *H. australis* and *H. teleus* have bidirectional biotic effects on each other's ranges. Two previous studies have modeled the ranges of these two species: an older conservation assessment with very coarse environmental data (Anderson and Martínez-Meyer 2004), and a more recent study with finer resolution data that addressed model evaluation for small sample sizes (Shcheglovitova and Anderson 2013). However, no recent conservation assessments have been done for these species using high-resolution environmental data and modern modeling methods that take biotic interactions into account.

As occurrence data for both species in this region are limited and suffer from sampling bias, IUCN range estimates that do not use SDMs for prediction of suitable areas will likely be vast underestimates. As mentioned above however, if SDMs are employed and occupied areas are predicted while ignoring inferred biotic interactions, AOO can be overestimated. It is clear that the range of one species should not be included in calibrating the SDM of the other when bidirectional effects are inferred (Anderson 2017). Here we use SDMs to derive EOO estimates for each species and then implement SVMs to mask out regions of the EOO predicted to be within the range of each species' parapatric congener. We use these results to make estimates of AOO, but make recommendations regarding threat status only for *H. teleus*. To our knowledge, no further information exists to determine the occupancy status within the modeled range of the species, and hence we report a gradient of estimates to be considered as upper and lower bounds of AOO (Section 3.2; IUCN 2019). We estimate AOO in different ways to account for this inherent uncertainty: (1) occupied grid cells, (2) SDM range prediction, (3) the range prediction masked using a SVM with spatial predictors only (occurrence locality coordinates), and (4) the same as (3) but also including environmental (SDM-derived) predictors (Fig. 1). For each approach, we estimate AOO bounds before and after considering current forest cover. We conclude by discussing the importance of this methodology for generating AOO gradients that consider uncertainty for low-data species and for making conservation decisions concerning assignments of threat categories.

Methods

Occurrence data

Heteromys australis ranges from Venezuela and Panama in the north to northwest Ecuador in the south, but localities from Ecuador and southwest Colombia (16 known occurrences) are widely separate from those in northern Colombia, likely due to sampling bias (Anderson and Jarrín-V. 2002). Here, we model this region of the species' distribution, which is common practice in conservation, e.g., country-specific red-listing

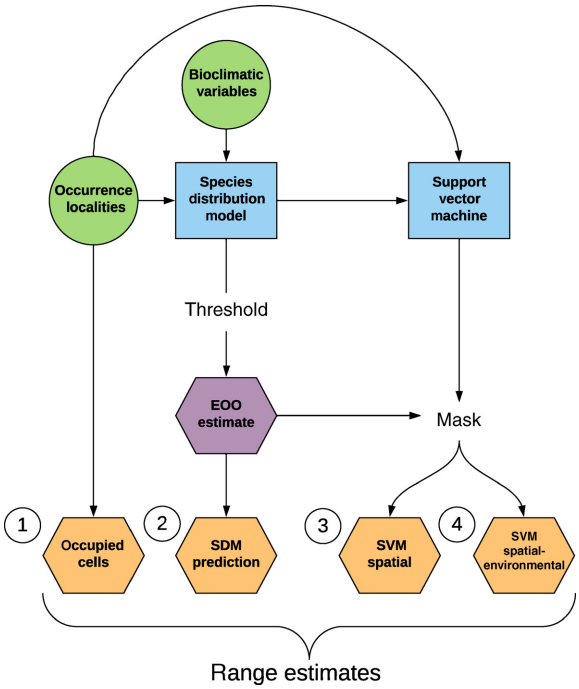

FIG. 1. In this flowchart of experimental design, data sources (green circles) and models (blue rectangles) undergo operations (i.e., threshold or mask) to produce extent of occurrence (EOO) estimates (purple hexagon) and area of occupancy (AOO) estimates (orange hexagons). All AOO estimates were calculated before and after considering current forest cover (not shown).

projects (Brito et al. 2010), as georeferenced occurrence data of high quality is not currently available for the more northerly populations. *Heteromys teleus*, with seven known occurrences, is restricted to central-western Ecuador (Anderson and Jarrín-V. 2002). Both species are found only in evergreen forest, but as a strong precipitation gradient exists in western Ecuador, *H. australis* is found mainly in wet and unseasonal areas and *H. teleus* in drier and more seasonal areas (Anderson and Martínez-Meyer 2004).

We updated occurrence data sets for each species based on museum specimens for a study area that encompasses the region west of the crest of the western Andes extending from southwestern Colombia (considered here as the area south of Tumaco Bay) through western Ecuador. We thoroughly vetted each set of geographic coordinates for each of the new localities. We obtained geographic coordinates from primary sources (documented on specimen tags or found in collectors' field notes) when available, and verified these by consulting topographic maps and contacting collectors. When coordinates were not available from these primary sources, we georeferenced localities using topographic maps based on the elevation and verbatim locality descriptions from specimen tags and collectors' field notes. We then combined these newly vetted occurrence data with previously published records (Anderson and Jarrín-V. 2002) and compiled an expanded data set to serve as input data for building SDMs (detailed in Appendix S1).

In addition to the 23 collection localities from this region reported by Anderson and Jarrín-V. (2002), we examined and report 34 specimens in total of the genus *Heteromys* from Ecuador, representing 16 additional collection localities (Appendix S1) and including some inferred hybrids (based on morphological characteristics). These derive from specimens in natural history museum collections, and we made identifications by examining morphological characters (Anderson and Jarrín-V. 2002), either in person or via photograph. We also considered measurements of hind-foot (recorded on specimen tags), a diagnostic trait for distinguishing *H. teleus* from *H. australis* (Anderson and Jarrín-V. 2002).

Species distribution models

For predictor variables, we selected a subset of eight bioclimatic variables from the WorldClim 2.0 data set (Fick and Hijmans 2017) at 30 arcsecond resolution ($\sim$1 km at the equator) that we hypothesized to be most closely associated with the distributions of these two species in this region, in accordance with the Red List Guidelines (section 4.10.7; IUCN 2019). These variables were mean diurnal range (bio02), temperature seasonality (bio04), minimum temperature of coldest month (bio06), mean precipitation of wettest month (bio13), mean precipitation of driest month (bio14), precipitation seasonality (bio15), mean precipitation of warmest quarter (bio18), and mean precipitation of coldest quarter (bio19). We chose WorldClim 2.0 because its broad temporal coverage (1970–2000) was appropriate for building SDMs with our occurrence data, which was collected over a wide temporal range (ranging from 1912 to 2016 with mean 1984). These variables, which relate to seasonality and climatic extremes, differentiated well between the wet and unseasonal north and the drier and more seasonal south; additionally, they provided information regarding differences between warmer lowlands and cooler highlands. We reduced this subset further by removing variables with high collinearity to aid in interpretation (Appendix S2).

Besides the occupied grid cells approach to estimating AOO, the others either represent SDM predictions or their post-processed output (Fig. 1). We carried out the same SDM procedure for both species using Wallace 1.0.6 (Kass et al. 2018), a modular and interactive ecological modeling application launched via the R package wallace. This procedure included spatial thinning of occurrence data to reduce clustering caused by sampling bias (package spThin; Aiello-Lammens et al. 2015) and iterative building of Maxent 3.4.1 (Phillips et al. 2017) models with settings that ranged from simple to complex, each with jackknife (leave-one-out) cross validation for small sample sizes (package ENMeval; Muscarella

et al. 2014). We selected optimal models sequentially by first choosing those that accurately predicted the most withheld occurrence localities (lowest average omission rate). We calculated omission rate based on the minimum training presence (MTP) threshold, which is the lowest predicted suitability value associated with the occurrence localities used to train the model. To break any ties, we then chose those models with the best discriminatory ability on the withheld occurrences (highest average test AUC, or area under the receiver operating characteristic curve). Although there are problems with interpreting AUC in absolute terms as a measure of accuracy for presence/background models (Lobo et al. 2008), it is a valid metric to compare model settings for a single species across the same study extent (Peterson et al. 2011). We then made binary range maps that predict presence and absence across a study extent by thresholding the SDM predictions of suitability for each species by the respective MTP value. All data preparation and analysis was done using the R programming language v3.5.1 (R Core Team 2018). More details on these methods can be found in Appendix S2.

Estimating extent of occurrence

For each species, we made estimations of EOO based on the binary range maps derived from the continuous SDM predictions. Standard practice for delineating EOO, which must be a boundary that encompasses all known, inferred, or projected sites of occurrence (section 4.9; IUCN 2019), is to draw a convex hull around all occurrence localities. But for poorly sampled species, including areas in the EOO that are predicted by SDMs as highly suitable can result in better representations of the true spatial extent of their ranges (Syfert et al. 2014, IUCN 2019). Although techniques have been developed to derive thresholds for producing binary range maps that best avoid overprediction (Syfert et al. 2014), in keeping with the definition of EOO, which states it must include all marginal localities (IUCN 2019), we chose to threshold using the MTP value. We then delineated convex hulls around the binary range predictions to create the EOO estimates. For areal comparison with another commonly implemented approach, we also generated EOO estimates using convex hulls around the occurrence localities (IUCN 2019).

Support vector machines

In order to better estimate the bounds of AOO (i.e., the occupied areas), we masked out inferred biotically unsuitable areas from the EOO by classifying the range of each species over the combined study extent using SVMs. These models can be used as simple classifiers to distinguish one data class from another. They differentiate data classes by finding the best-fitting boundary line (or lines) in predictor space, focusing only on those "vectors" (points) with the greatest "support," or those that are closest to others of a different class (Drake et al. 2006). This leads to a classification of grid cells within the combined study extent, indicating which species is more likely to be present. In geographic space, we made spatial classifications of each species' range with two SVMs fit with different predictor variables: one using the occurrence localities coordinates as predictors (spatial) and the other using these coordinates along with the SDM-derived suitability predictions for each species (spatial-environmental).

For each approach, we fit SVMs with the response variable as species identity and either the "spatial" or "spatial-environmental" predictor variables described above. We tuned the models with the Gaussian radial basis function for geometric ranges of the settings C (from 2^{-5} to 2^{15}) and gamma (from 2^{-15} to 2^{3}) and used 10-fold random cross validation for model evaluation (Hsu et al. 2003). The setting C (cost) specifies how strict the boundary should be at tolerating misclassifications: a lower value results in a simpler function and thus a smoother boundary, whereas a higher value results in a more complex function that prioritizes correct classifications. The setting gamma specifies how far away any particular point can be to influence the definition of the boundary: a lower value gives more influence to farther points whereas a higher value gives more influence to closer points. The settings with the lowest classification error after cross validation were chosen as optimal. As the random cross validation is stochastic, we ran 100 model replicates and chose the most frequently selected optimal settings for the final models. We fit all SVMs using the tune.svm() function from the R package e1071 (Meyer et al. 2017).

Within each species' estimated EOO, we used the SVMs to classify which cell belonged in which species' range, then masked out the areas classified to be in the range of the parapatric congener. As a qualitative assessment of ecological realism, we also plotted inferred hybrid occurrence localities on each support vector machine classification to determine the proximity of each to the contact zone. R functions for tuning the SVM model with/without SDM predictions and for making a raster prediction for the SVM can be found in Data S1: [rangeSVM.R].

Estimating area of occupancy bounds

We made estimates of the bounds of AOO by masking the estimated EOO using the four approaches outlined above (Fig. 1). For all approaches, we projected the binary rasters to UTM 17S (m) and resampled to 2 km resolution (using nearest neighbor interpolation for categorical data) before calculating area, in accordance with the IUCN suggestions for AOO (IUCN 2019). The occupied grid cells approach simply takes the area of all grid cells that overlap with occurrence localities, and the generic IUCN recommendation is 2×2 km; for this, we masked out all grid cells that did not overlap with at

least one occurrence locality. For the SDM prediction approach, we used the MTP-thresholded binary prediction (also used to delimit the EOO by convex hull). For the SVM approaches, we masked the binary SDM prediction by removing areas classified by the SVMs as within the range of the parapatric congener. To evaluate these AOO estimates, for each species we determined how many occurrence localities were misclassified. To avoid confusion, here we use "misclassification" to refer to localities of one species falling into the AOO bound estimate of another, whereas "omission" refers to a single species' model failing to predict occurrence localities of that species.

We calculated each AOO estimate before and after masking out grid cells that lack current broadleaf evergreen forest cover. We derived forest cover from the European Space Agency Climate Change Initiative (ESA CCI) Land Cover 2.0.7 data set, a global 300-m resolution categorical raster based on the UN Land Cover Classification System (European Space Agency 2018), for the most recent year available at the time of the analysis (2015). Although the SDMs were calibrated with climatic data from 1970–2000 for occurrence data mostly originating from this time period, we expected that any recent changes to these species' ranges likely resulted from anthropogenic changes to forest cover, thus, we aimed to use the most recent forest cover data to estimate these species' current range areas using the SDM predictions as baselines. We subsetted the forest cover raster to broad-leaved evergreen forest (value 50), cropped it to the shared extent, then projected it to UTM 17S and resampled to 2-km resolution. Finally, for each species we compared the areas of each of the resulting eight AOO estimates and calculated the percent difference between each and the SDM prediction approach (the highest areal estimate). Additionally, we derived the spatial similarity between each pair of SDM-based AOO estimates (with binary values) by calculating spatial overlap using Schoener's *D*. All raster operations were performed with the raster (Hijmans 2017) and dismo (Hijmans et al. 2017) R packages.

Results

Occurrence data

Of the 34 total *Heteromys* specimens newly examined for Ecuador and southwestern Colombia, we identified 24 as *H. australis* (representing 10 unique collection localities at the 1-km^2 scale), 6 as *H. teleus* (3 unique collection localities), and 3 as likely *H. australis* / *H. teleus* hybrids (2 unique collection localities). The specimens from the latter 2 localities displayed strange mixes of characters not found in any other specimens throughout Ecuador or southwestern Colombia (Appendix S1). Finally, one additional specimen, the only representative of a single collection locality, could not be identified beyond genus due to skull damage and is designated *Heteromys* sp. (Appendix S1). Addition of these new unique collection records increased the *H. australis* records from 16 to 26, and those of *H. teleus* from 7 to 10. Spatial thinning reduced these numbers to 20 for *H. australis* and 9 for *H. teleus*, and these occurrence data sets were used for analysis.

Some newly documented occurrence localities were proximal to (<2 km from) protected areas in Ecuador (Fig. 2). Based on the original occurrence locality data sets before the addition of those records in Appendix S1, both species have localities inside or proximal to protected areas in southern Colombia and Ecuador (Fig. 2; Ministerio del Ambiente de Ecuador 2019, Sistema de Información Ambiental de Colombia 2019). *Heteromys australis* was found in Los Cedros (n = 1, year = 1993) and close to Asociación Agrícola Carchi Imbabura (n = 1, year = 2000) in Ecuador, as well as Maindes and Cuenca Alta del Río Nembí (n = 1, year = 1995) in Colombia. *Heteromys teleus* was found in Cordillera Chongón-Colonche (n = 2, year = 1923), and close to Hacienda La Perla (n = 1, year = 1990) and Río Lelia (n = 1, year = 1996) in Ecuador. With the addition of the records from western Ecuador in Appendix S1, *H. australis* was also found in Cerro Golondrinas (n = 3, year = 2016) and Milpe Pachijal (n = 1, years = 2001–2002), while *H. teleus* was found in Tulipa Pachijal and San Francisco (n = 1, year = 2007).

SDMs and extent of occurrence estimates

The Maxent SDMs we selected for both species had settings different from default and had flexible, linear responses (hinge features; Appendix S3: Table S1). A total of six predictor variables were used for modeling, as we removed bio13 and bio15 from the data set due to high correlations with the other variables. The *H. australis* model was less complex (Appendix S3: Table S1) and had positive responses for mean diurnal temperature range (bio02), temperature seasonality (bio04), and precipitation of coldest quarter (bio19) (Appendix S3: Fig. S1). In contrast, the *H. teleus* model was more complex (Appendix S3: Table S1), with positive responses for precipitation of driest month (bio14), precipitation of warmest quarter (bio18), and precipitation of coldest quarter (bio19; Appendix S3: Fig. S1). The *H. teleus* SDM omitted proportionally fewer withheld occurrence localities on average (OR_{MTP} = 0.111) than that of *H. australis* (OR_{MTP} = 0.150). Across each species' study extent, the suitability predictions for *H. australis* were less variable (cloglog range, 0.43–0.75) than those of *H. teleus* (cloglog range, 0.22–1.00; Fig. 3), likely because there are greater differences in seasonality within the study extent of *H. teleus* (Fig. 3; Anderson and Martínez-Meyer 2004).

The EOO estimates for both species generated using SDM predictions were approximately triple the area of those based on occurrence localities. The convex hull of the binary SDM prediction (bounded by the study

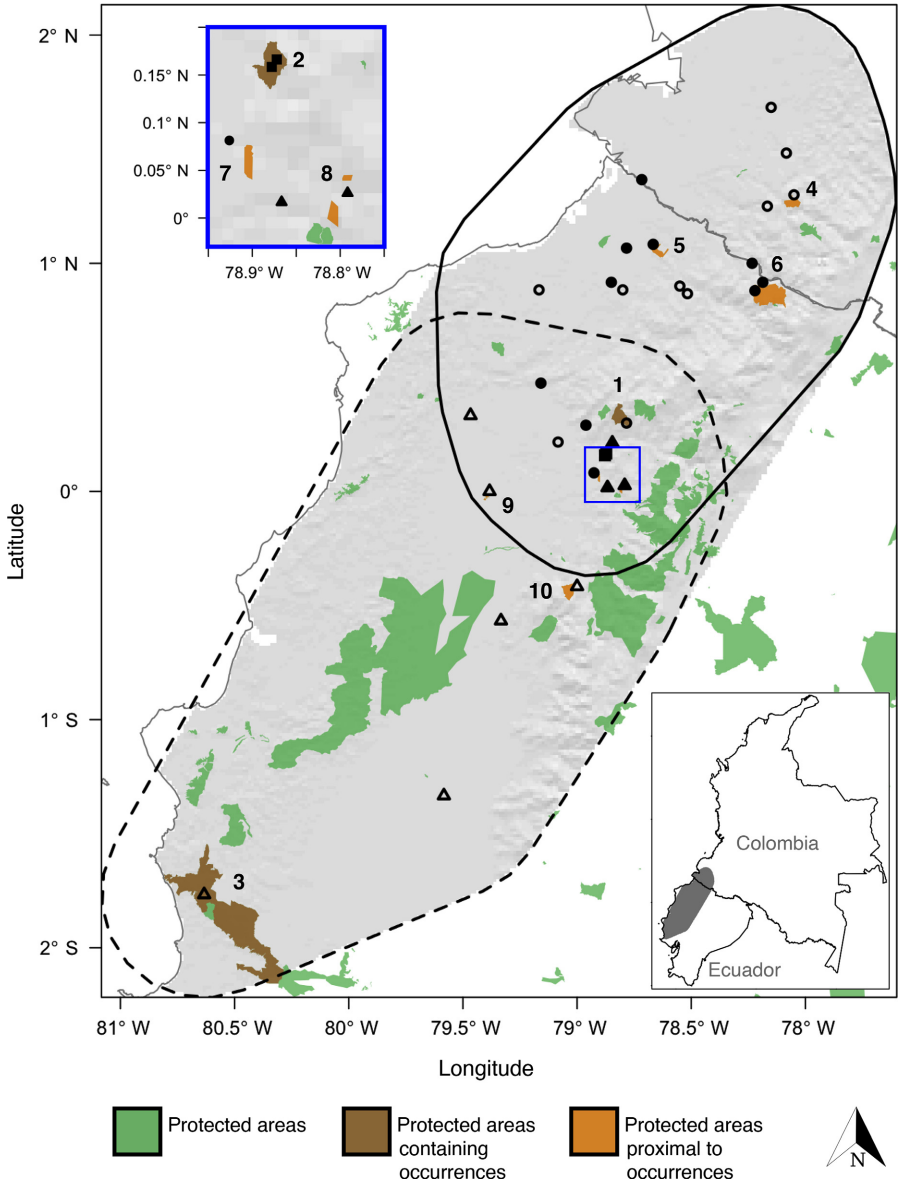

FIG. 2. Total occurrence localities (before thinning) for *Heteromys australis* (circles), *H. teleus* (triangles), and inferred hybrids (squares); filled shapes represent localities first reported in this study. Study extents are depicted as solid lines for *H. australis* and dotted lines for *H. teleus*, with the combined study extents represented by elevation hillshades. This region is also displayed for reference in gray with the political boundaries of Colombia and Ecuador (bottom right). Protected areas (Ministerio del Ambiente de Ecuador 2019, Sistema de Información Ambiental de Colombia 2019) are depicted in green, with those referred to in the text containing occurrence localities in brown and those proximal (<2 km) to localities in orange. Protected areas are labeled as follows (1–3 contain occurrences, 4–10 are proximal): (1) Los Cedros, (2) Mashpi, (3) Cordillera Chongón-Colonche, (4) Cuenca Alta del Río Nembí and Maindes, (5) Asociación Agrícola Carchi Imbabura, (6) Cerro Golondrinas, (7) Milpe Pachijal, (8) Tulipa Pachijal and San Francisco, (9) Hacienda La Perla, (10) Río Lelia.

extent, which excluded large areas likely beyond dispersal barriers) was 36,843 km^2 for *H. australis* (portion of range in southwestern Colombia and northern Ecuador; see *Methods*) and 54,985 km^2 for *H. teleus*, compared with 11,664 and 18,686 km^2 for the convex hulls of the occurrence localities, respectively. For *H. australis*, the EOO estimate excluded northwest areas with low mean diurnal temperature range (bio02), and areas in the east and southwest with drier conditions in the coldest quarter (bio19). For *H. teleus*, the EOO estimate included most of the area within the study extent, and excluded only the driest southern areas and those on the western coast that are driest in the warmest quarter (bio18). Both estimates included small areas of ocean, which are

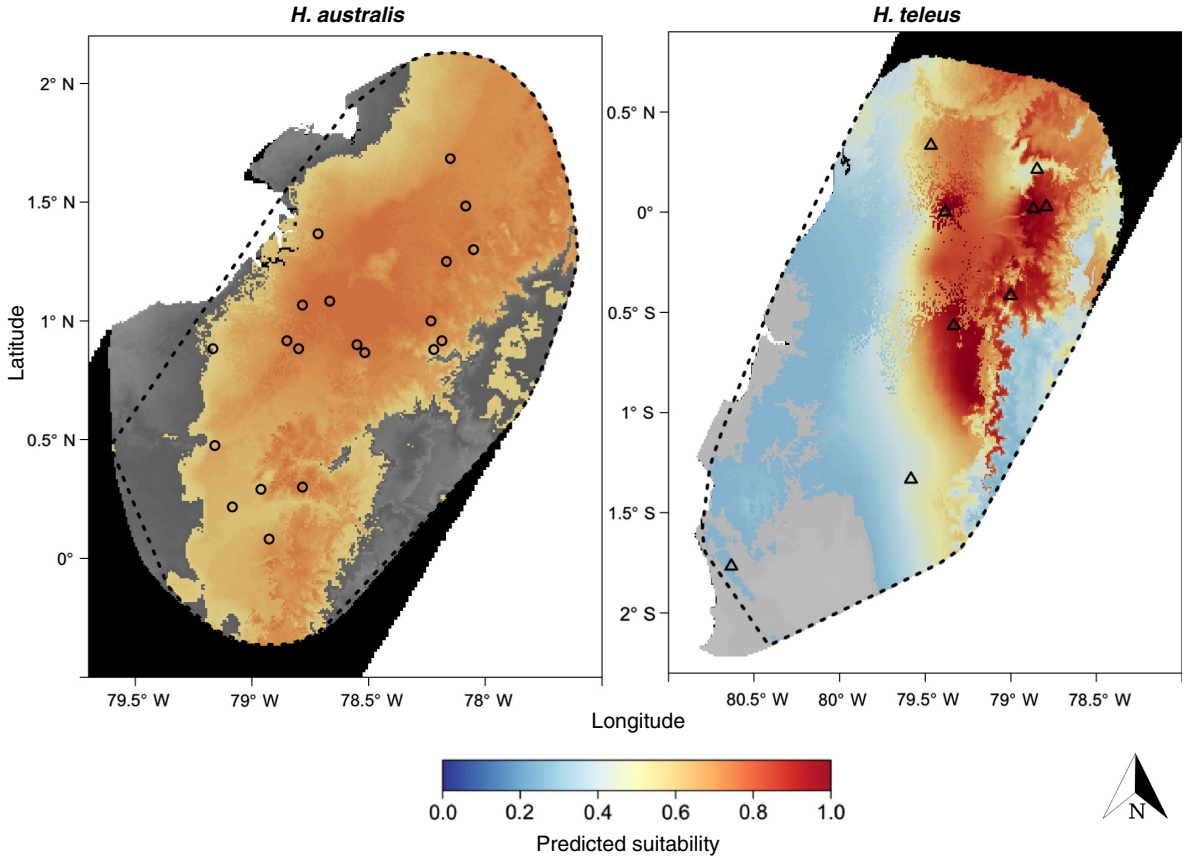

FIG. 3. Maxent species distribution model (SDM) continuous suitability predictions (cloglog transformation) thresholded by the minimum training presence (MTP) for *H. australis* and *H. teleus* on a scale from blue (low) to red (high). The estimation of extent of occurrence (EOO), a convex hull around the thresholded SDM prediction, is displayed as a dashed line, and occurrence localities for *H. australis* and *H. teleus* are depicted as circles and triangles, respectively. Areas outside the MTP threshold are displayed in a gray elevation hillshade, while areas outside the study extent used for model training are displayed in black.

relatively negligible and also in agreement with how the IUCN defines EOO (IUCN 2019).

Tuned support vector machine settings and classifications

The SVM settings that were optimal based on random cross validation were similar for both spatial and spatial-environmental. Both had a cost of 32, but gamma was higher for the spatial SVM (0.5) than for the spatial-environmental SVM (0.03), indicating the latter had greater weight on localities farther from the boundary. Both models omitted only one occurrence locality per species near the contact zone. The distance from the omitted *H. australis* locality to the closest cell centroid classified for this species was greater for the spatial SVM (1129 m) than the spatial-environmental SVM (745 m). Notably though, this same difference for the omitted *H. teleus* locality was more pronounced (spatial, 8,278 m; spatial-environmental, 3,363 m). Within the contact zone, the spatial-environmental SVM followed areas predicted to be climatically suitable for each species in addition to their occurrence coordinates (Fig. 4), making the classifications more ecologically realistic. In further support of this, the two inferred hybrid occurrence localities fell closer to the border between the two ranges predicted by the spatial-environmental SVM than by the spatial SVM (Fig. 4).

Area of occupancy estimates

The occupied grid cells approach had by far the lowest AOO estimate (Appendix S3: Fig. S2), while the SDM prediction had the highest, constituting the lower and upper bounds of AOO estimates for each species (Table 1). Both species had a single occurrence locality associated with lack of current forest cover, but the corresponding proportion of unforested localities for *H. teleus* (10%) was higher than that for *H. australis* (4%). The estimate of the spatial-environmental SVM was higher than that of the spatial SVM for *H. australis*, though the inverse was true for *H. teleus* (Table 1). Overall, considering only currently forested areas (56% of the

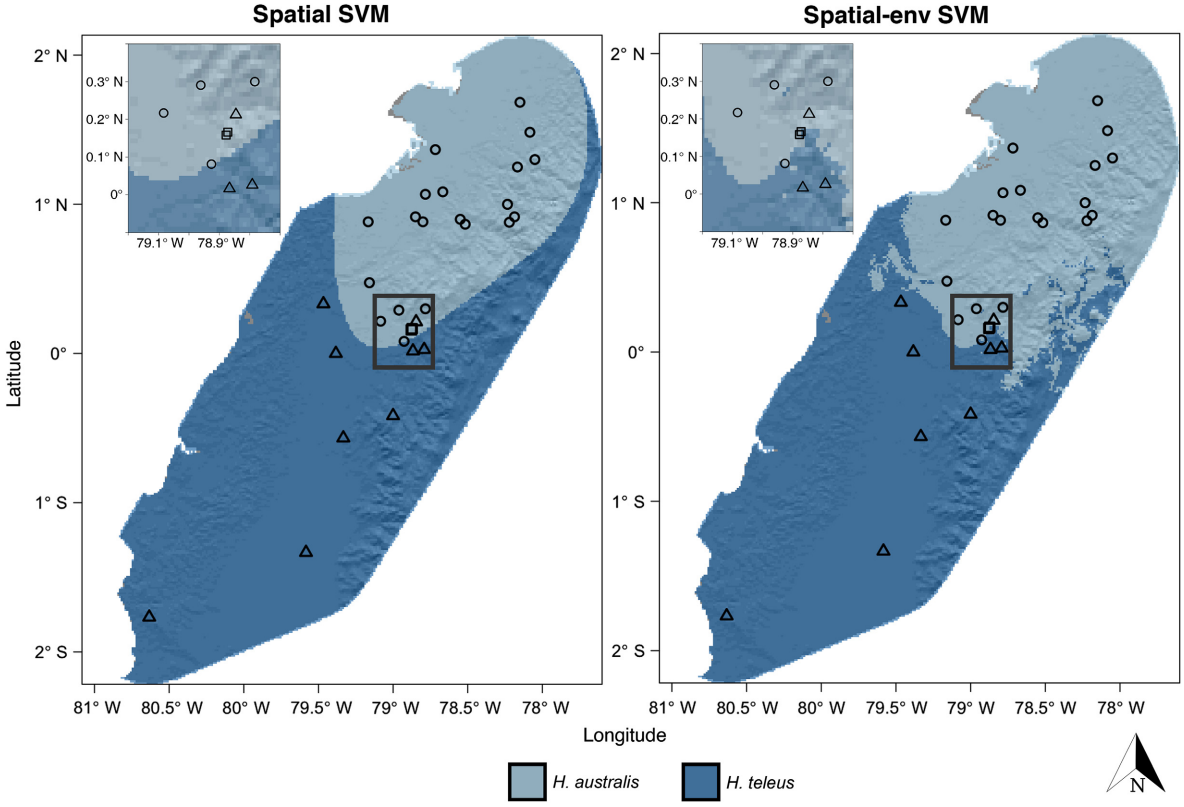

FIG. 4. Support vector machine classifications for *H. australis* (light blue) and *H. teleus* (dark blue) over the shared region on an elevation hillshade. The "spatial SVM" refers to the model fit with species occurrence coordinates only, and the "spatial-environmental SVM" to that fit with both coordinates and continuous SDM suitability predictions. Occurrence localities for *H. australis* (circles), *H. teleus* (triangles), and inferred hybrids (squares) are shown for reference. The upper-left box shows a zoomed-in view of the location of the inferred hybrids.

TABLE 1. Area of occupancy (AOO) estimates (km^2) for each approach, before and after considering current forest cover.

	Before		After	
Species and approach	Total area (km^2)	Total SDM prediction area (%)	Forest area (km^2)	Total SDM prediction area (%)
Heteromys australis				
Occupied grid cells	104	0.38	100	0.37
SDM prediction	27,280	100	24,100	88
Spatial SVM	22,572	83	20,288	74
Spatial-environmental SVM	24,472	90	21,696	80
Heteromys teleus				
Occupied grid cells	40	0.09	36	0.08
SDM prediction	46,360	100	25,064	54
Spatial SVM	40,652	88	19,812	43
Spatial-environmental SVM	39,000	84	18,544	40

Notes: Percent of total species distribution model (SDM) prediction area (before considering forest cover) is also reported to provide comparisons with the highest areal estimations.

shared region) reduced the AOO estimates minimally for *H. australis* but by more than one-half in some cases for *H. teleus* (Table 1). The SDM prediction estimates (thresholded by MTP) had no omissions by definition, but they resulted in some occurrence localities predicted to be in the congener's range (three for *H. australis* and five for *H. teleus*). These could be interpreted as misclassifications, or alternatively as areas of possible sympatry. In contrast, both SVM approaches had one omission (and thus one misclassification) per species.

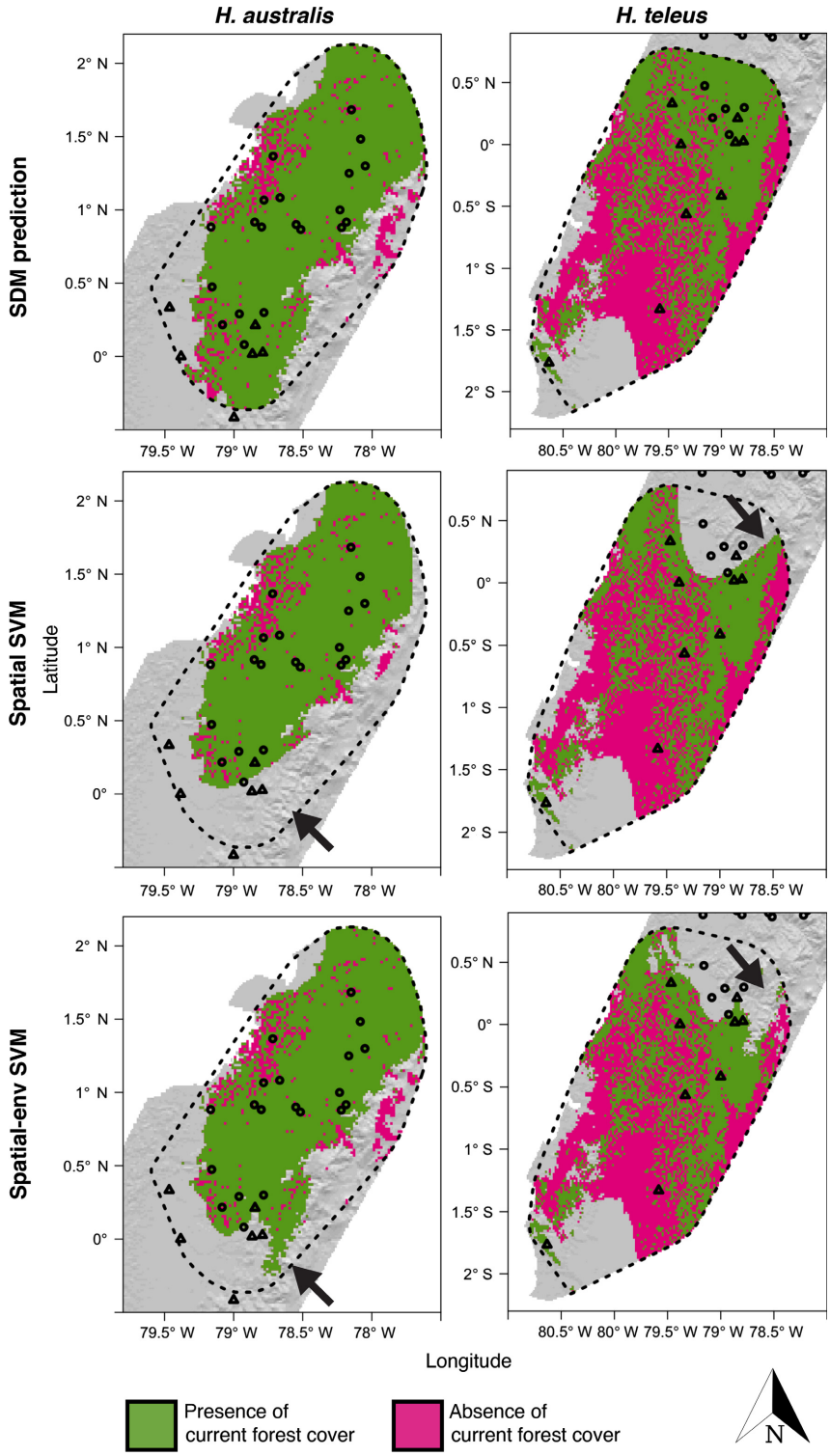

FIG. 5. Estimations of area of occupancy (AOO) for each species (geographic coordinate system) showing areas with current broadleaf evergreen forest cover (green) and those without (pink) over a gray elevation hillshade, representing areas outside the respective estimates of AOO. The dashed line represents the estimated extent of occurrence (EOO), and occurrence localities are shown for *H. australis* (circles) and *H. teleus* (triangles). Black arrows denote cool and mesic areas excluded from the spatial SVM estimate of *H. australis* but included in the spatial-environmental SVM estimate (or the inverse for *H. teleus*). Projected from UTM 17S (projected: meters) to WGS84 (geographic).

Although spatial overlap was high among the SDM-based approaches (more so for SVM-derived estimates; Appendix S3: Table S2), there was substantial variation at the contact zone boundary (Fig. 5). The SDM prediction estimate extended well into the known and inferred portions of the congener's range for each species, whereas the SVM-derived estimates were more restrictive around the contact zone and thus matched ecological expectations better. In comparison with the spatial SVM, the spatial-environmental SVM estimate for *H. australis* included a cool and mesic area in the center-east portion of the contact zone, which was classified as within the range of *H. teleus* by the spatial SVM approach (Fig. 5). The environmental characteristics of this area align best with the associations of *H. australis*, a designation made successfully by the spatial-environmental SVM.

Discussion

Support vector machine approaches improved AOO bound estimates

Using range area estimates derived from species distribution models (SDMs) to determine IUCN Red List threat categories can address issues of incomplete sampling, but we demonstrate here that also accounting for the spatial effects of inferred biotic interactions can result in more realistic AOO estimates even for a data-poor species. Using more accurate AOO estimates that better reflect the focal species' ecology, as well as considering different estimates as bounds for low-data species, can lead to better prioritization of critical habitat, selection of reserves, and estimates of impact from future climate and land use change (Guisan et al. 2013). Using SDMs allowed us to include areas in the EOO (within expert-delimited areas of likely dispersal) that lacked known occurrence records yet had high predicted suitability, which in turn expanded the potential extent of AOO. Additionally, we employed support vector machines to remove areas predicted to be in the range of each species' parapatric congener. Although we did not consider this in our study, it is important to note that EOO estimates could also be reduced by removing biotically unsuitable areas (section 4.9; IUCN 2019). Both SVM approaches predicted fewer localities of each species to be in their congener's range than the unprocessed SDM range prediction did, and we found that the SVM approach using spatial and SDM-derived predictor variables (spatial-environmental) resulted in delineations of the contact zone that aligned better with our ecological expectations. More occurrence data and consideration of other environmental predictor variables would likely improve range predictions within this contact zone, but we were still able to make sensible range-limit predictions for these data-poor species. Other factors not directly considered in this study such as the removal of occurrence outliers, increased precision in georeferencing, or different SDM or classification approaches may also change the AOO estimates, resulting in different bounds. Though these issues fall beyond the scope of the current product, future studies should evaluate their relative impact. Regardless, the point remains that removing biotically unsuitable areas with prior ecological knowledge using an operational approach should result in more conservative AOO estimates that are more ecologically realistic. We suggest that these methods should be attempted for other systems where competitive exclusion is inferred to assess how generally applicable they may be.

These analyses also indicate priority areas for future sampling to improve distributional estimates for these two spiny pocket mouse species. This is particularly true for the provinces Esmeraldas (west) and Pichincha (central) where no records in close proximity between these species exist. For *H. teleus*, there are several priority areas for new sampling efforts based on the current occurrence data and our SDM results. One is the Cordillera Chongón-Colonche in the far southwest, where two unique occurrence localities from 1923 are currently known (Anderson and Jarrín-V. 2002). Although this area is heavily deforested, some remaining forest still exists there, and confirmation of a possibly disjunct and imperiled population is important for conservation. If it is found that this population has since been extirpated, the EOO based on the occurrence localities for *H. teleus* would shrink by more than one-half, from 18,686 to 8,234 km^2. Another is the coastal plain in central-west Ecuador, which has been heavily developed for oil palm and banana plantations (Anderson and Jarrín-V. 2002), as it is feasible that populations exist there in forest patches. The last is southern Ecuador west of the crest of the Andes, as the southern limit of the range of *H. teleus* is currently unresolved, and discoveries that extend the range farther south would be of great conservation interest. However, modeled suitability and the presence of forest does not necessitate the existence of populations, which can be in decline or absent due to insufficiently large patch size or other anthropogenic disturbances not assessed (such as hunting or extraction of plant resources). Potential for the discovery of new populations and range limits makes sampling efforts in these areas worthwhile.

New geographic estimates and associated threat categories

These results provide new geographic estimates for conservation, which we propose should lead to IUCN Red List updates for one species. Although *H. australis* is currently labeled Least Concern (Anderson et al. 2018), and our analysis considers only part of its range, the range estimates we present can be informative for its distribution in southwestern Colombia and Ecuador. Future efforts should focus on refining the georeferences for the existing occurrence localities farther north, as

well as identifying museum specimens or collecting field specimens for gaps in the known range. A climate-based SDM that additionally includes *H. australis* occurrences outside of the region considered in this study may lead to discrepancies in range estimates due to an expansion of the environments considered by the model. Nonetheless, based on our current knowledge of the species, we still expect to observe a similar degree of overlap between the range estimates of these two species that SVMs can help address.

In contrast, *H. teleus* (whose entire known range is considered here) is labeled Vulnerable D2 due to a shortage of occurrence data (previously there were five to seven reported localities, meeting the restriction of five or fewer) and an estimated "13–19% suitable habitat remaining" (Naylor and Roach 2018). We updated the occurrence data set for *H. teleus* from 7 to 10 verified unique localities, which now further exceeds the requirement for Vulnerable D2 (≤5) but still makes the sub-requirement for Vulnerable B1 or B2 (≤10). The current IUCN listing reports EOO to be 46,156 km^2, which is based on an estimation using the existing IUCN range map (Shelby McCay, *personal communication*); this is close to our SDM-based EOO estimate (54,985 km^2). Although both these estimates exceed the threshold for Vulnerable B1 (<20,000 km^2), our lower-end EOO estimate based on the occurrence localities (18,686 km^2) would qualify.

The AOO estimates we present, on the other hand, are the first for *H. teleus*. As stated earlier, because we currently lack data to determine the occupancy status of modeled range areas for *H. teleus*, we present these AOO estimates as bounds and not as a single definitive calculation. Assuming sub-requirements for threat categories are met, the occupied grid cells estimate (36 km^2; low bound of estimates) would correspond to Endangered B2 (<500 km^2), and consideration of the estimates based on the SDM prediction (25,064 km^2; upper bound of estimates) and both SVM approaches (18,544–19,812; mid-range estimates) could correspond to Near Threatened given the uncertainty necessary for this attribution (section 10; IUCN 2019). Due to the sparse sampling in this tropical region (relative to 4 km^2 grid cells), we interpret that the occupied grid cells estimate is a strongly biased underestimate. Considering the continuing loss of forest cover in this region of Ecuador (Van der Hoek 2017), the assumptions concerning the geographic range category B subrequirements and future threatened status seem likely. The differences between the AOO bound estimates considering forest cover from the SDM prediction and SVM approaches are not large enough to result in a difference of threat category. Nevertheless, as the SVM-derived estimates were lower (9–14%; Table 1), they are more likely to result in higher threat categories for *H. teleus* if applied on data sets with less total forest cover (with present or future estimates). As mentioned earlier, if future field sampling determines the *H. teleus* population in the Cordillera de Chongón-Colonche is extirpated, the resulting EOO would shrink considerably and result in lower AOO estimates. Regardless, the problem remains that although we have realistic estimates for upper bounds of AOO (especially those from the SVMs), we still lack those for the lower bounds. Based on the above conclusions concerning IUCN threat categories and the associated uncertainties, we recommend that the listed AOO for *H. teleus* include the bounds we estimated. We propose conducting intensive fieldwork in a randomized sample of sites from within the spatial-environmental SVM estimate to determine a prevalence value, which could then be multiplied with the upper bound AOO estimates to narrow them further.

Considering gradients of AOO for IUCN Assessments

Especially for data-poor species with perceived extinction threats that lack an AOO estimate, such as the two species in this study, an AOO gradient from lower (restrictive) to upper (expansive) bounds based on different approaches with varying levels of uncertainty and ecological realism can be informative for conservation (sections 3.2.2 and 4.10.8; IUCN 2019). The lowest (and unrealistic) bound should usually be the occupied grid cells estimate, except in extreme cases with near-perfect sampling. All approaches that use either SDMs or expert maps in some way should fall between the extreme bounds of the gradient, assuming these AOO estimates are subsets of the EOO (the highest possible bound). Techniques that subset an expert map or SDM prediction based on ancillary information, such as known elevational limits (Harris and Pimm 2008), habitat estimates derived from remote sensing data products (Brooks et al. 2019), or inferred biotic interactions as in this study will result in estimates that are presumably far greater than the occupied grid cells estimate. Unmodified SDM predictions, particularly those thresholded by the minimum training presence value (MTP), will result in even higher estimates. Most expert maps will also likely fall in a similar position on the gradient, as they tend to be general in nature and thus suffer from high commission errors unless modified (Rondinini et al. 2006). Finally, the highest (yet unrealistic) bound on the AOO gradient should be the unmodified EOO estimate delineated around known occurrence localities or additionally including inferred or projected occurrences based on estimated available habitat (Brooks et al. 2019). The most realistic AOO for a majority of species likely falls somewhere between the marginal bounds of this gradient. In accordance with the Red List Guidelines, we recommend using the lowest estimate that is also realistic (section 3.2.4; IUCN 2019), such as those from the SVM approaches and additional consideration of forest cover presented here.

AOO estimates that account for important spatial restrictions, such as those imposed by biotic interactions, can lead to more accurate range estimates that better inform biogeographic studies as well as conservation

decisions such as planning long-term monitoring, assessing current or future extinction risk, and managing reserve systems. If species' ranges are incorrectly estimated, sparse funds could be misdirected, conservation actions could be negatively affected, and the species' survival could even be impacted. More broadly, we recommend this approach for other species that have parapatric ranges apparently impacted by biotic interactions, particularly for data-poor species of concern.

Conclusions

Despite growing evidence that biotic interactions help to shape species' ranges at the macroscale (Araújo and Luoto 2007, Wisz et al. 2013), the explicit, operational exclusion of inferred biotically unsuitable areas in estimates of AOO for IUCN conservation assessments is rarely, if at all, practiced. This is true despite the necessity for AOO to represent areas with good evidence of actual occupancy, and certainly not areas associated with the presence of a parapatric congener. Although we have reasons to hypothesize that competitive exclusion is driving parapatry between our two study species, confirming this mechanism is extremely difficult given the scarcity of field observations and difficulty of manipulative experiments for these nocturnal and elusive rodents (Koplin and Hoffmann 1968, Anderson et al. 2002), as well as their rarity even in museum collections. Given that confirming biotic interactions in general is difficult for most species based on macroscale data (Morales-Castilla et al. 2015), predictions for threatened species that account for inferred interactions should be compared with others in a gradient of different AOO estimates to address the inherent uncertainty in assigning areal values to ranges of low-data species.

Given that the competitive interactions between our two study species are indeed inferred, we thus acknowledge that other factors may be responsible for the current distributional patterns we observe. However, regardless of whether or not the underlying mechanism is biotic in nature, the results of the SVM approaches align better with our expectations regarding the spatial extents of these species' ranges based on the limited but well-georeferenced occurrence data available. Additionally, as mentioned above, we advocate for the collection of data to determine species prevalence, which would facilitate estimation of more realistic lower bounds of AOO. For species pairs or larger groups of species that are spatially structured by strongly supported biotic interactions, these techniques can help remove areas that are biotically unsuitable from AOO estimates. Nonetheless, for systems that are hypothesized to be spatially structured by interactions (i.e., *not verified* but with strongly suggestive evidence) such as the two species in this study, the SVM-derived estimates can make important contributions to more realistic upper bounds of AOO.

In addition to accounting for competition between parapatric species, we envision that the technique we demonstrate could be applied to other kinds of inferred interactions that result in similar effects on ranges. One example would be if the presence of a generalist parasite or parasitoid results in such high levels of mortality for a host species that it cannot sustain populations. If sites of absence for the parasite/parasitoid exist (in addition to sites of presence), SVMs could be used to mask out areas where it occurs from the AOO estimate of the host species. However, this technique can only be used when each species considered can subsist without the other, and hence would not work for most specialist parasites or parasitoids. Our aim here was not to propose SVMs as the optimal technique for considering biotic interactions in determining species ranges and calculating AOO estimates, but to more generally advocate for automated methods that can classify species ranges based at least partially on the positions of occurrence localities for the purpose of masking out areas of inferred biotic unsuitability. Future work would further benefit biogeography and conservation biology by comparing among methods to determine which are optimal for which systems and research goals.

Acknowledgments

J. M. Kass, R. P. Anderson, and S. F. Burneo conceptualized the project. Acquisition and vetting of the occurrence data was done by N. Tinoco, S. F. Burneo, and R. P. Anderson. Georeferencing and cataloging of the occurrence data was done by S. I. Meenan and J. M. Kass with oversight by R. P. Anderson. J. M. Kass conducted the analysis with conceptual guidance by R. P. Anderson. J. M. Kass drafted the manuscript with critical edits by S. I. Meenan and R. P. Anderson. All authors contributed to and approved the final manuscript for publication. In addition to funding from the City College of New York, the Graduate Center of the City University of New York, and the Pontificia Universidad Católica del Ecuador, this work was supported by the U.S. National Aeronautics and Space Administration (#80NSSC18K0406) and National Science Foundation (DBI-1661510, including Research Experiences for Undergraduates participation by S. I. Meenan). S. I. Meenan received additional funding from the City College Fellowship. New occurrence records for *Heteromys* in Ecuador derived in part from a grant from the Theodore Roosevelt Memorial Fund, American Museum of Natural History (to R. P. Anderson). We thank curators and collections staff at the following institutions for assistance and/or access to specimens under their care: Departamento de Ciencias Biológicas, Escuela Politécnica Nacional, Quito; Instituto Nacional de Biodiversidad, Quito (formerly Museo Ecuatoriano de Ciencias Naturales); Museo de Zoología, Pontificia Universidad Católica del Ecuador, Quito; and Natural History Museum, London (formerly British Museum [Natural History]). We would like to thank H. R. Akçakaya, J. D. Anadón, C. Babich Morrow, M. E. Blair, A. C. Carnaval, B. E. Gerstner, E. E. Johnson, L. L. Manne, A. Paz, G. E. Pinilla-Buitrago, M. W. Tingley, and two anonymous reviewers for their comments that helped improve this manuscript. C. Boada-Terán, S. D. Claxton, D. P. Gillman, R. D. Jarrín-E., and N. Rana assisted with compilation of the occurrence data and/or preliminary analyses.

Literature Cited

Aiello-Lammens, M. E., R. A. Boria, A. Radosavljevic, B. Vilela, and R. P. Anderson. 2015. spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography 38:541–545.

Anderson, R. P. 2015. Family Heteromyidae Gray, 1868. Pages 51–58 *in* J. L. Patton, U. F. J. Pardiñas, and G. D'Elía, editors, Mammals of South America, Volume 2. University of Chicago Press, Chicago, Illinois, USA.

Anderson, R. P. 2017. When and how should biotic interactions be considered in models of species niches and distributions? Journal of Biogeography 44:8–17.

Anderson, R. P., and P. Jarrín-V. 2002. A new species of spiny pocket mouse (Heteromyidae: *Heteromys*) endemic to western Ecuador. American Museum Novitates 2002:1–26.

Anderson, R. P., and E. Martínez-Meyer. 2004. Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. Biological Conservation 116:167–179.

Anderson, R. P., A. T. Peterson, and M. Gómez-Laverde. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. Oikos 98:3–16.

Anderson, R. P., D. Tirira, and R. Samudio 2016. *Heteromys australis* (errata version published in 2017). The IUCN Red List of Threatened Species 2016:e.T10005A115095587. https://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T10005A22223850.en

Araújo, M. B. et al. 2019. Standards for distribution models in biodiversity assessments. Science Advances 5:eaat4858.

Araújo, M. B., and M. Luoto. 2007. The importance of biotic interactions for modelling species distributions under climate change. Global Ecology and Biogeography 16:743–753.

Bateman, B. L., J. Vanderwal, S. E. Williams, and C. N. Johnson. 2012. Biotic interactions influence the projected distribution of a specialist mammal under climate change. Diversity and Distributions 18:861–872.

Brito, D., R. G. Ambal, T. Brooks, N. De Silva, M. Foster, W. Hao, C. Hilton-Taylor, A. Paglia, J. P. Rodríguez, and J. V. Rodríguez. 2010. How similar are national red lists and the IUCN Red List? Biological Conservation 143:1154–1158.

Brooks, T. M. et al. 2019. Measuring Terrestrial Area of Habitat (AOH) and its utility for the IUCN Red List. Trends in Ecology & Evolution 34:977–986.

Drake, J. M., C. Randin, and A. Guisan. 2006. Modelling ecological niches with support vector machines. Journal of Applied Ecology 43:424–432.

European Space Agency. 2018. Land Cover Climate Change Initiative (CCI) Climate Research Data Package, version 2.0.7. http://maps.elie.ucl.ac.be/CCI/viewer/download.php

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 4315:4302–4315.

Fivaz, F. P., and Y. Gonseth. 2014. Using species distribution models for IUCN Red Lists of threatened species. Journal of Insect Conservation 18:427–436.

Fortin, M. J., T. Keitt, B. Maurer, M. Taper, D. M. Kaufman, and T. Blackburn. 2005. Species' geographic ranges and distributional limits: pattern analysis and statistical issues. Oikos 108:7–17.

Freeman, B. G., and N. A. Mason. 2015. The geographic distribution of a tropical montane bird is limited by a tree: Acorn Woodpeckers (*Melanerpes formicivorus*) and Colombian Oaks (*Quercus humboldtii*) in the Northern Andes. PLoS ONE 10:e0128675.

Gaston, K. J. 2009. Geographic range limits of species. Proceedings of the Royal Society B 276:1391–1393.

Guisan, A. et al. 2013. Predicting species distributions for conservation decisions. Ecology Letters 16:1424–1435.

Gutiérrez, E. E., R. A. Boria, and R. P. Anderson. 2014. Can biotic interactions cause allopatry? Niche models, competition, and distributions of South American mouse opossums. Ecography 37:741–753.

Harris, G., and S. L. Pimm. 2008. Range size and extinction risk in forest birds. Conservation Biology 22:163–171.

Hijmans, R. J. 2017. raster: Geographic Data Analysis and Modeling. R package version 2.6-7. https://CRAN.R-project.org/package=raster

Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith. 2017. dismo: Species Distribution Modeling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo

Hof, A. R., R. Jansson, and C. Nilsson. 2012. How biotic interactions may alter future predictions of species distributions: future threats to the persistence of the arctic fox in Fennoscandia. Diversity and Distributions 18:554–562.

Hsu, C. W., C. C. Chang, and C. J. Lin. (2003) A practical guide to support vector classification. https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

IUCN (2019) Guidelines for Using the IUCN Red List Categories and Criteria, Version 14. IUCN Standards and Petitions Subcommittee. http://www.iucnredlist.org/documents/RedListGuidelines.pdf

Kass, J. M., B. Vilela, M. E. Aiello-Lammens, R. Muscarella, C. Merow, and R. P. Anderson. 2018. *Wallace*: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. Methods in Ecology and Evolution 9:1151–1156.

Koplin, J. R., and R. S. Hoffmann. 1968. Habitat overlap and competitive exclusion in voles (*Microtus*). American Midland Naturalist 80:494–507.

Lobo, J. M., A. Jiménez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17:145–151.

Marcer, A., L. Sáez, R. Molowny-Horas, X. Pons, and J. Pino. 2013. Using species distribution modelling to disentangle realised versus potential distributions for rare species conservation. Biological Conservation 166:221–230.

Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. 2017. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.6-8. https://CRAN.R-project.org/package=e1071

Ministerio del Ambiente de Ecuador (2019) Bosques y vegetación protectora. http://mapainteractivo.ambiente.gob.ec/portal

Morales-Castilla, I., M. G. Matias, D. Gravel, and M. B. Araújo. 2015. Inferring biotic interactions from proxies. Trends in Ecology & Evolution 30:347–356.

Muscarella, R., P. J. Galante, M. Soley-Guardia, R. A. Boria, J. M. Kass, M. Uriarte, and R. P. Anderson. 2014. 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.

Naylor, L., and N. Roach. 2018. *Heteromys teleus*. The IUCN Red List of Threatened Species 2018:e.T136333A22223930. https://doi.org/10.2305/IUCN.UK.2018-1.RLTS

Noon, B. R., L. L. Bailey, T. D. Sisk, and K. S. McKelvey. 2012. Efficient species-level monitoring at the landscape scale. Conservation Biology 26:432–441.

Ocampo-Peñuela, N., and S. L. Pimm. 2014. Setting practical conservation priorities for birds in the western Andes of Colombia. Conservation Biology 28:1260–1270.

Peterson, A. T., A. G. Navarro-Sigüenza, and A. Gordillo. 2018. Assumption-versus data-based approaches to summarizing species' ranges. Conservation Biology 32:568–575.

Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura, and M. B. Araújo 2011. Ecological niches and geographic distributions (MPB49), Volume 56. Princeton University Press, Princeton, New Jersey, USA.

Phillips, S. J., R. P. Anderson, M. Dudík, R. E. Schapire, and M. E. Blair. 2017. Opening the black box: an open-source release of Maxent. Ecography 40:887–893.

Pouteau, R., J. Y. Meyer, and B. Stoll. 2011. A SVM-based model for predicting distribution of the invasive tree *Miconia calvescens* in tropical rainforests. Ecological Modelling 222:2631–2641.

R Core Team 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org

Roberts, D. L., L. Taylor, and L. N. Joppa. 2016. Threatened or data deficient: assessing the conservation status of poorly known species. Diversity and Distributions 22:558–565.

Rondinini, C., K. A. Wilson, L. Boitani, H. Grantham, and H. P. Possingham. 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecology Letters 9:1136–1145.

Shcheglovitova, M., and R. P. Anderson. 2013. Estimating optimal complexity for ecological niche models: A jackknife approach for species with small sample sizes. Ecological Modelling 269:9–17.

Sistema de Información Ambiental de Colombia 2019. Reservas forestales protectoras nacionales. http://www.siac.gov.co/catalogo-de-mapas;downloaded04-10-2019

Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecology Letters 10:1115–1123.

Syfert, M. M., L. Joppa, M. J. Smith, D. A. Coomes, S. P. Bachman, and N. A. Brummitt. 2014. Using species distribution models to inform IUCN Red List assessments. Biological Conservation 177:174–184.

Urbina-Cardona, J. N., and O. Flores-Villela. 2010. Ecological-niche modeling and prioritization of conservation-area networks for Mexican herpetofauna. Conservation Biology 24:1031–1041.

Van der Hoek, Y. 2017. The potential of protected areas to halt deforestation in Ecuador. Environmental Conservation 44:124–130.

Wisz, M. S. et al. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biological Reviews 88:15–30.

Zhang, X., and A. C. Vincent. 2018. Predicting distributions, habitat preferences and associated conservation implications for a genus of rare fishes, seahorses (*Hippocampus* spp.). Diversity and Distributions 24:1005–1017.

Supporting Information

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/eap.2228/full

Data Availability Statement

Data are available on Figshare: https://doi.org/10.6084/m9.figshare.12725873